UNIVERSITY OF SOUTHAMPTON

Neurofuzzy and SUPANOVA Modelling of the Processing-Property Relationships of Aerospace Al-Alloys

by

Oliver Paul Femminella

A thesis submitted for the degree of
Doctor of Philosophy

in the
Faculty of Engineering and Applied Science
School of Engineering Sciences

September 2000

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

Faculty of Engineering and Applied Science

School of Engineering Sciences

Doctor of Philosophy

Neurofuzzy and SUPANOVA Modelling of the Processing-Property Relationships of Aerospace Al-Alloys

by Oliver Paul Femminella

Whilst direct physical understanding of materials behaviour that may be obtained from experimental and theoretical investigations is essential to materials development, the complex, multivariate nature of commercial processing routes may preclude the use of such explicit understanding for real materials/process development and optimisation. The development of process models which scale up to industrial processing environments thus remains elusive. In recent years there has been a growing interest in applying neural networks as alternative approaches in modelling material properties. A significant drawback however of general neural network architectures remains the lack of transparency in the modelling process.

Neurofuzzy networks and the SUPANOVA technique comprise two parsimonious adaptive modelling approaches, the former combining well established neural-type learning algorithms with the transparent knowledge representation of fuzzy systems, the latter emerging from recent advances in statistical learning theory and support vector methods for regression.

This thesis has investigated the performance of such techniques in modelling physical and tensile properties of Al-Mg-Li powder metallurgy and wrought Al-Zn-Mg-Cu alloy systems, from compositional and processing information. Prior system knowledge was employed in the form of physically motivated transformations and initialising (neurofuzzy) network structures.

By adapting their structure to infer the nature of the processing-property relationships contained in the data, both adaptive methods determined a number of non-linear dependencies, resulting in more appropriate models compared with multiple linear regression analyses. The data sets were seen to be representative of experimental, small and large scale processing conditions, which then reflected the different predictive performances exhibited by the adaptive methods.

Metallurgical understanding, conditioning and regression diagnostics, allowed a greater understanding of the statistical properties of the data, permitting an enhanced interpretation and validation of the models identified by the adaptive methods, understanding the empirical results in light of the representativeness of the set of input variables, sample sizes and data weaknesses characterising the data. Generally, the dependencies inferred by the adaptive methods were seen to be consistent with metallurgical understanding, a number of which suggesting interesting interdependencies, particularly an interaction between the Magnesium content of Al-Zn-Mg-Cu alloys and the age-hardening behaviour. Overall, the results gave a clear indication of the benefits associated by performing statistical analyses on experimentally designed data sets and highlighted the problems of modelling observational data.

Contents

Acknowledgements

There a many people that have made this thesis possible, none more so than my main supervisors Dr. P.A.S. Reed and Dr. I. Sinclair; I am deeply grateful for all their support and encouragement during the course of this work, and their patience in reading draft manuscripts.

I would also like to acknowledge Dr. M. Brown for his supervision in the initial development of this work and in the later stages of my work Prof. C.J. Harris.

The financial support of the Faculty of Applied Science and Federal Mogul Technology (Mr. J. Woodthope in particular) is gratefully acknowledged, as are the various suppliers of the data sets. These include Mr. A. Tarrant at Aerospace Metal Composites, U.K., Drs. J. Newman, S. Waters and Mr. A. Morris at British Aluminium Plate and finally Dr. P. Pitcher in the Mechanical Sciences Sector (MSS) Defence and Evaluation Research Agency (DERA).

I would like to thank Dr. M.J. Starink for his contribution made in Chapter 8, with whom I have enjoyed collaborating with, and which I believe constituted the second significant development in the successful continuation of this work.

I am grateful to a number of people in the ISIS research group, both past and present, for a number of useful MATLAB scripts, but particularly Dr. S.R. Gunn for making the SUPANOVA C++ code available. I thank all other colleagues both in the Materials and ISIS research groups, with whom I have had constructive criticism, and have made Southampton an interesting place to work.

Finally, I would like to thank my mother and father, who have always been immeasurably encouraging and supportive throughout my university career. It is to them that I dedicate this thesis.

Nomenclature

The following list includes the main terms used throughout this thesis. A number of additional terms are however identified and defined at specific points in the text, but are not generally repeated.

General Physical Metallurgy

$\sigma_{0.2}$	-	0.2% proof stress.
E	-	Young's modulus of elasticity.
σ_{el}	-	electrical conductivity.
uts	-	ultimate tensile strength.
$\%el.$	-	percentage elongation.
K_{Ic}	-	plane strain fracture toughness.
ΔH_{sol}	-	enthalpy of formation.
k_B	-	Boltzmann's constant ($13.8 \times 10^{-24} J/K$).
ν	-	Poisson's ratio.

General Data Modelling

N	-	number of samples (or observations).
$\mathcal{D}_N$	-	training data set (sample) comprising N data pairs (samples).
	=	$\{(\mathbf{x}_i, y_i)\}_{i=1}^N$
$\mathbf{x}$	-	vector of input samples.
	=	$[x_1, \ldots, x_n]^T$.
$\mathbf{y}$	-	vector of output samples (target values).
	=	$[y_1, \ldots, y_n]$.
$\mathbf{w}$	-	vector of regression coefficients (or weights).
$\mathbf{X}$	-	matrix of input vectors.
	=	$[\mathbf{x}_1, \ldots, \mathbf{x}_p]$.

$\hat{y}(\mathbf{x}, \mathbf{w})$	-	model output for the inputs in $\mathbf{x}$ and parameter vector $\mathbf{w}$.
$\mathbf{w}_{mp}$	-	maximum posterior estimate of the weights.
n_x	-	number of input variables in the model.
p	-	number of parameters in the model.
df	-	number of degrees of freedom in the model.
J_N	-	Mean Square Error (MSE) cost function.
	=	$\frac{1}{N}\sum_{i=1}^{N}[y_i - \hat{y}_i(\mathbf{x}, \mathbf{w})]^2$.
$\hat{\sigma}_N^2$	-	biased estimate of the variance of the assumed additive noise component.
	-	This corresponds to the Mean Square Error.
$\hat{\sigma}_{\text{df}}^2$	-	unbiased estimate of the variance of the assumed additive noise component.
	=	$\frac{1}{N-\text{df}}\sum_{i=1}^{N}[y_i - \hat{y}_i(\mathbf{x}, \mathbf{w})]^2$.
$\hat{\sigma}_{\text{df,std}}^2$	-	unbiased estimate of the variance of the assumed additive noise component determined on the output standardised to have zero mean and unit variance.
$\hat{\sigma}_w^2$	-	variance of the estimated model parameters.
$\sigma_{\hat{y}}^2$	-	variance of the assumed Gaussian noise on the model's output due to the variance in the model's weights. This is used to define error bars.
τ	-	signal-to-noise parameter.
$\mathbf{r}_{XX}$	-	matrix of correlation coefficients between input variables, with elements r_{xx}.
$\mathbf{r}_{YX}$	-	vector of correlation coefficients between the output and the input variables, with elements elements r_{yx}.
$C(\mathbf{X})$	-	condition number of the design matrix $\mathbf{X}$.
μ_k	-	k^{th} singular value of $\mathbf{X}$.
μ_{max}	-	largest singular value of $\mathbf{X}$.
μ_{min}	-	smallest singular value of $\mathbf{X}$.
η_k	-	k^{th} condition index of $\mathbf{X}$.
π_{jk}	-	$(j,k)^{th}$ variance-decomposition proportion: proportion of the k^{th} regression coefficient associated with the j^{th} component of its decomposition.
$\mathbf{H}$	-	hat matrix.
h_{ii}	-	i^{th} diagonal element of the hat matrix $\mathbf{H}$.
e_i	-	i^{th} error (or residual) between the observed value and the model's output.
	=	$y_i - \hat{y}_i(\mathbf{x}, \mathbf{w})$.
e_i^*	-	i^{th} semistudentised residual.
r_i	-	i^{th} studentised residual.
C_i^2	-	Cook's distance measure for the i^{th} observation.
H_i^2	-	Hadi's overall potential measure for the i^{th} observation.

Neurofuzzy Networks

$\mu_A(x)$	-	a fuzzy set, where A represents the fuzzy label.
$\mu_{\mathbf{A}^i}(\mathbf{x})$	-	i^{th} multivariate fuzzy set, the vector $\mathbf{A}$ represents the fuzzy labels. This also represents the basis functions in the neurofuzzy network.
A_k^i	-	fuzzy input labels.
B^j	-	fuzzy output labels.
c_{ij}	-	a rule confidence.
y_j^c	-	centre of the j^{th} fuzzy output set.
$a(x)$	-	j^{th} univariate B-spline basis function.
$\mathbf{a}(\mathbf{x})$	-	vector of membership function outputs.
p	-	number of multivariate fuzzy membership functions (different antecedents) in a neurofuzzy network (rule-base). This is also equal to the number of network weights.
U	-	number of subnetworks.
λ	-	vector of knots representing a B-spline fuzzy variable.
r	-	number of univariate membership functions defined on a variable.
p_u	-	number of multivariate basis functions in the u^{th} subnetwork. This also corresponds to the number of weights in the subnetwork.
k	-	order of a B-spline fuzzy membership function.
$\mathbf{A}$	-	$[N \times p]$ solution matrix.
$\mathbf{R}$	-	$[p \times p]$ autocorrelation matrix.
$\mathbf{p}$	-	the $[p \times 1]$ cross-correlation vector.
$\mathbf{S}$	-	the $[N \times N]$ smoother matrix.
s_{ii}	-	i^{th} diagonal element of the smoother matrix $\mathbf{S}$.
F_m	-	failure margin in the model construction termination criteria.
f_{tol}	-	forward tolerance in the model construction termination criteria.
b_{tol}	-	backward tolerance in the model construction termination criteria.

Regularisation

λ	-	regularisation coefficient.
J_R	-	regularised cost function.
$E_D(w)$	-	sum of squared error $(N \times MSE)/2$.
$E_w(w)$	-	penalty term.
$E_w^u(w)$	-	local penalty term.
$\mathbf{K}$	-	$[p \times p]$ curvature matrix.
$\mathbf{K}_i$	-	local $[p_i \times p_i]$ curvature matrix.

Probabilities

$P(\cdot)$	-	probability.
$p(\mathbf{x})$	-	probability density function (p.d.f.).
$p(\mathbf{x}, y)$	-	joint probability density function.
$p(y\|\mathbf{x})$	-	conditional density.
$\mathcal{X}$	-	set of input vectors.
	=	$(\mathbf{x}_1, \ldots, \mathbf{x}_N)^T$.
$\mathcal{Y}$	-	vector of output samples (target values).
	=	$(y_1, \ldots, y_N)$.
$\mathcal{H}$	-	represents model structure and type of regulariser used.
α	-	reciprocal of the variance of assumed Gaussian prior distribution.
β	-	reciprocal of the variance of assumed Gaussian likelihood function, i.e. the distribution of the assumed additive Gaussian noise.
$P(\mathbf{x}, \alpha)$	-	prior for the weights.
$P(\mathbf{x}, \alpha_u)$	-	local prior for the weights in a subnetwork.
$P(\mathcal{Y}\|\mathcal{X}, \mathbf{w}, \alpha, \beta, \mathcal{H})$	-	likelihood function.
$P(\mathcal{Y}\|\mathcal{X})$	-	evidence for the output data.
$P(\mathbf{w}\|\mathcal{D}_N, \alpha, \beta)$	-	posterior p.d.f. for the weights.
$P(\alpha, \beta)$	-	prior distribution for the hyperparameters.
$P(\alpha, \beta\|\mathcal{D}_N, \mathcal{H})$	-	posterior p.d.f. for the hyperparameters α and β for a given model structure and type of regulariser used.
Z_D	-	normalisation factor for the likelihood function.
Z_W	-	normalisation factor for the prior p.d.f.

Support Vector Machines

$R[f]$	-	expected risk functional or average loss.
$R_{emp}[f]$	-	empirical risk functional.
L_{quad}	-	quadratic loss function.
L_ϵ	-	ϵ-insensitive loss function (equivalent to the Laplacian loss for $\epsilon = 0$).
$K(x, y)$	-	general kernel function.
ξ	-	slack variables.
C	-	smoothing parameter.
α	-	dual variables or Lagrange multipliers.
b	-	bias term.
h	-	Vapnik-Chervonenkis dimension.
λ	-	sparseness parameter.
c	-	vector of coefficients associated with terms in an ANOVA basis.
S2tol	-	Stage II sparse selection threshold.

The Al-Mg-Li Powder Metallurgy System

PV_c	-	forging temperature.
$PV_{b/a}$	-	proprietary measure of the extent of mechanical alloying.
pca	-	process control agent.

The Al-Zn-Mg-Cu Experimental System

x_{Aa}	-	atomic concentration (at.%) of element Aa.
$x_{Aa,w}$	-	weight percentage (wt.%) of element Aa.
$x_{Cu,\alpha}$	-	atomic fraction of Magnesium dissolvable in the Al-rich phase.
$x_{Mg,\alpha}$	-	atomic fraction of Copper dissolvable in the Al-rich phase.
x_S	-	atomic fraction of the S phase.
$x_{\eta'}$	-	maximum atomic fraction of the η' phase.
$x_{Mg,xs}$	-	atomic fraction of excess Magnesium left in solution after complete formation of the main precipitates.
t	-	age-hardening heat treatment time.

The 7x75 Al-Alloy System

Q	-	quarter thickness position in a plate form.
C	-	mid-thickness position in a plate form.
.3W	-	1/3 width position in a plate form.
.5W	-	mid-width position in a plate form.
LT	-	longitudinal test orientation.
TL	-	long-transverse test orientation.
SL	-	short-transverse test orientation.
LT_ϵ	-	strain in the long-transverse direction.
L_ϵ	-	strain in the longitudinal (or rolling) direction.
CRS	-	cross rolling strain.
GR	-	gauge reduction.
T_f	-	final temperature of the plates exiting the hot rolling mill.
ST_t	-	solutionising heat treatment time.

Acronyms

AA	-	Al-Alloy.
AMN	-	Associative Memory Network.
ANOVA	-	ANalysis Of VAriance.
ARD	-	Automatic Relevance Determination.
ASMOD	-	Adaptive Spline Modelling of Observation Data.
BMLP	-	Bayesian Multi-Layer Perceptron.
CG	-	Conjugate Gradient.
DM	-	Data Mining.
ERM	-	Empirical Risk Minimisation.
FS/BE	-	Forward Selection/Backward Elimination.
GP	-	Guinier-Preston (zones).
IACS	-	International Annealed Copper Standard.
IADS	-	International Alloy Designation System.
I/M	-	Ingot/Metallurgy.
KDD	-	Knowledge Discovery in Databases.
LOOCV	-	Leave-One-Out Cross-Validation.
MA	-	Mechanically Alloyed.
MLP	-	Multi-Layer Perceptron.
MLR	-	Multiple Linear Regression.
MSE	-	Mean Squared Error.
PFZ	-	Precipitate Free Zone.
P/M	-	Powder/Metallurgy.
RBF	-	Radial Basis Function.
SCC	-	Stress Corrosion Cracking.
SFE	-	Stacking Fault Energy.
SLT	-	Statistical Learning Theory.
SRM	-	Structural Risk Minimisation.
SSSS	-	Super-Saturated Solid Solution.
SUPANOVA	-	SUpport vector Parsimonious ANOVA.
SVD	-	Singular Value Decomposition.
SW	-	Stepwise.
VC	-	Vapnik-Chervonenkis (dimension).
VIF	-	Variance Inflation Factor.
SV	-	Support Vectors.
SVM	-	Support Vector Machines.
TMP	-	Thermomechanical Processing.

Chapter 1

Introduction

1.1 Background

The aerospace industry demands for increased performance in mechanical and physical properties of structural components have greatly benefitted in recent decades by the development of new alloy series and processing routes. Whilst traditionally, significant developments have been attained by some form or other of "enlightened empiricism", more recently the ability to exert a greater control over industrial processing conditions and advances in thermodynamic and thermomechanical modelling have enabled both development and optimisation of new alloys, through the formulation of tighter compositional ranges, thermal treatments and mechanical processing, with microstructural characterisation remaining at the heart of recent advances. In addition, the greater ability to attain controlled microstructures allows minimisation of the variability in properties, which may result in substantial economic savings.

However, although extensive knowledge of the physical metallurgy of Al-alloys has been built-up over the years, quantitative understanding of commercial processing-property relationships remains limited. This may be attributable to the underlying complexity and interdependence of many of the microstructural evolution processes and performance micro-mechanisms in commercial materials, and the complexities of the industrial production environment.

The uncertainty as to which processing parameters/variables of industrial alloy production routes control the properties developed has resulted in processing specifications rather than reliable process models. Developing a theoretical or physically based model for a particular processing route from first principles is very demanding and unlikely to adequately characterise industrial processing conditions. Generally, these are often limited to describing one particular step in the process (e.g. casting, ageing, etc.) and frequently do not predict the final properties but intermediary variables, describing the microstructural development attained at a particular stage in the alloy fabrication. Models which do predict alloy properties seldom make use of production data to describe the industrial processing conditions. Thus, the development of accurate models for multi-stage commercial processes remains arduous.

Developing an empirical model directly from *bulk* or *raw* process data may be an effective alternative, enabling the development of more useful models in a minimal amount of time. In large scale processing conditions, although the variability between production runs for the commercial production route of a particular alloy may generally represent a limited range of microstructural conditions, there will be significant advantages of being able to identify the processing conditions responsible for the resultant property scatter.

1.2 Neural networks

Neural networks comprise a large number of statistical models which have been applied to modelling and classifying non-linear, interdependent and noisy data sets that are not fully understood and for which analytical solutions are lacking. Of the different classes of neural networks, multi-layered perceptrons (MLPs) (Haykin 1998) are the most widely used. Probably the main reason for their widespread use is that the underlying relationships do not have to be explicitly stated, rather these are learnt from the data, and hence may significantly outperform conventional techniques when prior system knowledge is poor.

In recent years, interest in applying neural and other adaptive modelling techniques has arisen in the materials science and engineering field, being proposed as alternative modelling approaches to standard regression analyses, semi-empirical and physical-based modelling approaches. Recently, in a preface to a special issue on the application of neural network analyses in materials science, Fujii and Bhadeshia (1999) stated that neural networks (particularly with reference to MLPs) are not black box models:

> *The function and its associated coefficients are precisely defined and hence transparent to observation; there are no black boxes!*

The general definition of *black box* is that the characteristics of the construct/box are known, but it's "black" because the internals are unspecified or not understood by the person looking at the box. Thus, in the construction of a black box model no physical knowledge is used, the model structure chosen is a parameterised function which is used to fit the observations. The definition of whether a model is a black box is then in the majority of the cases subjective, depending on whether the person is the designer or user, and so different perspectives can be associated with training algorithms, architectures and methods used for incorporating/extracting the knowledge stored in the model.

Whether or not neural networks constitute black box modelling approaches, one significant drawback of general architectures remains the lack of transparency in the modelling process and this has hindered more widespread use of the techniques. For the functional dependencies inferred by MLPs to be visualised and subsequently validated in light of physical-based system understanding, requires the generation of a number of artificial data sets, whereby the model is queried through varying one input variable at a time, whilst all other inputs are held at

constant values. At best this allows a very local form of transparency and a full understanding and validation of how the model generalises throughout input space is unfeasible.

As such, neural networks remain difficult to validate against physical understanding, despite the significant effort recently directed towards both including/extracting physical knowledge from these systems (Tresp et al. 1993; Shavlik and Towell 1994) and visualising the functions computed (Plate et al. 1998). Subsequent chapters will review several techniques which have been proposed to infer the *relevance* of the input variables, the most noteworthy of which being the automatic relevance determination (ARD) model.

1.3 Neural modelling in materials science

Materials science is a field where process/structure-property relationships are established or examined critically, and as such the use of black box or other empirical modelling approaches may be questioned as they offer no immediate physical insight. As a result, they are often regarded with suspicion by physical metallurgists. Industrialists on the other hand who routinely use regression and statistical techniques may be less reluctant to make use of them.

In materials science and engineering, neural networks have been applied to a range of systems and data sets, either collated from the scientific literature, generated experimentally or obtained from commercial alloy producers. Sumpter and Noid (1996) and Bhadeshia (1999) have reviewed recent applications of neural networks in materials science and engineering.

In the large majority of these studies, the predictive ability of neural techniques has generally been seen to outperform both physical-based models and simple linear regression analyses. For example, Thomson et al. (1999) after having subselected a managable number of input variables from a large database, compared the performance of a Bayesian MLP (which made use of a Markov Chain Monte Carlo method) with that attained by a simple linear regression analysis in the prediction of weld metal chemistry. The neural network was seen to exhibit a better empirical performance when the input-output dependencies were approximated by non-linear terms; in the presence of simple physical relationships, the performance of the neural network did not provide advantages over a linear regression analysis.

In an increasing number of studies, e.g. (Badmos et al. 1998; Badmos and Bhadeshia 1998), that address model validation other than in terms of predictive ability alone, through interpreting and validating the relationships in terms of metallurgical understanding, the physical trends inferred have been reported as being in general agreement with metallurgical understanding, with only a few studies reporting novel and/or conflicting relationships.

A number of simple studies (Song et al. 1995; Seibi and Al-Alawi 1997; Schmidt 1998; Zhang et al. 1999) have perhaps applied neural techniques in naive and misleading ways, and often neural networks are used as black boxes, generally by employing a large number of input variables (51 (Fujii et al. 1996) and 108 (Singh et al. 1998)), with little prior data analysis and pre-processing of the data. Furthermore, in a number of studies, it remains difficult to infer the

advantages in the application of neural modelling techniques, when a comparison with other modelling approaches, e.g. multiple linear regression analysis is lacking.

Whilst several classes of neural networks are being used, the most significant studies and results have been achieved through the application of modelling approaches which make use of Bayesian inference techniques, whereby the network learning process is given a probabilistic interpretation. These comprise the Bayesian MLPs (BMLPs), and Gaussian process (GP) modelling approaches. In BMLPs, the evidence framework and ARD model enable the selection of an appropriate model and the significance of each input variable to be assessed and subsequently compared with metallurgical understanding.

Although it may be argued that Bayesian learning and sophisticated regularisation approaches enable models to deal with a large number of inputs, the justification for the inclusion of a very large number of input variables, particularly where there is little prior analysis of how well the data is distributed in input space may be questionable. It may not only unduly increase the complexity of the models, increasing the risks of inferring non causal dependencies and overfitting the training data set, but makes the validation of the relationships inferred more problematic. For instance, in repeating the analysis originally performed by Gavard et al. (1996) which used BMLPs, the results obtained by Bailer-Jones et al. (1999) using a GP model on a subset of the original number of input variables was seen to exhibit improved performance, with more plausible models obtained.

Neural network applications in the modelling of complex materials characteristics have included studies in which phase transformations (Gavard et al. 1996; Vermeulen et al. 1998; Bailer-Jones et al. 1999), tensile properties (Singh et al. 1998; Jones 1997), fatigue crack propagation (Fujii et al. 1996; Schooling 1997; Schooling et al. 1999) of steels and Ni-base superalloys, were modelled as a function of compositional levels and processing conditions. Until recently comparably fewer studies have been performed on Al-alloy systems. In these light alloys, neural approaches have been applied in the prediction of the location and volume fraction of porosity in Al castings (Huang et al. 1998), the effects of deformation and solid solution time on the ageing dynamics in AA7175 (Song et al. 1995), and in modelling microstructural parameters (Sabin et al. 1997; Bailer-Jones et al. 1997; Clinch et al. 2000). An increasing number of studies are seen to be investigating the applicability of neural approaches, including recurrent neural networks, in modelling and optimisation of both cold rolling mill processes (Cho et al. 1997; Larkiola et al. 1996; Singh et al. 1998) and thermomechanical processing of materials (Sabin et al. 1998; Chun et al. 1999; Shercliff and Lovatt 1999). Comparisons between neural and fuzzy approaches in the areas of material processing and manufacturing have been presented in some recent publications (Cios et al. 1995; Kandola et al. 1999; Arafeh et al. 1999).

1.4 General approach

The empirical modelling design cycle defined in Figure 1.1 shows the general methodology used in this work, whereby the process of learning from data is seen as a continuous process in which the data set selection and analysis, use of prior knowledge sources and the construction and validation of a model are part of an iterative and interactive modelling cycle.

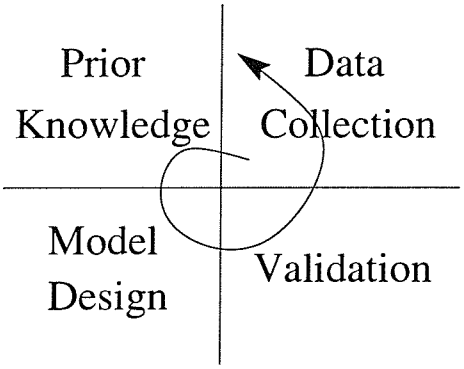


Figure 1.1: Empirical model design cycle.

Model validation by assessing predictive ability and validating the statistical inferences attained from physical/process understanding is critical in developing reliable models of a system. This requires a model that attempts to learn the relationships and trends contained in the data, in a manner that is easier to interpret and validate from use of domain expertise. This may be more desirable than an inherently less comprehensible model, even at the expense of slightly inferior predictive ability. Neurofuzzy networks and the recently developed SUPANOVA technique, comprise two adaptive modelling approaches which appear to be well suited to a *parsimonious* modelling approach. Both techniques will be reviewed in subsequent chapters, discussing strengths as well as limitations. The main aims of this thesis are then to assess of the perfomance of these two techniques in the modelling processing-property relationships of several Al-alloy data sets.

Modelling *per se* is only part of the process of discovering informative patterns or relationships which may be contained in the data. The validation of the model, whether by use of an independent test set and/or using metallurgical understanding of how processing variables are expected to influence the behaviour of the material under investigation is seen to be a fundamental step if meaningful inferences are to be obtained. Understanding the structure of the models inferred in terms of the statistical properties of the data set, may explain poor empirical performances attained by the neurofuzzy and SUPANOVA modelling approaches. For these purposes, a number of statistical methods have been used.

1.5 Thesis overview

As a result of both the breadth of the systems and properties investigated and the statistical/modelling techniques applied, this thesis represents a broad interdisciplinary scientific contribution, in which some fundamental metallurgy and data modelling material has been reviewed. A reader familiar with the principles of materials science and engineering, general linear regression analysis, data pre-processing issues and/or adaptive modelling techniques may wish to start directly at Chapter 7.

Chapter 2: Physical metallurgy and processing of aerospace Al-alloys

General physical metallurgy issues and the processing route used in the production of wrought high-strength Al-alloys are reviewed in this chapter. The focus is primarily on how metallurgical features and the physical and mechanical properties of heat-treatable Al-alloys are affected by the individual processing stages.

Chapter 3: Data modelling

This chapter serves as a general introduction to the basic assumptions and concepts of learning from data. A brief discussion of data quality issues and general modelling objectives is then intended to place the work of subsequent chapters into perspective. Approaches to parameter estimation and model selection, general approaches which are used for controlling the complexity of parametric models, and methods which are widely used to assess the generalisation performance of a model are outlined.

Chapter 4: Data pre-processing, sensitivity analysis and conditioning diagnostics

As data is seldom readily amenable to a meaningful statistical analysis, in particular for reliable knowledge extraction, this chapter introduces some elementary pre-processing steps that benefit the learning process and statistical inferencing. Two sources of data weaknesses, which can severely affect the estimation of parameters in linear models are then discussed. Established conditioning diagnostics are presented, which upon detecting the nature of any data weakness can help in the interpretation of poor inferences attained in linear regression analyses. In small data sets, the influence exerted by individual observations can be unduly high and be seen to exert a significant effect on the parametric estimates. For the purposes of detecting non-trivial influential data pairs, in high-dimensional settings, two measures of influence are presented.

Chapter 5: Neurofuzzy networks

This chapter summarises the developments achieved in recent years by several researchers of a particular class of neurofuzzy networks (or systems). A brief introduction to fuzzy systems

leads to the description of the modelling framework developed and the fundamental theory upon which these networks are based. Training of neurofuzzy networks is then considered in terms of weight and model structure identification. The model construction algorithm used to determine additive models is outlined together with the hypothesis testing framework and model complexity measure used to determine a model structure supported by the available data. The extension of Bayesian inference techniques within MacKay's (MacKay 1992a) evidence framework is then presented, focusing on the local regularisation approaches defined for additive neurofuzzy models by Bossley (1997). The derivation of error bars is discussed and the interpretation of neurofuzzy models as linear smoothers allows the influence measures defined in Chapter 4 to be extended to these networks. The chapter concludes with an illustrative example, whereby the neurofuzzy networks are used to model a non-linear analytical expression (equation for the plastic zone size at the crack tip).

Chapter 6: Statistical learning theory and support vector methods for regression

Recent developments in statistical learning theory and the development of support vector methods for regression have given rise to a promising subfield of machine learning. Support Vector Machines (SVM) offer an alternative approach to problems in which general neural network approaches have until recently prevailed. This chapter introduces the theory developed by Vapnik (1998) for statistical estimation from small samples. Vapnik's structural risk minimisation principle is seen to provide a general method which can be used for the purposes of model selection and complexity (capacity) control. The chapter then provides a brief introduction of SVM methods for regression and describes the SUPANOVA (SUpport vector Parsimonious ANOVA) modelling technique developed by Gunn and Brown (1999).

Chapter 7: Data modelling of processing-property relationships in an Al-Mg-Li powder metallurgy alloy system

The first of the data sets investigated comprised a small set of production runs of an Al-Mg-Li power metallurgy alloy system. After a short introduction to the metallurgy and strengthening mechanisms of mechanically alloyed Al-Mg(-Li) materials and a data analysis of the data set, multiple linear regression (MLR), neurofuzzy and SVM approaches were used in the modelling of three tensile properties: 0.2% proof stress ($\sigma_{0.2}$), ultimate tensile strength (uts), and percentage elongation ($\%el.$), using compositional information and processing variables. Results were compared both in terms of the approximation abilities and the structural relationships determined by the adaptive methods. Both conditioning and regression diagnostics were used as a further means of interpreting the results.

Chapter 8: Data modelling of structure-properties of experimental trials in the Al-Zn-Mg-Cu alloy system

In the second of the data sets investigated, comprising a small experimental series of compositional-ageing time alloy variants of the Al-Zn-Mg(-Cu) alloy system, two properties were investigated: $\sigma_{0.2}$ and electrical conductivity (σ_{el}). Prior to the application of the various data modelling techniques, the precipitation sequences governing these alloys are briefly covered, discussing how the microstructure determined by compositional and thermal treatments is expected to influence the properties developed. In addition to determining models from the original data set, which included compositional information and heat treatment times, two other data sets based on general rule-of-thumb understanding and a physically-based (microstructural) set of data transformations were defined. Models for these three data sets were then determined using the same methodologies as those used in Chapter 7 and results validated. The effect of the data transformations were also assessed by means of initialising neurofuzzy models with structures determined from metallurgical understanding and knowledge gained from the initial modelling results.

Chapter 9: Knowledge discovery and data mining of 7xxx series Al-alloy production databases

High-strength 7xxx series (Al-Zn-Mg-Cu) Al-alloys are seen to account for an ever increasing volume of structural material sold to the aerospace industry. The large scale commercial production of these alloys means that production databases store a large volume of processing information pertaining to the different stages of the manufacturing of wrought products. In this, the final statistical analysis performed, the feasibility of obtaining "useful" models of fracture toughness (K_{Ic}) levels in terms of different test conditions and variations in the processing variables is investigated. As the original data sets included a very large number of fields and the data required a considerable amount of pre-processing, the investigation was placed in a knowledge discovery in databases (KDD) context, and the modelling *per se* in a data mining (DM) perspective.

Chapter 10: Conclusions and future work

Finally, the results and main themes which have emerged from the different analyses performed are discussed and some general conclusions are drawn and suggestions for future work are outlined.

1.6 Contributions

Given the large interest that adaptive modelling techniques, MLPs in particular, have experienced in both steel and Ni-base superalloy property modelling and alloy development, this thesis has been directed towards extending the application of adaptive methods to the modelling of physical and tensile properties of Al-alloys. Thus, the main contributions of this work are seen as the following:

- The application and assessment of *transparent* modelling paradigms, namely neurofuzzy networks and SUPANOVA in the modelling of process-property relationships in several Al-alloy systems.
- Statistical techniques, typically used in applied statistics, namely conditioning and regression diagnostics have been used as complementary data analysis tools, allowing a more rigorous understanding of the data than that attained from the application of neural-type modelling approaches alone.
- Influence measures previously defined for linear models have been extended for the class of additive neurofuzzy networks considered in this work.
- Quantities based on stoichiometry and inspired by physical-based transformations have been derived and an interaction between the Magnesium content and the age-hardened condition of the alloy has been proposed as being a significant interdependency in influencing the microstructural development of Al-Zn-Mg-Cu alloys.
- KDD and DM are techniques which will experience increasing importance in the assessment of the production data. This work has introduced the KDD and DM techniques as useful approaches in the statistical investigation of large commercial processes.

This work has contributed to the following list of journal publications:

- Femminella O.P., Starink M.J., Brown M., Sinclair I., Harris C.J. and Reed P.A.S. Data pre-processing/model initialisation in neurofuzzy modelling of structure-property relationships in Al-Zn-Mg-Cu alloys. *ISIJ International*, vol.39, no.10, pages 1027-1037, 1999.
- Femminella O.P., Starink M.J., Gunn S.R. and Reed P.A.S. Neurofuzzy and SUPANOVA Modelling of Structure-Property Relationships in Al-Zn-Mg-Cu alloys. In *Proceedings of the 7^{th} International Conference on Aluminum Alloys (ICAA-7)*, Charlottesville, VA, April. *Materials Science Forum*, volumes 331-337, pages 1255-1260, 2000.
- Christensen S.W., Kandola J.S., Femminella O.P., Gunn S.R., and Reed P.A.S., and Sinclair I. Adaptive numerical modelling of commercial aluminium plate performance,

In *Proceedings of the 7th International Conference on Aluminum Alloys (ICAA-7)*, Charlottesville, VA, April. *Materials Science Forum*, volumes 331-337, pages 533-538, 2000.

Chapter 2

Physical Metallurgy and Processing of Aerospace Al-Alloys

2.1 Introduction

Structural materials used in the aerospace industry require a certain balance of physical and mechanical properties to enable their safe and efficient use. These properties can be arbitrarily separated into those that affect the structural performance *ab initio* and those limiting continuing service performance. Among the former, density, strength, stiffness are key properties, amongst the latter, resistance to fatigue and corrosion are important examples (Peel and Gregson 1995). Typical structural components include fuselage skin and stringers, wing spars, upper wing structures and lower wing skin, each of which require a particular balance of service performance properties, depending on the structural section of the airframe.

The ability to predict and control microstructure is central to the development of new materials and achieving desired levels of performance. The performance requirements for airframe components are reviewed by Starke and Staley (1996), in terms of the structure-property relationships, describing the background and drivers for the development of modern Al-alloys to improve performance. Developments in aerospace structural materials rely upon enhancement of a balance of properties, rarely upon the improvement and optimisation of a single characteristic, thus necessitating an understanding of the interdependencies and factors governing the balance of such properties. At present, major developments in aircraft structural components are concentrating on improvements and optimisation in properties at reduced manufacturing costs (Rendigs 1997; Warner et al. 1997).

Despite the significant advances that have been made in the physical understanding of microstructure and mechanical properties of aluminium based materials, commercial alloy and processing route development has until recently been driven by empiricism. However, it has become recognised that future alloy design can no longer be performed using such purely empirical and/or experimental approaches.

Combining knowledge acquired from scientific studies and industrial experiments has replaced the "trial and error" approach to alloy development. This has led to reduced development timescales and advances in the understanding of the relationships between composition, processing, microstructural characteristics and material properties, resulting in improvements in the properties attained for aircraft applications. Although a design of experiments approach to optimisation can be very efficient and cost effective (Sigli et al. 1996), it depends critically on the selection and range of the process conditions varied, and may be criticised[1] in that it gives little insight into the metallurgical phenomena occurring.

It is recognised that both a quantitative knowledge and greater understanding of thermo-mechanical process parameters and related metallurgical effects is required in order to be able to extrapolate and generalise to different processing routes and alloy variations. The integration of the different sub-models is recognised as a powerful tool for industrial product optimisation. For example, Vatne et al. (2000) have combined models for work-hardening, deformation texture evolution and recrystallisation. The underlying complexity and interdependence of many microstructural evolution processes and mechanical property micro-mechanisms, hinders the development of reliable *process models*[2] of multi-stage production routes and optimal methods of integrating existing models for commercial processing stages. Thus, there remains a lack of understanding of how to combine metallurgical models, which describe overall alloy behaviour, and the typically insufficiently quantified descriptions of microstructural features.

Thermodynamic calculation of phase diagrams has been a possibility since the 70's due to enhanced thermodynamic modelling and computational capabilities. However approaches based on analytical equations derived from equilibrium phase diagram analysis are of limited quantitative use when nucleation and growth kinetics are of primary importance during solidification, e.g. see Dubost (1993) for a review. In more recent years considerable advances have been made in the metallurgy of multi-component alloy systems and in the implementation and use of equilibrium phase diagrams using simple analytic criteria and advances have been made in nucleation and growth kinetics modelling.

The predictive capabilities of the modelling framework developed by Pechiney, for example, rely on existing understanding of physical phenomena, availability of physical models and on empirical parameters determined from both experimental and industrial data. This framework has led to the development of Al-alloys 7050A and 7040, used for ultra thick plate applications and the new upper wing skin alloys 7349 and 7449 (Warner et al. 1997; Shahani et al. 1998).

Although not offering immediate physical insight, statistical and adaptive methods, of-

[1] In addition to being expensive, time consuming and limited to the equipment and the specific alloys tested, Hirsch et al. (2000) criticises trial and error approaches. Results tend to be ambiguous, the limited set of trials impairing reliable inferencing, with the possibility that significant variables have not been measured.

[2] Shercliff and Ashby (1990) define a process model as a mathematical relation between process variables (e.g. alloy compostion, heat-treatment temperature and time, etc.) and the alloy properties (e.g. strength), based on physical principles (thermodynamics, kinetic theory, dislocation mechanics and so on).

ten regarded with deep suspicion or even hostility by physical metallurgists, are attracting significant interest due to their ability to infer trends from the data of multivariate systems. Industrialists on the other hand who are used to empirical methods may be less reluctant to make use of them. In industrial settings, adaptive methods may attain useful generalisation abilities, providing a means to optimise without time-consuming experiments by inferring trends in multivariate, noisy, non-systematic, and complex data.

In this chapter the physical metallurgy of conventional Al-alloys used in aerospace applications is reviewed. A description of the fabrication of wrought Al-alloys will address the multi-stage production methods in terms of processing conditions and metallurgical features. An understanding of the interactions between process variables and practical production limitations is fundamental for the validation of the dependencies and relationships that may be inferred from process data. Assessing the empirical performance of data-driven modelling techniques in terms of the statistical characteristics of the processing variables and integrity of the data available also requires this fundamental understanding.

2.2 Physical metallurgy of aluminium alloys

The following sections provide an overview of the physical basis of the properties of heat-treatable Al-alloys, with particular attention to the physical and mechanical properties and the processing route used to attain the properties required for aerospace applications. More extensive reviews of the physical metallurgy and processing of Al-alloys can be found in (Hatch 1984; Polmear 1981; Mondolfo 1976; Gregson 1995).

2.2.1 Principles of age-hardening

Heat-treatable Al-alloys contain elements that decrease in solubility with decreasing temperature, at concentrations that exceed their equilibrium solid solubility at room and ageing temperatures. This decrease from appreciable solid solubility at elevated temperatures to relatively low solid solubility at low temperatures is a fundamental characteristic that provides the basis for substantially increasing the hardness/strength of Al-alloys by solution heat treatment and subsequent precipitation ageing. A typical multi-stage age-hardening process involves the following thermal treatments:

- a solution treatment at a relatively high temperature within the solid state to dissolve the alloying elements,
- quenching, usually to room temperature, to obtain a supersaturated solid solution (SSSS) of both elements and vacancies,
- ageing treatment(s) used to achieve a controlled decomposition of the SSSS so that a finely dispersed precipitate is obtained.

The complete decomposition of the SSSS is usually a complex process which may involve several stages (Polmear 1981; Starke 1989).

Typically, coherent Guinier-Preston (GP) zones and a semi-coherent intermediate precipitate may precede the formation of the final incoherent equilibrium precipitate during ageing. GP zones are fine, clusters of the precipitating elements having the same crystal structure as the matrix. In general, heat treatments are designed to produce a very high density of fine GP zones or precipitates that strongly interact with dislocations, thereby increasing the yield strength of the alloy. Homogeneous precipitation involving GP zones depends upon the ageing temperature and the concentration of vacancies which will in turn be affected by the solution treatment temperature and cooling rate during the subsequent quench. The presence of such vacancies has a marked effect on precipitation kinetics and strengthening potential: several specific interactions between vacancies and various solute atoms influence the ageing kinetics, making the effects of even quite minor alloying additions important. In 7xxx series alloys, Mg content is believed to play a significant role in this process: because of its large atomic diameter, Mg-vacancy interactions may enhance the retention of excess vacancies during the quench (Hatch 1984).

The development of precipitate free zones (PFZ) at grain boundaries after a slow quench and subsequent ageing may be attributed to the depletion of solute atoms near grain boundary particles (i.e. segregation to the grain boundaries) and/or the migration of vacancies to boundaries during the quench. Processing procedures are frequently employed to minimise grain boundary precipitation and PFZ formation, as these two phenomena generally have a deleterious effect on the properties of Al-alloys as they can become regions of strain localisation, leading to premature fracture.

The prediction of precipitate phases in Al-alloys has been the subject of considerable study. Homogeneous nucleation theory may be used to explain the kinetic origin of precipitation from solid solution and classical solution thermodynamics used to derive equations that examine multiphase equilibrium and stability (Mondolfo 1976). As will be discussed in section 2.3.1, many studies have focused on relating precipitates to different strengthening mechanisms, which in addition to terms involving precipitate size spacing and volume fraction, include terms such as the elastic moduli of metastable phases (Embury et al. 1989).

2.2.2 Ageing condition

Varying the state of ageing is a simple means by which strength and balance of properties may be modified. Alloys which are allowed to develop and stabilise their strength at room temperature are said to be naturally aged, whilst alloys aged at temperatures above room temperature are said to be artificially aged. During artificial ageing the strength of the alloy initially increases as the number and size of the precipitates increases. Eventually, strength levels reach a maximum value, corresponding to the peak-aged condition and any further overageing leads

to the coarsening of the precipitates and a reduction in strength. Due to the nature of the hardening precipitates, many variables are significant in high-temperature ageing of Al-alloys. Since many heat treatable alloys exhibit age hardening at room temperatures after quenching, a slow heating rate up to the ageing temperature may effectively act as a two step treatment. The sensitivity exhibited by certain alloys to natural ageing intervals, heating rate and two-step ageing sequences have been attributed to the critical GP zone size formation. Modifying soak times can compensate for precipitation during heating and for the effects of soaking at temperatures above or below the nominal ageing temperature (Gregson 1995).

The rate and extent of strengthening during precipitation heat treatments can be considerably increased in some alloys by cold working after quenching, whereas other alloys show little or no added strengthening upon working. These improvements are commonly attributed to a combination of cold work and precipitation changes and due to nucleation of additional precipitate particles at dislocations.

2.2.3 Grain structure

During ingot casting, small additions of Ti and B are used as grain refiners. Boron is typically added to enhance the refining effect of Ti, the main effect of B being to reduce the solubility of Ti in liquid Al, enabling grain refinement at lower Ti levels. However, in excessive amounts, B can promote the formation of inclusions. The addition of Ti and B have little effect on changes in grain size that occur during or as a result of thermomechanical processing. The grain size of final wrought products is strongly influenced by the amount of work received prior to the solution treatment, generally decreasing as the amount of work increases. Annealing practice and rate of heating to the solution heat treating temperature can also affect grain size in recrystallised materials, a small grain size being promoted by a fast heating rate and a short time at the annealing temperature. Grain size control has many purposes, which include ensuring good stress corrosion cracking (SCC) resistance and high fracture toughness. Anisotropic mechanical properties may arise from the crystallographic textures developed within the material from working and annealing practices.

Recrystallisation processes in wrought materials are commonly stimulated by high dislocation densities in the vicinity of coarse constituent particles, but retarded by small particles which impede the nucleation and growth of recrystallising grains. Alloys which have received a certain amount of cold work may recrystallise completely during solution heat treatment, as the stored energy of deformation may be sufficient to overcome the pinning effect of small particles on grain boundaries. The decrease in dislocation density caused by recovery-type annealing produces a decrease in strength and other property changes (e.g. increase in electrical conductivity). In general, a greater amount of cold work reduces the time and temperature for recrystallisation. Complete recovery from the effects of cold working is obtained only with recrystallisation. Due to the heterogeneous distribution of dispersoid particles, the morphology

of the deformed grains may be retained in the final recrystallised product.

2.2.4 Constituent particles

Constituent particles (or coarse intermetallics) varying in size from 0.5 to 10μm, are commonly formed by non-equilibrium micro-segregation and liquid-solid eutectic reaction occurring during ingot solidification.

Insoluble constituent particles typically contain impurity elements, such as Fe and Si, while soluble constituents consist of equilibrium and/or metastable intermetallic compounds of the major alloying elements. Homogenisation and solutionising treatments are used to dissolve these soluble constituents (see section 2.4). A high density of dislocations develops in the vicinity of these particles during processing, providing nucleation sites for recrystallisation. In that they generally provide nucleation sites and propagation paths for low energy failure, constituent particles have a strong effect on the fracture behaviour of Al-alloys.

The size and size distribution of constituent particles are controlled by the rate of ingot solidification, the chemical composition, and the extent and nature of bulk deformation. Particle size decreases as solidification rate increases, impurity levels decrease, and as the amount of deformation increases. Generally, the insoluble constituent particles are coarsest and most heterogeneously distributed in thick plate and are finer and more homogeneously distributed in thin sheet. Hamerton et al. (2000) for example describe work conducted to improve understanding of the development of the coarse particles population through different processing stages.

2.2.5 Dispersoid particles

Small submicron (0.05$-$0.5μm) particles form during ingot homogenisation (or other high temperature treatments), by the solid state precipitation of compounds containing elements having modest solubility and which diffuse slowly in solid aluminium. Once formed these particles (e.g. Al_6Mn, $Al_{20}CuMn$, $Al_{12}Mg_2Cr$, Al_3Zr) resist both dissolution or coarsening during subsequent lower temperature thermal treatments. The primary role of these particles is in retarding recrystallisation and grain size control during processing and subsequent heat treatments, although they may also exert an important influence on certain mechanical properties through their effects on both the response of the alloy to ageing treatment and on the dislocation substructures imparted by plastic deformation. By promoting a homogeneous slip character, these particles may prevent dislocation pile-ups at grain boundaries. The effectiveness of a particular dispersoid in controlling the grain structure depends on its size, spacing and coherency, the former largely controlled by heat treatment time and temperature.

It is well established that the dispersoids are not homogeneously distributed in commercial wrought material, but can be observed in long thin bands and layered in the deformed grains: this results from segregation during solidification and the nature of most deformation processes.

2.3 Properties of heat-treatable Al-alloys

Properties of high-strength alloys are seen to be controlled by a complex interaction of the microstructural features. Chemical composition and processing are seen to control the microstructure and thus the physical and mechanical properties of heat-treatable Al-alloys. The following sections discuss the microstructural characteristics, processing and, where present, the relationships that are used to determined the physical and mechanical properties of high-strength Al-alloys that will be the subject of analysis in subsequent chapters.

2.3.1 Strengthening mechanisms

In commercial high-strength alloys, many strengthening mechanisms are operating simultaneously, the magnitude of the individual contributions depending on alloy composition, processing and fabrication (e.g. wrought, mechanically alloyed, etc.).

Recent work on modelling the strength of precipitation hardened Al-alloys (Shercliff and Ashby 1990; Gomiero et al. 1992; Hornbogen and Starke 1993; Starink et al. 1999; Starink et al. 2000) has provided quantitative relationships and strengthening contributions, establishing a set of comprehensive equations for the different contribution of the main hardening mechanisms, typically including the type, size, volume fraction, shape, coherency and distribution of precipitates. Different approaches have combined the contribution of separate strengthening mechanisms, such as simple additive or superposition laws. In multi-component precipitate systems and in systems where competitive mechanisms are present, more complex models, e.g. Pythagorean addition rules (Embury et al. 1989), laws of mixtures (Deshamps and Brechet 1999) have been found to be more appropriate in approximating empirical data. For example Hornbogen and Starke (1993) provide the following equation for the different contributions to high yield strength Al-alloys accounting for the main hardening mechanisms:

$$\sigma_y = \underbrace{\sigma_\perp}_{\text{pure Al}} + \underbrace{\Delta\sigma_s}_{\substack{\text{solid solution,}\\ \text{coherent particles}}} + \Big(\underbrace{\Delta\sigma_D^2}_{\substack{\text{dislocation,}\\ \text{sub}-\text{boundaries}}} + \underbrace{\Delta\sigma_p^2}_{\text{hard particles}} \Big)^{\frac{1}{2}} + \underbrace{\frac{k_y(\Delta\sigma_B)}{S_B}}_{\text{grain boundaries}}$$

However, the direct applicability and translation of the relationships described by theoretical and semi-empirical models developed for a particular strengthening mechanism, into practical plant process guidelines remains arduous.

In the following sections, the main strengthening contributions present in Al-alloys are briefly described. In Chapters 7 and 8 a more detailed treatment and discussion of the main strengthening effects in wrought Al-Mg-Zn-Cu wrought and mechanically alloyed Al-Mg-Li materials will be considered.

2.3.1.1 Grain-size effects

The proof stress (flow stress at a given strain), σ_p, as given by the Hall-Petch equation is commonly seen to depend inversely on the average grain diameter, d:

$$\sigma_p = \sigma_0 + \frac{k}{\sqrt{d}} \tag{2.1}$$

where σ_0 is a constant depending on dislocation interactions with the lattice, solute atoms and statistically stored dislocations (i.e. lattice resistance to deformation), and k is a constant associated with the stress concentration required to generate dislocations from the grain boundary or activate/unpin dislocation sources on the opposite side of the boundary. Thus, the greater the number of grain boundaries, the more difficult plastic deformation becomes and increased strength levels can be obtained by decreasing the grain size. Subgrain boundaries act in a similar way, but are less effective slip boundaries.

Generally, in age-hardenable alloys, grain size effects are not a major strengthening contribution, although they can become important at very fine grain sizes. Grain size contributions are seen to provide a significant strengthening mechanism in mechanically alloyed materials.

2.3.1.2 Solid solution strengthening

Strengthening from solute atoms arises either from differences in atomic size between the solute and solvent atoms and/or from differences in elastic modulus. In the case of aluminium, the size effect is considered the dominant source of solute strengthening. Solid solution strengthening contributions may be significant in underaged tempers. Theoretical approaches (Embury et al. 1989; Gomiero et al. 1992) express the increase in shear yield stress in terms of a solute atom mismatch parameter and other material variables (e.g. Burger's vector, shear and combined modulus, atomic concentration of the solute, etc.). Alloying additions that significantly increase solid solution strength of Al-alloys are Cu, Mg and Si. An important role of solute atoms is with respect to the modification of the work-hardening behaviour at large strains. Solid solution additions may also lower the stacking fault energy (SFE) of Al, enhancing planar slip.

2.3.1.3 Precipitation hardening

High strength levels developed by heat treatable Al-alloys are primarily due to the interaction of dislocations with precipitates that are formed during the ageing treatment. Strength levels can be optimised by maximising those elements that participate in the ageing sequence and can be put in solid solution at the solutionising temperature. The maximum solubilities of these elements should not be exceeded, since excesses will tend to form coarse particles that do not add significantly to the strength, adversely affecting other properties.

In theoretical approaches, precipitation hardening mechanisms are usually described in terms of the interaction of a dislocation with an array of point obstacles. Precipitation hardening

mechanisms reviewed by Martin (1998) and Ardell (1985) are considered to arise from several physical mechanisms including chemical (surface), stacking-fault, modulus, coherency, order and Orowan, e.g. see (Kovacs et al. 1977; Melander and Persson 1978) for some examples of these contributions. A comprehensive overview of physical-based models developed for each of the above mechanisms is given in (Embury et al. 1989).

In some systems hardening mechanisms may be operating simultaneously, in others one mechanism may dominate the precipitate-dislocation interaction. Orowan strengthening may be an important mechanism in overaged materials; coherency strengthening mechanisms are considered to play an important role in alloys aged to peak strength (Ardell 1985). Characterisation of the dislocation-particle interactions of commercial Al-alloy systems requires the quantification of several parameters. This is made difficult by the complex sequence of precipitation reactions that may occur in Al-alloy systems.

2.3.2 Additional factors

The effect of constituent and dispersoid particles on strength may be particularly reflected in their effects on the degree of recrystallisation and development of crystallographic textures, which impart an anisotropy of properties. Coarse intermetallic compounds have little direct effect on attainable strength levels, but can cause a marked loss in ductility. Processing-induced gradients in through-thickness composition and quench rate variations can significantly affect the through-thickness constituent and precipitate particle distribution, influencing plate properties (Miller et al. 2000). Most high-strength plate products exhibit a characteristic "W-shaped" through-thickness pattern of strength in the long-transverse direction: strength being lowest at the quarter thickness positions. This pattern has been attributed to the competing effects of texture strengthening, compositional variations and quench effects (Staley 1992; Chakrabarti et al. 1996).

2.3.3 Fracture toughness

In general, the toughness of Al-alloys decreases as the strength is increased by heat treatment. A microstructural condition favouring inhomogeneous slip (shearable strengthening precipitates in under and peak aged conditions) may have deleterious effects on toughness as associated strain concentrations may act as void nucleation sites. Once peak strength is reached for a particular alloy, some improvement in toughness may be obtained by overageing. In terms of tensile properties, changes in toughness with ageing have additionally been correlated with varying strain hardening capacity by a number of authors (Garrett and Knott 1978; Chen and Knott 1981).

The analysis of fracture toughness may be complicated by the presence of three types of particles in commercial alloys and the fact that grain and subgrain boundaries may also directly influence failure modes. The spatial distribution of all these (i.e. boundaries and particles) may

have direct influences on behaviour, the implications of which may be difficult to assess given the intimate relationships that exist between the different microstructural features during processing. Starke and Staley (1996) summarise and classify the contribution of microstructural features to toughness by means of a "toughness tree" diagram.

The effect of dispersoids on toughness is complex, as they can have both positive and negative influence (Garrett and Knott, 1978). Dispersoids may either cause a high energy absorbing fracture mode by suppressing recrystallisation and/or grain growth, or cause void sheet formation by nucleating microvoids at the particle-matrix interface. The detrimental effects due to microvoid nucleation may be reduced by decreasing the dispersoid size and improving particle-matrix interface bonding strength, both of which may be achieved by using Zr containing dispersoids instead of Mn or Cr containing dispersoids. The incidence of decreasing toughness with overageing has been identified with a change in fracture mode from predominantly transgranular to intergranular (Kirman 1971). Transgranular fracture was suggested to nucleate at dispersoid particles (decohesion of the particle/matrix interface), whilst intergranular fracture proceeded by the nucleation of voids at grain boundary $MgZn_2$ particle/matrix interfaces (Kirman 1971). As such, different influences of dispersoid distribution on toughness may be expected as overageing proceeds.

The effects of constituent particles (insoluble and soluble) and hardening precipitates are generally well known and understood (Kirman 1971; Hahn and Rosenfield 1975; Staley 1976). The primary insoluble Fe and Si intermetallics (Al_7Cu_2Fe, Mg_2Si and $(Fe,Mn)Al_6$) are readily fractured during primary processing or under service loading, providing crack nucleation sites and paths for low-energy crack propagation. In particular, Fe containing particles appear to be more detrimental to toughness than those containing Si (Van Stone et al. 1974). For higher fracture toughness levels to be attained, it is desirable to simultaneously decrease the size and increase the spacing of the larger void nucleating particles (Van Stone et al. 1974). Decreasing particle sizes increase the stress/strain levels necessary to cause void nucleation as well as decreasing the initial size of a void, thus permitting more void growth before coalescence.

The effect of fine precipitates on toughness is not simple as slip character invariably varies with the strength of the alloy. Fine-scale strengthening precipitates may have a beneficial effect on the fracture toughness by increasing the deformation resistance and decreasing the crack opening displacement, or a detrimental effect by promoting strain localisation and premature crack nucleation, facilitating void coalescence and final fracture (Ludtka and Laughlin, 1982).

Ehrstrom et al. (1996) showed that in order to model anisotropic toughness, distributions of particles should be taken into account, and in so doing significant improvement over Hahn and Rosenfield (1968) type models (see below) have been obtained. Ludtka and Laughlin (1982) suggest the difference in fracture toughness behaviour of 7xxx series alloys is dependent on the coarseness of matrix slip and the strength differential between matrix and PFZ, the latter considered as a measure of the tendency toward intergranular fracture, explaining the dependency between strength and toughness as a function of both ageing time and temperature.

With decreasing quench rate, grain boundary precipitation was seen to become coarser and the PFZ formed during ageing at grain and subgrain boundaries increases in width, giving reduced levels of toughness for a given ageing treatment. Morere et al. (2000) have reviewed parameters and models which have been proposed in the literature to characterise intergranular ductile fracture mechanisms, which may depend on void initiation at boundary precipitates, strain localisation within soft PFZ and stress/strain concentration at slip band/grain boundary intersections, all of which are strongly dependent on the ageing conditions.

The effects of quench rate and recrystallisation on toughness, whereby slack quenches reduce toughness, as do recrystallised structures (Staley 1976; Thompson 1975) are generally accepted and known to be interrelated. Increased recrystallisation levels and/or slow quench rates primarily compromise fracture resistance in wrought materials by promoting grain boundary failure (Thompson 1975; Ludtka and Laughlin 1982; Dorward and Beerntsen 1995). The effect of quench rate on the toughness of 7xxx series alloys can be influenced by the degree of recrystallisation, the loss in toughness with decreasing quench rate being substantially higher for recrystallised products (Dorward and Beerntsen 1995)[3]. Toughness was seen to decrease almost linearly with an increasing degree of recrystallisation, the effect being greatest in rapidly quenched plates, recrystallisation having less effect on slowly quenched material.

In Morere et al. (2000), quench rate and recrystallisation effects have been studied for constant yield strength levels in AA7010-T76 plate as a function of underlying material anisotropy at relatively low recrystallisation levels representative of commercial plate product forms. The high quench sensitivity exhibited in the SL orientation for unrecrystallised microstructures was attributed to the synergistic location of intermetallics and coarse boundary precipitates in slow quenched material.

Sugamata et al. (1993) review the different models describing the effects of precipitate shearing and bypassing, strain hardening, and precipitate sizes and volume fractions on K_{Ic} that have been proposed in the literature to characterise the dependencies between microstructural features and mechanical parameters on fracture toughness.

In terms of quantitative modelling, it is clear from the above that fracture toughness of high strength Al-alloys depends on many parameters, including flow strength, work hardening rate, slip character, dispersoid content, intermetallic content, grain structure and grain boundary structure (Garrett and Knott 1978). Based on the concept that the critical step in the fracture process being coalescence between the cracked particles Hahn and Rosenfield (1975) developed a model which correlated the spacing of cracked constituent particles to plane strain fracture

[3]A multiple regression analysis of the fracture toughness data, using an F-test of significance and hypothesis testing methodology, yielded the following model:

$$K_{Iv} = 14.1 + 7.63\ln Q - 0.041 R_x \cdot \ln Q$$

with a correlation coefficient of 0.95 and an estimated standard error of 1.5 MPa/$\sqrt{\text{m}}$, where R_x is the percent recrystallisation and Q the quench rate.

toughness:

$$K_{Ic} = \sqrt{2\sigma_y E\lambda}$$

λ is the nearest neighbour spacing of the void initiating particles. Crack extension is considered to occur when the extent of the heavily deformed region approaches the interparticle spacing which approximates to the ligament width λ_{crit}, this may be shown to give:

$$K_{Ic} \approx \left[2\sigma_y E \left(\frac{\pi}{6}\right)^{\frac{1}{3}} D\right]^{\frac{1}{2}} f_v^{-\frac{1}{6}} \tag{2.2}$$

where D is the diameter of the large inclusion particles, f_v is their volume fraction, σ_y and E are the yield stress and Young's modulus. This model has been shown to give a reasonable prediction of the effect of the volume fraction of inclusions for constant yield strength and constant particle size in a number of systems, but does not agree with experimental results concerning the influence of yield stress and particle size (Van Stone and Psioda 1975). In an earlier model, Hahn and Rosenfield (1968) considered the effects of strain hardening coefficient and yield stress on toughness, particularly in terms of the increased flow localisation, and hence crack tip "damage" that occurs with decreasing work-hardening rate. Garrett and Knott (1978) reviewed the derivation of this model leading to the relationship:

$$K_{Ic} \approx \sqrt{\frac{2CE\varepsilon_c^* \sigma_y n^2}{1-\nu^2}} \tag{2.3}$$

where C is a constant, ε_c^* is the critical crack tip strain at which unstable propagation occurs (ε_c^* is taken to be a function of the volume fraction of void nucleating particles), n is the work hardening exponent and ν is Poisson's ratio.

The predicted $n\sqrt{\sigma_y}$ dependency of fracture toughness for a constant distribution of particles has been shown to provide a reasonable description of toughness behaviour as a function of ageing between under- and overaged conditions for several Al-based alloys. Peak aged alloys, having lower n values compared to under and overaged conditions, will then exhibit low toughness levels due to an increased tendency for strain concentration reducing the crack tip opening strain before fracture. The effect of increasing alloy purity or optimising other aspects of the microstructure may then be observed through effects on ε_c^*. It has been suggested that the volume fraction of dispersoid particles may result in easier, and perhaps earlier void sheet formation and therefore lower toughness (Garrett and Knott 1978; Van Stone and Psioda 1975). Garrett and Knott (1978) consider the role of small particles in terms of their effect on the slip process, i.e. on the flow stress and hardening characteristics of the matrix, rather than in their effect on void initiation and coalescence. Chen and Knott (1981) have studied the effect of dispersoid particles on fracture, indicating that strain localisation within shear bands in the plastic zone ahead of the crack tip could lead to decohesion of the interface between the matrix and dispersoids. They proposed the use of a critical stress criterion to describe the decohesion

of the interface between the matrix and the dispersoids, and by using the same description of the plastic zone as Hahn and Rosenfield (1968) derive the following relationship:

$$K_{Ic} = \sqrt{CE\sigma_c\sigma_y n^2 \frac{\lambda}{d}} \tag{2.4}$$

where σ_c is the critical dispersoid-matrix decohesion stress, λ is the dispersoid spacing and d is the dispersoid diameter. In terms of this model, and the others noted above, it should be recognised that multiple failure mechanisms may occur in commercial products, with, for example, Gokhale et al. (1998) developing a multiple-mechanism model for toughness in partially recrystallised 7xxx series Al-alloys, relating fracture toughness to a variety of features such as the degree of recrystallisation, grain size of the recrystallised grains, the thickness of recrystallised regions, the total surface area of the constituent particles per unit volume, and microstructural anisotopy.

In summary, the fracture toughness of particle-strengthened Al-alloys may be optimised if the particles are small, non-shearable, and strongly bonded to the matrix. Alloy design trends for high fracture toughness have focused on reducing the volume fraction of void nucleating particles, by minimising Fe and Si contents, optimising solute content to prevent the formation of coarse primary phases, and using Zr as the dispersoid forming addition, to control the grain size and degree of recrystallisation. Whilst models for fracture toughness have been successfully applied in the analysis of selected aspects of toughness behaviour, the complexity and interdependencies involved in microstructure-toughness relations are problematic for development of comprehensive process-based models.

2.3.4 Stress-corrosion cracking

In Al-alloys stress corrosion cracking (SCC) (the deterioration of a material under the combined action of a stress and corrosive environment) almost invariably occurs at or near grain boundaries. The initial stages of precipitation hardening decrease SCC resistance whilst overageing beyond peak hardness improves SCC resistance (see Figure 2.1). There are differences in opinion concerning the relative importance of electrochemical and microstructural aspects of the failures. Most proposed mechanisms are variations of two basic theories - crack advance by anodic dissolution, or by hydrogen embrittlement. Although SCC eludes a complete mechanistic interpretation, it is clear that it is a stress-activated process. It is also evident when examining the properties of a wide range of Al-alloys that neither the toughness nor strength alone control their susceptibility to SCC (Speidel 1975). Controlling the chemical composition of the matrix and the precipitates is important in controlling SCC, the details of which may be complex in ternary and higher-order alloys (Speidel and Hyatt 1972). Variation in SCC resistance in 7xxx series alloys is furthermore a function of the rate of quenching from the solution treatment temperature (Speidel and Hyatt 1972), with poorer SCC resistance being seen in thick/slow quench materials.

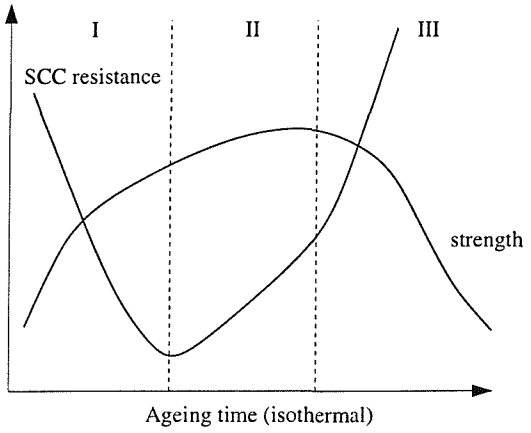


Figure 2.1: Relationship between strength and SCC resistance during ageing of high-strength Al-alloy (Speidel and Hyatt 1972).

Considerable effort has been directed towards understanding SCC mechanisms in Al-alloys, predominantly in terms of microstructural features such as PFZs, matrix and grain boundary precipitate characteristics, and solute concentrations in the vicinity of grain boundaries (Adler et al. 1972; Polmear 1981). The localisation of strain at either grain boundaries or inside the grains has been considered partially responsible for sensitivity to SCC (Sarkar et al. 1981).

Reduced susceptibility to SCC has been achieved through control of both composition and heat-treatment processing, e.g. addition of Zr to replace Cr and Mn (reducing the quench sensitivity of the alloys) and development of two step heat-treatments. An unrecrystallised elongated grain structure is also considered desirable for SCC resistance.

The quench rate and Cu content are known to play important roles in determining the SCC resistance. The effect of the former are markedly different for high and low Cu concentrations in Al-Zn-Mg alloys: the stress corrosion performance of low Cu alloys benefiting from as slow a quench rate as is consistent with strength requirements (Thompson 1975). In Deshais and Newcomb (2000) Cu and Fe-rich constituent particles were seen to increase the SCC susceptibility of 7xxx series Al-alloys. Sarkar et al. (1981) have related SCC kinetics for Al-Zn-Mg-Cu alloys with the deformation behaviour and electrochemical characteristics in terms of the Cu content.

SCC susceptibility in Al-Zn-Mg-Cu alloys is strongly influenced by grain boundary precipitation and has been related to the grain boundary segregation of Mg (Hepples 1987; Scamans et al. 1987), although no studies of Mg as a function of ageing time have been correlated with SCC resistance. Chen et al. (1977) and Viswanadham et al. (1980) observe that at the grain boundaries not all the Mg is present as the $MgZn_2$ intermetallic, a considerable fraction being present in solution at grain boundaries. Increased susceptibility to SCC for low Zn:Mg ratios was attributed to this free Mg. A high resistance to SCC develops upon overageing as

the free Mg is reduced and coarse grain boundary precipitates are developed.

Direct relationships between electrical conductivity (σ_{el}) and SCC of overaged 7xxx series Al-alloys have been established (e.g see Tsai and Chuang (1996)), with SCC resistance being found to increase with increasing σ_{el}. As such, minimum electrical conductivity levels are included in many corrosion resistant temper specifications. Conductivity increases with an increasing degree of ageing (Hepples, 1987) and it is related to changes in the type and volume fraction of matrix precipitates. The size and distribution of the precipitates are also significant, being responsive to changes in composition and thermal condition. Using both electrical resistivity and thermoelectric power techniques Vooijs et al. (2000) obtained qualitative and quantitative information about precipitation processes. Quenching an alloy after a solution treatment generally results in low σ_{el}, as a large amount of the alloying elements present are retained in solid solution. Depleting the solid solution of constituents, ageing and annealing all increase σ_{el}. Electrical conductivity thus provides a means of quantifying the changes in solute distribution.

2.4 Wrought alloy production

An understanding of the typical stages involved in the production of wrought alloys will be seen to be important in Chapter 9, where an understanding of commercial fabrication practices will be a mandatory step in data mining of process data. The typical stages involved in the production of wrought plate, from initial alloying at the melt, through to final heat treatment can be summarised as:

- ingot casting
- homogenisation
- hot rolling
- solutionising
- quenching
- cold working
- ageing

These stages are schematically illustrated in Figure 2.2. Depending on the particular alloy composition and extent of TMP, each stage may influence the balance of properties attained. Table 2.1, adapted from (Staley 1992), summarises the metallurgical features that are developed at each stage of the production of wrought plate products.

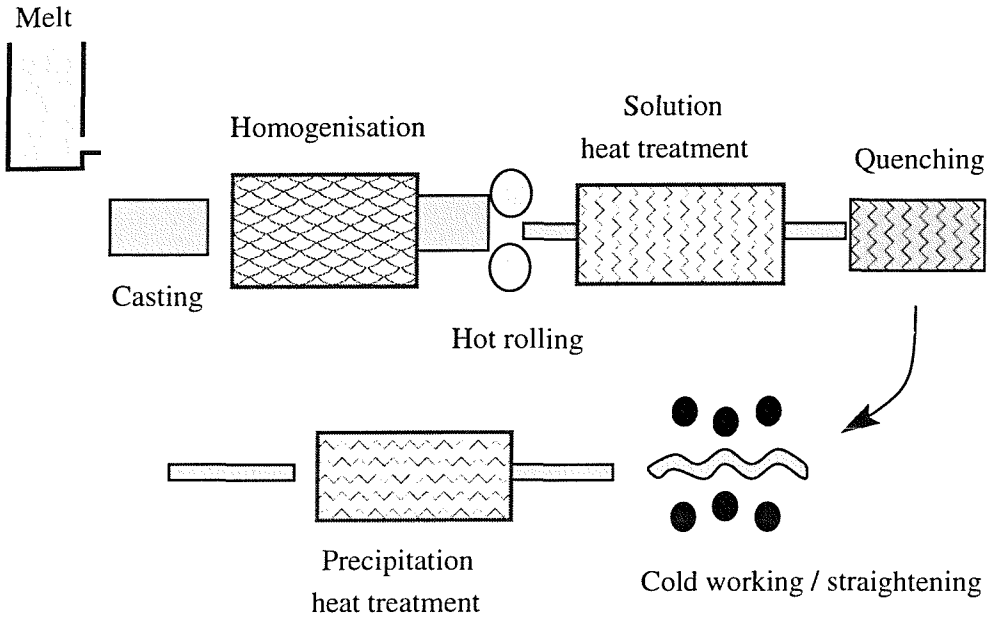


Figure 2.2: Schematic diagram of the stages involved in the commercial production route of the fabrication of wrought plate products used in aerospace applications.

Metallurgical Feature	Processing Step
constituents	solidification
dispersoids	homogenisation
grain structure	hot-rolling
dislocations	cold working
precipitates	final ageing treatment

Table 2.1: Development of metallurgical features.

2.4.1 Ingot casting

The production process begins with melting nominally pure aluminium ingots, together with production scrap off-cuts and quantities of alloying metals. During non-equilibrium freezing two main structural developments occur: an uneven distribution of constituents (both micro- and macro-segregation) and the formation of non-equilibrium phases. Most elements which have a very low solid solubility in aluminium are found to be segregated during casting and can be responsible for the formation of second phase particles which cannot be easily removed during subsequent processing. The sizes of the constituent particles developed are related to the solidification rate, with coarser constituents located at the centre of an ingot where the cooling rate is slowest. Control of grain structure in ingot casts is achieved by providing the right combination of nucleation and growth conditions. In section 2.2.3 it was seen that additions

of Ti and B as grain refiners are used for such purposes. The ingot structure developed upon solidification can considerably influence the properties attained in the final product. Analytical modelling of DC cast ingot solidification may be limited by the complexity of the process and by the accurate estimation of coefficients (Granger 1989).

2.4.2 Homogenisation

The as-cast ingots are heated up and soaked for a few hours in furnaces (soaking pits) at temperatures in the range of 400-500°C. The purpose of this pre-heating stage is to dissolve the soluble constituent particles formed during the ingot casting and alleviate micro-segregation, relying on the diffusion of alloying elements at the scale of the cast dendrite structure. Also an improved workability of the ingot can be attained.

While commercial homogenising treatments are effective in homogenising the distribution of some alloying additions (e.g. Cu, Mg, Si and Zn) they cannot fully eliminate the segregation of other elements (e.g. Cr and Fe) and so the as-cast structure of some elements is retained after this thermal treatment. Homogenisation is particularly important for the higher-strength alloys as it serves to precipitate submicron dispersoid particles. The homogenisation treatment which is a solution treatment for the former may act as a precipitation treatment for the latter. Time, temperature and heating rate will control to a certain extent the final particle distribution, although significant heterogeneity is retained from the casting process.

2.4.3 Hot rolling

Ingots are hot-worked to breakdown the as-cast structure with the intent of achieving refined grain structure, as well as reduced constituent particle size and spatial distribution. During the hot-deformation process (Zaidi and Wert 1989) ingots of up to 500mm may be rolled down to plate/sheet forms with typical thicknesses ranging from hundreds of mm to a few mm.

Aluminium dynamically recovers during this hot deformation, producing a network of subgrains, their size being influenced by alloying additions and by the deformation variables (Jonas et al. 1969). In hot-rolling plate, the degree of working is uneven throughout the thickness, decreasing from surface to centre. However, thicker products are usually fabricated at high temperatures to minimise flow stress and will have a low driving force for recrystallisation. Also, thick sections will typically experience less deformation, and hence will not undergo bulk recrystallisation, retaining an elongated grain structure.

2.4.4 Solution heat treatment

During solution heat treatment, material is heated to within a few degrees of the temperature at which incipient melting occurs and held at such a temperature for a predetermined length of time. During the solutionising heat treatment, the stored energy of deformation is given the

opportunity to decrease by static recovery and/or recrystallisation. Reduced solution treatment temperatures and times may give a reduction in both the solute and in the number of vacancies retained in solid solution upon quenching. The time required to heat a load to the treatment temperature increases with section thickness and furnace loading. The soak times for wrought alloys should be designed to take into account the normal thermal lag between furnace and part and the difference between surface and centre temperatures.

2.4.5 Quench

Quench sensitivity of 7xxx series Al-alloys is a major concern in the aerospace industry, where the drive towards producing thick plate and the desire for reducing quench-induced internal stresses result in relatively slow cooling rates at plate mid-planes. A slow quench not only promotes solute precipitation, it also permits vacancies to diffuse. A balance must be obtained between the need to quench sufficiently quickly to retain most of the hardening elements in solution and the need to minimise residual stress and distortion in the parts being quenched. The magnitude of the residual stresses is determined by the severity of the quench and section thickness. The extent to which slower quench rates can be tolerated is controlled by the *quench sensitivity* of the alloy. During a slow quench, extensive nucleation of precipitates may occur on various microstructural features, particularly grain boundaries and dispersoids, resulting in a decrease of the solute available for fine-scale precipitation and therefore a reduction in the hardening potential. Of the dispersoids commonly present in the high-strength alloys, the Cr-containing phases ($Al_{12}Mg_2Cr$) are thought to cause the greatest quench sensitivity, whereas the Zr-containing phases (Al_3Zr) have a much less deleterious effect.

The decrease in quench rate from surface to centre and the solute depletion at the mid-plane and solute enrichment at the quarter through to surface will impart a gradient in properties. By considering homogeneous nucleation theory and kinetics of precipitation it is possible to calculate the fraction of solute precipitated and thus predict the effects of precipitation during the quench on properties. Evancho and Staley (1974) introduced quench factor analysis as a method of interrelating the cooling rate, section thickness of a part, and alloy precipitation kinetics, developing analytical expressions to represent time-temperature-property (TTP) C-curves. Thus, quench factor analysis has been used by Staley (1987) to describe how precipitation during the quench affects the development of properties of Al-alloys during subsequent ageing treatments. Quench factor analysis can be used to predict the strength in Al-alloys, from experimental and analytical cooling curves. Bates and Totten (1988) devised a quench factor, Q, to interrelate quenching variables, part section size and transformation rate data, providing a variable indicating the extent to which a solution-treated Al-alloy can be strengthened by subsequent ageing. Quench factor and heat transfer analyses have been combined to develop quenching practices providing optimal combinations of residual stresses and strength after ageing (Archambault et al. 1980).

The effect of quenching conditions on fracture characteristics are more complex than for strength. Rapidly quenched, precipitation hardened Al-alloys generally fracture in a trans-granular manner by the coalescence and growth of microvoids. Decreasing the quench rate increases the amount of grain boundary precipitate which, in turn, increases the tendency for low toughness, intergranular fracture to occur.

Several material and quenchant characteristics influence the rate of heat removal from a part during quenching, and hence cooling rates, e.g. see (Bates and Totten 1988). In commercial spray quenching, the quench rate is a function of the coolant flow rate (or spray/nozzle pressure), i.e. the velocity and volume of water per unit time of impingement on the workpiece. The rate of travel of the workpiece through the quenching zones will also be an important variable.

2.4.6 Cold working and stretching

The residual stresses introduced during the quench may cause a certain amount of buckling. To reduce this problem, prior to stretching, wrought plates are commonly passed through a leveller. By subsequently stretching the plates to permanently increase their length, typically in the range $1.5-8\%$, the internal stresses are almost entirely relieved and the plate is flattened and straightened to its final dimensions. When cold working is followed by age-hardening heat treatments, the precipitation effects on strength may be accentuated. In some alloy systems, a degree of work hardening may be retained in the formed product.

2.4.7 Precipitation heat treatment

In this, the final stage of the age hardening process responsible for the high strength levels of 7xxx series Al-alloys, batches of plates are held at pre-specified temperatures (up to 200°C, $+/-$ a few °C), below the meta-stable solvus of the desired strengthening phase. Some of the loss in strength resulting from slow quenching can be minimised by decreasing the ageing temperature to maximise homogeneous nucleation.

2.5 Advances in thermodynamic and physical-based modelling

Efforts to model phenomena such as solidification, homogenisation, and segregation rely heavily on the availability of high quality thermodynamic and phase diagram data. The potential for using phase diagrams in process modelling is beginning to be realised with the development of thermodynamically calculated phase diagrams, e.g. see (Saunders, 1996; Kolby, 1996). Phase diagram, solidification and long-range diffusion routines have been integrated in a modelling framework under development by Pechiney for heat-treatable Al-alloys (Sainfort et al. 1997). Classical laws for homogeneous nucleation, growth and coarsening of spherical precipitates (and co-precipitation) are used to describe the kinetics of precipitation of the most

stable precipitates and to predict their volume fraction, density and radius (Sigli et al. 1998; Sigli 2000). The proprietary methods used require fewer adjustable parameters than more fundamental thermodynamics approaches.

To produce a valuable commercial tool, such modelling must be integrated through an entire production process (e.g. casting, homogenising, hot rolling, solutionising, quenching, stretching, ageing) to yield a final microstructure, which is then used to predict properties of interest. Final properties may be determined by combining physically-based models with processing parameters measurable at the industrial scale and parameters describing the alloy microstructure such as chemical composition, volume fraction and nature of precipitates, precipitate and grain geometry/dimensional parameters. Models may also be used to take into account heterogeneities within the material due to macrosegregation, as well as temperature and deformation gradients occuring during secondary processing (Sainfort et al. 1997).

Modelling of commercial thermomechanical processing (TMP) and subsequent materials' properties, may be broken down into an understanding of temperature, strain and strain rate evolution, the microstructural response of the material to a given thermomechanical history and structure-property relationships for the performance characteristics of interest. The temperature and strain/strain rate distributions associated with any given TMP may be further dependent on previous microstructural developments via changes in the constitutive equations of the underlying material. Basic quantitative process models of commercial TMP stages and associated microstructural evolution have been put forward in recent years (Sellars 1987; Shercliff and Ashby 1990; Sellars 1992; Sigli et al. 1996), often relying on a high degree of empiricism and extensive *calibration* data for specific alloy systems. As noted in section 2.1, this may be attributed to the complex nature of commercial TMP and difficulties in simulating industrial scale processes in laboratory experimental trials.

Microstructural models developed from a physical basis are likely to provide improved predictive capabilities for alloy processing such as rolling, extrusion and forging when compared to empirical approximations (Furu et al. 1996), as the lack of the fundamental mechanisms in the latter is not suitable for alloy development. However, as a greater physical detail is integrated into the physical model, the number of adjustable parameters increases, compromising the predictive capabilities of physical-based modelling approaches in industrial settings (Furu et al. 1996). Shercliff and Lovatt (1999) have reviewed physical-based, semi-empirical and adaptive approaches (BMLPs, recurrent neural networks and Gaussian process models) used in modelling microstructural evolution during hot deformation, in predicting flow stresses during deformation and the subsequent annealing behaviour of Al-Mg alloys, suggesting that although adaptive methods do not provide any physical insight into the deformation processes, their main use, in addition to providing valuable predictive abilities (with associated measures of uncertainty, i.e. error bars), is seen as methods for optimising time-consuming experimental work and sufficient in themselves for industrial applications.

2.6 Conclusions

From the alloy fabrication route described in this chapter, it is seen that a large number of processing conditions and practices may, depending on the processing history, influence the final properties developed.

Both experimental and industrial studies have contributed significantly in the development and optimisation of many processing conditions which are now widely employed in commercial settings. However, as Staley (1992) discusses, theoretical modelling approaches provide only the broadest guidance to the metallurgist interested in selecting alloying levels and processing conditions in order to obtain a useful balance of properties, controllable in large scale commercial processes. Industrial processing routes may gather a potentially very large amount of process data which will be complementary to the scientific and theoretical research studies. Typically, in these settings, empirical curve fitting is used without regard to mechanisms, and although providing useful and practical equations for the processing conditions considered, as discussed in section 2.1, such analysis in many cases provides no insight into the underlying mechanisms responsible for the levels of properties attained.

The use of adaptive data modelling techniques such as neural networks and support vector machines may offer significant advantages over simple *ad hoc* regression analysis, both in terms of system description and predictive modelling, whereby the application of advanced learning methods are able to determine a model that is well supported by the information contained in the available data.

An inadequate description of the processing conditions in terms of process variables and microstructural features will be reflected in a sub-optimal process description. Consequently, data-driven modelling techniques will lack the empirical knowledge required to obtain a good description of the process. Under these circumstances, models will exhibit poor generalisation performance and inferring the significance of certain processing variables will at best be problematic and any inferences should be viewed with caution. In such settings a good understanding of the production route, in particular the historical microstructural developments, is a mandatory step in validating the modelling results, both in terms of the empirical performance attained but also in terms of the nature of the dependencies inferred.

Chapter 3

Data Modelling

3.1 Introduction

In any physical process in which processing conditions vary, it is of interest to understand how these affect the system's behaviour. In many real world systems there will be uncertainty as to the true causes (variables) responsible for the behaviour observed, and there will rarely be a simple relationship between explanatory and response variables, as the relationships of most physical processes will be too complex to be described by means of simple analytical expressions. In addition, the data will generally be characterised by measurement errors. It becomes desirable then to *approximate* the true unknown function from the available data, augmenting the empirical observations with prior knowledge when possible. This chapter presents a general introduction to data modelling, focusing on a number of topics including linear regression, complexity control, model selection and Bayesian methods. Subsequent chapters will discuss data integrity issues and the two adaptive modelling techniques comprising the main modelling approaches used in this work.

3.2 Function approximation

The task of function approximation can be viewed as the process of *learning*, i.e. estimating unknown input-output dependencies of the process which are responsible for generating the observations. Figure 3.1 shows a simplified view of statistical process modelling. The objective of learning is therefore to predict an output y from an input vector $\mathbf{x}$, where the pair $(\mathbf{x}, y)$ is drawn from some unknown joint probability distribution $p(\mathbf{x}, y)$, assumed to be independently drawn and identically distributed. This assumption allows a distribution of y for a given $\mathbf{x}$.

The process is assumed to be characterised by the vector $\mathbf{x}$ of random samples, drawn from an unknown density $p(\mathbf{x})$. For an input vector $\mathbf{x}$, the process produces an output value y according to the unknown conditional density $p(y|\mathbf{x})$. Typically, the process will be characterised by experimental measurement errors and often will have unmeasured inputs.

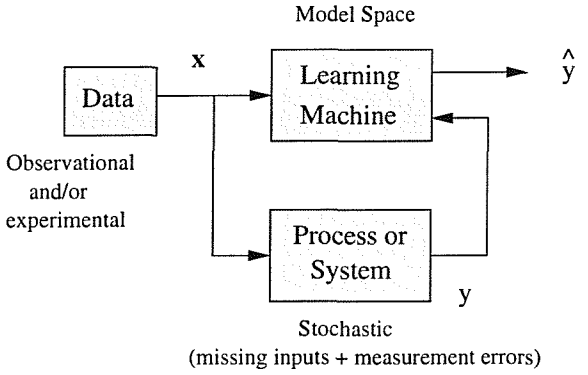


Figure 3.1: Schematic diagram of the statistical process modelling problem (Cherkassky and Mulier 1998).

A *training set*, $\mathcal{D}_N = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_N, y_N)\}$, is a collection of N observed $(\mathbf{x}, y)$ pairs containing the response observed y for each input $\mathbf{x}$, the learning problem then being that of determining a function $f(\mathbf{x})$ from this training set, so that $f(\mathbf{x})$ approximates the desired response y. The function $f(\cdot)$ is generally chosen to minimise some objective function (or model performance criteria) such that the model is taught in a way that the error $(y - \hat{y})$ is sufficiently small.

The different approaches to the problem of modelling a probability density function $p(\mathbf{x})$ broadly fall into either *parametric*, *non-parametric* or *semi-parametric* methods (Bishop 1995). In parametric methods, the form of the functional relationship between the response and the explanatory variables is assumed to be known and so the modelling process is then limited to a parameter estimation task. When *a priori* models of the process are unavailable, semi-parametric and non-parametric methods are adopted as these methods respectively assume a very general class of functional forms or none whatsoever. Many neural network classes (Haykin 1998) can be regarded as semi-parametric methods in that they allow the number of adaptive parameters to be increased/decreased, this flexibility enabling to approximate a wide range of functions.

Computational methods of model construction have proliferated in recent years. Friedman (1994) reviews the underlying principles of many approaches which have been developed in the fields of applied mathematics, statistics, pattern recognition and machine learning, placing them in a common perspective. Methods that have become popular over the past two decades include tree based methods (Breiman et al. 1984), generalised additive models (Hastie and Tibshirani 1990), generalised linear models (McCullagh and Nelder 1989), regression splines (Friedman 1991) and several classes of neural networks, fuzzy systems (Wang 1992) and hybrid paradigms combining fuzzy-type reasoning within neural network architectures and learning methods (Brown and Harris 1994; Jang et al. 1997). The application of Bayesian inferencing techniques within neural networks (MacKay 1992b; Neal 1996) has been a recent significant development. Finally, despite being originally proposed nearly thirty years ago, sup-

port vector machines (SVM) (Vapnik 1995) (see Chapter 6) have only recently seen significant implementations.

3.3 Observational and experimental data

Before addressing statistical and computational aspects of data modelling it is of value to discuss the nature of typical data sets. Understanding the nature of the data plays a critical role in any subsequent inferences. As Snee (1983) notes, modelling observational data, as opposed to data generated from carefully designed experiments is, at best, a risky business. Data collected without the aid of statistical design criteria may be suspect for a variety of reasons, including errors in both explanatory and response variables, near-linear dependencies amongst the former, inadequate sampling of the experimental region, etc. Box et al. (1978) discuss the hazards of performing a regression analysis on what the authors term "happenstance" data, such as ranges and distributions of input variables limited by process control, semiconfounding effects (i.e. change in one processing condition accompanied by a corresponding change in another), nonsense correlations (often attributable to unmeasured variables), etc.

The $(\mathbf{x}, y)$ pairs must be drawn from the same process, e.g. if y is the value obtained from a tensile test performed on a specimen taken from a certain plate, then all x's, comprising the complete $(\mathbf{x}, y)$ instance should be determined for the same plate. If instead $(\mathbf{x}, y)$ is the collation of historical data pertaining to the same type of alloy, but with the x's obtained from different plates (despite having the same alloy designation and nominal processing conditions), statistical inferences will be questionable. The *integrity* of the data set appears an obvious requisite, but many industrial and commercial alloy property databases are gathered in an unsystematic manner.

Depending on the gathering process, data may be viewed as either *experimental* or *observational.* Although processing conditions are taken to be specified design variables, in many cases it is more appropriate to consider them as random variables and assume a joint distribution on y and $\mathbf{x}$.

The data analyst should have some understanding of the likely interactions and effects of the variables recorded. Variables included in the data set may describe a broad range of processing conditions, characterised by different statistical properties: some may be continuous, while others will assume a finite set of values (which may assume a natural ordering). Certain variables may be labels, e.g. tensile test directions: long-transverse (LT), transverse (TL) and short-transverse (SL). Statistical properties will reflect the origin and nature of the variables: typically, *continuous* variables will be measurements of processing conditions (e.g. the temperature of a hot mill, solutionising times), or physical characteristics (e.g. weight percentages of alloying elements, physical properties, mechanical responses), whereas *discrete* variables will typically provide information identifying the alloy types, test conditions, (e.g. environment, test positions, etc.).

The observed quantities may finally be either *response*, *intermediate* or *explanatory* variables (intermediate variables are responses to some variables and explanatory to others), thus yielding a hierarchy between the variables.

3.3.1 Data quality

The quality of a data set plays a significant role in any form of empirical modelling, since the main limitation in extracting reliable and meaningful knowledge from the data is ultimately related to its richness and information content. In the neural network literature the phrase *garbage in - garbage out* is used to emphasise the importance of having a good data set. The quality of the data set can be assessed in terms of the statistical properties of the variables (e.g. a variable that does not experience much variation in values will be uninformative) and the representativeness of the set of variables in describing the process. The main difficulty in learning from data is in separating the structural information from the noise. Finally, the data set may have limited depth, with important variables missing.

If data is to be collected purposefully for a statistical investigation, from which reliable inferences are to be attained, then experimental design criteria should be employed whenever possible. Unfortunately, due to process limitations, constraints and costs, employing such design criteria is not always feasible.

Chapter 4 will review diagnostic statistics useful in identifying potential sources of modelling problems. The rest of this section discusses the major reasons why data quality is of such concern and attempts to clarify practical reasons behind the generation of a poor data set.

3.3.1.1 Sampling distributions

Data is seldom recorded for the purposes of statistical analyses: the data set may be historical, recorded solely for process control and product traceability. Process requirements may reflect the distibution of the data in input space, generating dependencies amongst process variables. For instance, master alloys rich in certain alloying elements (e.g. Mg, Zn), may contain other elements (e.g. Fe, Mn, Si), giving rise to correlations between compositional elements. Similarly, requirements to dilute or refine the melt, say, due to unacceptable levels of certain compositional elements (e.g. Fe, Ni), will affect compositional levels of many other elements, also generating dependencies among compositional elements. The effect here is the inability to isolate the distribution of certain variables with respect to others (e.g. Fe and Si both of which are impurities in Al-alloys, will be present in similar quantities: trying to limit the Si content while not affecting Fe may be problematic). It then becomes problematic to infer which properties are affected for example by Fe, Mn or both.

In general, the sampled distribution $p(\mathbf{x})$ of many observational process variables will be influenced by the process development history and design criteria pursued, often resulting in a poor coverage of the input space. As Hocking (1983a) notes, important processing conditions

will be varied only over a limited range because of the desire to retain tight process control procedures. Standard regression analyses will conclude such variables to be statistically "not significant".

In addition, there may be data deficient regions of the input space attributable to either the lack of, or poor design criteria, for instance changing processing conditions through a concomitant change in more than one process variable. Also, dependencies may arise due to other processing requirements, e.g. thicker plates will experience different processing conditions (e.g. longer soak times, quench rates, amount of deformation).

It should be clear that as well as differing in size, commercial databases gathered without the aid of sound design criteria will exhibit substantially different statistical properties to experimental and scientific data sets.

3.3.1.2 Process variables

As the input quantities present in the data set will be used to describe the process, and in many cases used to determine a predictive model, it is important that these variables are representative of the system. Unfortunately, in many systems a number of important features can remain unobserved. There may be several reasons for missing variables, including cost, proprietary information and a limited understanding of the process. In addition, it may be problematic to characterise certain microstructural features (e.g. PFZs, volume fractions, etc.). In large databases on the other hand, typically containing a very large number of fields (variables), valuable information may be contained in a small subset of these. Attempting a statistical analysis of such data, combined with lack of process understanding will result in a greater risk in failing to identify significant variables, increasing the risk of inferring noncausal dependencies as well as an increased unnecessary computational effort.

3.3.1.3 Noise levels

Process variables may be characterised by different noise levels with some affected by higher noise levels than others. Likewise the observed response will also have a noise component.

All the above will affect the entire knowlege discovery and modelling processes, limiting the reliability of statistical inferences.

3.4 Modelling objectives

Prior to any statistical analysis, the main objectives of the investigation should be identified and prioritised; this will help in the choice of an appropriate modelling approach. Generally, there will be different reasons for modelling data. For instance, the sole purpose of the analysis may be simply to explore the data distributions, identifying anomalies in the control of particular

processing conditions. On the other hand, the primary objective may be predictive ability, i.e. developing a model of the process so that future outcomes of the response variable may be predicted from a set of input variables. A greater process understanding may be desirable, the objective being to infer which input variables have the strongest effects on the response and the nature of the dependencies. Finally, it may be desirable to perform some form of data compression, summarising the data by means of summary statistics. The following sections discuss the principal attributes that are sought in this work.

3.4.1 Generalisation (prediction)

Generalisation[1] is generally considered the most important property of a model. The model should not only provide an adequate fit to the exemplar training data, it should generalise well to previously unseen, or new observations (patterns) obtained from the same process that was responsible for generating the training set.

Three conditions are typically necessary (although not sufficient) for a model to have good generalisation (Sarle 1997). Firstly, the inputs should contain sufficient information pertaining to the response, so that structural relationships between inputs and outputs exists. Secondly, the function to be learnt should be, in some sense smooth: a *small* change in the inputs should produce a *small* change in the outputs. Thirdly, the training set should be a sufficiently large and representative sample of the population to which to generalise to. The importance of this last condition is related to the fact that there are two different types of generalisation: *interpolation* and *extrapolation*. The former applying to cases that are more or less surrounded by nearby training cases; everything else should be regarded as extrapolation, particularly cases that are outside the range of the training data. Interpolation can often be done reliably, whereas extrapolation is notoriously unreliable, emphasising the importance of having sufficient training data to avoid the need for extrapolation.

Good generalisation to unseen data will be possible when the model captures the underlying relationships contained in the data, rather than fitting the idiosyncrasies present in the training set (noise contribution). Typically, for parametric models, generalisation ability reaches an optimum value for a particular degree of complexity of the model, generally occurring when the number of degrees of freedom in the model (section 3.7) is relatively small compared to the size of the data set. Recent developments in the neural network field show that it is not so much the number of weights but the relative magnitude of these that is important (Bartlett 1997) and by the application of Bayesian methods (Neal 1996) it has also been shown there is no statistical need to limit the complexity of the model, thus models can have more parameters than the number of data points (Neal 1994; Neal 1996; Williams 1997).

[1]Typically defined as the ability of the system model to respond appropriately to input conditions which are not contained in the training set.

3.4.2 Interpretability (description)

In many situations, it is desirable to *learn* the trends and patterns that the learning paradigm (model) has itself learned. In such cases, it is important for the framework to be in some sense *transparent*, so that the knowledge stored in the model may be interpreted and validated. It may be desirable to infer which variables explain the variation in the observed values of the process response and the nature of the dependencies characterising the explanatory and response variables.

Unfortunately general neural network architectures may well have no affinity to the physical laws of the system being modelled, with network weights having no relation to physical parameters. As such neural network models are in many cases used as a means to a predictive end.

In general, accuracy and interpretability may be conflicting attributes, with some paradigms more suited than others in achieving a useful balance. Predictive abilities will only be achieved if an appropriate model is inferred; if the inherent complexity of the process is inadequately approximated, this will not only compromise its generalisation capabilities, but result in misleading inferences.

3.4.3 Uncertainty

Generally, regression analyses will lead to some form of inferencing. This may be either in terms of the model structure identified, predictions, value of the estimated parameters, generalisation performance, noise levels, etc. The model should reflect the uncertainty associated with these inferences, which in many cases will be based on finite sample sizes, with some measure of uncertainty (or confidence). Uncertainty is dealt with in different ways depending on the properties of the modelling technique.

Often regression analyses may be performed for two purposes: prediction and intervention. The former is feasible if it can be assumed that the same dependencies that were present when the data was gathered, are also present when predictions are made. To intervene and change a system's behaviour, causal relationships are necessary. However, when modelling historical data, correlation may indicate causation, although this may not necessarily always be the case. Box et al. (1978) assert that to safely infer causality, natural or historical happenings of the data cannot be relied upon.

3.5 Multiple linear regression

The classical (or normal) linear regression model and its extension to more than one input variable, multiple linear regression (MLR) remain commonly used regression methods. Many textbooks (Draper and Smith 1981; Fox 1984; Weisberg 1995; Neter et al. 1996) provide a

comprehensive introduction to regression analysis. Unfortunately, there are several factors limiting the applicability of simple regression analysis, mainly the recognition that many real world processes depend on many variables in a more complex, often non-linear way.

The principal purpose of regression analyses is generally to *learn*, i.e understand the relationships between input variables and the response variable. However, the presence of a statistical relation between x and y does not imply in any way that y depends causally on x (Box 1966). Thus, caution is needed when drawing conclusions from a regression analysis. The simplest and most straightforward way of approaching these problems, is in using domain knowledge when this is available, by means of selecting a good set of explanatory variables and using similar expertise to validate the results.

The basic form of the MLR model is expressed as:

$$y = f(\mathbf{x}, \mathbf{w}) = w_0 + w_1 x_1 + \cdots + w_p x_p = \sum_{j=0}^{p} w_j x_j \tag{3.1}$$

where $x_1, \ldots, x_p$ are the set of *explanatory* (*independent*, *predictor* or *covariate*) variables, $w_0, \ldots, w_p$ the set of adjustable *weights* (*parameters*), and y the *response* (or *dependent*) variable. In equation 3.1 w_0 is the *bias* (intercept) term, associated with which is a dummy vector x_0 of length N, fixed at 1, where N is the number of observations (training patterns). The functional form $f(\cdot)$ indicates the type of input-output relationship, which in equation 3.1 is linear with respect to the weights and the response variable.

Acknowledging the fact that there will be some noise present in both explanatory and response variables, and expressing equation 3.1 in matrix notation:

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \mathbf{e} \tag{3.2}$$

where $\mathbf{X}$ is an $[N \times (p+1)]$ matrix of input variables of full rank, $\mathbf{y}$ an $[N \times 1]$ response vector, $\mathbf{w}$ a $[(p+1) \times 1]$ vector of regression coefficients and $\mathbf{e}$ an $[N \times 1]$ vector of additive noise components[2] that are assumed to be independent of x and normally distributed with mean $E\{e_i\} = 0$ and variance σ^2. Graphical methods are commonly employed (Weisberg 1995) for detecting non-normality of the error distribution assumptions. Carroll and Ruppert (1988) present an analysis of regression data when some of the above assumptions are violated, discussing suitable methods (e.g. data trasformation and robust estimation procedures) for dealing with such data sets.

When the data lacks an orthogonal design factor space, a number of problems arise due to this non-orthogonality of the data. The effects on inferences will be discussed in section 3.12, while computational and statistical issues are the subject of Chapter 4.

[2]Generally $\mathbf{e}$ is dismissed as a random variable with properties described above, however it should be understood to describe the effect of all unmeasured variables.

3.6 Parameter estimation

Given a data set of size N, denoted by $\mathcal{D}_N = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, and a functional form for the model, $f(\cdot)$, in conventional maximum likelihood estimation a single set of values for the weights is determined by minimisation of a suitable error fuction. The Bayesian approach however considers a probability distribution over weight values, representing the relative degrees of belief in different values for the weight vector (Bishop 1995). In Bayesian estimation, the parameters $\mathbf{w}$ are then seen to be treated as random variables, with a probability density function (p.d.f.) $P(\mathbf{w})$, before the data has been observed, representing the most likely distribution of the weights before the data has been seen. Once the data has been observed, the p.d.f. for the weights can be converted to a posterior distribution through use of Bayes' theorem:

$$P(\mathbf{w}|\mathcal{D}_N) = \frac{P(\mathcal{Y}|\mathcal{X}, \mathbf{w})P(\mathbf{w})}{P(\mathcal{Y}|\mathcal{X})} \tag{3.3}$$

where $\mathcal{X}$ denotes the set of input vectors $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ and $\mathcal{Y}$ are the target values $(y_1, \ldots, y_N)$. Form the above, the *a posterior* p.d.f. is seen to be composed of three different p.d.f.'s: 1. the *likelihood* function $P(\mathcal{Y}|\mathcal{X}, \mathbf{w})$, a conditional p.d.f. of the data given the parameters, 2. the *prior* $P(\mathbf{w})$, and 3. the *evidence* $P(\mathcal{Y}|\mathcal{X})$, which is simply a normalisation coefficient:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{evidence}}$$

This posterior distribution can then be used to evaluate model predictions on new data. Instead of evaluating the full posterior distribution for $\mathbf{w}$, a common approach is that of finding the weight vector that corresponds to the maximum of the posterior probability[3]. The value for $\mathbf{w}$ for which the observed data has the highest probability is obtained by maximising the likelihood function $P(\mathcal{Y}|\mathcal{X}, \mathbf{w})$ with respect to $\mathbf{w}$.

As the amount of training data increases, the maximum likelihood solution will provide a good approximation to the most probable solution. For small data sets, the prior will have an important role in determining the most probable solution. If no information about the weight vector is known $P(\mathbf{w})$ should represent complete ignorance, which can be modelled by a prior distribution with a very large variance. Then, the a posterior p.d.f. will be swamped by the likelihood function and the maximum of the posterior will be equivalent to maximum likelihood (ML) estimation, equivalent to finding the parameter vector, $\mathbf{w}$, that minimises:

$$J_N = \frac{1}{N}\sum_{i=1}^{N}[y_i - \hat{y}_i(\mathbf{x}, \mathbf{w})]^2 \tag{3.4}$$

[3]The full Bayesian approach determines the posterior probability distribution of the model's output by integrating over the weight space (i.e. over all possible values of the parameters, weighted by their posterior distribution).

This cost function is referred to as the mean square error (MSE) and is a biased estimate of the variance ($\hat{\sigma}_N^2$) of the assumed additive noise component[4]. For a model that is linear in its weights this corresponds to the solution:

$$\mathbf{w} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \tag{3.6}$$

The least squares method provides unbiased point estimators of the parameter vector $\mathbf{w}$, that have minimum variance among all unbiased linear estimators.

3.7 Model selection

The major problem in the use of parametric models is that of determining an appropriate model structure. A model that is not sufficiently flexible may fail to learn the underlying function (*underfitting*), whereas a model that is too flexible may model the noise contribution together with the true underlying function (*overfitting*). Overfitting can easily lead to predictions that are far beyond the range of the training data.

From the finite data set $\mathcal{D}_N$, the objective is then to find a model $\hat{y}(\mathbf{x};\mathcal{D}_N)$ such that the expected error across *all* data sets of size N:

$$E_{\mathcal{D}_N}\left[\left(\hat{y}\left(\mathbf{x};\mathcal{D}_N\right) - f(\mathbf{x})\right)^2\right]$$

is minimised. The objective of model identification is therefore to identify the $\hat{y}\left(\cdot\right)$ such that the MSE across these data sets is minimised.

It should be recognised that there may be several "good" models which can be identified, with a similar number of weights, that give a similar fit to the training data. Box (1976) states that *"all models are wrong"* since it is impossible to obtain a completely "correct" model of the process and so an economical (parsimonious) description of the phenomena should be sought.

3.7.1 The bias/variance trade-off

In practice, $E_{\mathcal{D}_N}\left[\left(\hat{y}\left(\mathbf{x};\mathcal{D}_N\right) - f(\mathbf{x})\right)^2\right]$ cannot be measured, but gives rise to the fundamental bias/variance decomposition of generalisation error (Geman et al. 1992), describing the sources of modelling errors:

[4]For linear models, *p degrees of freedom* (df) will have been used in determining the mean $\hat{y}(\mathbf{x},\mathbf{w})$. Thus, the variance of the noise on the training data (MSE) will be a biased estimate. An unbiased estimate of this variance will be:

$$\hat{\sigma}_{\mathrm{df}}^2 = \frac{1}{N-\mathrm{df}}\sum_{i=1}^{N}\left[y_i - \hat{y}_i(\mathbf{x},\mathbf{w})\right]^2 \tag{3.5}$$

$$E_{\mathcal{D}_N}\left[\left(\hat{y}\left(\mathbf{x};\mathcal{D}_N\right)-f(\mathbf{x})\right)^2\right]= \tag{3.7}$$
$$\underbrace{E_{\mathcal{D}_N}\left[\left(\hat{y}(\mathbf{x};\mathcal{D}_N)-E_{\mathcal{D}_N}\left[\hat{y}\left(\mathbf{x};\mathcal{D}_N\right)\right]\right)^2\right]}_{variance}+\underbrace{\left(E_{\mathcal{D}_N}\left[(\hat{y}(\mathbf{x};\mathcal{D}_N))\right]-f(\mathbf{x})\right)^2}_{bias^2}$$

The **bias** measures the extent to which the average (over all data sets $\mathcal{D}_N$ of size N) model differs from the true function $f(\mathbf{x})$, thus representing a consistent tendency towards over or underestimation. The **variance** measures the extent to which the model $\hat{y}\left(\mathbf{x};\mathcal{D}_N\right)$ is sensitive to the particular data set used by measuring the expected error between the average model and a model identified on a single data set instance. A model which is too simple will exhibit a high bias term, while one which has too much flexibility in relation to the particular data set will have a high variance.

In Figure 3.2 polynomials of differing orders are fitted on a data set comprising 25 training samples, generated by a third order polynomial on which Gaussian noise was added. The linear model will exhibit low data set sensitivity but a high MSE, systematically under and overestimating the true value of the response. The highest order polynomial will exhibit a very low MSE but extreme data set sensitivity, also entailing a poor generalisation ability. The polynomial approximation in Figure 3.2 (b), achieves a reasonable approximation to the training data and amongst the three models, will be expected to have the best generalisation performance.

It is then clear that using equation 3.4 as the objective (cost) function to be minimised, is non-optimal, as a model will always be seen to reduce the MSE by including additional degrees of freedom. The variance component of a model is seen to be an increasing function of the number of variables and/or parameters added to the model. Best generalisation will be obtained when a compromise between the conflicting requirements of low bias and low variance are met: by reducing the number of redundant parameters in the model, the bias and variance contributions can be drawn towards their minimum. In order to find this optimal balance between bias and variance, a way of controlling the complexity of the model is required, for example in neural networks, this can be controlled by changing the number of hidden nodes and hence the number of adaptive parameters in the network. Many of the criteria and objective functions used to evaluate and compare models are based on empirical error based measures (section 3.8.1) or on criteria in which the objective function penalises models with a large number of parameters (section 3.9).

The extent to which the model fits the data should be traded off against its complexity. However, the addition of a variable or extra flexibility, may decrease the squared bias component, but this decrease may be small relative to the increase in the variance. Statistical theory will show in subsequent sections that the variance of an estimate is affected by its complexity. Statistical inferencing from a simpler model, compared to a more complex one with the same generalisation ability, will be preferable and is generally referred to as **Occam's razor**.

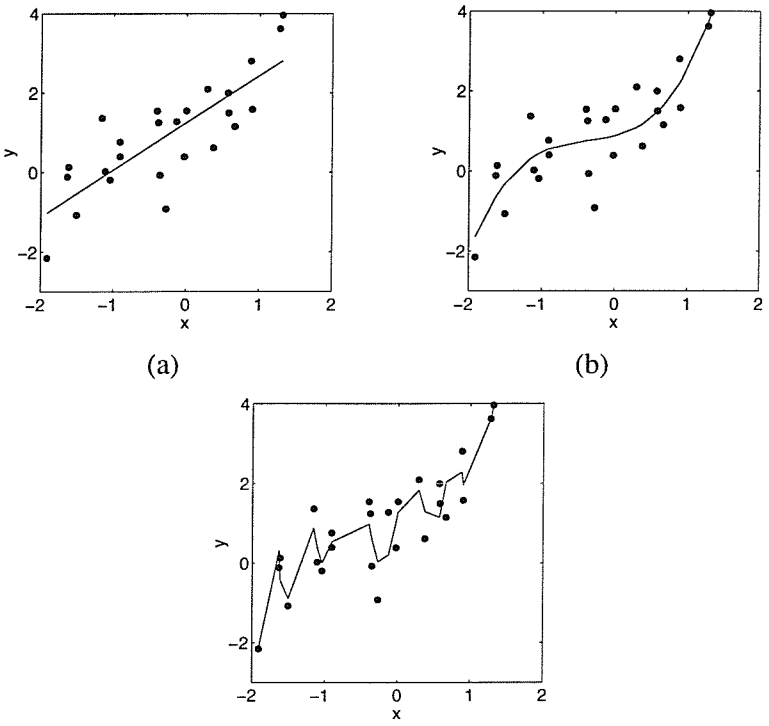


Figure 3.2: Illustration of the bias/variance trade-off: (a) underfitting (high MSE, high bias), (b) adequate approximation to the underlying function, (c) overfitting (low MSE, high variance).

3.8 Assessing generalisation performance

As will be discussed in more detail in Chapter 6, minimising equation 3.4 is also referred to as *empirical risk minimisation.* However, the model which minimises J_N does not necessarily minimise the *generalisation error*, the minimum of the risk functional (section 6.2.1) over the full distribution of possible inputs and their corresponding outputs.

The *prediction risk* (Moody 1994), defined as the expected performance of an estimator in predicting new observations (equation 3.8), cannot be computed directly but must be estimated.

$$P = E\left\{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\left(\mathbf{x}, \mathbf{w}\right)\right)\right\} \tag{3.8}$$

The simplest and most commonly employed method for obtaining an unbiased estimate of the prediction risk (generalisation performance), is to reserve part of the data as a *test set*, which must not be used in any way during training[5]. The disadvantage of reserving part of the data in this way is that it reduces the amount of data available for both training and validation. Even

[5]If $\mathcal{M}$ models are trained on the same data set and a *validation set* is used to select the best network topology, a third data set, the test set, should be used to obtain an unbiased estimate of the generalisation error for the selected model.

in cases where the data set is large, it is desirable to use as much data as possible for training, since the estimation error associated with the model variance becomes worse as the training set size is reduced. Furthermore, holding out a small and unrepresentative set of independent test samples will provide noisy estimates of the generalisation performance.

In these situations, alternative approaches that enable the estimation of the prediction risk trom the training data alone must be considered. Resampling schemes such as cross-validation (Stone 1974) and bootstrapping (Efron and Tibshirani 1993) and statistical significance (or model complexity) measures allow all the data to be used during training.

A simple statistic which is commonly used for obtaining an (unbiased) estimate of σ^2 is *leave-one-out cross-validation* (LOOCV), where each one of the observations are in turn removed from the training set and the i^{th} observation estimated from all the other $N-1$ observations:

$$LOOCV = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i^{(i)}\right)^2 \tag{3.9}$$

3.8.1 Model performance measures

The MSE is an estimator of the accuracy of the model, and as such is often employed within statistical significance (SS) measures. There are many analytical functions which have been proposed in the literature, typically in the form:

$$SS = MSE \times f(N,p) \tag{3.10}$$

where p is the number of weights and N is the number of training patterns.

Typically, such analytic model selection criteria have been developed in statistics using asymptotic (large-sample) theory.

Gunn et al. (1997) assessed the performance of several model selection criteria in the framework described in Chapter 5. The measure derived from the *structural risk minimisation* (SRM) principle (Vapnik 1995) was seen to provide a reliable method of matching the complexity of the model to the amount of training data available. The SRM measure was seen to differ from the other statistical significance measures in that it placed an upper limit on p, whilst the other measures did not place an upper limit on the number of weights (potentially allowing for over-fitted models to be selected) or exhibited an asymptotic behaviour for $p = N$. The SRM measure thus limits the number of weights in the network to be less than the number of training patterns.

Cherkassky et al. (1999) using the complexity measure proposed by Vapnik-Chervonenkis (VC) theory (section 6.2.2) also found it significantly outperformed other model selection criteria for linear and penalised linear estimators, particularly for small sample sizes. As will be seen in Chapter 6, statistical learning theory provides analytical generalisation bounds that can be used for estimating the prediction risk.

The above results suggest that the model complexity measure derived from VC-theory is superior to classical methods for model selection, in terms of selecting the models with the lowest worst-case prediction risk, and the lowest variability of risk estimates[6], despite a slight underperformance observed for large sample sizes under low-noise settings.

The disadvantage of all these measures is that they are inherently dependent upon the MSE as a measure of model suitability, and as such rely on a biased estimate of the variance.

3.9 Complexity control

In (semi)-parametric models there are two methods which attempt to limit the complexity of the approximating function, these are regularisation and model selection. The complexity control approaches used within the neural network field are architecture selection, regularisation, early stopping (Wang et al. 1994) and training with noise (Reed et al. 1995).

In the neural network literature, model selection generally refers to selecting an appropriate network architecture, i.e. choosing a network with j hidden layers and h nodes, and redundant structure is typically removed by eliminating connections between nodes by means of pruning methods. Even in regularised networks, the presence of irrelevant inputs may still affect the generalisation performance and so the use of soft or hard feature selection remains desirable. Van de Laar et al. (1997) proposed an input selection employing partial retraining (Van de Laar 1999) to remove in an iterative way the least relevant input variables from a trained neural network. Thodberg (1993) and MacKay (1995) have implemented two different pruning methods within MacKay's Bayesian evidence framework.

By using multiple weight decay parameters, one associated with each input variable, the *automatic relevance determination* (ARD) technique (MacKay 1995; Neal 1996) performs a soft-feature selection by shrinking the weights associated with certain inputs to zero. ARD can then be used to perform a hard feature selection, as the inspection of the hyperparameters leads to selection of a subset of variables which can then be used for training a new network.

Within the statistics literature, model selection is generally synonymous with variable selection, i.e. selecting an appropriate number of variables to be included in a model.

In Chapters 5 and 6, it will be seen that in the search for an appropriate description of the system, both the neurofuzzy and support vector techniques seek a parsimonious description of the system, adapting their structure and number of variables present in the model.

3.9.1 Model selection

In variable selection strategies, it is desirable to determine whether the number of input variables can be reduced without an appreciable loss (or even a gain) in predictive ability, so

[6]The performace of classical methods was seen to be greatly affected by random variability of (small) training samples, whereas VC-based model selection was found to be very insensitive in this respect.

that subsequent inferences may then concentrate on a reduced set of variables. This may be desirable for example in lowering the costs associated with data collection.

From a statistical viewpoint, selection of a smaller set of explanatory variables finds its motivation in variance reduction (i.e. greater confidence) and a parsimonious modelling approach. Since it is unfeasible to search for all possible models in high-dimensional problems, alternatives are to use prior knowledge and select a particular model structure comprising the subset of variables believed to be important, or to apply automatic model search algorithms.

Various (sub-optimal) methods have been proposed for evaluating only a small number of models, generally refining the structure of the model in an iterative manner driven by heuristic searches and/or using the criteria discussed in section 3.8.1. Typical subset selection techniques are generally based on either forward selection, backward elimination, best subsets and stepwise (Hocking 1976; Miller 1984). The deficiencies of stepwise methods have been discussed in several papers (Hocking 1976; Berk 1978). Model construction procedures have been criticised on many occasions, the most common being that neither will assure, except for the very simplest of cases, that the "best" model is identified. The instability of heuristic searches is well recognised (Breiman 1996b). Subset selection procedures suffer from an instability with respect to small perturbations in the data: the removal or addition of a datum or variable from the data set may drastically change the model structure and weights. Variables that appear important, or conversely, appear to have little value based on standard indicators may in fact appear so because of a single influential observation.

The performance of widely used subset selection criteria are assessed in several studies (Breiman and Spector 1992; Breiman 1992). It is of some concern that many criteria have been empirically found to perform poorly for model selection purposes, common pitfalls being in the selection of models of a too high dimensionality, resulting in over-optimistic estimates of the variance. Breiman and Spector (1992) observed that leave-one-out was inferior to 10-fold cross-validation in the selection of the best subset dimension, interpreted as a consequence of instability (Breiman 1996a).

Two main problems are that stepwise procedures imply an order of importance to the variables, an ordering that may be misleading (it is not uncommon to find that the first variable added is superfluous in the presence of other variables), and in the case of early termination, the procedure may fail to detect important (or important combinations of) variables.

The problem of variable selection is often confusing and aggravated when collinearities are present. If the collinearity is inherent in the system being investigated, then one of the variables involved in a near-linear dependency might be eliminated. On the other hand, the collinearity might be induced by the sample size. In this case, subset results may lead to misleading conclusions and model misspecifications.

Derksen and Keselman (1992) presented results from a Monte Carlo study on the frequency with which "authentic" and noise variables were selected by iterative model selection algorithms. Their results indicated that the degree of correlation between the explanatory

variables influenced the frequency with which the former were included in the model, while the number of candidate input variables affected the number of the latter type which were included in the model. Sample size was seen to be of limited importance in determining the number of authentic variables present in the final model. Vafaie and DeJong (1993) also attributed the high variance exhibited by *greedy search* (iterative) algorithms to feature interdependencies, and proposed a more robust approach by employing a genetic algorithm (GA), despite being less efficient when there are few or no intercorrelations between variables.

In the context of feature selection Jensen and Cohen (2000) discuss how in multiple comparison procedures (i.e. iterative/incremental algorithms), the failure to take into account and adjust for the number of candidate refinements considered was responsible for attribute selection errors, overfitting and oversearching[7].

From the above, it emerges that iterative procedures can result in inadequate representations of the data. However it should be acknowledged that iterative model contruction algorithms were not designed to find "best" models; but rather to select subsets from data sets "padded" with extraneous variables.

3.9.2 Regularisation

Regularisation approaches are common to many disciplines, particular forms of which are known as weight-decay in the neural network literature (Krogh and Hertz 1992), ridge regression (Hoerl and Kennard 1970) by statisticians and the Levenberg-Marquardt optimisation method (Gill et al. 1993). Draper and Van Nostrand (1979) discuss the two circumstances in which ridge regression is appropriate, these are in a Bayesian and a restricted least squares formulation, both of which involve the inclusion into the problem of some external information. Ridge regression was originally proposed as a method for solving badly conditioned linear regression problems[8]. In the presence of multicollinearity, regularisation techniques result in biased estimates of the parameters which however exhibit a smaller variance than least squares estimators. The regularisation performed in ridge regression can be interpreted as augmenting the data with dummy observations (Allen 1974), which improve the condition of the learning task[9]. From a computational viewpoint, regularisation involves constraining or penalising the solution of the parametric estimation problem in order to improve generalisation. The regularised cost function is generally formulated as:

$$J_R = J_N + \lambda E_w \tag{3.11}$$

[7]Jensen and Cohen (2000) overfitting *per se* arises because attempting a large number of models leads to a high probability of finding a model that fits the training data purely by chance.

[8]Badly conditioned here reflects difficulties in performing the inverse of $\mathbf{X}^T\mathbf{X}$.

[9]The ridge estimator is equivalent to a least-squares estimator in which the data has been augmented by a fictitious set of points, such that the response is zero and a diagonal matrix is added to $\mathbf{X}^T\mathbf{X}$. The possibility of actually collecting additional data, which would improve the stability of $\mathbf{X}^T\mathbf{X}$ is thus implied.

where E_w is the regularisation (or penalty) term resulting from the introduction of a prior and λ is a (shrinkage or regularisation) parameter used to determine the amount of regularisation performed, effectively controlling the bias/variance trade-off.

The regularisation term may take several different forms, with common choice of priors including p.d.f.'s which make small weights and/or small model output curvature highly probable, corresponding to the regularisers:

$$E_w = E\left[|\mathbf{w}|\right]$$

and

$$E_w = E\left[\left|\frac{d^2\hat{y}}{dx^2}\right|\right]$$

referred to as zero and second order regularisation, respectively. An analytically attractive form of the regulariser is one which is a quadratic function of the weights:

$$E_w = \mathbf{w}^T\mathbf{K}\mathbf{w} \tag{3.12}$$

as for generalised linear models this produces a cost function which can be solved by simple linear optimisation techniques. The solution to the resulting cost function (equation 3.11) is found by differentiating with respect to the weights and setting to zero, giving the solution:

$$\mathbf{w} = \left(\mathbf{X}^T\mathbf{X} + \lambda\mathbf{K}\right)^{-1}\mathbf{X}^T\mathbf{Y}$$

Zero order regularisation[10] results in weight estimates given by the solution of equation 3.9.2 with $\mathbf{K}$ the identity matrix, i.e. $\mathbf{K} = \mathbf{I}$.

In general, regularisation methods may considerably improve upon ordinary least squares regression when the data is sparse and/or noisy, or when the predictor variables are highly collinear (Frank and Friedman 1993). In studying the effects of instability on predictive loss and on the bias and error of the prediction error estimates Breiman (1996a) observed that subset selection attained high predictive losses, while ridge regression failed in problems where there was a mixture of large and small coefficient values. Ridge estimation was found to be superior to variable selection procedures when all but a few of the parameters were nearly zero and the rest large, whereas in the presence of only a few nonzero coefficients, subset selection attained good performance. Thus, the belief that ridge regression is always better than least squares is unjustified. Draper and Van Nostrand (1979) condemned a mechanical application of ridge regression and Smith and Campbell (1980) provide a strong critique on its use. Marquardt and Snee (1975) suggest that it is only appropriate to use variable selection in the presence of multicollinear data after having biased the estimators. Breiman (1995) proposed the nonnegative garrote as an intermediate technique to subset selection and ridge regression.

Although ridge regression was not designed for the purpose of variable selection, it should be realised that there is an inherent deletion of variables (namely, those for which the coefficients are shrunk to zero by λ).

[10]Known as ridge regression in statistics and weight-decay in neural networks.

3.10 Bayesian methods

In section 3.6 Bayesian estimation was seen to provide powerful technique for performing statistical inferences. Central to Bayesian methods is the assumption that alternative models and their competing assumptions are often plausible. Rather than making an estimate based on a single model, several models can be considered and an estimate obtained as the weighted average of the estimates given by the individual models. Hence, in Bayesian model averaging, priors are attached to the models rather than just the model parameters.

The correct Bayesian approach is to make use of the complete set of models, where the predicted outputs are obtained by performing a weighted sum over the predictions of all the models. More probable models therefore contributing more strongly to the predicted output. However, in certain systems it may be essential to recognise that different interpretations describe the data equally well and that procedures which force a single choice of model may be potentially misleading. Bayesian model averaging does not lead to a simple model. This may not be important for certain applications but is when description and interpretation of the model structure is desired. Models with low posterior probabilities may be discarded to keep the problem manageable and then a weighted sum of the remaining competing models is taken. Model expansion approaches, advocated by Draper (1995) and the concept of an Occam's window (Madigan and Raftery, 1994) may be pursued.

3.10.1 Evidence framework

The *evidence framework* was first proposed by Gull (1989) as an approximate method for performing Bayesian inference, and has been subsequently applied to neural networks (MacKay 1992a; MacKay 1992b) providing a unifying theoretical treatment of learning in neural networks. Its practical benefits include principled methods for determining optimal regularisation coefficients and methods for both soft feature selection and model comparison[11]. It also provides a framework for using committees of networks.

The maximum evidence framework looks for the simplest model which can be used to represent the data and so Occam's razor is implicit in this approach. While the full Bayesian method produces a posterior distribution for the weights (or a predictive distribution for the outputs) which can be very complicated, the evidence framework relies on the approximations that the *hyperparameters*, α and β, are optimised rather than integrated over and that

[11] In Bayesian model selection, the evidence is obtained by multiplying the best fit likelihood by the Occam factor. In the evidence framework, the Occam factor is used as a measure of complexity of the model, but unlike the VC dimension it is related to the complexity of the predictions that the model makes and is therefore dependent not only on the number of data points and parameters in the model but also on the prior probabilities that the model assigns to them. However, in assessing the performance of the evidence framework on a number of classification problems, Penny and Roberts (1998a) came to the conclusion that the use of the evidence in model selection was only tenable when the number of training examples exceeded the number of network weights by a factor of five or ten.

the posterior distribution for the weights is approximated by a Gaussian, centred on a (local) maximum of the distribution, thus making analytical derivations tractable, so that predictions can be summarised by a point estimate and corresponding error bar.

The evidence approach has been questioned by several researchers for not being statistically correct since the correct Bayesian treatment for the hyperparameters involves *marginalisation*, i.e. integration over all possible values. In the evidence framework, this integration is approximated and the hyperparameters are fixed to their most probable values. Methods which integrate over the hyperparameters analytically have been proposed (Buntine and Weigend 1991; Wolpert 1993; Williams 1995), and the approximation strategies of Buntine and Weigend (1991) and MacKay (1992a) for handling hyperparameters have been compared in (MacKay 1999).

Empirically, the correlation between the evidence and generalisation error is often good, but a theoretical connection between the two is not established.

3.11 Characterising uncertainty

From previous sections it emerges that there are two sources of uncertainty: structural and parametric. The uncertainty associated with parameter estimates and how this propagates through to the model predictions $\hat{y}$ should be understood. For the MLR model specified in equation 3.1, the variance of the estimated regression parameters $\mathbf{w}$ can be estimated from the variance-covariance $(\mathbf{X}^T\mathbf{X})$ matrix:

$$\sigma_w^2 = \hat{\sigma}^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tag{3.13}$$

The uncertainty in the weights are obtained from multiplying the variance estimate by the diagonal elements of $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$, from which a simple statistic such as

$$\tau_k = \left| \frac{w_k}{\sigma_{w_k}} \right| \tag{3.14}$$

may be informative, and in the case of a small value should draw our attention towards understanding the source of such uncertainty.

Structural uncertanty which will be greatest in small sample sizes is often not accounted for, with inferences proceeding as if the structural and error distribution assumptions were known to be correct. Draper (1995) discusses the propagation of model uncertainty, in terms of Bayesian approaches that are becoming feasible through recent computational advances.

It is common practice to use computationally intensive model selection algorithms to search for a single optimal model from a large number of candidates. Chatfield (1995) discusses model uncertainty and model selection biases which result from data-dependent specification searches. Model selection biases result when a model is formulated and fitted to the same data, resulting in biased and over-confident inferences. Bayesian model averaging approaches then provide a natural way of accounting for model uncertainty.

MacKay (1995) notes that the practice of deliberately using a simple model when data is sparse is one of the reasons for dangerously overconfident predictions.

3.11.1 Error bars

In addition to estimating the overall generalisation error, it is often useful to be able to estimate the accuracy of the model's predictions on individual cases, through means of prediction versus target scatterplots, augmenting the model estimates with confidence or prediction intervals. Error bars are a means to represent model output uncertainty with respect to an input. In regions where the training data is noisy and/or sparse, error bars should reflect uncertainty and therefore should be large. Conversely, in regions where training data is populating the input space well, a good approximation to the data can be inferred and a relatively high confidence in the output can be placed, resulting in small error bars on the prediction. Thus, derivation of error bars is a valuable property of the model, giving a measure of confidence of how well the model performs in different regions of the input space. For the MLR model, the variance of the estimate of the mean response, $\hat{y}_i$, can be shown (Neter et al. 1996; Penny and Roberts 1998b) to be given by:

$$\sigma_{\hat{y}}^2 = \hat{\sigma}^2 \mathbf{x}_i^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{x}_i \tag{3.15}$$

or expressed as a function of the variance-covariance matrix of $\hat{\mathbf{w}}$:

$$\sigma_{\hat{y}}^2 = \mathbf{x}_i^T \sigma_w^2 \mathbf{x}_i$$

The uncertainty associated with the prediction from a new observation, $\mathbf{x}_i$, is given by:

$$\sigma_{y_i}^2 = \hat{\sigma}^2 + \hat{\sigma}^2 \mathbf{x}_i^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{x}_i \tag{3.16}$$

then the standard deviation σ_{y_i} can be directly interpreted as the magnitude of the error bar.

3.12 Knowledge extraction

It is often desirable to infer qualitative knowledge as well as quantitative estimates.

If $\mathbf{X}$ has orthogonal columns, the effects of individual variables are clear, as the tasks of estimation and model selection are greatly simplified. Unfortunately with undesigned experiments the columns of $\mathbf{X}$ are rarely orthogonal and there may be the presence of near-linear dependencies amongst the explanatory variables. In such cases, it is generally difficult to assess the effects of individual variables on the output response. In such cases, as will be discussed in Chapter 4, an eigenvalue examination should be an integral part of the regression analysis and direct towards suitable "remedies" (e.g. subset selection, regularisation methods, data augmentation). Collinearity, also discussed in the following chapter may be responsible for unnecessarily high variances and nonsensical values in some of the regression coefficients. The resulting model estimates may be unreliable, especially outside the immediate neighbourhood of the training data.

3.12.1 Input relevance

The notion of *relevance*[12] is subjective. For example, the conclusion that variables omitted by subset selection procedures are not relevant and do not have a dependency with the response variable may be unsatisfactory and potentially misleading. In iterative model construction algorithms the order of inclusion/deletion of variables in a model should be understood not to relate to variable importance, as it is possible that the first variable included in a forward selection procedure is the first variable deleted in a subsequent backward elimination procedure.

Within the neural network literature, several measures have been proposed to give an indication of the contribution of each input to the response (Garson 1991; Milne 1995; Gedeon 1997), based on an analysis of weight magnitudes (sums of products of normalised weights). Sarle (1997) with a simple example has however illustrated how these measures can lead to incorrect conclusions.

Wichmann and Bartlett (1997) have proposed the use of a general regression neural network as a means of ranking inputs with respect to their importance (based on LOOCV), which was then used as a subset selection technique from which subsequent analyses (e.g. network training) can be performed. Van de Laar et al. (1999) propose partial retraining of the neural network in order to determine the relevance of inputs of a trained model.

The most significant method developed for the purposes of inferring the relevance of the input variables, remains the ARD framework (MacKay 1995; Neal 1996). Interpretation of the hyperparameter values allows the influence of a particular input to be assessed. Importantly, ARD avoids the use of iterative pruning strategies.

However, Penny and Roberts (1998a) found ARD to be useful in networks with many hidden units and in data sets containing many irrelevant variables, while Kandola et al. (1999) have observed that ARD results tend to be sensitive to the initial values of the hyperparameters.

3.13 Prior knowledge

Prior knowledge can be effectively employed as a means of reducing the variance of the estimates (if this reduction is not overshadowed by a comparable increase in the bias). Presceening the input variables, initialising a model or formulating a set of fuzzy rules may aid in directing the iterative model construction procedures towards better described regions of the model space. In section 3.6, it was seen that in Bayesian inferencing, the prior $P(\mathbf{w})$ p.d.f. is used to encode prior knowledge about the parameter vector $\mathbf{w}$, before the data is observed.

[12] Van de Laar et al. (1999) defines the relevance $\mathcal{R}_i$ of an input x_i as the difference in performance on a given task with and without x_i, given all other input variables:

$$\mathcal{R}_i = \mathcal{P} - \mathcal{P}_{(i)}$$

The blind use of priors should however be viewed as a black box approach to regression problems (Hocking 1983b).

3.14 Data mining and knowledge discovery in databases

The requirement for appropriate techniques able to extract meaningful knowledge from large databases emerges from the fact that many companies are confronted with the challenge of handing an ever-increasing amount of data. As a result it is becoming more difficult to access the information that is contained in the data. Knowledge discovery in databases (KDD) is an emerging multi-disciplinary field of research that combines techniques from machine learning, pattern recognition, statistics, exploratory data analysis, expert systems, databases, and visualisation to automatically extract concepts, interrelationships, and patterns of interest from large databases. The basic tools used to extract patterns from data are referred to as *data mining* methods, while the process surrounding the use of these tools (including pre-processing, selection, and transformation of the data) and the interpretation of patterns into "knowledge" is the KDD process.

A commonly used definition of data mining, given by Fayyad et al. (1996) is:

> *the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.*

The knowledge discovery process involves a series of iterative stages:

1. **Data selection** Develop an understanding of the application domain, the relevant prior knowledge, and the objectives of the analysis. Selecting a data set or focusing on a subset of variables or data samples, upon which knowledge discovery is to be performed.

2. **Data cleaning** Basic operations such as the removal of outliers, deciding strategies for handling missing data fields.

3. **Data representation and coding** Find useful features to represent the data, using dimensionality reduction or transformation methods to reduce the effective number of variables or to find invariant representations for the data.

4. **Data mining** Choosing appropriate methods to be used for searching for patterns in the data.

5. **Validation** Interpreting the knowledge inferred, which may involve visualisation of the data and patterns inferred.

Steps 1-4, can all be viewed as data *pre-processing* steps and will be discussed in the following chapter.

The relative importance of prediction and description for particular data mining applications can vary considerably. In the context of KDD, description tends to be more important than prediction, making the Bayesian model averaging approach less appealing, and is in contrast to pattern recognition and machine learning applications where prediction is often the primary goal. Elder and Pregibon (1996) give good statistical perspectives of KDD and review developments that have characterised the last few decades.

3.15 Conclusions

This chapter has introduced the general concepts, assumptions and problems in the modelling of empirical data. The major conceptual limitation of all regression techniques is that one can only ascertain possible relationships, but never be sure about underlying causal mechanisms. Subsequent chapters will discuss in greater depth some of the topics introduced, such as model selection, regularisation and data pre-processing. In any regression analysis it is useful to understand the nature of the data and the limitations that a poor data set imposes. This will be the subject of the next chapter.

Chapter 4

Data Pre-Processing, Sensitivity Analysis and Conditioning Diagnostics

4.1 Introduction

Previous chapters have emphasised that a significant objective of regression analyses is data interpretation. It has been discussed how drawing inferences from observational data may be affected by several potential sources of misinterpretation. As such, it is important to determine the statistical properties and integrity of the available data. In this chapter statistical techniques, namely sensitivity measures and conditioning diagnostic procedures, are reviewed for such purposes.

It seldom occurs that data obtained from a process or retrieved from large databases, is readily amenable to a statistical analysis. Often the form in which the data is obtained prevents a straightforward knowledge extraction. Identifying missing values, inconsistencies, outliers as well as defining sensible training sets and appropriate transformations of the data using appropriate forms of pre-processing will enhance the structural integrity of subsequent inferences. Data pre-processing is then an important but often neglected stage in the modelling process, the nature of which is determined by the integrity and quality of the data, the complexity and nature of the system, and the type of inferences sought.

Sensitivity and diagnostic procedures are concerned with the factors which characterise the conditional match of the model to the data. It seems desirable when a single model is determined from the data to see how sensitive any conclusions are to the model assumptions. These not only determine whether a collinearity or conditioning problem exists but can be used to identify the number of variates involved, determining the effect of individual or a subset of the observations. Detecting the presence of dependencies amongst the explanatory variables will help to understand the nature of the data and dependencies inferred.

4.2 Data pre-processing

From section 3.14 it was seen that there are several stages involved in pre-processing the data for a successful application of data mining (modelling) techniques and subsequent knowledge extraction. The objectives of the pre-processing stage being to detect and rectify problems and/or to apply transformations to the data.

When processing large quantities of data it becomes essential to automate procedures, to avoid possible error-prone manual sifting of the database entries. The following sections discuss the different forms of pre-processing, drawing examples from the analyses of subsequent chapters.

4.2.1 Data set selection

Selection of a data set comprising a representative set of explanatory and response variables requires domain knowledge as large commercial databases may store records pertaining to different systems (e.g. age-hardenable, not age-hardenable alloys), product forms (cast, wrought), processing (single, double ageing sequences), etc. This simple pre-screening of the variables may reduce the risks of inferring non-causal dependencies.

It is seldom that a first analysis answers all the questions sought, since in the first definition of the data set significant variables may be missing. Unsatisfactory results may indicate the absence of important explanatory variables. In some cases, it may be possible to go back and retreive these from the database, in others a new set of observations may be required. In certain systems though, important physical variables will remain inaccessible.

In these circumstances it may be possible to characterise unmeasured features, by one or more of the variables that are contained in the data set. Considering the situation where process understanding suggests a set $\mathbf{x}_a$ of explanatory variables to be related to the response variable, y. However, when some variates in $\mathbf{x}_a$ are inaccessible or characterisation is problematic, this may require the use of a less optimal explanatory vector, $\mathbf{x}_b$, (which may or may not comprise some of the variates present in $\mathbf{x}_a$), whose characterisation is less problematic.

The following example illustrates the above situation: it is well known that grain size information (e.g. morphology, distribution, etc.) is significant in strengthening certain alloy systems. Hence, a direct quantification of such microstructural features will have significant explanatory abilities for the tensile properties. However, while in scientific studies such features often form part of the alloy characterisation, in industrial settings their measurement is seldom accomplished. The variance in the material's tensile properties will then have to be explained through other variables characterising the microstructure, which may exhibit both higher experimental scatter (and a less direct relationship to the output value). The question is then whether for example compositional information (e.g. weight percentages of Fe, Mn, Si) together with processing variables can adequately characterise the microstructure, i.e. the degree of recrystallisation, which will then affect the tensile properties.

As discussed in section 3.4, the development of a predictive model that generalises sensibly relies critically on the availability of a set of both representative and informative features. In view of the suboptimality of iterative model construction algorithms, it is generally advisable to pre-screen variables present in the original data set/database and retain only those which according to prior knowledge are expected to be related to the response variable.

4.2.2 Data cleaning

For meaningful analyses to be performed, the data set should be free of nonsensical values or inconsistencies. It is the purpose of data cleaning routines to sift the data set for erroneous entries, typically introduced during data gathering. While some forms of data corruption may be easily detected by a simple inspection of the data distributions and removed prior to the analysis (e.g. domain inconsistencies, missing or zero entries, etc.), others arise only after a first analysis has been undertaken (e.g. from inspection of the residuals), and are associated with errors in the information storage (e.g. typos, duplicate records). The removal (de-duplication) of replicate cases is advisable in frameworks in which the complexity of the approximating function is bound by the sample size.

In certain cases, it may be appropriate to remove certain outliers. Learning algorithms (see Chapter 6) based on ideas drawn from robust statistics (Huber, 1981) have been developed, offering robustness in the presence of outliers.

4.2.3 Data representation and coding

An appropriate representation of the variables may enhance the interpretability of the data, facilitating the data mining process and knowledge extraction.

In certain cases, it will be necessary to re-define the data format: in *flattening*, a data field with cardinality c is replaced by c fields, as illustrated in Table 4.1 (see also Chapter 9). Table 4.2 shows two alternative ways of coding a categorical variable (i.e. label). The advantage of the latter is in representing the labels with a single dummy variable (whilst the former uses three dummy variables), but may impose an inappropriate ordering of the attributes. It should be recognised that the usual definitions of smoothness do not apply to variables that assume unorderable categorical values. B-splines (section 5.3.1) however provide a way of coding c ordinal variables into fewer than c variables, intrinsically performing a dummy coding (whereby the attributes are defined on a fuzzy variable) (van Rijckevorsel 1988). Concepts of smoothness appropriate for categorical variables are discussed by Friedman (1993).

A simple alternative (and sometimes more meaningful approach) for regression analyses in the presence of categorical variables is to partition the data set into several smaller subsets to be modelled independently, or to use an appropriate algorithm (Breiman et al. 1984).

PLATE NO	···	HEATTYPE	···	HEATTIME
47294	···	⋮	···	⋮
47294	···	PT	···	102
47294	···	⋮	···	⋮
48239	···	H	···	500
48239	···	⋮	···	⋮
48239	···	ST	···	470
48239	···	⋮	···	⋮
48239	···	PT	···	104
⋮	···	⋮	···	⋮
49610	···	H	···	200

PLATE NO	···	H	ST	PT
⋮	···	⋮	···	⋮
47294	···	520	395	102
48239	···	500	470	104
48239	···	700	472	97
⋮	···	⋮	···	⋮

Table 4.1: Example of flattening, whereby the information contained in the HEATTYPE and HEATTIME fields is replaced in the new set of defined fields H, ST and PT.

TL	LT	SL
1	0	0
0	1	0
0	0	1

(a)

DIRECTION
-1
0
+1

(b)

Table 4.2: Examples of: (a) 1-of-C coding of a categorical variable DIRECTION that assumes three different labels; (b) representing (ordering) the labels of the same variable, by means of a single variable (LT=1, TL=0, SL=-1).

4.2.4 Data transformation

There may be several reasons why it is appropriate to perform a transformation of the data set prior to the regression analysis. These may be motivated by the inclusion of domain knowledge, representational issues, data conditioning, or limitations of the learning paradigm. The following sections discuss transformations which have been part of the analysis of subsequent chapters.

4.2.4.1 System transformations

Statistical techniques will be most effective when combined with prior system knowledge. Often a non-linear transformation of a variable can increase the smoothness of the functional relationship as well as representing dependencies in a form simpler to validate (e.g. $\log x_i$, $\sqrt{x_i}$). It can be argued that modelling techniques should not require such a "linearisation" of the data, however limitations imposed by different modelling techniques (e.g. curse of dimensionality, complexity control, etc.) should be recognised. Multivariable transformations,

for instance of the effects of certain variables, x_1 and x_2, may be more appropriately expressed in terms of their ratio x_1/x_2 and/or their sum $x_1 + x_2$.

Transformations induced by physical-based understanding (Chapter 8), can augment the data or define more informative features, which may lead to improved generalisation performance.

4.2.4.2 Standardising the data

Standardising[1] the input and/or target variables tends to improve the numerical condition of the optimisation problem, making the training process better behaved.

Standardising (or normalising) the output variable and the k set of variables to have zero mean and unit variance (also referred to as the *correlation transformation*) is given by:

$$\frac{y_i - \mu_y}{\sigma_y} \qquad \frac{x_{ik} - \mu_{x_k}}{\sigma_{x_k}}$$

where μ_y and μ_{x_k} are the means of the quantities and the standard deviations σ_y and σ_k are given by:

$$\sigma_y = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \mu_y)^2} \qquad \sigma_{x_k} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_{ik} - \mu_{x_k})^2} \qquad k = 1, \ldots, p$$

By forming the matrix $\mathbf{X}^T\mathbf{X}$, and the vector $\mathbf{YX}$, the simple correlation coefficients r_{ij} between explanatory variables and those between the response variable and explanatory variables respectively, are readily summarised:

$$\mathbf{r}_{XX} = \begin{bmatrix} 1 & r_{1,2} & \cdots & r_{1,p-1} \\ r_{2,1} & 1 & \cdots & r_{2,p-1} \\ \vdots & \vdots & & \vdots \\ r_{p-1,1} & r_{p-1,2} & \cdots & 1 \end{bmatrix}, \quad \mathbf{r}_{YX} = \begin{bmatrix} r_{y,1} \\ r_{y,2} \\ \vdots \\ r_{y,p-1} \end{bmatrix}$$

The standardised weights can then be shown to be:

$$\mathbf{w} = \mathbf{r}_{XX}^{-1}\mathbf{r}_{YX}$$

which allow a meaningful comparison of the magnitude of the individual weights. When the variates are linearly related, the correlation coefficient will be a measure of the degree of relationship present. Limitations and misinterpretations of the simple correlation coefficients are discussed in (Neter et al. 1996). Inspecting the distribution of the data will allow a better interpretation of the correlation coefficients.

[1] Standardising a vector usually refers to subtracting a measure of location (centering) and dividing by a measure of scale (scaling).

For neurofuzzy networks (Chapter 5), the data is scaled to lie in the interval $[-1, 1]$, by means of the following standardisation:

$$x_{norm} = (x - x_{centre})\,(2/x_{width})$$

with:

$$\begin{aligned} x_{centre} &= (x_{max} + x_{min})\,/2 \\ x_{width} &= (x_{max} - x_{min}) \end{aligned}$$

In the application of the support vector methods (Chapter 6) the data is scaled to lie in the [0,1] interval.

Scaling can help in controlling numerical problems associated with the inversion of $\mathbf{X}^T\mathbf{X}$, as well as expressing the data in a form that lends itself to more straightforward interpretation. Centering is advised as it eliminates potential collinearities arising from the origins of the variables. Together then, centering and scaling are viewed as removing "non-essential" ill-conditioning in $\mathbf{X}$, reducing the variance inflation in the coefficient estimates.

4.3 Collinearity and short data

Together, *collinearity* and *short data* constitute the sources of *data weakness* in the estimation of linear models by least squares. Data weaknesses are characteristics of the data that rob them of the information needed for statistical analysis to proceed in some dimensions with adequate precision (Belsley 1991).

Intuitively, the potential harm that results from collinear data can be understood by realising that a collinear variate, being nearly a linear combination of other variates does not provide information that is very different from that already inherent in these others. In many systems, collinearity is a natural flaw in the data set, resulting from physical or processing constraints and as such it can be difficult to infer the separate influence of such explanatory variables on the response. Small sample sizes, unrepresentative data sets, misspecified models and many other factors can contribute to unexpected coefficient signs and magnitudes.

The different characteristics and weaknesses of a data set which can arise as categorised by Belsley (1991) are shown in Figure 4.1:

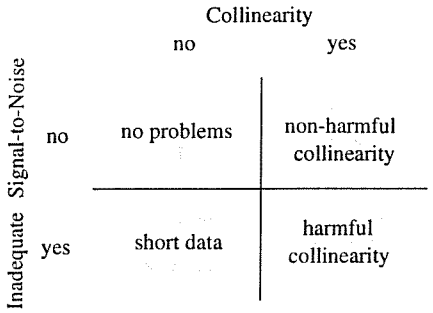


Figure 4.1: Collinearity - short data

I Represents the ideal situation for both system identification and parameter estimation: neither collinearity nor inadequate signal-to-noise are present.

II Collinearity is present but does not result in inadequate signal-to-noise: the ill effects of collinearity are mitigated by the presence of relatively small σ^2, and/or long $||\mathbf{x}_i||$.

III There is no collinearity in $\mathbf{X}$ but an inadequate signal-to-noise is present. In this situation, while data weaknesses exist, collinearity is not the culprit. Poor estimates and predictive abilities are attributable to what Belsley calls short data, and occurs when a variate $\mathbf{x}_i$ has little length, $||\mathbf{x}_i||$, so that the term $w_i x_i$ adds little to the overall determination of y_i.

IV The joint occurrence of collinearity and inadequate signal-to-noise constitutes the situation of *harmful collinearity*.

From the above, collinearity is seen not to be the only problem afflicting the quality of the data. In section 3.11 a simple *signal-to-noise parameter*[2] was suggested (equation 3.14). Parameters associated with estimates that are small in comparison to their variances (e.g. $\tau_k < 2$) are unlikely to be well determined. Low signal-to-noise values can indicate the presence of inflated variances, reducing the power of statistical significance and hypothesis tests. Although useful for detecting the presence of weak data, simple parameters such as τ cannot determine whether the cause is attributable to collinearity or short data. Collinearity diagnostics (section 4.3.4) by contrast cannot detect the presence of weak data, since they ignore the parameters while trying to assess the information inherent in the data (Smith and Campbell 1980), but can determine whether an already detected data weakness is due to collinearity or short data. A detailed treatment of signal-to-noise measures, including conditions establishing when harmful collinearities prevail is discussed by Belsley (1991).

While not affecting signal-to-noise, linear transformations of the data can mitigate ill-conditioning, i.e. they cannot remove data weaknesses but can only alter its form.

[2]It is seen that τ bears a superficial resemblance to the $t-$statistic. Alternatively, a measure defined as the inverse of τ, the *coefficient of variation* can be defined.

4.3.1 Ill-conditioning

Least squares solutions given by equation 3.6 are said to be *ill-conditioned* if small relative changes in the $\mathbf{x}$'s (and y's) can result in large relative changes in the estimates $\mathbf{w}$. The severity of the ill-conditioning can be characterised by $C(\mathbf{X})$, the *condition number* of the design matrix, defined as:

$$C(\mathbf{X}) = \frac{\mu_{max}}{\mu_{min}} \tag{4.1}$$

where μ_{max} is the largest singular value of the spectral decomposition of $\mathbf{X}$ and μ_{min} its smallest singular value[3]. Thus, the degree of ill-conditioning depends on how small the minimum singular value is relative to the maximum singular value. As noted in section 4.2.4.2 standardising the data is seen to improve the condition number.

Often $\mathbf{X}^T\mathbf{X}$ is ill-conditioned (near singular) due to either a poor distribution of input data and/or model mismatch. Ill-conditioning then makes it hard to efficiently determine the optimal weight vector and can make the learning algorithm slow to converge[4].

Ill-conditioning is generally due to excessive degrees of freedom in a model, i.e. redundant parameters. In neurofuzzy networks, a common cause of redundant parameters is poorly excited basis functions which produce zero (or significantly small) diagonals in the autocorrelation matrix.

4.3.2 Sources of collinearity

The nature and sources of collinearity that are typically encountered in the analysis of small, multivariate, observational data have been summarised in several studies (Gunst 1983; Mason and Gunst 1985) as the following:

- Model specification (constraints). The correctness of the model does not preclude the occurrence of collinearities.
- Population characteristics that restrict variate values due to some inherent characteristics of the process. In many cases, this type of collinearity cannot be eliminated and as such,

[3]The singular value decomposition (SVD) of $\mathbf{X}$ is given by:

$$\mathbf{X} = \mathbf{U}\Lambda\mathbf{V}^T$$

where $\mathbf{U}$ is a $[N \times p]$ orthonormal matrix whose columns represent the p eigenvalues associated with the p largest eigenvalues of $\mathbf{X}\mathbf{X}^T$, $\mathbf{V}$ is a $[p \times p]$ orthonormal matrix whose columns represent the eigenvectors of $\mathbf{X}^T\mathbf{X}$, and Λ is a diagonal matrix whose diagonal contains the singular values of $\mathbf{X}$, which are generally assumed to be ordered, i.e. $\Lambda = diag(\mu_1, \ldots, \mu_p)$.

[4]Collinearity is thus concerned with the existence of near linear relationships among a set of variates whereas conditioning is centred on the sensitivity of a given relation to perturbations in the data. Collinearity among the covariates can result in sensitivity (ill-conditioning) in the least-squares estimates to changes in the data, the converse however need not be necessarily true, i.e. ill-conditioning does not necessarily imply a collinear problem. Thus, collinearity and ill-conditioning should not be used synonymously to denote a particular form of data weakness.

can seriously limit regression analyses and subsequent inferences. Typically associated with observational rather that experimental data. Transformations may mitigate this type of collinearity, removing "non-essential" ill-conditioning (section 4.2.4.2).

- Sampling deficiencies of a particular data set, not expected to occur in similar data sets. Due to the substantial biases introduced, subset selection procedures are not to be recommended, the most satisfactory solution being data augmentation, i.e. collecting more data or including a priori knowledge e.g. Bayesian and biased estimation.

- Overdefined models, where the number of parameters exceed the number of training samples ($N < p$). Such collinearities introduced are a direct consequence of the rank of $\mathbf{X}$.

- Collinear-influential observations. Either removing the outliers or using robust regression methods will mitigate or even remove the collinearity problem.

The above categorisation of the sources of collinearity are not mutually exclusive. An inadequate sampling may yield an outlying observation, thus inducing the collinearity problem. Some of the problems described above can be eliminated by simply re-defining the training set, whilst others necessitate use of appropriate statistical techniques.

4.3.3 Ill-effects of collinearity

Although collinearities need not always be harmful or hinder the regression analysis, they may still have some adverse effects regardless of problems attributable to short data, particularly in a qualitative interpretation of regression results. In the presence of collinearity among the columns of $\mathbf{X}$, the common interpretation of regression coefficients as indicating the change in the expected value of the response variable when the given predictor is varied while all other predictor variables are held constant is not fully applicable. Other "ill-effects" (Belsley 1991; Neter et al. 1996) that should be acknowledged include:

1. Large sampling variability of the estimated regression coefficients, when a particular variable and/or observation is added or removed. As noted previously, the presence of collinearity can severely reduce the power of standard statistical tests of significance.

2. Marginal contribution of a particular covariate in reducing the MSE.

3. Estimated regression coefficients become more imprecise as more inputs are included in the model, consequently large variance (i.e. wide confidence intervals) are observed for the regression coefficients. Parameter estimates may result in an algebraic sign that is opposite of that expected from physical understanding (e.g. increasing instead of decreasing trends inferred).

4. Stability in the precision of the estimated mean response, despite the inflated variance in the regression coefficients.

4.3.4 Diagnosing collinearity

Some of the ill-effects discussed above are however not necessarily due to the presence of collinearity. If collinearity is to be blamed, it must it be shown to be adversely affecting the estimates. As Gunst (1983) observes, the unexpected occurrence of coefficient estimates and large standard errors is not necessarily due to collinearities among the predictor variables.

In order to assess conditioning meaningfully, the data must be in a form that possesses structural interpretability itself. Belsley (1984) argued that mean-centering can remove information needed to assess conditioning correctly, masking the role of the constant term in any underlying near-dependencies, producing misleading diagnostics.

The absence of high correlations among covariates cannot be viewed as evidence of the absence of a collinearity problem. A single measure will rarely fully characterise the nature and effects of collinear predictor variables, and some measures will be more interpretable and informative than others (Belsley et al. 1980). Some measures will be appropriate for assessing the sensitivity of least squares estimates to minor perturbations of the input data (condition indices), others more readily measure the effects of collinearity on the variances of the estimators (variance inflation factors), and others still help to identify the nature of the collinearities (variate correlation coefficients, eigenvalues and eigenvectors of $\mathbf{X}^T\mathbf{X}$). As a result, a large number of studies have been directed towards assessing the relative merits (and deficiencies) of various diagnostic measures. The following sections present some useful conditioning diagnostics.

4.3.4.1 Variance inflation factors

Hoerl et al. (1986) draw attention to the trace of $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$ as one of the best measures of the degree of collinearity[5]. The variance inflation factor of the k^{th} regression coefficient, VIF_k is defined as the diagonal element of the matrix $\mathbf{r}_{XX}^{-1}$[6]:

$$VIF_k = \mathbf{r}_{xx}^{-1} \tag{4.2}$$

then for the k^{th} term in a MLR model, equation 3.13 can be written as:

$$\sigma_{w_k}^2 = \hat{\sigma}^2 VIF_k \tag{4.3}$$

The variance, or confidence interval of a regression coefficient is therefore directly proportional to the collinearity of x_k with the other x's in the model, as measured by VIF_k (Montgomery

[5]The k^{th} diagonal entry in $\mathbf{H} = \mathbf{X}^T\mathbf{X}$ is the variance for x_k. The sum of the variances on the diagonal of $\mathbf{H}$, the *trace* of the matrix, $tr\ (\mathbf{H})$ will be the total variance.

[6]Provided the columns of $\mathbf{X}$ are scaled to have unit norm.

and Peck 1992). Each VIF is then a measure of the collective impact of the simple correlations on the variance of each weight in the model and the largest VIF can be used to indicate the severity of the collinearity[7]. Berk (1977) shows that the VIF is a lower bound on the condition number of $\mathbf{X}$. Scaling is seen not to affect the VIF's, whilst centering will.

The limitations of the VIF's for detecting collinearity are discussed in (Belsley 1991; Neter et al. 1996).

4.3.4.2 Collinearity diagnostic based on the eigensystem

When the source of the collinearity is not clear from the VIF's, the eigensystem of the correlation matrix can be studied. In particular, the eigenvalues can be used to form a set of condition indices that allow determination of the strength and number of near dependencies, with both eigenvalues and eigenvectors forming a set of variance-decomposition proportions that allow determination of variate involvement.

The condition indices (determining the number of collinear relations) As each near linear dependency will give rise to a *small* singular value of $\mathbf{X}$ (or eigenvalue of $\mathbf{X}^T\mathbf{X}$), the number of collinear relations present can be determined by establishing the number of *small* singular values present. In section 4.3.1, it was seen that the severity of the ill-conditioning can be characterised by the condition number. However, used by itself it has limited value as a collinearity diagnostic (Snee 1983) since it applies to the whole model and does not identify which terms are contributing to the collinearity.

It is useful to determine the set of *condition indices* of $\mathbf{X}$, defined as:

$$\eta_k \equiv \frac{\mu_{max}}{\mu_k} \quad k = 1, \ldots, p \tag{4.4}$$

where the μ_k's are the singular values of $\mathbf{X}$. Thus, there will be as many near dependencies among the columns of $\mathbf{X}$ as there are large condition indices.

The variance-decomposition (determining variate involvement) Performing the singular value decomposition, $\mathbf{X} = \mathbf{U}\Lambda\mathbf{V}^T$, the variance-covariance matrix and the corresponding weight variances can be written as:

$$\sigma_w^2 = \hat{\sigma}^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1} = \hat{\sigma}^2 \mathbf{V}\Lambda^{-2}\mathbf{V}^T \tag{4.5}$$

From this the variance of the k^{th} regression coefficient w_k can be written as:

$$\sigma_w^2 = \hat{\sigma}^2 \sum_{j=1}^{p} \frac{v_{kj}^2}{\mu_j^2} \tag{4.6}$$

[7] A maximum VIF value in excess of 10 is often taken as an indication that collinearity may be unduly influencing the least squares estimates (Snee 1983; Neter et al. 1996), larger values indicating that it is appropriate to consider the use of biased regression techniques.

where the μ_j's are the singular values of $\mathbf{X}$ and $\mathbf{V} \equiv (v_{ij})$. Thus, the variance associated with the k^{th} weight is decomposed into a sum of components, each associated with one (and only one) of the p singular values of $\mathbf{X}$. In the presence of *small* singular values, a high proportion of the variance of two or more coefficients will be concentrated in components associated with the same (small) singular value, providing evidence that the variates corresponding to those coefficients are involved in the near dependency corresponding to the particular singular value. Thus, not all the VIF's will be equally affected by a small eigenvalue, but the effect will be dependent on the v_{ij} values (Gunst 1983).

Belsley (1991) determines the variance-decomposition proportions by letting:

$$\phi_{kj} \equiv \frac{v_{kj}^2}{\mu_j^2} \qquad \text{and} \qquad \phi_k \equiv \sum_{j=1}^{p} \phi_{kj} \qquad k = 1, \ldots, p$$

Then the variance-decomposition proportions are given by:

$$\pi_{jk} \equiv \frac{\phi_{kj}}{\phi_k} \qquad k, j = 1, \ldots, p$$

The $(k, j)^{th}$ *variance-decomposition proportion* is defined as the proportion of the variance of the k^{th} regression coefficient associated with the j^{th} component of its decomposition obtained in equation 4.6. These proportions can be summarised in the $\Pi-$matrix, where each row corresponds to a singular value (μ_j), or equivalently its corresponding condition index (η_j). Generally, the rows are ordered so that condition indices are in increasing order. Note that the columns of π's should sum to unity.

Condition	Proportions of			
Index η	$\sigma^2_{w_1}$	$\sigma^2_{w_2}$	$\cdots$	$\sigma^2_{w_p}$
η_1	π_{11}	π_{12}	$\cdots$	π_{1p}
η_2	π_{21}	π_{22}	$\cdots$	π_{2p}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
η_p	π_{p1}	π_{p2}	$\cdots$	π_{pp}

4.3.5 The hat matrix

From section 3.6 the vector of output estimates can be written as:

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y} \tag{4.7}$$

where:

$$\mathbf{H} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T \tag{4.8}$$

is termed the *hat* matrix. This can be interpreted as the orthogonal projection operator on the column space of $\mathbf{X}$ (Weisberg 1995). The *leverage* of the i^{th} observation may be defined as the magnitude of the derivative of the i^{th} predicted value with respect to the i^{th} response value ($\partial\hat{y}_i/\partial y_i = h_{ii}$). Then the diagonal elements of the hat matrix,

$$h_{ii} = \mathbf{x}_i \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \mathbf{x}_i^T \tag{4.9}$$

directly indicate the sensitivity of the i^{th} estimate, $\hat{y}_i$, to changes in the i^{th} observed response, y_i. Similarly, elements h_{ij} of $\mathbf{H}$ have a direct interpretation as providing a measure of leverage (or *influence*) exerted on $\hat{y}_i$ by y_j. Two significant properties of $\mathbf{H}$ are:

$$0 \leq h_{ii} \leq 1 \qquad \text{and} \qquad \sum_{i=1}^{N} h_{ii} = p$$

from which it is seen why Hoaglin and Welsch (1978) suggest an observation to have high leverage if $h_{ii} > 2p/N$.

Considering the case in which one of the $h_{ii} = 1$, then $\hat{y}_i = y_i$, i.e. the model has learnt the i^{th} datum value exactly, effectively allocating a parameter to the i^{th} observation.

The h_{ii} values may be used to identify hidden extrapolations for unseen $\mathbf{x}_p$ vectors: if the leverage of these is within the range of h_{ii} values of the training data set, no extrapolation is involved.

4.3.6 Analysis of residuals

The analysis of residuals can provide useful information, regarding the adequacy of the model and the constancy of the variance of the error terms as well as providing information about outliers. The *residual* is the difference between the observed value y_i and the fitted value $\hat{y}_i$, $e_i = y_i - \hat{y}_i$, not to be misinterpreted as the unknown, true error.

Semistudentised (equation 4.10) and studentised (equation 4.11) residuals (Belsley 1991; Neter et al. 1996) are a means of standardising the residuals, observations with large values of r_i will be outliers in the Y-space.

$$e_i^* = \frac{e_i}{\sqrt{MSE}} \tag{4.10}$$

$$r_i = \frac{e_i}{\sigma_{e_i}} = \frac{e_i}{\sqrt{MSE\,(1 - h_{ii})}} \tag{4.11}$$

Modified residuals can allow some understanding as to the effect of the inclusion of a variable in a regression model. At the k^{th} stage of an iterative model construction, a plot of the residual e_i as a function of the input variables will elucidate residual information:

$$r_{x_j|k} = \mathbf{y} - \mathcal{M}_k(\mathbf{x}, \mathbf{w})$$

where $\mathbf{y}$ is the vector of observed responses and $\mathcal{M}_k(\mathbf{x}, \mathbf{w})$ the corresponding vector of estimates obtained from the model.

4.4 Outliers and influential cases

From the above, it is seen that a linear regression model may be interpreted as a smoothed representation of the data, that captures global and essential features of the data. However as Chatterjee and Hadi (1986) note, this view is not always appropriate since salient features of the model can be dominated by a single observation, particularly when the data is poorly distributed and conditioned. The analysis of outlying observations can lead to diagnosing model inadequacies, such as the omission of an important variable and/or other model misspecifications. In some cases influential observations may contribute more to the understanding of the process under study than the rest of the data combined. An *influential observation* (Mason and Gunst 1985) refers to an outlier whose inclusion in a data set substantially changes regression coefficient estimates, predicted responses, or the results of inferential procedures. An observation may influence (some or all) regression results if it is an:

- outlying response value,
- high leverage point in the factor space,
- combination of both.

When reasoning about the nature of outlying cases background knowledge should be used in order to distinguish between outliers that represent interesting/unusual cases, i.e. "phenomena of interest" (Wu et al. 1997) and outliers induced by large experimental/observational errors (including outliers resulting from typos).

When two or more explanatory variables are included in the regression model, identification of outlying cases by simple graphical means becomes difficult and appropriate sensitivity measures should be used. Measures of influence which study whether a particular case is influential are typically based on omitting a single case and measuring the influence this has on the quantities of interest, e.g. change in individual or all response estimates, $\hat{y} - \hat{y}_{(i)}$, parameter estimates $\mathbf{w} - \mathbf{w}_{(i)}$. Chatterjee and Hadi (1986) review measures proposed in outlier analysis. Measures based on the change in the condition of the hat matrix resulting from the deletion of an observation can discover collinearity-creating or masking observations. Hoaglin and Welsch (1978) have drawn attention to the effect of removing a high-leverage observation in increasing the variance of the estimated parameters in the model, proposing a sensitivity analysis based on the h_{ii}'s for detecting high-leverage design points, and studentised residuals r_i for diagnosing discrepancies in the response values. A collinearity-influential case however is not necessarily an influential observation as established by some measures of influence.

4.5 Measures of influence

Because high-leverage observations tend to have small residuals, examination of the residuals alone may not detect influential observations. In such cases analytical measures can be used.

4.5.1 Cook's distance measure

Cook's distance (Cook 1979) measures the influence of the i^{th} case on all N response estimates[8]. This measure has the advantage that it can be calculated without having to repeatedly re-estimate the weights after each deletion, but can be estimated by an algebraic expression involving the h_{ii}'s:

$$C_i^2 = \frac{\sum_{j=1}^{N} \left(\hat{y}_j - \hat{y}_{j(i)}\right)^2}{p\,MSE} = \frac{r_i^2}{p}\left(\frac{h_{ii}}{1-h_{ii}}\right) \tag{4.12}$$

A large value of C_i^2 indicating that the i^{th} case exerts a strong influence on the estimates. From equation 4.12, the i^{th} case can be determined as being influential by having:

- a large residual and only a moderate leverage
- a large leverage with only a moderate sized residual
- both a large residual and a large leverage

While equation 4.12 gives useful information about the model identified, Leger and Altman (1993) argue that despite being computationally demanding, an unconditional approach (specifying an equivalent C_i^2 measure based on re-determining the model with the i^{th} case omitted) would lead to greater insight about the influence of individual observations on the model construction process.

4.5.2 Hadi's overall potential measure

Hadi (1992) proposes the development of an overall potential influence measure, H_i^2, as other influence measures proposed highlight observations which influence a particular inference, failing to identify cases which may influence other least-squares results. Hadi's measure is based on the fact that potentially influential observations are outliers in either the X-space, the Y-space, or both:

$$H_i^2 = \frac{p}{(1-h_{ii})}\frac{d_i^2}{1-d_i^2} + \frac{h_{ii}}{1-h_{ii}} \tag{4.13}$$

[8]Cook's distance measure can equivalently be written in terms of the effect of the deletion of the i^{th} case on the p weights in the model:

$$C_i^2 = \frac{\left(\mathbf{w}-\mathbf{w}_{(i)}\right)^T \mathbf{X}^T\mathbf{X}\left(\mathbf{w}-\mathbf{w}_{(i)}\right)}{p\,MSE}$$

where p is the number of parameters in the model and $d_i^2 = e_i^2/e^Te$ is the square of the i^{th} normalised residual. A comparison between C_i^2 and H_i^2 shows that while the former is a multiplicative function of the residual and the leverage value, the latter is an additive function of the two.

Generally, graphical methods are used to infer the influence of particular cases in relation to others. Cut-off points should be used with caution and values attained from one diagnostic measure should not be compared with others.

If a data set contains a single outlier or influential observation, the problem of identifying such cases is relatively simple. However, if the data set contains more than one of the above cases, the problem of identifying these observations becomes more difficult. This is due to masking and swamping effects. *Masking* occurs when an outlying subset is undetected, generally due to the presence of another, usually adjacent, subset. *Swamping* occurs when observations are incorrectly identified as outliers because of the presence of another, usually, remote, subset of observations. Several procedures for the identification of multiple outliers in linear models have been proposed, generally based on robust estimation methods or fuzzy clustering strategies (Atkinson 1986; Hadi and Simonoff 1993; Seaver et al. 1999).

4.6 Conclusions

From preceding chapters it is clear that knowledge discovery is an iterative and interactive process, whereby the data may be revisited frequently, subjected to various forms of pre-processing, elucidating and improving data quality/information content and consequently assessing the reliability of the statistical inferences drawn.

This chapter has presented methods, namely conditioning diagnostics and sensitivity analyses which allow the data-model (mis-)match to be investigated. The detection and understanding of the causes of any data weaknesses can then lead to appropriate analytical solutions.

For instance it may be that through experimental design criteria more informative samples can be gathered, which may address the data weaknesses inherent in the original data set.

Once the data weaknesses have been acknowledged, appropriate "remedies" should be employed. For instance, a single observation may be instrumental in determining a model structure. Chatterjee and Hadi (1988) studied the impact of simultaneously omitting a variable and an observation in linear regression whilst Hoeting et al. (1996) proposes a method for simultaneous variable selection and outlier identification, the latter based on computing posterior model probabilities.

The decrease in parameter sensitivity that can result from introducing prior information suggests that anomalous data can be confounded with problems arising from collinearity. In (Belsley 1991), it is seen that collinearity can disguise anomalous data, as indicated by the increased coefficient sensitivity, when data conditioning was improved through the use of prior

knowledge.

Biased estimators, as discussed in Chapter 3, are ways of dealing and understanding certain collinearity problems, for example the ridge trace in ridge regression estimates provides a graphical account of the effect of collinearity in the coefficient estimates (Gibbons and McDonald 1984). These methods however, may not be effective alternatives when collinearities are outlier-induced.

Chapter 5

Neurofuzzy Networks

5.1 Introduction

Neurofuzzy networks (or systems) have evolved over recent years as researchers have tried to combine the linguistic, vague representation of a fuzzy system with the structural and learning attributes of neural networks. The algorithms have predominantly been developed by researchers in the fields of system identification and control (Brown and Harris 1994; Jang et al. 1997) as it proved natural to use both expert, fuzzy knowledge for either initialising or validating a trained system, and neural-type learning algorithms which made the best use of the available training data. This combination was proposed as an efficient way to allow systems to overcome the explicit dependence on either an expert or the availability of numerical data.

This chapter describes a class of neurofuzzy networks used in subsequent chapters in modelling several properties of Al-alloys. The material presented reflects the progress made in the neurofuzzy framework initially formalised by Brown and Harris (1994), and subsequently extended by Bossley (1997) and ultimately by Gunn et al. (1997), in the development of parsimonious system modelling. Some fundamental concepts will be described together with the important links that characterise neurofuzzy networks.

5.2 Fuzzy systems

A fuzzy system is a non-linear model whose behaviour is described by a set of easily interpretable rules, representing an alternative approach to neural networks in system modelling. Unlike typical neural network architectures, fuzzy systems reason in a seemingly natural manner, by the application of a series of rules. A fuzzy system generally consists of a rule-base composed of vague production rules such as:

$$\text{IF } Cu \text{ is } high \text{ AND } ageing\ time \text{ is } medium \text{ THEN } proof\ stress \text{ is } high$$

The rules are generally linguistic representations and since the knowledge in a fuzzy system is stored as a set of interpretable rules, such systems are said to be *transparent.* Transparency is arguably the most important element in any fuzzy system, as the designer can understand how different fuzzy rules interact and adjust the process accordingly.

Often fuzzy systems are identified from heuristics and limited empirical knowledge. However, such an approach may result in inadequate models. This has motivated the development of adaptive fuzzy systems which adjust their rule base parameters via heuristic training rules. Both approaches have been heavily criticised, in terms of formulation of inadequate models to which no standard system identification mathematical analysis can be applied. Furthermore, there has been a notable lack of rigorous theory to expain how these systems generalise, and also to provide insights into the relative merits of different implementation strategies.

5.2.1 Fuzzy sets

To represent linguistic statements (such as x is *small*), Zadeh (Zadeh 1965; Zadeh 1973) introduced the concept of a *fuzzy set.* A fuzzy set, A, is a collection of elements defined in a *universe of discourse*, labelled X. Mathematically, this can be written as:

$$\mu_A(x) : X \rightarrow [0, 1]$$

where A is the fuzzy label, set or linguistic variable, describing the variable x. As an extension to Boolean logic, $\mu_A(x)$ represents the *grade of membership* of x belonging to the fuzzy set A. Thus, fuzzy logic generalises the concept of a classical set by allowing its elements to have *partial* membership, the degree to which x belongs to A is characterised by a *membership function*, $\mu_A(x)$. By specifying membership functions, the vague fuzzy labels are given a precise definition. Each linguistic term is represented by a membership function and the set of all of these terms determines how an input variable is represented in the fuzzy system.

An example of the representation of a fuzzy variable, *ageing time*, in terms of three fuzzy sets *short*, *medium* and *long* is shown in Figure 5.1. At any given value of the variable, membership of the three sets adds to unity. For example, at a value of 0.25, the membership of the set *short*, μ_{short} (*ageing time*), is 0.5, of the set *medium* is 0.5 and of the set *long* is zero. The point at which the membership of the fuzzy sets comes to zero (at a value of 0.5 in Figure 5.1) is called a *knot*, and is likely to represent a change in the trend between input and output.

Fuzzy sets enable a system to generalise locally between neighbouring rules, so that similar inputs will produce similar outputs.

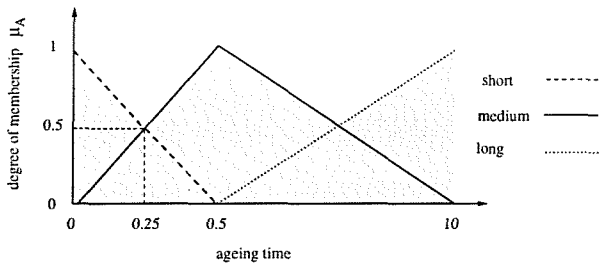


Figure 5.1: Schematic representation of a fuzzy variable

5.2.2 Fuzzy rules

For an n-dimensional input, $\mathbf{x} = (x_1, \ldots, x_n)$, single output fuzzy system, the rule base is composed of a sequence of fuzzy rules formally defined as:

$$r_{ij} : \text{IF } \overbrace{x_1 \text{ is } A_1^i \text{ AND } x_2 \text{ is } A_2^i \text{ AND } \ldots \text{ AND } x_n \text{ is } A_n^i}^{antecedent} \text{ THEN } \overbrace{y \text{ is } B^j}^{consequent} \; c_{ij} \qquad (5.1)$$

The terms A_k^i, $k = 1, \ldots, n$ and B^j are linguistic (input and output) variables which represent vague terms such as *small*, *medium* or *large* and associated with each rule is a variable $c_{ij} \in [0, 1]$ that denotes the *confidence* in the rule being true. The rule maps the antecedent, formed by the intersection (the fuzzy AND operation) of n linguistic statements x_k is A_k^i, to the consequent formed by a single univariate linguistic statement y is B^j (describing the system's output), through the fuzzy implication operator (IF (.) THEN (.)). A confidence of $c_{ij} = 0$ means that the rule will never fire whereas if $c_{ij} > 0$ the rule will partially fire when the input is a partial member of the antecedent. The r_{ij}^{th} rule can be written more concisely as:

$$\text{IF } \left(\mathbf{x} \text{ is } \mathbf{A}^i\right) \text{ THEN } \left(y \text{ is } B^j\right) \; c_{ij}$$

where $\mathbf{A}^i$ is the i^{th} multivariate fuzzy set formed from the fuzzy intersection (AND) of the individual univariate fuzzy sets. Multivariate fuzzy membership functions are formed by taking the tensor product of n univariates, resulting in different rule antecedents.

In general, a number of production rules are required to adequately describe the relationship between input and output and these are connected together using the fuzzy union (OR) operator to form the fuzzy algorithm. A system with m outputs can be implemented as m fuzzy algorithms each with n inputs and one output.

Fuzzy rules provide a finite, vague description of the continuous input-output mapping, hence they can be used to either initialise or validate the model. The fuzzy rules provide a point-to-point mapping, the shape of the membership function determines how the network interpolates between these points.

5.3 Neurofuzzy systems

Neurofuzzy systems (Brown and Harris 1994; Jang et al. 1997) are a particular fuzzy system where the fundamental components of the fuzzy system are constrained, enforcing a mathematical structure. A neurofuzzy system combines the positive attributes (principally, learning and function approximation abilities) of a neural network with those of a fuzzy system representation. This provides a transparent framework for representing linguistic rules with well defined modelling and learning characteristics. Whilst neurofuzzy techniques provide quantitative predictions of complex multivariate systems, dependencies on individual input variables can remain transparent since underlying relationships contained in the data are (modelled) represented by easily understood rules.

Generally, neurofuzzy networks are designed by first training a network on numerical data, then validating and correcting/extending its behaviour using fuzzy rules. The fuzzy interpretation plays a small (but significant) role, whereas the major part is the extraction of structural knowledge from the numerical data. When the basis functions are allowed to be distributed arbitrarily across the input space, as occurs in many radial basis function (RBF) networks (Jang et al. 1997), it is difficult to produce an easily understood fuzzy interpretation; however, if they are constrained to lie on a lattice (Bossley 1997), the concept of a fuzzy variable has a natural interpretation, the network's generalisation capabilities are transparent (enabling improved validation to take place) and simple fuzzy rule-bases can be generated.

In lattice based systems, the fuzzy membership functions and the fuzzy operators will be seen in section 5.3.3 to be defined in such a manner that there exists a direct invertible relationship between the fuzzy rule-base and an associative memory network (AMN). Brown and Harris (1994) have shown that when B-splines are used to implement multivariate fuzzy membership functions, the output of a neurofuzzy system is given by:

$$y(\mathbf{x}) = \sum_{i=1}^{p} \mu_{\mathbf{A}^i}(\mathbf{x})\, w_i \tag{5.2}$$

where $\mu_{\mathbf{A}^i}(\mathbf{x})$ is the membership function of the multivariate fuzzy input set $\mathbf{A}^i$, and w_i is the corresponding weight. Therefore, the network's output is seen to be a weighted linear combination of the input membership functions.

5.3.1 B-splines

B-splines represent a method for approximating general multivariate functions in the form of polynomial segments defined over a grid partitioning of the input space. Figure 5.2 shows univariate B-splines of orders $k = 1, \ldots, 4$. In each partition the output is approximated by a polynomial specified by the order of the univariate B-splines, joined together to form a continuous and smooth function over the input domain. On each interval, k coefficients are necessary to represent a polynomial of order k, and so k B-spline weights (basis functions)

contribute to the output. A series of these univariate basis functions are defined on the complete

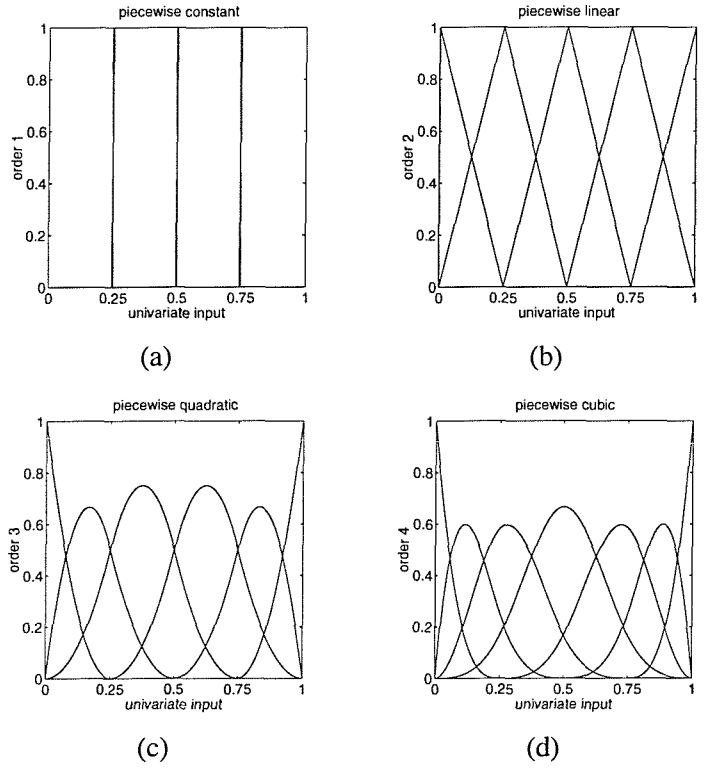


Figure 5.2: Univariate B-spline basis functions of order 1-4.

input space of each input variable. The shape of a univariate B-spline is defined by its *order*. The shape, position and number of these basis functions determines the accuracy and flexibility of the resulting model, the degree of smoothness depending on the polynomial degree of the B-spline. The set of univariate B-splines are defined by a knot vector, $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_{r+2k})^T$, where k is the order of the splines, and r the number of univariate basis functions defined on the axis. The input domain of a set of univariate basis functions is given by $[\lambda_k, \lambda_r]$ giving a total of $(r - 2k)$ interior knots, λ_k and λ_r are known as the boundary knots. Univariate B-spline outputs are evaluated by the simple, stable recursive relationship:

$$a_{j,k}(x) = \frac{x - \lambda_j}{\lambda_{j+k} + \lambda_j} a_{j,k-1}(x) + \frac{\lambda_{j+k+1} - x}{\lambda_{j+k+1} - \lambda_{j+1}} a_{j+1,k-1}(x)$$

where $a_{j,k}$ represents the j^{th} basis function with order k.

Generally, it is advisable to limit the order of the B-splines to as low an order as is possible that produces an acceptable fit to the data, since high-order basis functions are generally too flexible, causing the network to overfit the data, reflected by a badly conditioned learning problem.

Multivariate basis functions are constructed by tensor multiplication of the univariate basis functions. Each univariate and consequent multivariate basis function is *tensor* multiplied with every other one defined on the remaining variables, resulting in a lattice partitioning of the

input space. The number of multivariate basis functions produced from this operation is an exponential function of the number of inputs:

$$p = \prod_{i=1}^{n} p_i$$

where p_i is the number of univariate basis functions defined on the i^{th} input. This property is known as the *curse of dimensionality* (Bellman 1961) and is common to many modelling techniques. As a result the cost, in terms of the number of basis functions required to implement a network increases exponentially as the input space grows linearly. This is the one major disadvantage of B-spline networks, limiting the use of such networks to problems of low dimensions (i.e. $\leq$4).

5.3.2 B-spline networks

B-spline networks have been widely used as surface fitting algorithms in graphical applications, a task comparable to modelling a continuous input-output mapping. In these networks, B-spline basis functions are employed to produce a continuous piecewise polynomial output and belong to the class of AMNs as they generalise locally (i.e. similar inputs map to similar outputs, whereas dissimilar inputs map to independent outputs). Since only the output layer weights are adapted, well established linear training algorithms, with provable behaviour characteristics, such as conjugate gradient (CG) or singular value decomposition (SVD), can be employed.

B-spline networks form the output by a weighted sum of multi-dimensional basis functions, given by:

$$y(\mathbf{x}) = \sum_{i=1}^{p} a_i(\mathbf{x}) w_i = \mathbf{a}^T \mathbf{w} \tag{5.3}$$

where $y(\mathbf{x})$ is the output, $\mathbf{a}$ is the vector of the multi-dimensional basis function outputs $(a_0(\mathbf{x}), \ldots, a_p(\mathbf{x}))$, when presented by the input vector $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{w}$ is the vector of the associated weights. The structure resulting from this type of network is shown in Figure 5.3. Since the representation given in equation 5.2 is equivalent to that of a B-spline network, where the multivariate basis functions, $a_i(\mathbf{x})$ are the multivariate fuzzy membership functions, univariate B-spline basis functions can be used to implement fuzzy linguistic terms such as *ageing time* is *short*. This direct equivalence enables mathematical and algebraic tools used to train and analyse B-spline networks to be applied to neurofuzzy systems, with the added advantage that behaviour of the model can be represented as a set of fuzzy rules.

The modelling ability of these networks is determined by the non-linear mapping performed by the p n-dimensional basis functions, with the fundamental non-linear function controlled by the set of adjustable weights.

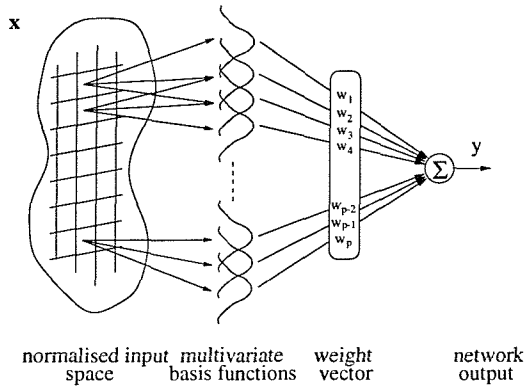


Figure 5.3: Representation of the structure of a B-spline network

5.3.3 Weights and rule confidences

From equation 5.2, the consequent and rule confidence corresponding to each rule antecedent is seen to be represented by the weight, w_i. If the fuzzy output membership functions are chosen as symmetrical B-splines of order $k \leq 2$, the following relationship has been shown (Brown and Harris 1994) to hold:

$$w_i = \sum_{j=1}^{q} c_{ij} y_j^c \tag{5.4}$$

where q is the number of fuzzy output sets, y_j^c is the centre of the j^{th} output set and where:

$$c_{ij} = \mu_{Bj}^k (w_i) \tag{5.5}$$

i.e. given a weight w_i, the corresponding rule confidence c_{ij} can be calculated by evaluating the membership of the weight of the j^{th} fuzzy output set. Each weight will then generate a unique rule confidence vector. This unique invertible relationship allows the transformation from a network representation to the fuzzy rule base, and vice versa with no loss of information.

5.3.4 Additive neurofuzzy models

From section 5.3.1 it is seen that the application of neurofuzzy systems is limited to problems involving a small number of input variables by the curse of dimensionality. Thus to make high-dimensional approximation feasible some form of complexity reduction must be performed, i.e. the number of production rules must be reduced without affecting (and possibly improving) the quality of the approximation.

The fact that experts find it hard to correctly articulate rules that depend on a large number of inputs, coupled with the curse of dimensionality, has directed considerable effort (Brown et al. 1995; Mills et al. 1995; Bossley 1997) to develop neurofuzzy construction algorithms capable of determining their structure automatically from a training set. The algorithms investigated were driven by and embodied the following data modelling principles:

- **Principle of data reduction**: the smallest number of input variables should be used to explain a maximum amount of information.
- **Principle of network parsimony**: the best models are obtained using the simplest possible, acceptable structures that contain the smallest number of adjustable parameters. This is also referred to as Occam's razor.

The curse of dimensionality can be alleviated by exploiting structural information (such as *redundancy*) present in the data, whereby the function to be approximated is additively decomposed (globally partitioned) into a series of smaller submodels each of which can be viewed as a conventional neurofuzzy system:

$$y(\mathbf{x}) = \sum_{u=1}^{U} s_u(\mathbf{x}_u) \tag{5.6}$$

where $s(\cdot)$ is a conventional neurofuzzy system whose input vector $\mathbf{x}_u$ is a subset of $\mathbf{x}$.

In globally partitioned neurofuzzy networks, both network transparency and the use of simple linear training algorithms is retained. The construction procedures search for additive structural relationships that may exist in the training data using an ANalysis Of VAriance (ANOVA) decomposition:

$$f(\mathbf{x}) = f_0 + \sum_{i=0}^{n-1} f_i(x_i) + \sum_{i=0}^{n-1} \sum_{n-1}^{j=i+1} f_{i,j}(x_i, x_j) + \ldots + f_{0,1,\ldots,n-1}(\mathbf{x}) \tag{5.7}$$

where the function $f(\mathbf{x})$ is simply an additive decomposition of simpler subfunctions, in which f_0 is a bias component and $f_i(x_i)$, $f_{i,j}(x_i, x_j)$, ... represent univariate, bivariate, and higher order terms. For many functions certain interactions are redundant, and hence their associated subfunctions can be removed from the ANOVA decomposition in equation 5.7, resulting in a less complex, more parsimonious model.

The structure of an additive neurofuzzy network using second order basis functions is shown in Figure 5.4, in which the contribution of inputs $x_1, \ldots, x_4$ have been modelled by three separate subnetworks, two of which are given by piecewise linear approximations of x_1 and x_2 and the third subnetwork contribution is provided by a tensor product involving x_3 and x_4.

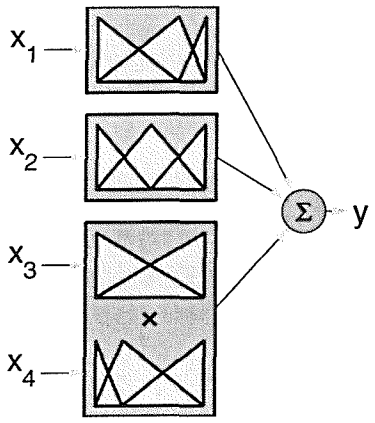


Figure 5.4: Representation of a neurofuzzy additive network structure

5.3.4.1 ASMOD algorithm

The Adaptive Spline Modelling of Observation Data (ASMOD) algorithm (Kavli 1993) searches for an appropriate ANOVA decomposition, from a representative training set.

Model construction can begin with an empty ASMOD model (containing no subnetworks) or it can be initialised from a set of fuzzy rules provided by an expert, specifying a set of statements describing the expected behaviour of the system. In the ASMOD algorithm, a succession of step refinements are performed on the model until an "optimal" model is obtained. The candidate refinements used and described below fall into either model *building* or *pruning* categories.

- **Univariate addition**: inclusion of a new input variable into the model is achieved by introducing a new one-dimensional submodel. Univariate models are only included into the existing model if the new input is not present in the existing model. To make the model construction process more efficient candidate univariate subnetworks are drawn from an external *store*. These external stores represent a set of univariate subnetworks of different fuzzy densities for each input variable. Each univariate subnetwork in the external store defined on the input variables is added to the current subnetwork to produce a new candidate model.

- **Tensor product**: an existing submodel is given the dependency on another input variable by allowing a tensor multiplication of this with a univariate subnetwork drawn from the external store. Tensor multiplication is only allowed if the new input variable appears in the current ASMOD model, but not in the subnetwork to which it is to be multiplied. By having an external store containing different fuzzy set densities, the candidate models are provided with a range of possible flexibilities.

- **Knot insertion**: modelling flexibility of a subnetwork can be increased by insertion of a new basis function. This is achieved by inserting a new knot into one of its univariate

knot vectors, creating a new set of fuzzy membership functions for the input variable. To reduce the number of candidate refinements, the locations of these new knots are restricted to lie halfway between existing interior knots present in the current subnetwork.

During model construction, redundant model flexibility (i.e. fuzzy rules) may be introduced in certain regions of the input space. This undue flexibility (complexity) may be reflected in local overfitting and may contribute to inflating the uncertainty in some rules and weights. Redundant flexibility (variance) may be removed by means of a set of model pruning refinements:

- **Subnetwork deletion**: superfluous submodels in the ANOVA decomposition are removed by deleting each of the subnetworks in the current model, producing a set of candidate evaluation models, each containing one less subnetwork than the current model.
- **Tensor split**: some input variable interactions in the current model may be redundant and/or some basis functions poorly conditioned. These are prevented by allowing the splitting of existing submodels into two new additive submodels.
- **Knot deletion**: redundant flexibility can be pruned by the deletion of input membership functions, by deleting a knot from one of the subnetwork knot vectors. Every interior knot in the entire ASMOD model is considered for knot deletion.
- **Reduce order**: the condition of the basis functions and risk of overfitting small data sets may be improved by reducing the current order of the basis functions.

The refinements described above are generally combined (e.g. (Friedman 1991)), to provide a coherent model search, into a series of pass structures. A typical *forward selection / backward elimination* (FS/BE) pass structure is:

pass
univariate addition
tensor product
tensor split
pass
subnet deletion
pass
knot insertion
pass
knot deletion
pass
reduce order

Thus, in the FS/BE iterative construction algorithm an additive decomposition is identified by using univariate additions and tensor product refinements. When input variable interactions and general ANOVA decomposition have been found, the specific structure of the subnetworks can be further refined by evaluating a series of knot insertions and deletions. Both building and pruning refinements are performed until no further model improvement can be determined.

5.4 Training neurofuzzy networks

Training neurofuzzy networks can be decomposed into their linear and non-linear components: a search technique is used to identify the structural non-linearities whereas standard linear optimisation algorithms are used to identify the weights. This corresponds to the two distinct optimisation problems of model identification and parameter estimation. The following sections describe the model search methodology, criteria employed and weight estimation.

5.4.1 Iterative model construction

In common with many other modelling techniques (e.g. determining the optimal number of nodes in the hidden layer of an MLP, choosing the polynomial order of the approximating function, etc.) an appropriate model structure is determined by searching for a model that achieves an "optimal" performance on the particular objective function.

The model should have sufficient adaptability to learn the structural information contained in the training data, but it should not be over-parameterised, as this will cause the model to fit the noise which is inherent in the data, and corresponds to selecting an appropriate size and structure for the model.

When system understanding is limited, automatic model construction algorithms are required to search for an optimal model. Since an exhausive search in model space becomes computationally unfeasible in high-dimensional approximations, it is desirable to evaluate only a limited number of models. Section 5.3.4.1 outlined a series of candidate refinements to be considered in the development of an appropriate model structure. Implementation of the ASMOD algorithm into the neurofuzzy framework can be summarised by the following iterative procedure:

- **Start**: model construction is initialised from an initial network structure, which may simply be an empty model or a model reflecting prior system understanding.
- **Evaluation**: depending on the *pass structure* employed in the model search and the n^{th} step in the model construction, a number of candidate refinements (indicated by the p^{th} pass) are considered for updating the current (n^{th}) model.
- **Estimation**: the set of candidate models defined by the p^{th} pass are trained on the training data using efficient matrix update formulae (Orr 1996) wherever possible to

speed up this computationally demanding process.

- **Selection**: the performance of all candidate models resulting from the p^{th} pass, including the current (n^{th} model), are evaluated. Model selection is then based on the network exhibiting the best performance measure (section 5.4.1.1).

- **Termination/Re-evaluation**: if the model construction has not reached the final pass, then either the search moves on to the successive pass (p^{th+1}) or a new set of candidate refinements defined by the (same) p^{th} pass are evaluated, depending on whether the n^{th+1} model resulting from the selection procedure is respectively, the same as the n^{th} model or one of the candidate refinements evaluated from the p^{th} pass (in which case $n^{th+1} \neq n^{th}$ model). If the final pass has been evaluated, with no development in the model (i.e. $n^{th+1} \equiv n^{th}$), the model search is terminated, otherwise a new set of candidate models defined by the final pass are evaluated.

As with all iterative searches, there is still no guarantee that an optimal model will be found. In section 5.4.1.2 criteria which negotiate local minima in the objective function, built into the construction algorithms are discussed. Marenbach and Brown (1997) proposed an alternative model search strategy based on genetic algorithms to successfully overcome such local minima.

5.4.1.1 Model performance measure

From the set of candidate models present at each step in the model construction, a model selection is performed. The selection must be such that a compromise between the accuracy and the complexity of the approximation is achieved, i.e. addressing the bias/variance trade-off discussed in Chapter 3.

As discussed in section 3.8.1 there are a number of techniques which try to address this problem, all of which try to minimise the prediction risk (equation 3.8). It was seen that a method that tries to achieve a balance between the quality of the approximation and the complexity of the approximating function of the final model is through use of statistical significance (SS) measures derived from information theory heuristics and which measure network parsimony. These combine the training MSE with some measure of the model's complexity, giving an expression of the form presented in equation 3.10.

In view of the performance and properties exhibited by the SRM principle, this was used as the model complexity measure throughout this work, yielding the following SS measure:

$$SS\,(\mathrm{MSE}, N, p) = \mathrm{MSE} \times \left[\frac{1}{1 - K_1 \sqrt{\frac{(1+p)\ln(2N) - \ln((1+p)!) + K_2}{N}}} \right] \tag{5.8}$$

defined in (Gunn et al. 1997), see also section 6.2.3.

5.4.1.2 Termination criteria

To terminate the model selection at each pass, termination criteria are built into model construction procedures. Two rules, a *failure margin* and a *tolerance* requirement are used to embody model search termination criteria (Gunn et al. 1997). The failure margin (F_m) effectively allows for the model search to escape from local minima during the refinement process, by allowing the model to look n steps ahead into the model structure (as defined by the value set for F_m). The second rule, defined by the forward (f_{tol}) or backward (b_{tol}) tolerance levels, places an emphasis on the parsimony of the network. In a model building stage a new refinement (which in general will increase the network size) will be required to reduce the model performance measure by a certain percentage, defined by f_{tol}, whereas during a model pruning stage a new refinement (which generally will reduce the network size) will be allowed to produce an increase in the model performance measure by a certain percentage, defined by b_{tol}. These may improve model construction by rejecting refinements for which the inclusion/retention of extra parameters yields only a marginal improvement in the model performance measure, thus acting as a further safeguard against constructing over-parameterised models.

The typical values for the tolerance levels and failure margin which have been used in previous studies have been retained in the present work and are summarised in Table 5.1. At

F_m	f_{tol}	b_{tol}
2	3%	− 1.5 %

Table 5.1: Model search termination criteria values employed.

each iteration of the ASMOD algorithm, the refinement producing the largest reduction in the statistical significance measure, subject to the tolerance requirements, is chosen. If a succession (defined by F_m) of optimal refinements fail to improve the statistical significance measure of the whole model, then the algorithm is terminated.

5.4.2 Parameter estimation

From section 3.6 it was seen that assuming a flat prior distribution for the weights, the most probable weights for a given model structure are found by the maximum likelihood (ML) estimates which minimise equation 3.4. This cost function can be expanded to a positive quadratic function of the weight vector, given by:

$$J_N = \mathbf{w}^T\mathbf{R}\mathbf{w} - 2\mathbf{w}^T\mathbf{p} + \frac{\mathbf{y}^T\mathbf{y}}{N} \tag{5.9}$$

where

$$\mathbf{R} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{a}_i\mathbf{a}_i^T = \frac{1}{N}\mathbf{A}^T\mathbf{A} \tag{5.10}$$

is the $[p \times p]$ autocorrelation matrix, $\mathbf{A}$ is the $[N \times p]$ solution matrix whose i^{th} row, $\mathbf{a}_i$, is the transformed input vector for the i^{th} input, and

$$\mathbf{p} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{a}_i y_i = \frac{1}{N}\mathbf{A}^T\mathbf{y} \tag{5.11}$$

is the p-dimensional cross-correlation vector. Equation 5.9 has its minimum at w^*, where each element of the gradient vector $\partial J/\partial \mathrm{w}$, is identically equal to zero (i.e. $\partial J_N/\partial w_i = 0$, $1 \leq i \leq p$):

$$\frac{\partial J_N}{\partial \mathbf{w}} = 2\mathbf{R}\mathbf{w} - 2\mathbf{p} = 0 \tag{5.12}$$

Then, J_N is minimised by the weight vector $\widehat{\mathbf{w}}$, which satisfies the system of linear equations, the *normal equations*, $\mathbf{R}\mathbf{w} = \mathbf{p}$, giving the least squares solution to the system, $\mathbf{A}\mathbf{w} = \mathbf{y}$. Thus, minimising the MSE across the training set reduces to identifying the optimal weight vector given by:

$$\mathbf{w}^* = \mathbf{R}^{-1}\mathbf{p} \tag{5.13}$$

which relies on the availability of computing the inverse of the autocorrelation matrix, $\mathbf{R}$, which is by definition a real symmetric positive semi-definite matrix. If $\mathbf{R}$ is *nonsingular* (rank $(\mathbf{A}) = p$), it can be inverted, the normal equations have a unique solution and the MSE has a global minimum in weight space. By contrast if $\mathbf{R}$ is *singular*, there exists an infinite number of solutions and the problem is said to be *underdetermined* (often the existence of a singular autocorrelation matrix is the result of a poorly distributed training set). In such cases, it is desirable to find the *minimum norm solution*, i.e. the unique solution which yields the smallest size weight vector $\|\mathbf{w}\|_2$ (small weight values mean that the network should generalise sensibly across its input space). The solution to the system $\mathbf{A}\mathbf{w} = \mathbf{y}$ can then be determined by performing the pseudo-inverse of $\mathbf{A}$:

$$\mathbf{A}^\dagger = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T = \mathbf{R}^{-1}\mathbf{A}^T$$

from which the least-squares solution is given by

$$\widehat{\mathbf{w}} = \mathbf{A}^\dagger\mathbf{y}$$

Generally methods based on the normal equations are computationally cheaper but round-off errors can be introduced in the formulation of $\mathbf{R}$, which for ill-conditioned problems produces large errors in the solution. Since B-splines are naturally sparse (the constant term for one submodel can be shifted by an arbitrary amount and the other submodels adjusted so that the output is equivalent) the autocorrelation matrix will be singular.

Parameter convergence is essential if the model is expected to generalise sensibly outside its training domain. The condition of the basis functions provides an important measure of how

well this is achieved. The magnitude of the weight vector error related to the output error is shown by Brown and Harris (1994) to be bounded by the following:

$$\frac{\|\epsilon_w\|}{\|\mathbf{w}\|} \leq \sqrt{C(\mathbf{R})}\frac{|\epsilon_y|}{|y|}$$

The significance of this equation stems from the fact that when the condition number of the set of basis functions is large, a small normalised output error does not necessarily imply that the normalisation error in the weight vector is small, i.e. output error tests are not sufficient to determine when to stop training, as the parameter errors could be very large.

A further result is that when the order of the B-splines is incremented by one, the corresponding condition of the set of basis functions increases by a factor of two (Brown and Harris 1994). Thus, the order of the basis functions should therefore be sufficiently high that the desired function can be modelled adequately, but it should also be as small as possible to keep the basis well conditioned. Using basis functions of as low an order as possible also reduces the computational cost of the algorithm and lowers the possibility of overfitting the data.

A detailed description and comparison of methods (both direct and indirect) that have been used to solve the linear optimisation problem in neurofuzzy systems is given in (Brown and Harris 1994; Bossley 1997). Among these SVD based and conjugate gradient (CG) methods were preferred as they possess several important properties. SVD is a direct method in which the pseudo-inverse is found in a well defined number and sequence of arithmetic operations; CG is an iterative method that optimises J_N without explicitly performing matrix inversion and is particularly well suited to large sparse systems. In CG, an initial solution is successfully improved by repeating a set of simple operations until an acceptable solution is obtained.

For singular matrices, both methods give the minimum norm solution (for CG by initialising weight values to zero) and so introduce uniqueness. The minimum norm solution becomes relevant if there are two or more subnetworks, or an input is included multiple times.

5.5 Regularised networks

Neurofuzzy model construction algorithms attempt to minimise structural redundancy by searching for a parsimonious model structure where weight identification was based on maximising the likelihood function in equation 3.3. However, iterative construction algorithms have been found to be prone to overfitting, due to the inclusion of redundant degrees of freedom during the model construction. In regions of the input space where there is redundant model structure, rule confidences will be inadequately identified by the data, often resulting in erratic output surfaces and nonsensical fuzzy rules. In neurofuzzy networks the principal reason for the presence of redundant parameters results from the structural symmetry enforced by the requirement for model transparency. As discussed in section 3.6, a solution to this is to assign a *prior* p.d.f., $P(\mathbf{w})$ on the value of the weights in order to regularise superfluous model structure.

By giving the network learning process a probabilistic interpretation (Bishop 1995), the error function is interpreted as minus the log likelihood for a noise model:

$$P(\mathcal{Y}|\mathbf{w},\beta,\mathcal{H}) = \frac{1}{Z_D(\beta)} \exp(-\beta E_D(\mathbf{w}))$$

where $Z_D(\beta)$ is a normalisation factor, $E_D(\mathbf{w}) = N/2J_N$ corresponds to the assumption of Gaussian noise on the output and β defines the noise level $\sigma_\nu^2 = 1/\beta$. Similarly, defining the regulariser in terms of a log prior probability distribution over the parameters:

$$P(\mathbf{w}) = \frac{1}{Z_W(\alpha)} \exp[-\alpha E_w(\mathbf{w})]$$

where $Z_W(\alpha)$ is a normalisation factor and the corresponding prior distribution is a Gaussian with variance $\sigma_w^2 = 1/\alpha$. Then, making normal regression theory assumptions, the maximum of the posterior estimate of the weights becomes the weight vector that minimises the cost function given by:

$$\begin{aligned} J_R &= \frac{\beta}{2}\sum_{i=1}^{N}(y_i - \hat{y}_i(\mathbf{x},\mathbf{w}))^2 + \alpha E_w(\mathbf{w}) \\ &= \beta E_D(\mathbf{w}) + \alpha E_w(\mathbf{w}) \end{aligned} \tag{5.14}$$

The ratio $\alpha/[N\beta]$ is the regularisation coefficient (λ), controlling the trade-off between the MSE and the regulariser, and so controls the bias/variance trade-off. The amount of regularisation performed will significantly affect the generalisation ability of the model and a good match must be established between the model's bias and variance. The function $E_w(\mathbf{w})$ is known as the regulariser, which penalises the MSE cost function by constraining the values of the weights, and in so doing controls redundant degrees of freedom. From section 3.9.2 the form of a second order regulariser should represent the expected curvature of the model's output across the input space (denoted by $\mathcal{D}$):

$$E_w(\mathbf{w}) = \int_{\mathcal{D}} \left| \frac{d^2\hat{y}(\mathbf{x},\mathbf{w})}{d\mathbf{x}^2} \right|^2 p(\mathbf{x})d\mathbf{x} \tag{5.15}$$

and in the limit the resulting curvature regularisation should produce models which are linear.

To form the curvature matrix $\mathbf{K}$ requires the second order derivatives of the model's output to be determined. Bossley (1997) approximated the expected value of the curvature for additive B-spline networks by taking the sum of the curvature at various points throughout the input space. To adequately regularise across the model's input space a grid was considered on this space by taking the tensor product of the centres of the univariate basis functions of the various subnetworks. The curvature is then calculated at the intersects of the grid, requiring the sum of $n_p = \prod_{u=1}^{U} p_u$ curvatures, where p_u is the number of multivariate basis functions in the u^{th} subnetwork. The curvature of the model at a given input was then considered as the sum of the

weights. Denoting the intersects of this grid by $\mathbf{g}_i$, the regulariser becomes:

$$\begin{aligned} E_w(\mathbf{w}) &= \frac{1}{2}\sum_{i=1}^{p}\left[\frac{d^2\hat{y}(\mathbf{g}_i,\mathbf{w})}{d\mathbf{x}^2}\right]^2 \\ &= \frac{1}{2}\sum_{i=1}^{p}\left[\sum_{u=1}^{U}\frac{d^2\hat{y}_u(\mathbf{g}_i,\mathbf{w}_u)}{d\mathbf{x}_u^2}\right]^2 \\ &= \frac{1}{2}\sum_{i=1}^{p}\left[\sum_{u=1}^{U}[\mathbf{k}_i^u]^T\mathbf{w}\right]^2 \end{aligned} \tag{5.16}$$

where $[\mathbf{k}_i^u]^T\mathbf{w}$ represents the curvature of the output of the u^{th} subnetwork at $\mathbf{g}_i$. As $\mathbf{w}$ is the concatenation of the subnetwork's weight vectors:

$$\begin{aligned} E_w(\mathbf{w}) &= \frac{1}{2}\sum_{i=1}^{np}\left[\mathbf{k}_i^T\mathbf{w}\right]^2 \\ &= \frac{1}{2}\sum_{i=1}^{np}\left[\mathbf{w}^T\mathbf{k}_i\mathbf{k}_i^T\mathbf{w}\right] \\ &= \frac{1}{2}\mathbf{w}^T\mathbf{K}\mathbf{w} \end{aligned} \tag{5.17}$$

where $\mathbf{k}_i$ is the concatenation of the subnetwork curvature vectors $\mathbf{k}_i^u$, and so $\mathbf{K}$ represents the sum of the curvature squared (evaluated at $\mathbf{g}_i$, the curvature is determined at the centres of the p_j subnetwork basis functions) for an additive model.

Applying this to conventional neurofuzzy models, modifies the cost function to be:

$$J_R = \frac{N\beta}{2}J_N + \frac{\alpha}{2}\mathbf{w}^T\mathbf{K}\mathbf{w} \tag{5.18}$$

and since J_N is the conventional MSE cost function (equation 5.9), this leads to the following quadratic cost function:

$$J_R = \frac{N\beta}{2}\left[\mathbf{w}^T\mathbf{R}\mathbf{w} - 2\mathbf{w}^T\mathbf{p} + \frac{\mathbf{y}^T\mathbf{y}}{N}\right] + \frac{\alpha}{2}\mathbf{w}^T\mathbf{K}\mathbf{w} \tag{5.19}$$

and differentiating this with respect to the weights and setting to zero, gives the solution:

$$\mathbf{w}_{mp} = \beta\left[\beta\mathbf{A}^T\mathbf{A} + \alpha\mathbf{K}\right]^{-1}\mathbf{A}^T\mathbf{y} \tag{5.20}$$

Thus by defining the regulariser as a quadratic function of the weights, for fixed α and β, the optimal weight vector can be found by simple linear optimisation techniques. The resulting prior distribution is a multivariate normal distribution with zero mean and covariance matrix $[\alpha\mathbf{K}]^{-1}$, and so the number of degrees of freedom identified by the data are controlled by the variances of the priors. A global prior across the complete weight vector assumes an equal degree of smoothness throughout the input space. For additive models it may be more appropriate to treat the weights of the individual subnetworks independently, since real non-linear

processes exhibit different degrees of smoothness. As such, different priors may be placed on different areas of the input space, giving a more flexible, *local* form of regularisation.

Thus for *local* regularisation, different priors are assigned to the different subnetworks, producing a posterior p.d.f. for the weights given by:

$$P(\mathbf{w}|\mathcal{D}_N, \alpha, \beta) = \frac{P(\mathcal{Y}|\mathcal{X}, \mathbf{w}, \beta, \mathcal{H}) \prod_{u=1}^{U} P_u(\mathbf{w}|\alpha_u, \mathcal{H})}{P(\mathcal{Y}|\mathcal{X}, \alpha, \beta, \mathcal{H})} \tag{5.21}$$

where α represents a U-dimensional vector consisting of the variances of the U priors. These priors are defined as normal distributed p.d.f.'s producing quadratic regularisers of the form:

$$E_w^u = \frac{1}{2}\mathbf{w}_u^T \mathbf{K}_u \mathbf{w}_u \tag{5.22}$$

and the maximum of the posterior weight estimate from this new posterior gives the following cost function:

$$J_R(\mathbf{w}, \alpha, \beta) = \beta E_D(\mathbf{w}) + \sum_{i}^{U} \alpha_u E_w^u(\mathbf{w}_u) \tag{5.23}$$

where the individual subnetwork's weights are regularised independently by their own regulariser, each with its own regularisation coefficient, thus accounting for different smoothness conditions on the subnetwork's output. Here $\mathbf{K}$ will be a block diagonal matrix comprising all the local prior's variances α_u's and $\mathbf{K}_u$ matrices:

$$\mathbf{K} = \begin{bmatrix} \alpha_1 \mathbf{K}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \alpha_U \mathbf{K}_U \end{bmatrix}$$

The extension of Bayesian regularisation techniques to the neurofuzzy framework, within an evidence framework, was developed by Bossley (1997) as a post-processing method of controlling redundant degrees of freedom in models determined by the iterative construction algorithms.

In the evidence framework, briefly discussed in section 3.10.1, the balance between the model's bias and variance will be controlled by the variances of the likelihood function and the prior p.d.f.'s, the reciprocals of which are referred to as the *hyperparameters*, i.e. β and α. From section 3.10.1 these control how much the weights are regularised, thus controlling the effective complexity of the model. The identification of the multiple hyperparameters can be found by maximising their posterior p.d.f.:

$$P(\alpha, \beta|\mathcal{D}_N, \mathcal{H}) = \frac{P(\mathcal{Y}|\mathcal{X}, \alpha, \beta, \mathcal{H}) P(\alpha, \beta|\mathcal{H})}{P(\mathcal{Y}|\mathcal{X}, \mathcal{H})} \tag{5.24}$$

where $\mathcal{H}$ is introduced to these p.d.f's to represent the model structure, and the type of regulariser used.

Assuming the prior p.d.f. for these hyperparameters to be uniform (representing no *a priori* knowledge) and since $P(\mathcal{Y}|\mathcal{X},\mathcal{H})$ is simply a normalising constant, the maximum posterior values for the hyperparameters can be found by maximising the likelihood term $P(\mathcal{Y}|\mathcal{X},\alpha,\beta,\mathcal{H})$. This is the *evidence* for the data given α and β, which is the normalisation coefficient in the posterior p.d.f. for weights given by:

$$\begin{aligned} P(\mathcal{Y}|\mathcal{X},\alpha,\beta,\mathcal{H}) &= \int_{\mathcal{W}} \text{likelihood} \times \text{prior}\, d\mathbf{w} \\ &= \int_{\mathcal{W}} P(\mathcal{Y}|\mathcal{X},\mathbf{w},\beta,\mathcal{H}) P(\mathbf{w}|\alpha,\mathcal{H}) d\mathbf{w} \end{aligned} \tag{5.25}$$

By using a Taylor series expansion of the cost function around the maximum of the posterior weight estimate, $\mathbf{w}_{mp}$, performing appropriate Gaussian integrals and taking the log of the resulting evidence, an equivalent evidence function can be determined:

$$\begin{aligned} E(\alpha,\beta) = -J_R(\mathbf{w}_{mp},\alpha,\beta) - \frac{1}{2}\log\det(\mathbf{H}) \\ + \frac{N}{2}\log\beta + \sum_{u=1}^{U}\left[\frac{p_u}{2}\log\alpha_u + \frac{1}{2}\log\det(\mathbf{K}_u)\right] \end{aligned} \tag{5.26}$$

The hyperparameters are found by maximising this evidence and results in a non-linear optimisation problem. By assuming the Hessian and the maximum of the posterior weight estimate to be stationary with respect to the hyperparameters, Bossely derived re-estimation formulae in order to determine the values of β and α which maximise the evidence:

$$\beta^{k+1} = \frac{N - \beta^k\, tr\left(\mathbf{H}^{-1}\mathbf{A}^T\mathbf{A}\right)}{2E_D(\mathbf{w}_{mp})} \tag{5.27}$$

$$\alpha_u^{k+1} = \frac{p_u - \alpha_u^k\, tr\left(\mathbf{H}^{-1}K_u'\right)}{2E_w^u(\mathbf{w}_{mp})} \tag{5.28}$$

where:

$$K_u' = \frac{\partial\mathbf{H}}{\partial\alpha_u}$$

where $\mathbf{H}$ is the Hessian of the cost function, given by $\mathbf{H} = \beta\mathbf{A}^T\mathbf{A} + \mathbf{K}$, and $\mathbf{w}_{mp}$ is the maximum of the posterior weight estimate determined for the current value of the set of hyperparameters. By constraining the hyperparameters such that $[\alpha_j/(N\beta)] \in (0,1]$ these were found to converge to a practical and consistent solution (Bossley 1997). However, as these are non-stationary, a maximum of the posterior weight vector is re-evaluated after each iteration.

By considering α and β as opposed to a single regularisation coefficient, λ, can give insight into the success of the final model, i.e. β gives an estimate of the noise variance and can be used to incorporate available *a priori* knowledge about the expected noise level. Also, Bayesian regularisation allows the values of the regularisation coefficients to be determined directly from the data, without the necessity of employing cross-validation procedures.

5.6 Linear smoothers

As neurofuzzy networks are linear in the weights, they can be represented as a *linear smoother*, where the fit to the data set is given by:

$$\widehat{\mathbf{y}} = \mathbf{S}\mathbf{y} \tag{5.29}$$

The matrix $\mathbf{S}$ is known as the *smoother matrix*, and for unregularised models is represented by:

$$\mathbf{S} = \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^T \tag{5.30}$$

Since the number of degrees of freedom is a measure of the number of independent parameters used to fit the data, for a linear smoother this is given by the sum of the eigenvalues of $\mathbf{S}$, i.e. $tr(\mathbf{S})$ (Buja et al. 1989). For neurofuzzy networks, $\mathbf{S}$ has r non-zero eigenvalues, where r is the rank of $\mathbf{A}$. Hence the degrees of freedom found in a conventionally trained neurofuzzy model[1] is given by the rank of $\mathbf{A}$:

$$\text{df} = rank(\mathbf{A}) = tr(\mathbf{S}) \tag{5.31}$$

As regularisation changes the form of this smoother, and consequently the number of degrees of freedom, the amended linear smoother matrix for a regularised neurofuzzy network becomes:

$$\mathbf{S}_{\alpha,\beta} = \beta\mathbf{A}[\beta N\mathbf{R} + \alpha\mathbf{K}]^{-1}\mathbf{A}^T \tag{5.32}$$

The number of degrees of freedom are hence controlled by the hyperparameters α and β. If we represent $[\beta N\mathbf{R} + \alpha\mathbf{K}]$ by $\mathbf{H}$, i.e. the Hessian of the cost function, the number of degrees of freedom of a regularised model is given by:

$$\text{df} = N\beta tr(\mathbf{R}\mathbf{H}^{-1}) \tag{5.33}$$

For linear smoothers the LOOCV variance estimate can be determined as

$$LOOCV = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i - \hat{y}_i(\mathbf{x},\mathbf{w})}{1 - s_{ii}}\right)^2 \tag{5.34}$$

where s_{ii} are the diagonal elements of the smoother matrix defined in equation 5.30 and so there is no need to remove the i^{th} data pair from the training set and re-estimate the weights, as it can be determined with simple matrix algebra[2].

[1] For additive unregularised networks the degrees of freedom can also more simply be determined as:

$$\text{df} = p - (U - 1)$$

[2] For linear models, LOOCV can be calculated analytically (Orr 1996) as:

$$LOOCV = \frac{\widehat{\mathbf{y}}^T\mathbf{P}\,(diag\,(\mathbf{P}))^{-2}\,\mathbf{P}\widehat{\mathbf{y}}}{N} \tag{5.35}$$

where $\mathbf{P} = \mathbf{I} - \mathbf{A}\mathbf{R}^{-1}\mathbf{A}^T$ is referred to as the projection matrix.

5.7 Error bars

In Chapter 3 a useful property of a model was identified to be the representation of output uncertainty, i.e. assessing the confidence of the model's own predictions with respect to an input. As the neurofuzzy systems are linear in the weights, error bars can be derived in a similar way to that presented in section 3.11. In neurofuzzy networks, the sensitivity of the output with respect to the weights is given by $\partial\hat{y}(\mathbf{x})/\partial\mathbf{w} = \partial\mathbf{a}(\mathbf{x})^T\mathbf{w}/\partial\mathbf{w} = \mathbf{a}(\mathbf{x})$.

As there will be uncertainty in both the estimate given by the regression function and the noise level on the output, the variance in the model's prediction for unregularised models will be given by:

$$\sigma_y^2 = \hat{\sigma}^2 + \hat{\sigma}^2\mathbf{a}^T(\mathbf{x})\mathbf{H}^{-1}\mathbf{a}(\mathbf{x}) \tag{5.36}$$

where $\mathbf{H} = \mathbf{A}^T\mathbf{A}$ is the Hessian[3] of the cost function, J_N (equation 5.9).

In networks regularised within a Bayesian framework, error bars will be comprised by a term comprising the posterior weight uncertanty (i.e. width of the posterior distribution of the network weights) and one which arises from the intrinsic noise in the data as determined by the hyperparameters. Under the assumption that the posterior in weight space can be approximated by a Gaussian, the contribution of the weight uncertainty to the variance in the prediction is given by:

$$\sigma_{\hat{y}}^2 = \mathbf{a}^T(\mathbf{x})\mathbf{H}^{-1}\mathbf{a}(\mathbf{x})$$

Then, the noise on the network's output will be a Gaussian p.d.f. with mean $y(\mathbf{x}, \mathbf{w}_{mp}) = \mathbf{a}^T(\mathbf{x})\mathbf{w}_{mp}$ and variance:

$$\sigma_y^2 = \frac{1}{\beta} + \mathbf{a}^T(\mathbf{x})\mathbf{H}^{-1}\mathbf{a}(\mathbf{x}) \tag{5.37}$$

where $\mathbf{H} = \beta\mathbf{A}^T\mathbf{A} + \mathbf{K}$ is now the Hessian of J_R (equation 5.19). However, as error bars will be large when the output is inferred from weights for which both the prior and the likelihood function have little confidence, there will be no indication of whether the model's output has been inferred from the regulariser as opposed to the data. Therefore, regularisation may lead to unfounded confidence in the model's output estimates.

[3]The Hessian matrix of a cost function J, denoted by $\mathbf{H}$, is defined as the second derivative of J with respect to the weight vector $\mathbf{w}$:

$$\mathbf{H} = \frac{\partial^2 J}{\partial\mathbf{w}^2}$$

Then the covariance matrix $\mathbf{H}^{-1}$ represents the uncertainty in the (maximum posterior) weight estimate. For directions in the weight space for which the variance is high the associated confidence in the weights will be low. To provide error bars on the output, this covariance needs to be translated from the weight space into the output space. As neurofuzzy networks are linear in the weights this is seen to be trivial.

When model outputs have been inferred entirely from the prior p.d.f. little confidence should be placed in the model's predictions. By approximating the posterior p.d.f. by the likelihood function, error bars can be derived entirely from the data as in maximum likelihood estimation (Bossley 1997), representing the confidence in the model's output given by the data:

$$\sigma_{\hat{y}}^2 = \mathbf{a}^T(\mathbf{x}) \left[\beta \mathbf{A}^T \mathbf{A}\right]^{-1} \mathbf{a}(\mathbf{x})$$

5.8 Network diagnostics

In section 4.5, two examples of diagnostics (influence measures) for linear models were considered. In a similar way, an appropriate Cook distance measure has been defined for generalised additive models (Hastie and Tibshirani, 1990), sensitivity measures can also be defined for additive neurofuzzy models, resulting in the following one-case deletion formulae, C_i^2 and H_i^2:

$$C_i^2 = \frac{e_i^2}{\mathrm{df}\, MSE} \left[\frac{s_{ii}}{(1 - s_{ii})^2}\right] \tag{5.38}$$

and

$$H_i^2 = \frac{\mathrm{df}}{(1 - s_{ii})} \frac{d_i^2}{1 - d_i^2} + \frac{s_{ii}}{1 - s_{ii}} \tag{5.39}$$

where s_{ii} are the diagonal elements of $\mathbf{S}$ and df, the degrees of freedom as given by equations 5.31 or 5.33 depending on whether the model has been regularised.

A comparison between the C_i^2 and the diagonal elements of the respective smoother matrices, ($\mathbf{H}$,$\mathbf{S}$) obtained on the BAP4SLC5W data set (Chapter 9) from a MLR model and an equivalent neurofuzzy model structure[4] are shown in Figure 5.5.

[4]The sensitivity results shown in Figure 5.5 were obtained from defining equivalent linear models (i.e. the neurofuzzy network was defined with two second order basis functions), including all 13 inputs in the BAP4SLC5W data set. In the MLR model the bias term was defined implicitly. For both models, df = 14 and hence $\hat{\sigma}_N^2 = 8.89$.

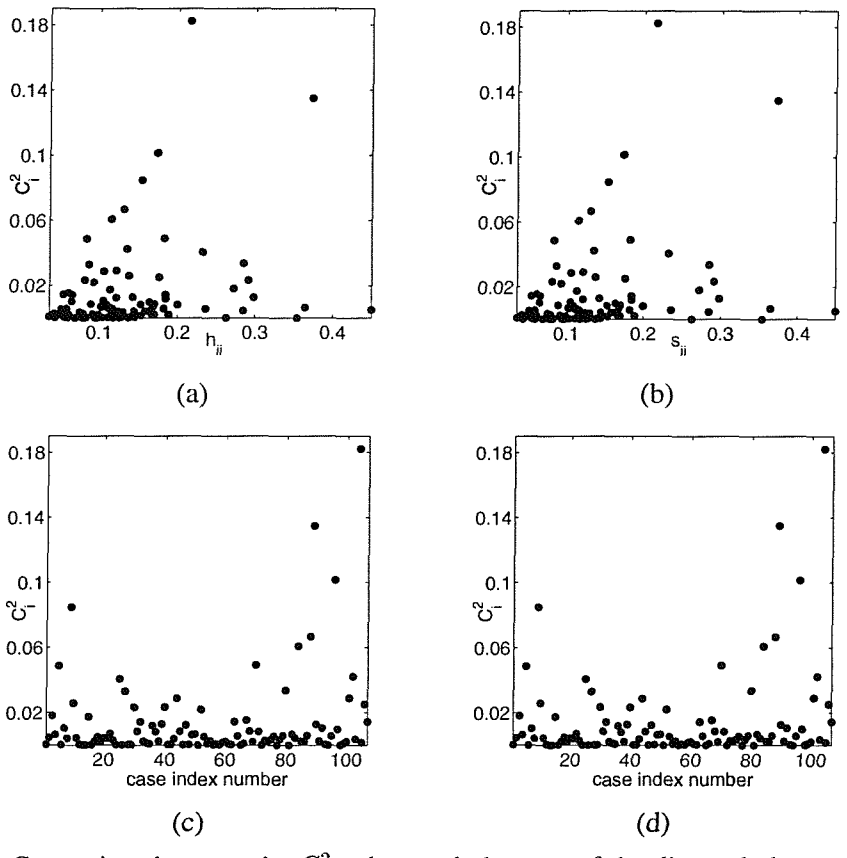


Figure 5.5: Comparison between the C_i^2 values and elements of the diagonal elements of the hat (smoother in the neurofuzzy model) matrix determined for a MLR model (a) and (c) with those derived form an equivalent unregularised neurofuzzy model (b) and (d).

5.9 Worked example: modelling the equation of the plastic zone size

For the purposes of illustrating the empirical capabilities of the neurofuzzy framework considered throughout this work, a worked example relevant to the materials application investigated is presented. A suitable test for the ASMOD construction algorithm would be to see whether a known analytical function can be approximated from a simulated data set generated from a known function.

The equation for the plastic zone size (PZS) at the crack tip, r_p, (i.e. boundary of the plastic zone) is derived from a simple elastic description of the crack tip stress field, given by:

$$\begin{aligned} r_p \;=\; \tfrac{1}{2\pi} \;[\; & K_I^2 \left(\cos\left(\tfrac{\theta}{2}\right)\right)^2 \left[(1-2\nu)^2 + 3\left(\sin\left(\tfrac{\theta}{2}\right)\right)^2\right] \\ & + K_I K_{II} \sin(\theta) \left[3\cos(\theta) - (1-2\nu)^2\right] \\ & + K_{II}^2 \left[3 + \left(\sin\left(\tfrac{\theta}{2}\right)\right)^2 \left[(1-2\nu)^2 - 9\left(\cos\left(\tfrac{\theta}{2}\right)\right)^2\right]\right] \;] \end{aligned} \tag{5.40}$$

with:

$$K_{\mathrm{I}} = \frac{1}{\sqrt{1+\alpha^2}} \text{ and } K_{\mathrm{II}} = \frac{\alpha}{\sqrt{1+\alpha^2}}$$

and where θ is the angle with respect to the crack tip, ν is Poisson's ratio of the material, K_{I} and K_{II} are respectively mode I and II stress intensity factors, and finally α defined as $K_{\mathrm{II}}/K_{\mathrm{I}}$ describes the elastic stress field (see e.g. (Ewalds and Wanhill 1984)).

Additive Gaussian noise was added to the inputs ($N(0, 0.001)$) and to the output values ($N(0, 0.01)$) obtained for the estimated radius of the plasic zone size determined from equation 5.40 from the noisy inputs.

Data was appropriately sampled and constrained to lie within sensible ranges: θ was uniformly sampled in the range $[0, 2\pi]$, ν was normally distributed around a typical value for metallic structures (0.33) and α was uniformly sampled in the range $[0, 1]$ (i.e. from pure K_{I} to $K_{\mathrm{I}} = K_{\mathrm{II}}$; thus for the same net crack tip driving force (or strain energy release rate), the $K_{\mathrm{II}}/K_{\mathrm{I}}$ ratio has been varied). A relatively small data set comprising 250 samples was generated, from which a training set of 200 randomly sampled data pairs was defined, leaving the remaining samples (50) as a validation set.

Starting from an empty model structure, the ASMOD algorithm defined in section 5.3.4.1 was left to determine an appropriate model structure by employing the model selection criteria outlined in section 5.4.1. The order of the basis functions was limited to be ≤ 3 while the data was normalised to lie between $[-1, +1]$.

In Figure 5.6 the refinements performed in the iterative search by the model construction algorithms can be seen. From the variables included in the final model (Figure 5.6 (h)), it is

seen that the behaviour of the crack tip stress field is characterised by α, with both K_{I} and K_{II} omitted from the model. The omission of these two variables is reasonable as they are part of the definition of α. The evolution of the iterative search in terms of the model performance measure (SS) and the training set MSE are illustrated in Figure 5.7. The corresponding subnetwork outputs for the final model are shown in Figure 5.8: the tensor subnetwork $\alpha \times \theta$ can be understood in terms of both the range of α's considered and the symmetry exhibited by θ. The scatterplot between model predictions and the corresponding targets (output vectors) are shown in Figure 5.9 for both training and test data sets, showing that good predictive performance has been obtained.

Finally, in Figure 5.10 the predictions obtained for the neurofuzzy model determined are compared with the true analytical values given by equation 5.40 as a function of θ and by setting the other inputs to arbitrary values. Clearly, for this simulated data set, the construction algorithms have been able to approximate the underlying analytical solution well.

5.10 Conclusions

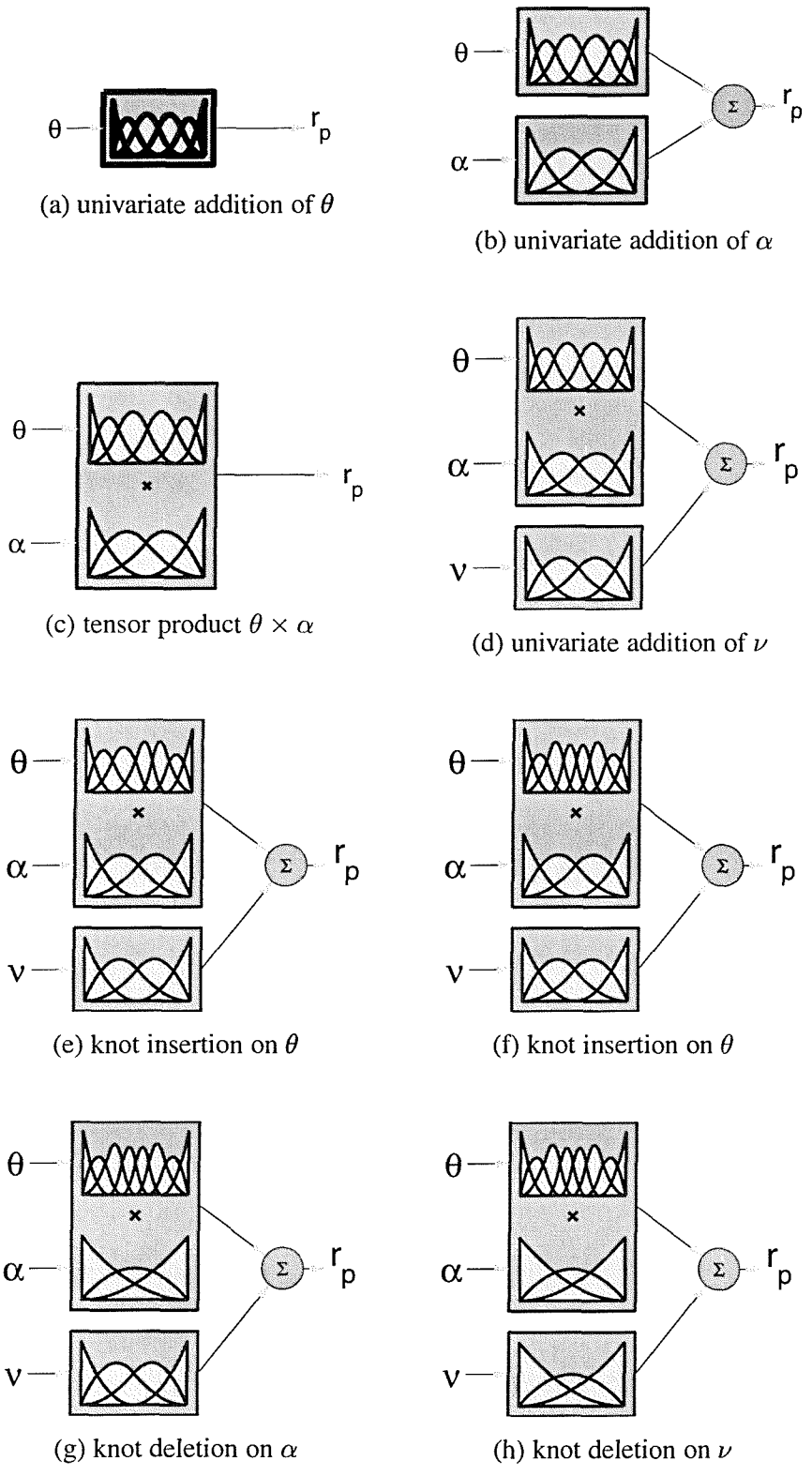
Two points, raised by Brown et al. (1996) are considered important in the present work: in the simplest situation, a network should be able to model a linear input-output relationship (which should be explicit in the network's structure) and the network's structure must be the simplest possible as this will enhance the transparency, enabling the validation of the dependencies inferred, and by limiting redundancy is likely to result in improved generalisation abilities.

As with any empirical modelling technique, the quality of the data (as discussed at length in Chapter 3) is important, both in terms of the availability of a sufficiently descriptive set of training samples, as well as a set of reliable explanatory input variables. In high-dimensional approximation however, regions of the input space will be poorly covered by data (a direct consequence of the curse of dimensionality). The additive decomposition used in this work can be an effective method of dealing with poorly distributed data sets. The main advantage of a global decomposition of the input space is in the attainment of simple local representations which can be used to explain the knowledge extracted from the training data, with the advantage of a fuzzy rule base interpretation, i.e. transparency and parsimony.

The advantages of the models being linear in the weights, not only allows efficient linear learning algorithms to be used to adapt the weights, but also allows straightforward generalisations of statistical methods developed for linear models to be implemented.

The construction algorithms may however still produce models that generalise poorly, which can be attributed to ill-conditioned basis functions, resulting from a poor distribution of the training data in intervals corresponding to certain basis functions. Local ill-conditioned basis functions are not picked up in the overall model's condition number due to continuity constraints imposed by the B-splines.

Bayesian regularisation techniques can be used as a post-processing method for control-



(a) univariate addition of θ

(b) univariate addition of α

(c) tensor product $\theta \times \alpha$

(d) univariate addition of ν

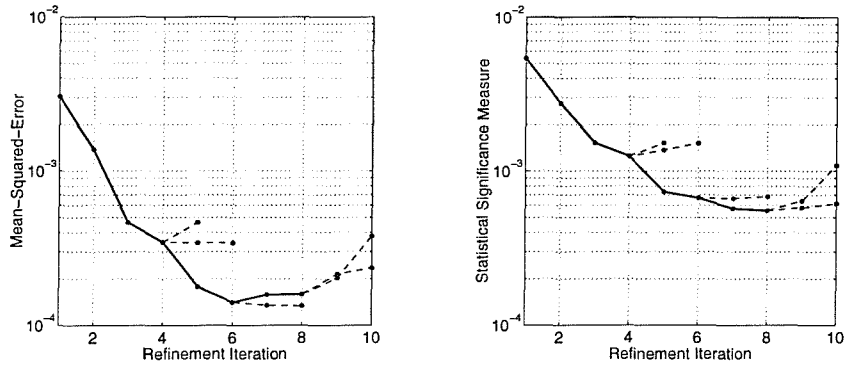
(e) knot insertion on θ

(f) knot insertion on θ

(g) knot deletion on α

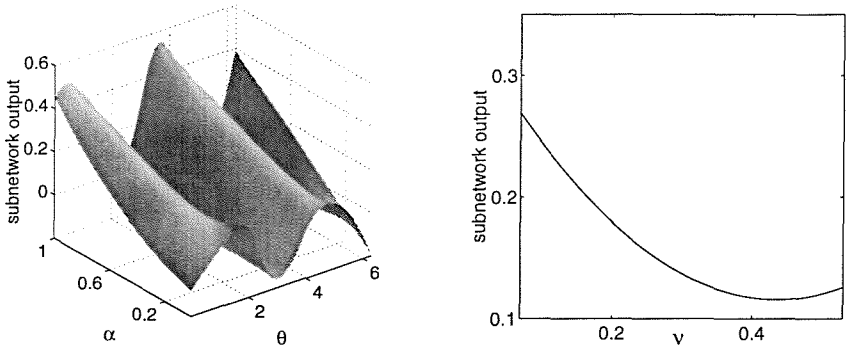
(h) knot deletion on ν

Figure 5.6: Iterative model refinements performed by the ASMOD algorithm in the modelling of equation 5.40. Subfigures (a) to (h) illustrate the refinements performed at each step in the model construction.



(a) (b)

Figure 5.7: Model performance measures at each step in the model construction: MSE (a), SS (b). The dashed line represents the refinements considered at particular stages of the model search which although may in some circumstances improve the fit to the training set, failed to satisfy the model selection criteria.



(a) (b)

Figure 5.8: Subnetwork responses corresponding to the final model (Figure 5.6 (h)) obtained by the ASMOD model construction algorithms: the $\alpha \times \theta$ tensor subnetwork (a), and the univariate subnetwork for ν (b).

ling excessive degrees of freedom present in the models identified by the iterative construction algorithms.

The implementation of Bayesian forms of inferencing (e.g. model comparison and regularisation) within iterative model construction procedures would result in a computationally very expensive training framework since the evidence and the optimal values for the hyper-parameters have to be determined for each candidate refinement considered. Although such inferencing may enable improved performance over maximum likelihood estimation, since larger models are likely to be determined (arising from a less severe form of variable selection) as a result of biasing the maximum likelihood weight values, the iterative nature of the model construction will not prevent local minima entrapment.

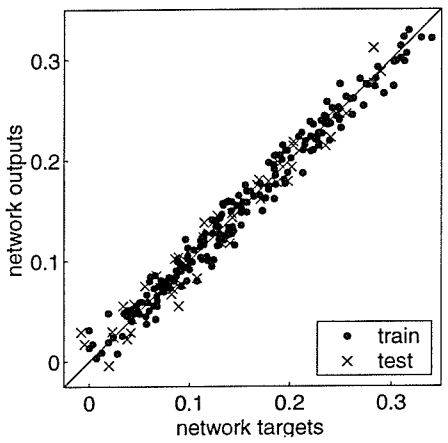


Figure 5.9: Scatterplot between model predictions and output targets.

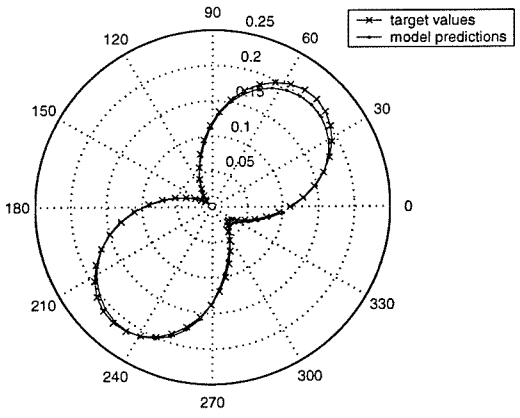


Figure 5.10: Comparison between the neurofuzzy predictions obtained as a function of the angle θ and for arbitrary values of the other variables ($\alpha = 0.5, \nu = 0.33$) and the true estimates of the underying function.

Chapter 6

Statistical Learning Theory and Support Vector Methods for Regression

6.1 Introduction

In recent years there have been significant advances in statistical learning theory and the development of support vector (SV) methods for both classification and regression (Scholkopf et al. 1999). Support vector machines (SVM) are an emerging technique, which has proven successful in many traditionally neural network dominated applications (Cherkassky and Mulier 1998) (see also references in (Burges 1998)). In this chapter, the fundamentals will be described upon which SV methods for regression are based. Both theoretical and implementation issues will be briefly described to enable a general understanding of the properties of these learning machines[1]. SVM properties and implementation techniques offer an interesting alternative to the neurofuzzy approach. The chapter will conclude by describing a recently proposed transparent modelling approach, SUpport vector Parsimonious ANOVA (SUPANOVA). In subsequent chapters this will be the principal SVM implementation of interest, which has been applied to the alloy systems under investigation and compared with the results achieved by the neurofuzzy framework.

6.2 Statistical learning theory

In contrast to the classical approaches developed for large samples and based on using various types of *a priori* information, *statistical learning theory* (SLT) (Vapnik 1998) describes statistical estimation for small data samples. SLT provides a quantitative characterisation of the

[1]Here learning machine refers to a family of functions $f(\mathbf{x}, \mathbf{w})$.

trade-off between the complexity of approximating functions and the quality of fitting to the training data.

6.2.1 Empirical risk minimisation

The goals of predictive learning are to estimate unknown dependencies in a class of approximating functions using the available data, with the optimal estimate corresponding to the minimum of the expected risk functional:

$$R[f] = \int L(y, f(\mathbf{x}, \mathbf{w}))\, p(\mathbf{x}, y)\, d\mathbf{x}dy \tag{6.1}$$

The objective is therefore to find a function $L(y, f(\mathbf{x}, \mathbf{w}))$ that minimises equation 6.1 in the case when the probability distribution $p(\mathbf{x}, y)$ is unknown but the sample $\{\mathbf{x}, y\}_{i=1}^{N}$ of observations drawn randomly and independently according to $p(\mathbf{x}, y)$ is available.

The *empirical risk minimisation* (ERM) principle seeks an estimate providing the minimum of the (known) empirical risk, as a substitute for the (unknown) true expected risk:

$$R_{emp}[f] = \frac{1}{N} \sum_{i=1}^{N} L(y, f(\mathbf{x}_i, \mathbf{w})) \tag{6.2}$$

However, approximations provided by the ERM principle for a given sample size N are always biased estimates of the "optimal" functions minimising the true risk (as the true risk does not depend on a particular sample), since for a given sample size it can be expected that $R_{emp}[f] < R[f]$, as the learning machine always chooses a function that minimises the empirical risk but not necessarily the true risk.

The ERM method is said to be consistent if it provides the sequence of functions $L(y, f(\mathbf{x}, \mathbf{w}))$, $N = 1, 2, \ldots$, for which both $R[f]$ and $R_{emp}[f]$ converge in probability to the minimal possible (for a given set of functions) value of risk as the number of samples grows infinite. The ERM method however does not contain any means of capacity control (besides choosing a smaller set of functions) which makes it very sensitive to overfitting and noisy data.

6.2.2 Structural risk minimisation

SLT does not rely on a priori knowledge about the problem to be solved and provides upper bound estimates on the expected risk. The minimisation of these bounds, which depend on both the empirical risk and the capacity of the function class, leads to the principle of structural risk minimisation (SRM) (Vapnik 1979).

The SRM inductive principle provides a formal mechanism for choosing an optimal model complexity for a finite sample. Unlike statistical methods, SRM provides accurate analytical estimates for model selection based on generalisation bounds.

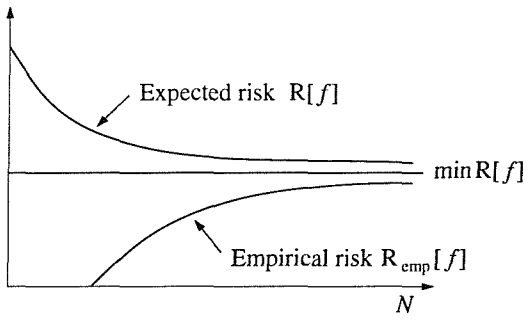


Figure 6.1: The learning process is consistent if both the expected risks $R[f]$ and the empirical risks $R_{emp}[f]$ converge to the minimal possible value of the risk $R[f]$. Estimates provided by the ERM principle should converge to the true (or best possible) values as the number of training samples grows large.

According to the SRM principle, solving a learning problem with finite data requires a priori specification of a structure on a set of approximating (or loss) functions. Under SRM the set S of loss functions $L(y, f(\mathbf{x}, \mathbf{w}))$, $\mathbf{w} \in \Omega$ has a *structure*, that is, it consists of the nested subsets (or elements) $S_k = \{L(y, f(\mathbf{x}, \mathbf{w})), \mathbf{w} \in \Omega_k\}$ such that

$$S_1 \subset S_2 \subset \ldots \subset S_k \subset \ldots$$

where each element of the structure S_k has finite Vapnik-Chervonenkis (VC) dimension h_k (Vapnik 1998). The VC dimension, h, is a scalar value that measures the *capacity* or expressive power of a set of functions realised by the learning machine[2]. By definition, a structure provides ordering of its elements according to their complexity:

$$h_1 \leq h_2 \ldots \leq h_k \ldots$$

SLT shows that it is imperative to restrict the class of functions that f is chosen from to one which has a capacity that is suitable for the amount of available training data. For a given set of observations, the SRM method chooses the element S_k of the structure for which the smallest bound on the risk is achieved.

To control generalisation in the framework of this paradigm, two factors are taken into consideration: the quality of the approximation of the data by the chosen function and the capacity of the subset of functions from which the approximating function was chosen.

The main difference between the SRM principle and statistical methods is in the control of a general capacity factor (VC dimension) instead of a specific one (e.g. number of parameters). An important feature of the SRM principle is that capacity control can be implemented in many different ways (using different types of structures).

[2]For linear methods, the VC dimension can be estimated as the number of degrees of freedom.

6.2.3 Generalisation bounds

The theory of uniform convergence in probability (Vapnik 1979) provides bounds on the deviation of the empirical risk from the expected risk. The upper bounds on the rate of uniform convergence of the learning process evaluate the difference between (unknown) true risk and the known empirical risk as a function of sample size N, properties of the unknown p.d.f. $p(\mathbf{x}, y)$, properties of the loss function, and properties of the set of approximating functions (Cherkassky and Mulier 1998). These bounds not only provide the main theoretical basis for the ERM inference, but also motivate the SRM method of inductive inference. Both the necessary and sufficient conditions of consistency and the rate of convergence of the ERM principle depend on the capacity of the set of functions implemented by the learning machine. With probability $1 - \eta$, the bounds are given by:

$$R[f] \leq R_{emp}[f] + \underbrace{\sqrt{\frac{h \ln\left(\frac{2N}{h} + 1\right) - \ln\left(\frac{\eta}{4}\right)}{N}}}_{\text{VC confidence}} \tag{6.3}$$

The right-hand side of equation 6.3 is termed the *risk bound* or *guaranteed risk*. Minimising the expected risk $R[f]$ requires the minimisation of the right-hand side of the inequality simultaneously over both terms, thus requiring the VC dimension to be a controlling variable.

From the above it is seen that, in order to achieve small expected risk (i.e. good generalisation performance), both the empirical risk and the h/N ratio have to be small. The first term in the above inequality depends on a specific function of the set of functions, while for a fixed number of observations the second term depends mainly on the VC dimension of the whole set of functions. Best generalisation performance is then attained by matching the machine capacity to the amount of training data available.

In the above inequality, for a fixed number of training examples N, the training error decreases as the capacity (or VC dimension) h is increased, whereas the confidence interval increases. Accordingly, both the guaranteed risk and the generalisation error go through a minimum.

A trade-off between the accuracy of the approximation of the training data and the capacity of the machine has to be made, as formalised in the SRM principle.

Before the minimum point is reached, the learning problem is *overdetermined* in the sense that the machine capacity h is too small for the amount of training detail. Beyond the minimum point, the learning problem is *underdetermined* because the machine capacity is too large for the amount of training data.

For large sample sizes, the value of the *confidence interval* (or VC confidence) becomes small, and the empirical risk (and hence the ERM principle) can be safely used as a measure of true risk. However, for small sample sizes[3], a small value of the empirical risk does not

[3] In (Vapnik 1998) the size N of the sample is considered to be small for the purpose of estimating functions, on the basis of the set of functions with VC dimension h if the ratio N/h is small, say $N/h < 20$.

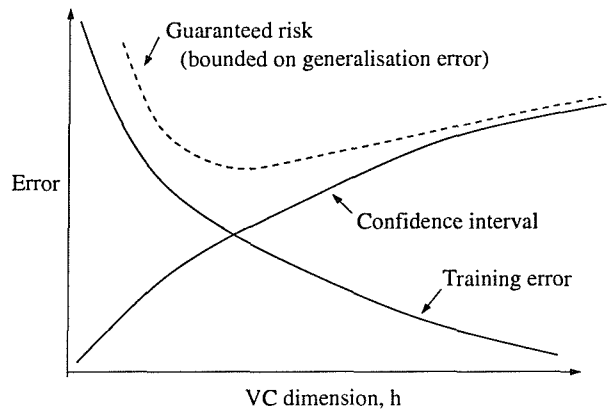


Figure 6.2: Relationship between training error, confidence interval and guaranteed risk. The bound on the risk is the sum of the empirical risk and of the confidence interval. The smallest bound of the risk is achieved on some appropriate element of the structure S defined on the set of functions.

guarantee a small value of the expected risk.

The above then leads to the following principle (Vapnik 1998) for controlling the generalisation ability of learning machines:

> *To achieve the smallest bound on the test error by controlling (minimising) the training error, the machine (the set of functions) with the smallest VC dimension should be used.*

The SRM inductive principle describes a general method of *capacity control*, whereby the amount of data is taken into account.

For the quadratic loss function, SLT provides an upper bound estimation for prediction risk:

$$\text{Prediction risk} \leq \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}_i\left(\mathbf{x},\mathbf{w}\right)\right)^2 \left(\frac{1}{1 - K_1\sqrt{\frac{h\ln(2N)\ln(h!)+K_2}{N}}}\right) \tag{6.4}$$

6.3 Support vector machines

The SVM is a universal constructive learning procedure (the SV method describes a general concept of a learning machine) based on SLT, the formulation of which embodies the SRM principle. The foundations of Support Vector Machines (SVMs) have been developed by Vapnik (Vapnik 1998). A schematic of the method is shown in Figure 6.3. The SVM maps the input vectors x into a high-dimensional *feature space* (Z) through some non-linear mapping chosen a priori. A linear approximation in the feature space with parameters w is used to determine the output. Thus, unlike conventional statistical and neural methods, the SVM approach does not attempt to control model complexity by keeping the number of features small.

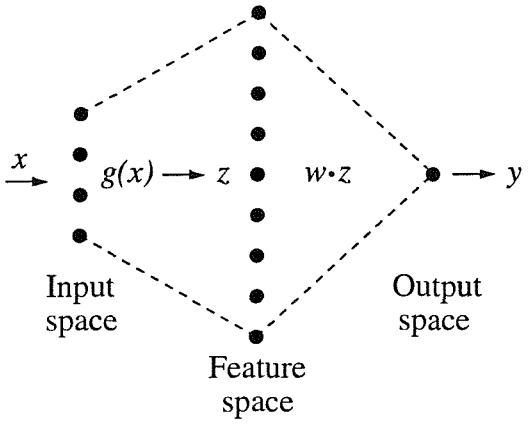


Figure 6.3: Illustration of the mapping implemented by the support vector machine.

The SVM overcomes two problems in its design (Vapnik 1998): the *conceptual problem* of how to control the complexity of the set of approximating functions in a high-dimensional space (in order to provide good generalisation ability) is solved by using penalised linear estimators with a large number of basis functions. The resulting SVM approach results in a constrained quadratic optimisation formulation of the learning problem. The *computational problem* of how to perform numerical optimisation (i.e. solve the quadratic optimisation problem) in a high-dimensional space is solved by taking advantage of the dual kernel representation of linear functions. In SVM, capacity control can be controlled effectively through the regularisation functional used (see section 6.4).

In summary, the SVM approach is seen to be characterised by the following:

- New implementation of the SRM inductive principle.
- Input samples mapped onto a high-dimensional (feature) space by using a set of non-linear basis functions (mapping) defined a priori.
- Linear functions with constraints on complexity used to approximate the data in high-dimensional space.
- Duality theory of optimisation used to make estimation of model parameters in a high-dimensional space computationally tractable.
- Characterisation of complexity independently from dimensionality.
- Non-linear feature selection.

6.3.1 Loss functions

Under conditions of normal additive noise, the quadratic loss function:

$$L_{quad} = (y, f(\mathbf{x}, \mathbf{w})) = (y - f(\mathbf{x}, \mathbf{w}))^2 \tag{6.5}$$

provides an efficient (best unbiased) estimator of the regression function.

However, least-squares estimators are sensitive to the presence of outliers and may perform poorly when the underlying distribution of the additive noise has a long tail. To overcome these limitations, a robust estimator that is insensitive to small changes in the model is required. In the ϵ-insensitive loss function, defined by:

$$L_\epsilon(y) = \begin{cases} 0, & \text{for } |f(\mathbf{x}, \mathbf{w}) - y| < \epsilon, \\ |f(\mathbf{x}, \mathbf{w}) - y| - \epsilon & \text{otherwise} \end{cases} \tag{6.6}$$

the loss is equal to 0 if the discrepancy between the predicted and the observed values is less than ϵ. Then, ϵ represents in some way the desired resolution.

The L_ϵ loss function is an attractive choice in the implementation of an SVM since choosing the value of ϵ controls the number of support vectors, introducing sparseness in the solution, unlike the quadratic cost function, where all the data points will be support vectors.

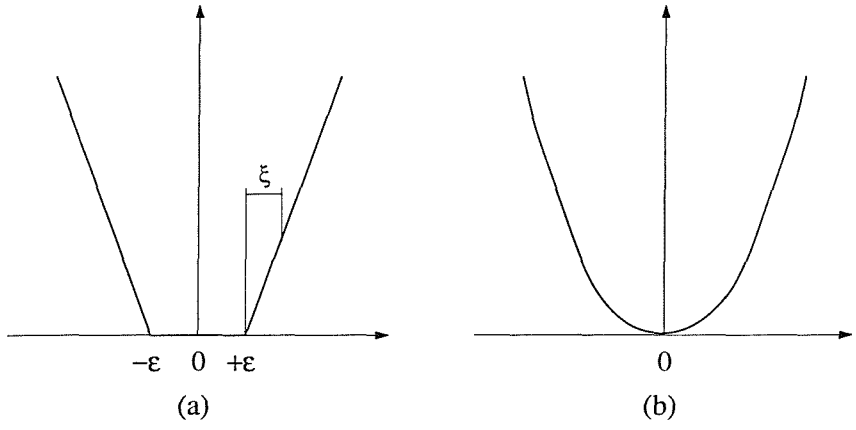


Figure 6.4: Typical loss functions used in the implementation of a SVM: (a) ϵ-insensitive and (b) quadratic loss functions.

6.4 Support vector regression

The generalisation of the SV method for function approximation as described in (Vapnik 1995; Vapnik et al. 1997) exploits the idea of computing a linear function in high-dimensional feature space, corresponding to a non-linear function in the input data space.

6.4.1 Linear support vector regression

In the SV method, estimation of the optimal (linear) regression function of the form:

$$f(\mathbf{x}, \mathbf{w}) = (\mathbf{w} \cdot \mathbf{x}) + b \tag{6.7}$$

with precision ϵ, may be determined by minimising:

$$\frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{N}|y_i - f(\mathbf{x}_i)|_\epsilon \tag{6.8}$$

Written as a constrained optimisation problem, this is equivalent to the problem of finding the pair $\mathbf{w}, b$ that minimises the quantity defined by the *slack variables* $\xi_i, \xi_i^*, i = 1, \ldots, N$ (the introduction of slack variables into the constraints, assumes an extra cost for errors). The optimal regression function is then given by the minimum of the cost functional:

$$\Phi(\mathbf{w}, \xi^*, \xi) = \frac{1}{2}||\mathbf{w}||^2 + C\left(\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N}\xi_i^*\right) \tag{6.9}$$

subject to the following constraints,

$$\begin{aligned}
&((\mathbf{w}\cdot\mathbf{x}_i) + b) - y_i \leq \epsilon + \xi_i \\
&y_i - (\mathbf{w}\cdot\mathbf{x}_i) + b \leq \epsilon + \xi_i^* \\
&\xi_i, \xi_i^* \geq 0
\end{aligned}$$

for all $i = 1, ..., N$.

In the above, C is the smoothing parameter which can be determined by resampling methods (e.g. cross-validation), and ξ, ξ^* are slack variables representing upper and lower errors on the outputs of the model (the second term, thus represents a bound on the training error, thus including an implicit form of regularisation). It has been shown that for certain kernels the parameter C can be directly related to a regularisation parameter (Smola et al. 1998).

The quadratic loss function produces a solution which is equivalent to zero order regularisation, where the regularisation parameter $\lambda = \frac{1}{2C}$.

6.4.2 Kernel functions

Generalisation of the SV regression method to non-linear regression is performed using kernel functions:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}_j) \tag{6.10}$$

where $K(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function performing the non-linear mapping into feature space. A mapping into a high dimensional space constructed by the use of reproducing kernels, enables operations (linear regression) to be performed in the input space rather than the potentially high dimensional feature space, addressing the curse of dimensionality.

To construct polynomials of degree d, one can use the following generating kernel:

$$K(\mathbf{x}_i, \mathbf{x}_j) = [(\mathbf{x}_i, \mathbf{x}_j) + 1]^d \tag{6.11}$$

Infinite splines (Vapnik 1998) comprise a practical kernel choice in the implementation of a SVM. An infinite spline kernel of order p, which passes through the origin, is defined on the interval $[0, 1)$ by:

$$K(u, v) = \int_0^1 (u - \tau)_+^p (v - \tau)_+^p \, \mathrm{d}\tau \tag{6.12}$$

In the case of a first order ($p = 1$) spline, the kernel is given by

$$K(u, v) = uv + \frac{uv}{2} \min(u, v) - \frac{1}{6} (\min(u, v))^3 \tag{6.13}$$

The formulation of equation 6.12 constrains the univariate spline to pass through the origin ensuring a unique expansion in the ANOVA term is obtained.

Multidimensional kernels can be obtained by forming tensor products of the univariate kernels (Vapnik 1998). A multivariate ANOVA spline kernel is given by the tensor product of univariate spline kernels plus a bias term,

$$K(\mathbf{u}, \mathbf{v}) = \prod_{d=1}^{n} \{1 + K(u_d, v_d)\} \tag{6.14}$$

For example, the ANOVA expansion for a three dimensional input vector is given by:

$$\begin{aligned} K(\mathbf{u}, \mathbf{v}) &= \prod_{i=1}^{3} \{1 + K(u_d, v_d)\} \\ &= 1 + g(u_1, v_1) + g(u_2, v_2) + g(u_3, v_3) \\ &+ g(u_1, v_1) g(u_2, v_2) + g(u_1, v_1) g(u_3, v_3) + g(u_2, v_2) g(u_3, v_3) \\ &+ g(u_1, v_1) g(u_2, v_2) g(u_3, v_3) \end{aligned}$$

For spline kernels, the computational complexity of the solution depends on the number of support vectors that are necessary to adequately approximate the desired function with ϵ-accuracy, rather than on the dimensionality of the space or on the number of knots.

6.4.3 Non-linear support vector regression

In the Lagrangian formulation of the problem, the constraints in section 6.4.1 are replaced by constraints on the Lagrange multipliers themselves, the training data only appearing in the form of dot products between vectors.

From the preceding sections it is clear that the implementation of a non-linear regression approximation to the data is achieved by replacing the dot product $(\mathbf{x}_i, \mathbf{x}_j)$ with $K(\mathbf{x}_i, \mathbf{x}_j)$.

For the case of using the ϵ-insensitive loss function, the solution to the constrained optimisation problem is obtained by maximising $W(\alpha, \alpha^*)$ with respect to the Lagrange multipliers:

$$W(\alpha, \alpha^*) = -\sum_{i=1}^{N} \epsilon_i (\alpha_i^* + \alpha_i) + \sum_{i=1}^{N} y_i (\alpha_i^* - \alpha_i) - \frac{1}{2} \sum_{i,j=1}^{N} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(\mathbf{x}_i, \mathbf{x}_j) \tag{6.15}$$

subject to a new set of constraints that include the parameter C:

$$0 \leq \alpha_i \leq C \ , \ i = 1, \ldots, N$$
$$0 \leq \alpha_i^* \leq C \ , \ i = 1, \ldots, N$$
$$\sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0$$

and satisfying:

$$\alpha_i \alpha_i^* = 0 \ , \ i = 1, \ldots, N. \tag{6.16}$$

Solving equation 6.15 subject to the constraints in equations 6.4.3 determines the Lagrange multipliers α_i, α_i^*, from which the regression estimate takes the form:

$$f(\mathbf{x}) = \sum_{SVs} (\hat{\alpha}_i - \hat{\alpha}_i^*) K(\mathbf{x}_i, \mathbf{x}) + \hat{b} \tag{6.17}$$

Due to the nature of the quadratic programming problem, only a number of coefficients $\alpha_i - \alpha_i^*$ will be non-zero. The input data points $\mathbf{x}_i$ associated with these are the *support vectors*.

For $\epsilon = 0$, the optimisation problem is simplified to:

$$\min_{\beta} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \beta_i \beta_j K(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^{N} y_i \beta_i \tag{6.18}$$

with constraints,

$$-C \leq \beta_i \leq C \ , \ i = 1, \ldots, N$$
$$\sum_{i=1}^{N} \beta_i = 0$$

Using a quadratic loss function, the solution is given by maximising the quadratic form:

$$W(\alpha, \alpha^*) = -\frac{1}{2} \sum_{i,j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{N} y_i (\alpha_i - \alpha_i^*) - \frac{1}{2C} \sum_{i=1}^{N} (\alpha_i^2 + \alpha_i^{*2}) \tag{6.19}$$

simplifying, yields:

$$\min_{\beta} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \beta_i \beta_j K(\mathbf{x}_i \cdot \mathbf{x}_j) - \sum_{i=1}^{N} y_i \beta_i + \frac{1}{2C} \sum_{i=1}^{N} \beta_i^2 \tag{6.20}$$

with constraints:

$$\sum_{i=1}^{N} \beta_i = 0 \tag{6.21}$$

The choice of the kernel function, together with the values of the parameters C and ϵ, determine the smoothness properties of the solution and should reflect prior knowledge of the data. The number of support vectors depends on both C and ϵ.

6.5 Support vector parsimonious ANOVA modelling

The SUPANOVA (SUpport vector Parsimonious ANOVA) technique (Gunn and Brown 1999) selects a parsimonious model representation by extracting a small set of terms from the complete ANOVA representation of equation 5.7. This is achieved by decomposing the non-linear modelling problem into three stages, as shown in Figure 6.5. This technique contrasts with the neurofuzzy and other parsimonious techniques, in that it aims to find a full model, from which a subsequent sub-selection of the significant terms is performed. Consequently, the sparse basis selection avoids problems associated with local minima entrapment during the construction process. For the purposes of determining the regularisation parameter C and

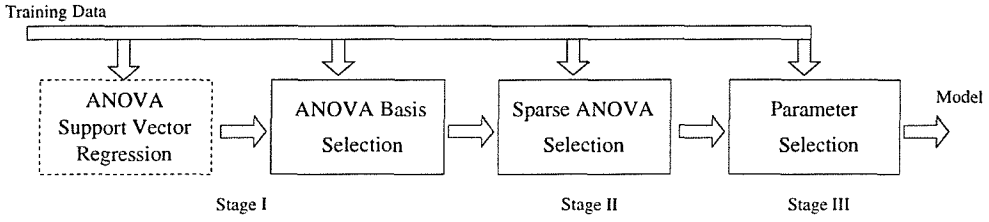


Figure 6.5: Schematic representation of the SUPANOVA technique: the first stage is used to select a complete ANOVA basis from which the second stage selects a sparse subset that maintains good accuracy. Finally, a new model is constructed using a sparse representation.

obtaining an estimate of the generalisation performance of the machine, the data is partitioned and resampled (using 8-fold cross validation) as illustrated in Figure 6.6.

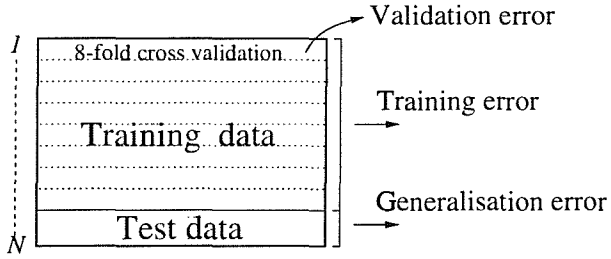


Figure 6.6: Illustration of the data partitioning and resampling performed within the SUPANOVA framework.

6.5.1 ANOVA basis selection

The initial stage in the SUPANOVA framework is to obtain a complete ANOVA basis that is a suitable representation for the model. This is implemented by approximating the data with a SVM with an ANOVA spline kernel. This produces a complete ANOVA expansion from which the 2^N components of equation 5.7 can be extracted.

The model will consist of a product of piecewise cubic splines with a finite number of knots located at a subset of the data points (support vectors).

At this stage, the estimate of the capacity control parameter C is determined using 8-fold cross validation, by searching for a local minimum of the validation error. The value of the C parameter being determined by a gradient descent algorithm (Press et al. 1992).

6.5.2 Sparse basis selection

Interpretation of the complete ANOVA representation obtained in the first stage is difficult. Furthermore the ANOVA expansion may determine an overly flexible approximation to the data. It is then desirable to select a sparse basis from an overdetermined one, by means of extracting the most significant terms in the ANOVA expansion. For such purposes, the SUPANOVA technique borrows a technique (Chen 1995) used in the wavelet community, whereby a trade-off between the error in the approximation and the sparseness of the representation is sought:

$$\min Error + \lambda\, Sparsity$$

with λ controlling the trade-off. To enforce continuity, the implementation of the technique retains the loss function that was employed in the basis selection stage.

The optimisation problem is given by

$$\min_{c} L\left(y, \Phi c\right) + \lambda S\left(c\right) \quad \text{subject to} \quad c_i \geq 0\,, \lambda \geq 0$$

where L is the loss function, Φ is the ANOVA basis obtained from the first stage, S is the sparseness measure and c is a vector of coefficients associated with each of the terms in the ANOVA basis. The goal in selecting a sparse representation is to minimise the number of non-zero coefficients, c_i.

By characterising sparseness by means of a p-norm:

$$S\left(c\right) = ||c||_p$$

with a value of $p = 1$, employed in basis pursuit techniques (Chen 1995), produces a sparse solution which can be implemented by solving a quadratic program.

The sparseness parameter λ is estimated by selecting the value giving the nearest loss to the error estimate obtained on the validation sets employed in the first stage.

6.5.3 Parameter selection

The two stages together can be considered as a sparse kernel selection method. In the final stage, the sparse kernel representation of stage II is deployed within an SVM to produce a sparse ANOVA model.

6.6 Conclusions

This chapter has presented a summary of the statistical learning theory adopted by the two techniques used in this work in the determination of parsimonious models from small data samples. The common interpretation of Occam's razor (i.e. in terms of a simplicity concept) given in previous chapters differs from the assertion that emerges from SLT (capacity concept). The distinction between the two concepts are highlighted in (Vapnik 1998).

Both neurofuzzy and SV regression methods realise the SRM principle (although there is no clear correspondence between SRM and greedy optimisation techniques (Cherkassky and Mulier 1998)). However, the two techniques differ in the implementation of the capacity control: by model selection and by regularisation. Under SRM a set of approximating functions is specified a priori, whereas under the greedy nature of model selection algorithms, approximating functions are considered as dictated by the data. In (Vapnik 1998) it is stated that while such an approach can be useful for data fitting and data analysis, there is no theory and little empirical evidence to suggest its usefulness as an inductive principle for predictive learning.

Training a SVM is seen to consist in a (computationally efficient) quadratic programming problem, guaranteed to find a global solution (minima of the error surface). This is in contrast to the case of general neural networks, which in (Vapnik 1998) are referred to as "not well-controlled learning machines". The implementation of the SUPANOVA framework avoids the problem of local minima entrapment which is seen to afflict the majority of model construction algorithms, including the ASMOD algorithm (used in the neurofuzzy approach).

Overall, the main features of SVM, e.g. embodiment of the SRM principle, robustness to outliers, overcoming the curse of dimensionality, etc., characterise the SVM as a powerful technique for determining models possessing good generalisation properties. In subsequent chapters the empirical performance of SVM will be investigated and compared with the semi-parametric nature of the neurofuzzy framework, on the materials data sets that are the focus of this work.

Chapter 7

Data Modelling of Processing-Property Relationships in an Al-Mg-Li Powder Metallurgy Alloy System

7.1 Introduction

In this chapter a commercial data set comprising a set of production runs of an Al-Mg-Li powder metallurgy (P/M) alloy system has been investigated. The manufacturers wished to assess how the tensile properties of the alloys fabricated were influenced by the particular processing conditions and compositional levels. An analysis of the data precedes the results of applying the neurofuzzy and SVM methods. The performance of the neurofuzzy framework is assessed and compared with a simple MLR analysis, the modelling inferences attained being assessed by both conditioning diagnostics and sensitivity analyses. The results of the SVM methods are then compared with those determined by the ASMOD construction algorithm in terms of the empirical approximation abilities and the ANOVA representation inferred by both the neurofuzzy and SUPANOVA frameworks, by means of multiple model runs determined from resampling the complete data set.

7.2 Powder metallurgy

Over the last two decades, P/M alloy fabrication has been considered as an alternative to traditional ingot metallurgy (I/M) routes in the production of specialised aerospace structural components. Compared to traditional I/M fabrication the advantages of a P/M route include the fabrication of (fine) composite materials with controlled microstructures, higher strengths

attained without a sacrifice of corrosion resistance (which accompanies the presence of precipitates), property isotropy, good elevated temperature properties and stability, and a potential for near-net-shape forming. Currently, the major disadvantage of the P/M route remains the costs associated with the production of alloys for a very small market. A review and description of P/M technology and the processing steps involved in the fabrication of high-strength Al-alloys can be found in several publications, e.g. see (Pickens 1981; Hatch 1984; Hildeman and Koczak 1989).

7.2.1 Mechanical alloying

The production of mechanically alloyed (MA) powders (Gilman and Benjamin 1983) consists of grinding together the constituent elements/master alloys as powders, subjecting these to large compressive forces in a high speed mill. Through control of repeated solid state (cold) welding and subsequent micro-fracturing of the constituent powder particles in the ball mill, the resultant fine powder particles attain the chemical composition of the original powder charge. Since the process takes place entirely in the solid state, limitations imposed by the phase diagrams (solubility limits) can be overcome, with the desired mixing of the elements achieved without melting taking place. Concurrently with the mechanical aspects of alloying, dissolution of solute elements is facilitated by heating and the presence of lattice defects and short diffusion distances (Gilman 1983).

In order to prevent the powders welding together to form one solid mass during mechanical alloying, an organic lubricant is added as a process control agent (*pca*), so that a dynamic balance between fracturing and welding can be established (welding tends to increase average particle size, while fracturing tends to decrease particle size). In delaying cold welding, the *pca* allows the powder particles to work harden and fracture.

The powder particle size will reach a steady-state distribution which is dependent upon the composition of the system and processing variables (e.g. powder size distribution, processing time, etc.) (Benjamin and Violin 1974; Gilman and Nix 1981) with much of the *pca* eventually trapped between the cold welded layers of the composite particles, decreasing the amount of *pca* available to delay cold welding. This steady-state processing is associated with reaching saturation hardness and constant particle size distribution, although structural refinement continues (Benjamin and Violin 1974), described by an approximate logarithmic function of time, depending on the mechanical energy put into the MA process and work hardening properties of the materials being processed. Aikin and Courtney (1993) have used a discretised form of a fission-fusion (i.e. fracture-welding) equation to model the particle size distributions during the mechanical milling process. Zhang et al. (1999) have used standard back-propagation training of an MLP to predict the amount of *pca* required for a particular mean particle size and milling duration[1].

[1]In this study 12 data pairs are used for training. Model selection and empirical performance of different

Once a stable particle size is achieved, the powder is de-gassed and consolidated using conventional P/M methods. The hot pressed billet may then be subsequently forged. Aluminium oxide (γ-Al_2O_3) particles, originating from prior particle boundaries of the original powder (i.e. the surfaces of the original aluminium/master alloy powders) and those generated by the oxidation of fractured particle surfaces exposed during MA are embedded into the matrix of the composite aluminium powders by the repeated fracturing and cold welding of the powder charge. During processing the *pca* breaks down to form carbon, hydrogen and oxygen. While the latter two elements are removed by subsequent degassing (hot pressing), carbon which is embedded amongst the powder particles during MA, forms aluminium carbide (Al_4C_3).

Particle fracturing will be assisted to some extent by the increased carbon and oxygen content resulting from the inclusion of the *pca*. The cold work imparted by ball-milling, much of which is retained after hot pressing, results in increased dislocation strengthening and upon consolidation a finer grain (and subgrain) size than obtained by working a larger ingot.

The high angle boundary pinning imparted by the dispersoids develops a highly stabilised microstructure inhibiting static recrystallisation and grain growth. Boundaries corresponding to the surface of the powder particles (prior particle boundaries) are seen to have a higher concentration of dispersoids, when compared with other boundaries.

7.2.2 Powder metallurgy Al-Mg-Li alloys

From the previous section, the MA process can be understood to enhance strength, ductility and toughness, through the attainment of a fine homogeneous grain structure and distribution of dispersoids. Optimising processing conditions and parameters at the powder processing, de-gassing and consolidation stages such that the strength is derived from intrinsic dispersoid strengthening effects, will minimise the loss in strength caused by subsequent processing (Donachie and Gilman 1983). Dispersoid and grain strengthening, avoids the problems associated with undesirable precipitates that may form during ageing of heat treatable alloys.

In the 5xxx series, the MA route has been used to produce Al-4wt.% Mg alloys, e.g. (Pickens et al. 1981a), combining the strengthening mechanisms described in section 7.2.3. The requirements of the aerospace industry for lightweight alloys subsequently resulted in the development of Li containing alloys, such as Al-alloy 5091 (or AL-905XL) (Pickens 1985), where careful selection of the Al-Mg-Li base composition enables a single phase matrix to be achieved. The primary reason for the addition of Li is to reduce density and increase modulus of the alloy (Last and Garrett 1996). The increase in modulus may be due to the presence of Li in solid solution or in oxides. Low Li levels ($< 1.5\%$) are chosen to avoid embrittling precipitates that may form during ageing[2]. Since these alloys do not precipitate a second phase,

network architectures was inferred from 4 data pairs (*through trial-and-error, the final network structure has 8 hidden neurons* (Zhang et al. 1999)).

[2]The IN-905XL Al-alloy having 1.5 wt.% Li showed significant age hardening, i.e. large increase in strength and a decrease in ductility (Weber and Phillips 1993), which led to the refinement of the Li content (1.3wt.%)

they do not soften by overageing or dissolution of precipitates, thus imparting microstructural stability. With an 8% density saving, IN-905XL has been proposed as a lightweight substitute for Al-alloy 7075-T73. The nominal composition of the Al-alloy 5091 which has found use in structural airframe applications (e.g. see (Smith 1992)) is given in Table 7.1. Weber and Phillips (1993) report standard deviations for strength for this alloy as 10-17 MPa, typical of I/M Al-alloys.

Mg	Li	C	O
3.7-4.2	1.2-1.4	1.0-1.3	0.20-0.70

Table 7.1: Composition ranges (wt.%) for AA 5091.

Dispersoids can be formed during the MA stage and/or consolidation and thermomechanical processing stages and may be present not only as carbides and Al oxides, but also in the form of Li or Mg oxides or combinations thereof (Sugamata et al. 1998). In the case of solute Mg co-existing with solute Li, the former tends to be internally oxidised to MgO in preference to Li (Sugamata et al. 1998). Intermetallic dispersoids may also be present, e.g. Al_3Mg_2 (Vitiaz et al. 1997). In 8090 P/M material, coarse lithium carbonate (Li_2CO_3) dispersoids were identified to form during the fabrication process (Hunt 1992). Preferably, the dispersoid levels are kept as low as is possible, consistent with desired strength. In particular, the oxygen content is generally controlled to be present in a small but effective amount (preferably not exceeding 0.4-0.5%, so that carbon levels exceed oxygen levels with a ratio greater than 2:1 (Donachie and Gilman 1986)) for increased strength and stability. Alloys with high oxygen concentrations (>1 wt.%) will exhibit poor ductility.

7.2.3 Strengthening mechanisms

In Al-Mg(-Li) alloys a combination of solid solution, dispersion, substructural (high dislocation density and fine subgrain size) and fine grain size are regarded as the operative strengthening mechanisms (Pickens 1981). The major strengthening effect has been attributed to the development of a submicron grain size, whilst solid solution and dispersion strengthening mechanism contributions to the yield strength are considered as important secondary additive components (Last and Garrett 1996)[3]. Alloying additions (Mg and Li) present in solid solution after powder processing will contribute a solution strengthening component. Due to its relatively high solubility Mg, is the most effective solution strengthener of the Al-matrix (Polmear

and specification of the Al-alloy AL-905XL. Despite a very slight supersaturation of Mg/Li in some grains, which upon extended ageing was seen to form a low volume fraction precipitate (probably δ'), Bridges et al. (1985) did not detect a hardening response upon ageing. Characteristics of age-hardening of Al-Mg-Li-C-O alloys have been studied by Papazian and Gilman (1990).

[3]In this particular study, the effect of Li, in particular how this alloying addition is distributed between the matrix and the oxides, remained unanswered.

1981). The Mg solute strengthening contribution in Al-alloys has been reported as approximately 18 MPa/wt% (Polmear 1981) and 13.8 MPa/wt% in the IN-905XL Al-alloy (Mukai et al. 1995). Benjamin and Schelleng (1981) report that in a MA Al-Mg alloy, a 4 wt.% Mg addition increased the yield stress by approximately 150 MPa, whilst Li has been found to give approximately a 24 MPa strengthening contribution per wt.% in binary Al-Li alloys (Noble et al. 1982). In SAP-type (sintered aluminium powder) materials, strengthening due to grain and subgrain effects and oxide dispersion contributions have been considered as additive (Benjamin and Bomford 1977)[4].

From section 7.2.1, the *pca*, as well as being a weld-controlling agent, is also a dispersoid contributing agent, facilitating the homogeneous dispersion of both carbides and oxides: the fine dispersion of γ-Al_2O_3 (Singer et al. 1980) and Al_4C_3 particles will provide the major dispersoid strengthening component (through an Orowan type mechanism).

Tensile properties are seen to be dependent on the volume fraction of the dispersoids efficiently embedded within the matrix. As many of the dispersoids are located on the grain boundaries (i.e. they do not contribute to the intragranular hardening mechanism), grain size is considered to provide the main contribution to $\sigma_{0.2}$ (Michot and Champier 1991). Results obtained for a pure Al-C-O alloy have showed a linear increase of the yield and tensile strengths with an increasing volume fraction of dispersoids, whilst ductility was seen to be a decreasing function of volume fraction (Michot and Champier 1991). Benjamin and Bomford (1977) noticed that the correlation between tensile strength and volume percent dispersoid was stronger for the combined volume fraction of Al_2O_3 and C than for Al_2O_3 alone. An empirical regression for the ultimate tensile strength used by Pickens et al. (1981b) establishes a 20 MPa increment per unit wt.% of C additions. Styles and Pitcher (1998) considered increasing carbon and solute contents to both result in almost linear increases in proof stress in a range of MA Al-Mg-Li-C MMCs, with Hall-Petch type relationships being used to predict the flow stress in these alloys. Mukai et al. (1995) have determined analytical expressions for the strengthening components present in the IN-905XL alloy, though experimental values agreed with the analytical models only over a limited range of alloys. England et al. (1988) assessed strengthening contributions due to grain boundary and dispersoid strengthening in P/M Al-Mg-Zr alloys using both Hall-Petch and Orowan equations; Pythagorean rules of addition yielded a better accuracy than simple linear additive components.

In a series of experimental trials performed on Al-Li-Mg alloys Narayanan et al. (1983) concluded that Mg additions contributed to a substantial solution strengthening component, whilst Li exhibited the least amount of solution hardening. Gilman et al. (1985) found that the tensile properties of three Al-4Mg-1.5Li alloys appeared similar to those of non-lithium containing Al-4Mg alloys. Narayanan et al. (1983) considered the fracture processes in P/M alloys to be dominated by the presence of inclusions and weak interparticle interfaces, which

[4]Typically, SAP alloys require greater oxide levels to achieve the same level of strength attained by MA materials, with the latter also exhibiting greater ductility and workability.

were believed to account for poor correlations between ductility and composition. A fractographic analysis identified deleterious features such as foreign particles that promoted crack nucleation and failure along prior powder particle boundaries (the latter generally recognised as a consequence of poor interparticle bonding, caused by insufficient break-up and redistribution of the oxide film originally present on powder particle surfaces). Improvements in ductility through a P/M approach cannot be achieved by grain refinement alone, but must be sought also through powder processing and consolidation procedures. Lower ductility will be exhibited if a critical amount of work is not achieved during forming as particle boundaries will not be bonded effectively (failure in these alloys has been proposed to initiate at oxides and carbides present on grain boundaries). Benjamin and Schelleng (1981) found that in Al-Mg alloys, both tensile and yield strengths were found to increase with increasing the Mg content, concomitant with a decrease in ductility.

Hence, it can be seen that although some physical understanding of composition/microstructure/property relationships exists, existing models of these relationships tend to be rather simplistic linear regressions, or based upon microstructural features that are difficult to assess, rather than simple processing variables.

7.3 The AA5091 data set

The data set investigated in this study was based on the specifications set out for the 5091 Al-alloy, whose nominal composition range was that given in Table 7.1. Design criteria moved away from these specifications as the development programme evolved, driven by a combination of trial and error, and cost constraints. The proprietary data set, provided by Aerospace Metal Composites (AMC) U.K., comprised six process variables, two of which described processing conditions (PV_c corresponding to forging temperature and $PV_{b/a}$ corresponding to a measure of the extent of mechanical alloying), while the remainder were weight percentages (wt.%) of the four compositional elements (Li, Mg, C, O) measured. In addition to these process variables, each alloy was designated with a *billet number* identifier, and, depending on the final product form, a label indicating whether the product was of *die forged* or *plate* type.

Tensile properties, 0.2% proof stress ($\sigma_{0.2}$), ultimate tensile strength (uts), percentage elongation ($\%el.$) and Young's modulus (E) were measured for each billet[5]. A differing number of tensile tests were repeated for each individual billet.

Although experimental in origin, the heuristic design criteria employed rendered the statistical properties of the data set much more characteristic of typical commercial alloy production runs, rather than those typically associated with carefully designed experimental programmes. Several alloys were observed to differ only in terms of a single process condition (PV_c), as the consolidated billets were sectioned and then forged at different temperatures. The data set

[5]Several modulus (E) tests were not carried out for several alloys and hence constituted missing entries.

contained four alloys with certain compositional elements omitted, intended to provide an indication of the baseline tensile properties attainable for particular boundary process conditions, e.g. commercial purity MA aluminium, etc.

7.4 Data pre-processing

The original supplied data set comprised a total of 264 data pairs (i.e. input and corresponding outputs), corresponding to 71 different billets, of which 67 were of *plate*, the remainder *die forged*. As the latter product type was seen to exhibit significant differences in tensile properties for identical processing conditions to those of an equivalent plate type, the analysis was limited to the plate data.

A de-duplication of the plate data set reduced the number of training patterns to 56 data pairs, corresponding to the different processing and compositional conditions. Mean tensile properties were determined for identical input patterns and used as the observed output values. The estimation of a *pure error* component, shown in Table 7.2, obtained from instances of multiple tests on individual materials provided an estimate of the experimental scatter in the tensile properties.

Property	$\sigma_{0.2}$ (MPa^2)	uts (MPa^2)	$\%el.$ ($\%^2$)
SSPE	1577	1154	4.66

Table 7.2: Summary of the pure error sum-of-squares (SSPE) components for the tensile properties considered.

For the MLR analysis the input variables were standardised to have zero mean, unit variance, while for the adaptive methods these were standardised to lie between ± 1 for the neurofuzzy framework and in the [0,1] interval for the SVM. Inferences were also obtained from standardisation of the output quantities (zero mean, unit variance). To preserve commercial confidentiality, all plots involving input variables are presented in normalised form, while predictions are presented in their original ranges.

7.5 Data analysis

As initial alloy production was based on AA5091, a number of billets show limited variance in the processing conditions. However, once the design criteria evolved, a wider range of processing conditions were used, as reflected in the statistical properties and data distributions shown in Figures A.1 to A.4. The mean and standard deviations for the process variables (inputs) and properties measured (outputs) are summarised in Tables A.1 and A.2, from which the limited variance in certain compositional elements (e.g. Mg, Li) is evident.

The values of the simple correlation coefficients, r_{ij}, among the variates are summarised below in the correlation matrices, $\mathbf{r}_{XX}$, $\mathbf{r}_{YY}$ and the vectors of correlation coefficients, $\mathbf{r}_{YX}$, for uts, $\sigma_{0.2}$ and $\%el$.

$$\mathbf{r}_{XX} = \begin{bmatrix} & PV_c & PV_{b/a} & \text{Li} & \text{Mg} & \text{C} & \text{O} \\ PV_c & 1 & .129 & -.500 & -.331 & .213 & .002 \\ PV_{b/a} & .129 & 1 & -.091 & .010 & .734 & .151 \\ \text{Li} & -.500 & -.091 & 1 & .738 & -.187 & .108 \\ \text{Mg} & -.331 & -.010 & .738 & 1 & -.155 & .080 \\ \text{C} & .213 & .733 & -.187 & -.155 & 1 & .415 \\ \text{O} & .002 & .151 & .108 & .080 & .415 & 1 \end{bmatrix}$$

$$\mathbf{r}_{YY} = \begin{bmatrix} & \sigma_{0.2} & uts & \%el. \\ \sigma_{0.2} & 1 & 0.967 & -0.659 \\ uts & 0.967 & 1 & -0.698 \\ \%el. & -0.659 & -0.698 & 1 \end{bmatrix}$$

$$\mathbf{r}_{YX} = \begin{bmatrix} PV_c & -.465 \\ PV_{b/a} & .504 \\ \text{Li} & .627 \\ \text{Mg} & .632 \\ \text{C} & .399 \\ \text{O} & .299 \end{bmatrix} \quad \mathbf{r}_{YX} = \begin{bmatrix} PV_c & -.439 \\ PV_{b/a} & .585 \\ \text{Li} & .489 \\ \text{Mg} & .572 \\ \text{C} & .445 \\ \text{O} & .280 \end{bmatrix} \quad \mathbf{r}_{YX} = \begin{bmatrix} PV_c & .246 \\ PV_{b/a} & -.488 \\ \text{Li} & -.201 \\ \text{Mg} & -.081 \\ \text{C} & -.589 \\ \text{O} & -.496 \end{bmatrix}$$

(uts) $(\sigma_{0.2})$ $(\%el.)$

For the reasons discussed in Chapter 4, the r_{ij} values are regarded simply as descriptive measures of the degree of linear association between the variates. The influence of the outliers corresponding to the boundary processing conditions and baseline alloys on the simple r_{yx} correlation coefficients is evident from the full residual plots. The effects (leverage) exerted by such cases on subsequent statistical inferences will be discussed in appropriate sections.

From the $\mathbf{r}_{XX}$ matrix, the largest correlations among the input variables are identified as those between $PV_{b/a}$ and C, and, Li and Mg, whilst from inspecting the $\mathbf{r}_{YX}$ vectors it is seen that correlations for Mg and Li with $\sigma_{0.2}$ are somewhat milder than those exhibited for uts, for which $PV_{b/a}$ exhibits a larger coefficient.

In order to elucidate some of the statistical properties of the data set, it is useful to understand the nature of the processing variables, how these were quantified and most importantly, justifications for their inclusion in subsequent analyses.

Forging temperature was characterised through measuring the temperature of the billets at which the material was pre-heated, and as such PV_c is a rough estimate of the actual temperature of the material during forging. A histogram for this variable (Figure A.1 (a)) shows

how PV_c exhibits little variation, with the large majority of the observations shared between two temperatures. The billet forging temperatures reflected preferred temperature ranges for Al-Mg-Li alloys (Pickens et al. 1981b; Gilman and Donachie 1987), reflecting optimum workability conditions without sacrificing strength.

A more accurate estimate of the forging temperature would be achieved by taking into consideration factors such as pre-heated temperature, transfer time, die temperatures, accounting for possible adiabatic heating, thermal mass of the object, final section thickness, etc. Typically, lower forging temperatures reduce the degree of grain growth and recrystallisation. In the alloy system considered, the submicron microstructure developed during the MA process is stabilised by the pinning effect of the dispersoids. From the full residuals shown in Figures A.5 and A.6 (a) it can be tentatively inferred that higher forging temperatures have an adverse effect on the strength levels obtained, consistent with our metallurgical understanding: as the forging temperature is decreased a finer grain size is retained and dislocations will no longer be annealed out. Process constraints (i.e. shaping and formability) issues, limit the practical use of low forging temperatures.

The second of the process variables, $PV_{b/a}$, characterises the amount of energy imparted to the powder during the MA: the extent of powder processing should be sufficient to attain a submicron microstructure and an efficient dispersion of carbides and oxides; suboptimal powder processing (underprocessing) will be reflected by poor tensile properties. As discussed in section 7.2.1, the addition of a carbon containing lubricant facilitates powder processing. Then, intuitively a dependency characterising $PV_{b/a}$ and C would not be surprising (Tarrant 1998). The extent of the correlation between these two variates for low values is clear from the corresponding pairwise scatterplots shown in Figure A.3. The lack of a dependency for both higher $PV_{b/a}$ and C levels suggests a change in powder processing conditions.

From Figures A.1 (b) and (e), it is seen that a significant number of billets have been fabricated using very similar powder processing conditions and a further comparison of the full residuals is shown in Figures A.5 (b) and (e), with Figures A.7 (b) and (e), revealing C and $PV_{b/a}$ to exhibit similar relationships with tensile properties. These scatterplots also show that comparable strength levels can be attained using lower carbon contents, though a further analysis of the data is required to elucidate the source of strengthening for such alloys (e.g. different strengthening mechanism, more efficient processing conditions, etc.). A close inspection of the data distributions reveals C to exhibit a greater scatter than $PV_{b/a}$. This may reflect C being inherently less controllable (i.e. having an inherently greater scatter). The similar dependencies exhibited with the tensile properties indicates that these two variables comprise similar information, and are descriptive of the same phenomena.

The effects of C additions on microstructural development may be understood in terms of two main effects: an extrinsic grain refinement effect and an intrinsic dispersoid strengthening contribution, both dependent on the efficiency of the powder processing (as characterised by $PV_{b/a}$). It was suggested (Tarrant 1998) that carbon concentrations below a certain level

contribute to the grain refinement, whilst in higher concentrations, dispersoid strengthening becomes more significant (grain refinement contributions for high C levels would then be comparable). It is clear that the efficiency of powder processing and in particular whether an optimal processing of the powder has been attained is a significant factor. However, only a full quantification of microstructural features such as grain size morphology, grain boundary features and dispersoid volume fractions for differing carbon levels would clarify this issue.

Process understanding (Tarrant 1998) indicated that the carbide dispersoids are not the most effective of the dispersion strengthening mechanisms attainable, though preferable to those induced by aluminium oxides, thus there is a manufacturing interest in processing conditions minimising C levels (providing these do not hinder the development of submicron grain sizes). Minimising carbon contents also reduces processing costs and results in less hazardous processing conditions[6].

Although higher C concentrations are seen to result in lower $\%el.$ values, this effect may be attributed to powder processing efficiency since $PV_{b/a}$ exhibits a similar, although weaker dependency, as reflected by the values of the correlation coefficients.

From the above, it becomes important to know how carbon concentrations are determined, for instance whether smaller amounts or less carbon rich *pca* are used, and establishing whether $PV_{b/a}$ determines the amount of *pca* (or vice versa). In particular, whether a similar amount of energy can be efficiently imparted during the powder processing, despite lower C levels (Figure A.3), and consequently whether work-hardening, grain refinement effects and comparable dispersoid volume fractions can be achieved. In addition the *pca* may be added at various times during the powder processing, based on conditions such as the ball-to-powder ratio, starting powder size, mill temperature, etc. However, since this information was unavailable it was problematic to establish whether all alloys have been processed in a similar way, in particular it is difficult to elucidate the effectiveness of varying the carbon levels, whilst imparting a similar amount of powder processing.

The distribution of both Li and Mg, although slightly broader than the nominal compositional range for AA5091, is seen to be limited with only a few alloys showing significant variations from mean values. Unsurprisingly, from an inspection of the data there is no tangible indication that these elements significantly affect tensile properties, although a comparison with the baseline alloys reveals that 5091 type alloys exhibit considerably higher strength levels. The cluster of outlying data evident in the distribution of Li were a result of a misjudgement in alloying levels. It is noteworthy that billet fabrication was characterised by a difficulty in controlling Mg levels.

An indication of oxide levels can be inferred through quantifying oxygen concentrations present in the alloys. However, due to costs associated with this characterisation, oxygen levels were not always measured and nominal values were used (Tarrant 1998). Unless a

[6]The fine powders required to give sufficiently small interparticle spacings present handling problems and are often pyrophoric. The problem of pyrophoricity being more severe when reactive alloying elements are present.

strong leverage is given to a limited number of observations at the higher O levels (in which case it could be speculated that a weak increasing dependency with strength and a decreasing trend for $\%el.$ are present) the residuals do not reveal any strong dependencies with the tensile properties. Any interactions characterising O and dispersoid-forming compositional elements may be explained by dispersoid formation during powder and subsequent processing.

The distribution and ranges of output properties investigated shown in Figures A.2 and A.4 show the wide range in tensile properties attained. The strong dependency between uts and $\sigma_{0.2}$ is evident, and the r_{yy} coefficients between these and $\%el.$, indicate a strong relationship between strength and ductility. Tests performed on very ductile specimens (exceeding 12.5%) had been recorded simply as 12.5 $\%el.$, due to equipment limitations, resulting in some biased measurements.

7.6 Multiple linear regression

From process understanding and the analysis of the data presented in the previous section, it was decided to perform a regression analysis retaining all six processing variables. The results of the MLR analysis for each of the tensile properties investigated are summarised in Table 7.3. From these results, it is seen that compared to the approximation obtained for uts, a considerably higher output variance is seen to characterise $\sigma_{0.2}$. A comparison of the proportions of explained variance determined on the normalised output quantities ($\hat{\sigma}^2_{\text{df,std}}$) shows that the linear models are capable of accounting for 85%, 81% and 54% in the variation of uts, $\sigma_{0.2}$ and $\%el.$ from the mean respectively.

	$\hat{\sigma}^2_N$	$\hat{\sigma}^2_{\text{df}}$	$\hat{\sigma}_{\text{df}}$	$\hat{\sigma}^2_{\text{df,std}}$	df	n_x
$\sigma_{0.2}$	1312	1500	38.37	0.19	7/6	6
uts	939	1073	33.66	0.15	7/6	6
$\%el.$	3.32	3.79	1.95	0.46	7/6	6

Table 7.3: Summary of the MLR results obtained for the three tensile properties. In these results $\hat{\sigma}_{\text{df}}$ is the unbiased root-mean-squared error and $\hat{\sigma}^2_{\text{df,std}}$ the variance estimate corresponding to a MLR performed on the output values normalised to have zero mean, unit variance.

Tables 7.4 to 7.6 summarise the standardised regression coefficients, associated parametric uncertainties and signal-to-noise estimates (τ) for each term in the model.

An inspection of the weights for the $\sigma_{0.2}$ model reveals increasing trends for all input variables, with exception of PV_c, which exhibits a decreasing trend. The largest weights are associated with the two process variables and Mg, while Li and O exhibit small magnitudes and smallest τ levels. The weight magnitude for C being of an intermediate value. As the weights typically exhibit similar uncertainties ($\hat{\sigma}_w$), the highest τ's are seen to correspond to terms with the largest weight magnitudes.

	w	$\hat{\sigma}_w$	τ
PV_c	−0.375	0.068	5.52
$PV_{b/a}$	0.437	0.091	4.80
Li	0.063	0.095	0.67
Mg	0.431	0.089	4.88
C	0.255	0.102	2.51
O	0.068	0.068	0.99

Table 7.4: Weights and parametric inferences inferred for the MLR model for $\sigma_{0.2}$.

	w	$\hat{\sigma}_w$	τ
PV_c	−0.310	0.061	5.19
$PV_{b/a}$	0.337	0.081	4.06
Li	0.293	0.085	3.30
Mg	0.351	0.079	4.64
C	0.300	0.091	3.41
O	0.064	0.061	1.12

Table 7.5: Weights and parametric inferences inferred for the MLR model for uts.

	w	$\hat{\sigma}_w$	τ
PV_c	0.294	0.107	2.76
$PV_{b/a}$	−0.217	0.143	1.52
Li	−0.220	0.149	1.50
Mg	0.146	0.138	1.05
C	−0.390	0.159	2.46
O	−0.289	0.107	2.71

Table 7.6: Weights and parametric inferences inferred for the MLR model for $\%el.$

Similar trends were exhibited for uts, though the magnitude of the corresponding weights (with the exception of O, which retains a very small value) are much more comparable, particularly the weight associated with Li. Thus, the signal-to-noise parameters are seen to attain comparable values.

For $\%el.$, weight values indicate increasing trends for PV_c and Mg, with the remaining variables characterised by decreasing trends, with C exhibiting the largest magnitude.

A comparison between the r_{yx} correlation coefficient between Mg and $\%el.$ and the corresponding weight in the MLR model, reveals how the formulation of a model can result in different inferences[7]. The small τ associated with the w_{Mg} term in the model (seen to be

[7]Note also that the weight values differ from the values of the simple correlation coefficients r_{yx}. The only

the smallest) shows how the Mg contribution is poorly determined. As such, whether reliable inferences can be obtained from the MLR models remains questionable: small τ levels indicate large uncertainty, which in the presence of very small weight magnitudes limits reliable inference.

To further understand the statistical properties of the data set and infer whether high parametric uncertainties and hence small signal-to-noise ratios are attributable to data conditioning problems, the condition of the design matrix, $\mathbf{X}$, was determined[8]. An SVD of $\mathbf{X}$ resulted in the singular values shown in Table 7.7, from which the smallest singular value, μ_6, is seen to be significantly larger that zero. Further, the condition number $C(\mathbf{X}) \simeq 3.5$, does not indicate the presence of any significant ill-conditioning and hence data weaknesses arising from near-linear dependencies among the variates. The variance-decomposition of $\mathbf{X}$, shown in Table 7.8, does however indicate some variate involvement in the weight uncertainties associated with $PV_{b/a}$ and C, although the small magnitude of the respective condition index, η_6, limits variance inflation in these weights.

Singular Values
$\mu_1 = 11.17$
$\mu_2 = 9.87$
$\mu_3 = 6.83$
$\mu_4 = 6.14$
$\mu_5 = 3.64$
$\mu_6 = 3.18$

Table 7.7: Singular values of the design matrix, $\mathbf{X}$.

Condition	Proportions of					
Index η	$\sigma^2_{w_{PV_c}}$	$\sigma^2_{w_{PV_{b/a}}}$	$\sigma^2_{w_{\mathrm{Li}}}$	$\sigma^2_{w_{\mathrm{Mg}}}$	$\sigma^2_{w_{\mathrm{C}}}$	$\sigma^2_{w_{\mathrm{O}}}$
1	0.061	0.022	0.044	0.040	0.028	0.004
1.132	0.013	0.055	0.028	0.035	0.045	0.084
1.637	0.003	0.129	0.001	0.012	0.002	0.601
1.819	0.779	0.002	0.014	0.144	0.006	0.000
3.067	0.108	0.075	0.791	0.498	0.136	0.096
3.515	0.036	0.718	0.122	0.270	0.782	0.215

Table 7.8: Condition indices and associated variance-decompositon proportions for $\mathbf{X}$.

way in which regression/correlation weights/coefficients will result in identical values is if the mean and standard deviation of both variates are identical (Edwards 1984).

[8]In this and all other subsequent conditioning diagnostics determined in subsequent chapters, inferences were drawn from normalised data sets.

The variance-decomposition proportions also revealed a marginal Li-Mg variate involvement (corresponding to the weight uncertainties associated with η_5), possibly induced by one or more of the outlying cases noted previously.

As the weight uncertainties associated with the variates do not appear to be associated with near-linear dependencies among the columns of $\mathbf{X}$, this suggests that other causes of data weakness (e.g. short data, high noise levels) are responsible for the large uncertainties characterising the parameters in the models.

7.7 Neurofuzzy data modelling

The neurofuzzy framework described in Chapter 5 was used to determine models for each of the tensile properties. Thus, the model construction pass structure, termination criteria and the statistical significance measure used are those outlined in respective sections. Due to the small sample size, the order of the B-spline basis functions was restricted to be less than or equal to two, in order to prevent the inclusion of severely ill-conditioned basis functions.

In an initial investigation, the complete data set (56 data pairs) was used in training (i.e. both model construction and parameter estimation), as it was considered inappropriate to partition the data into training and test sets since this would further limit the number of samples available during model identification. Furthermore, a single training-test split would yield sample sizes too small to provide a reliable estimate of the generalisation performance of the model.

The ASMOD algorithm, employing a FS/BE pass structure was used to search for an appropriate ("optimal") network structure starting from an empty model, reflecting no *a priori* beliefs as to the dependencies expected to be present in the data.

Once the construction algorithms determined a final model structure, leave-one-out cross validation (LOOCV) was used to obtain an unbiased variance estimate for the model structures inferred[9]. These estimates can then be compared with the variance estimates determined from the training set: if these are seen to significantly differ (e.g. the training estimates being much smaller than the LOOCV estimates) this strongly suggests a poor model and hence description of the process, i.e. poor generalisation abilities.

The models were then assessed in terms of the adjusted training set MSE (taking into account the number of degrees of freedom used), the number of variables (n_x) included in the model and the LOOCV estimates. These are summarised in Table 7.9.

A comparison between the empirical performance attained by the neurofuzzy models, with those obtained in section 7.6 shows how the iterative model construction algorithms have inferred models exhibiting comparable accuracies on the training data sets. Although using a

[9]However, as the model structures are inferred from the complete data set, the LOOCV variance estimates will inevitably be biased. In addition, whether such estimates can be considered as good indicators of the generalisation performance of the model, will rely on the data being a representative sample of the process.

	$\hat{\sigma}^2_N$	$\hat{\sigma}^2_{\text{df}}$	$\hat{\sigma}_{\text{df}}$	LOOCV	$\hat{\sigma}^2_{\text{df,std}}$	SS	df	n_x
$\sigma_{0.2}$	1354	1487	38.56	1710	0.171	4344	5	4
uts	985	1104	33.23	1266	0.155	3675	6	3
$\%el.$	4.07	4.30	2.07	4.53	0.528	10.3	3	2

Table 7.9: Summary of the performance obtained by the neurofuzzy models for the models determined for the tensile properties investigated.

similar number of degrees of freedom to those in the corresponding MLR models, the network structures for the strength models are seen to include only a subset of the inputs. The most severe subset selection is seen to be performed in determining a model for $\%el.$, whereby only two input variables are retained in the final model structure. It is then not surprising that the neurofuzzy models do not show any improvement over the MLR analysis.

From the network structures identified for the different tensile properties (Figures 7.1 to 7.3) and the corresponding subnetwork responses (Figures 7.4 to 7.6), it is seen that the models are comprised by simple linear subnetworks, except for the C term determined in the uts model, which is seen to be approximated by a piecewise linear fit.

As shown by the evolution in the model performance measures (MSE and SS) at each step in the model construction (Figures 7.7 to 7.9) corresponding to the refinements performed by the ASMOD algorithm for each of the tensile properties, the model searches exhibit both a different set and number of refinements. The set of refinements and corresponding model performance measures evaluated at each step in the model construction are also summarised in Tables A.3 to A.5 included in Appendix A[10].

PV_c and C are seen to have been retained in the final model structures for all three properties, Mg is present in the strength-related models and $PV_{b/a}$ present in only the $\sigma_{0.2}$ model. Neither Li or O are present in the final models, the former never considered as the best candidate refinement at any stage in the model constructions, whilst the contribution of the latter was attempted in the $\%el.$ model but failed to achieve a significant reduction in the SS measure.

Table A.3 reveals that the second refinement selected in the model construction of the uts model comprises a piecewise approximation with a high flexibility (store 2). Subsequently, during model pruning refinements, a knot deletion is seen to remove one of the internal knots. The inclusion of the C subnetwork with a relatively complex fit, from inspection of the corresponding modified residual, is seen to be influenced by a limited number of observations, which may have induced an overfitted model structure and affected the subsequent model search.

[10]These tables summarise the set of refinements considered at the n^{th} step in the model search, which are either accepted (in which case $n = n + 1$) or rejected. The set of refinements attempted over two steps ($F_m = 2$) are identified by the same letter. The refinement types are identified by: ua, tp, ts, ki, kd and ro and the number within the $(\cdot)$ indicates either the store (for ua, tp and ts) or the position along the subnetwork axis where a knot is placed/removed (for ki, kd).

However, it may be the case that these training data pairs constituted a set of alloys which exhibited anomalous uts levels. Section 7.7.1 will assess whether the final model is unduly influenced by these observations. The results obtained in subsequent analyses will reveal the sensitivity of the inferred model structures to small changes in the training data set.

From the model construction for $\%el.$ shown in Table A.5, it is seen that the inclusion of O was attempted over two steps, yielding a more accurate model but not giving a significant improvement on the SS measure. The omission of $PV_{b/a}$ suggests that carbon levels provided a greater variance in processing conditions and hence a greater description of the process.

The performance of the models on the individual observations, in terms of both training and LOOCV network estimates can be inferred from the target versus prediction scatterplots shown in Figure 7.10. The bias exhibited in the model predictions for $\%el.$ is evident: low ductility alloys are consistently overestimated, whilst high ductility alloys are consistently underestimated (the presence of this bias may be understood from inspection of the subnetworks and full residual plots for C and PV_c). The uncertainty in the model estimates (σ_{y_i}) are displayed from deriving error bars reflecting the uncertainty in the weights (i.e. second term on the r.h.s. of equation 5.36), as the term characterising the inferred output noise level was seen to overshadow the weight uncertainty term, making the error bars less informative.

From these scatterplots it is seen that there is limited variation between LOOCV and full training estimates (though this does not imply that the weights have low variance), which is not surprising as the model has a small number of degrees of freedom.

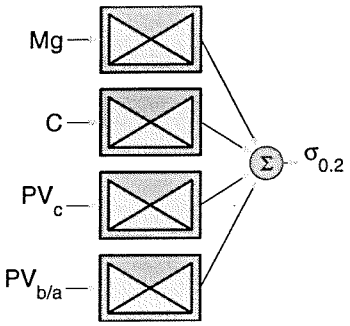


Figure 7.1: Network structure for the neurofuzzy model determined for $\sigma_{0.2}$.

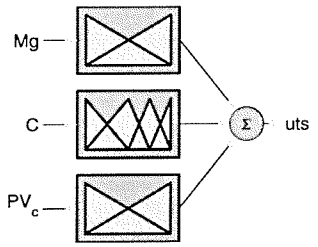


Figure 7.2: Network structure for the neurofuzzy model determined for uts.

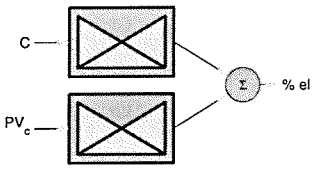


Figure 7.3: Network structure for the neurofuzzy model determined for $\%el$.

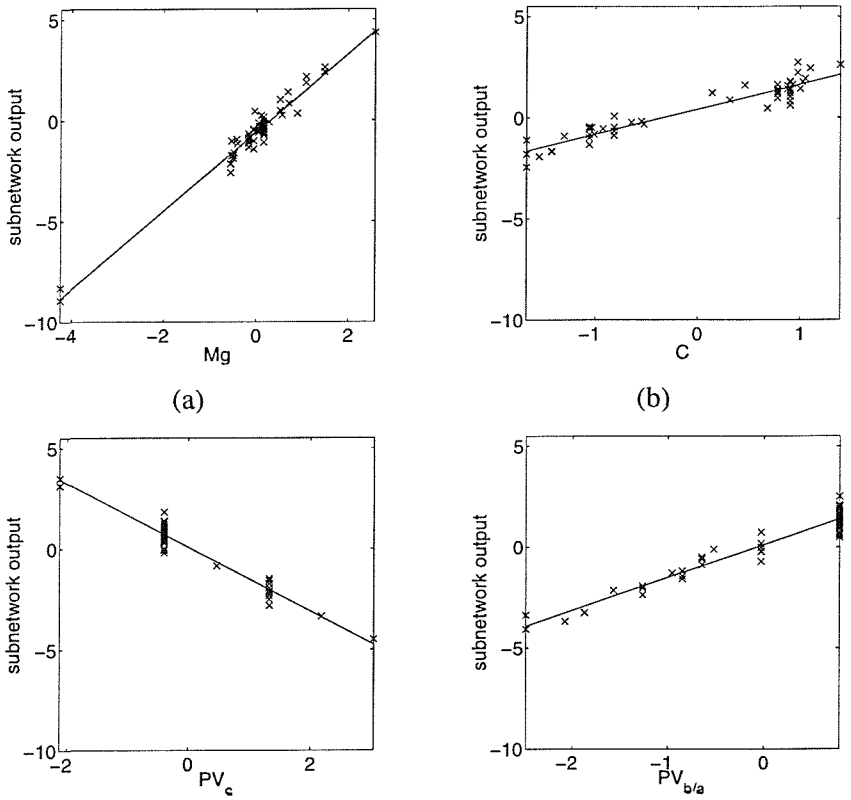


Figure 7.4: Subnetwork outputs for the neurofuzzy model determined for $\sigma_{0.2}$: Mg (a), C (b), PV_c (c) and $PV_{b/a}$ (d).

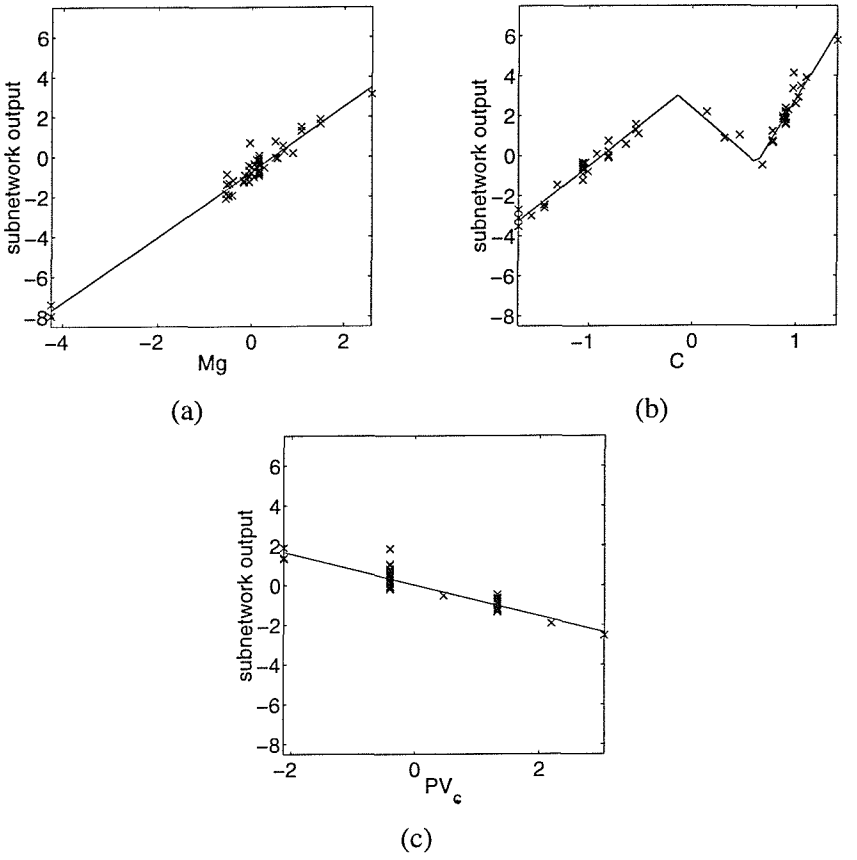


Figure 7.5: Subnetwork outputs for the neurofuzzy model determined for *uts*: Mg (a), C (b) and PV_c (c).

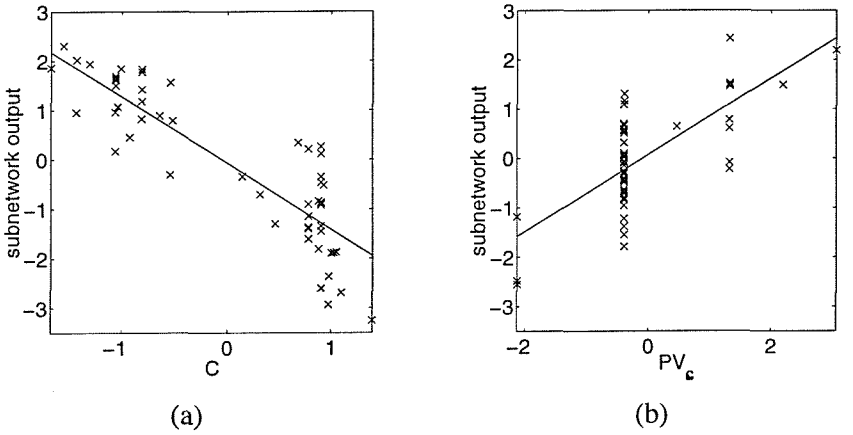
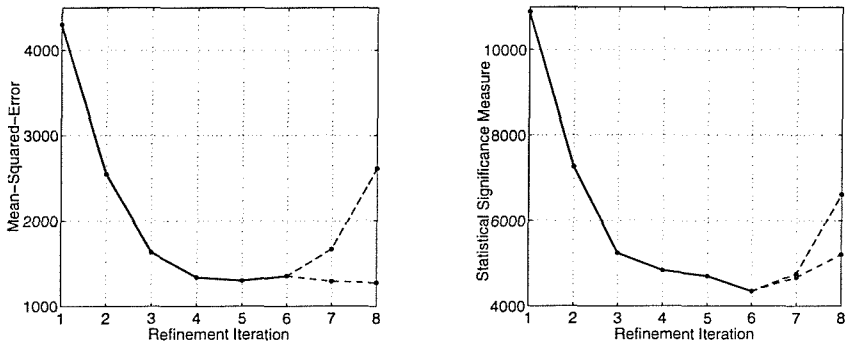
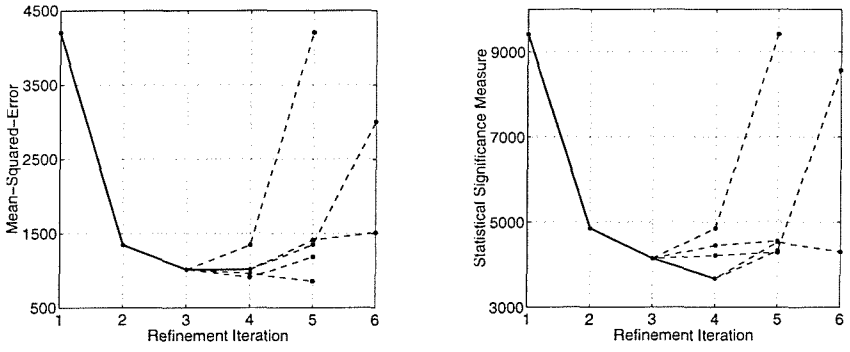


Figure 7.6: Subnetwork outputs for the neurofuzzy model determined for *%el.*: C (a), PV_c (b).



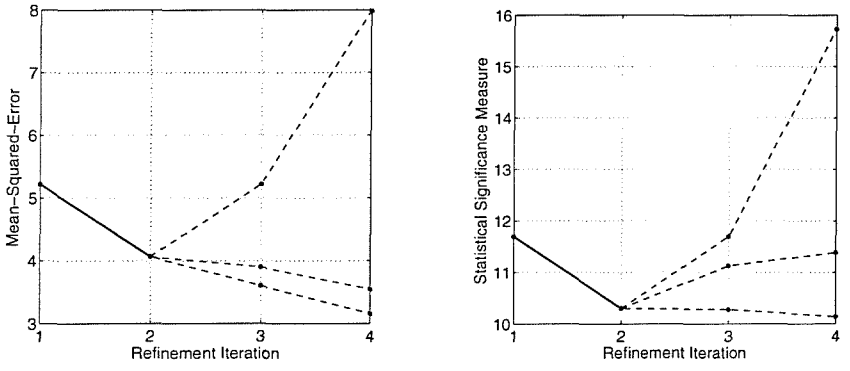
(a) (b)

Figure 7.7: Evolution of the training set MSE (a) and SS measure (b) during the iterative model search performed in determining an appropriate neurofuzzy network structure for $\sigma_{0.2}$.



(a) (b)

Figure 7.8: Evolution of the training set MSE (a) and SS measure (b) during the iterative model search performed in determining an appropriate neurofuzzy network structure for uts.



(a) (b)

Figure 7.9: Evolution of the training set MSE (a) and SS measure (b) during the iterative model search performed in determining an appropriate neurofuzzy network structure for $\%el$.

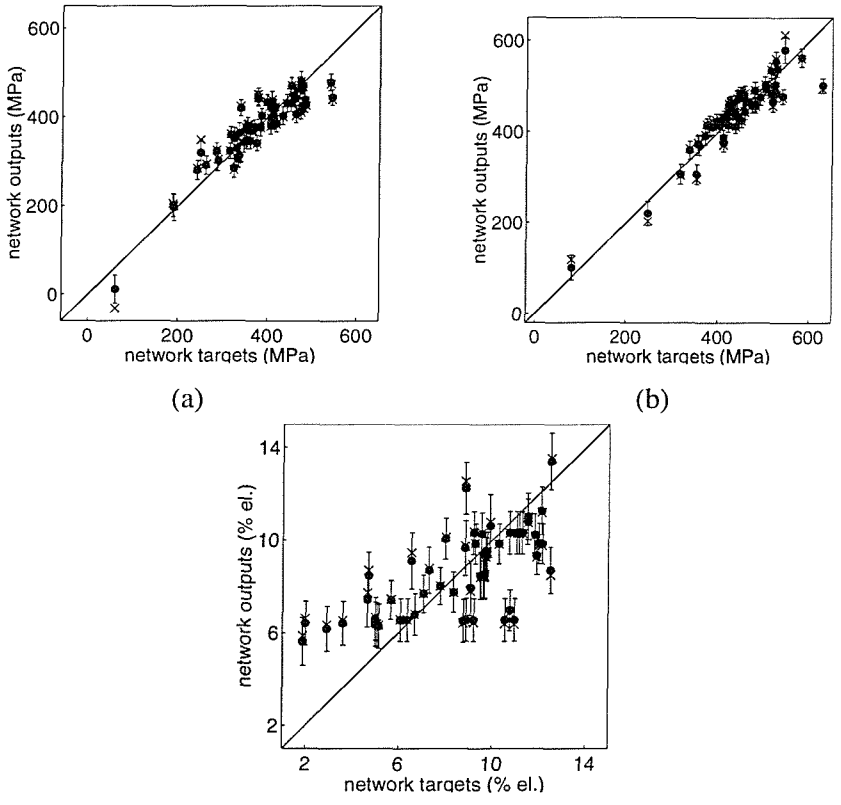


Figure 7.10: Prediction scatterplots for: 0.2% proof stress ($\sigma_{0.2}$) (a), tensile strength (uts) (b) and percentage elongation ($\%el.$) (c). Model estimates attained are shown for both the training ($\bullet$) and leave-one-out data ($\times$), the former augmented with $\pm 1\sigma$ error bars.

7.7.1 Influence analysis

Whilst the form of the approximations identified on C for all tensile properties is in general agreement with metallurgical understanding, e.g. (Vitiaz et al. 1997), validation of the extra flexibility introduced in the C subnetwork for the uts model remains problematic, as this piecewise approximation results from a poor model refinement induced by the poor distribution of the data. The intermediate piecewise approximation (exhibiting a decreasing trend) is seen to be determined by only four data pairs, and is inferred early on in the model search.

In order to infer whether these observations, and the baseline alloys induced a high influence in the results, the diagnostic measures for the neurofuzzy networks defined in section 5.8 were determined for the uts network upon training on the complete data set. The influence of the individual samples as established by C_i^2 and H_i^2 influence measures are shown in Figure 7.11.

From the index influence plots, both influence measures identify one case as being particularly influential, although the particular observation differs in each case. In addition a subset of the training samples are seen to have a moderately higher influence in a comparison to the remainder of the data. The proportional influence plots reveal that whilst H_i^2 determines the most influential case as having the largest residual, C_i^2 is seen to suggest the most influential observation as having a high leverage. Figure 7.11 (g) confirms that the observation with the largest C_i^2 corresponds to the case with the highest leverage, s_{ii}. In the response and modified residuals corresponding to the C subnetwork, Figures 7.11 (e) and (f) reveal how the subset of training patterns approximated by the intermediate piecewise term are seen to have comparable influences to other observations. Thus, although responsible for the piecewise approximation and hence influencing the model search, these four training samples are seen not to be influential cases, as determined by both C_i^2 and H_i^2. The justification for why these sensitivity measures fail to place a high influence to such cases may be understood in terms of both the low order of the basis functions and their local nature, reflected by the sparsity of $\mathbf{A}$, the solution matrix, from the presence of more than two univariate fuzzy membership functions defined on the fuzzy variable.

From the influence analysis, the baseline alloys are seen to have only a moderately higher influence.

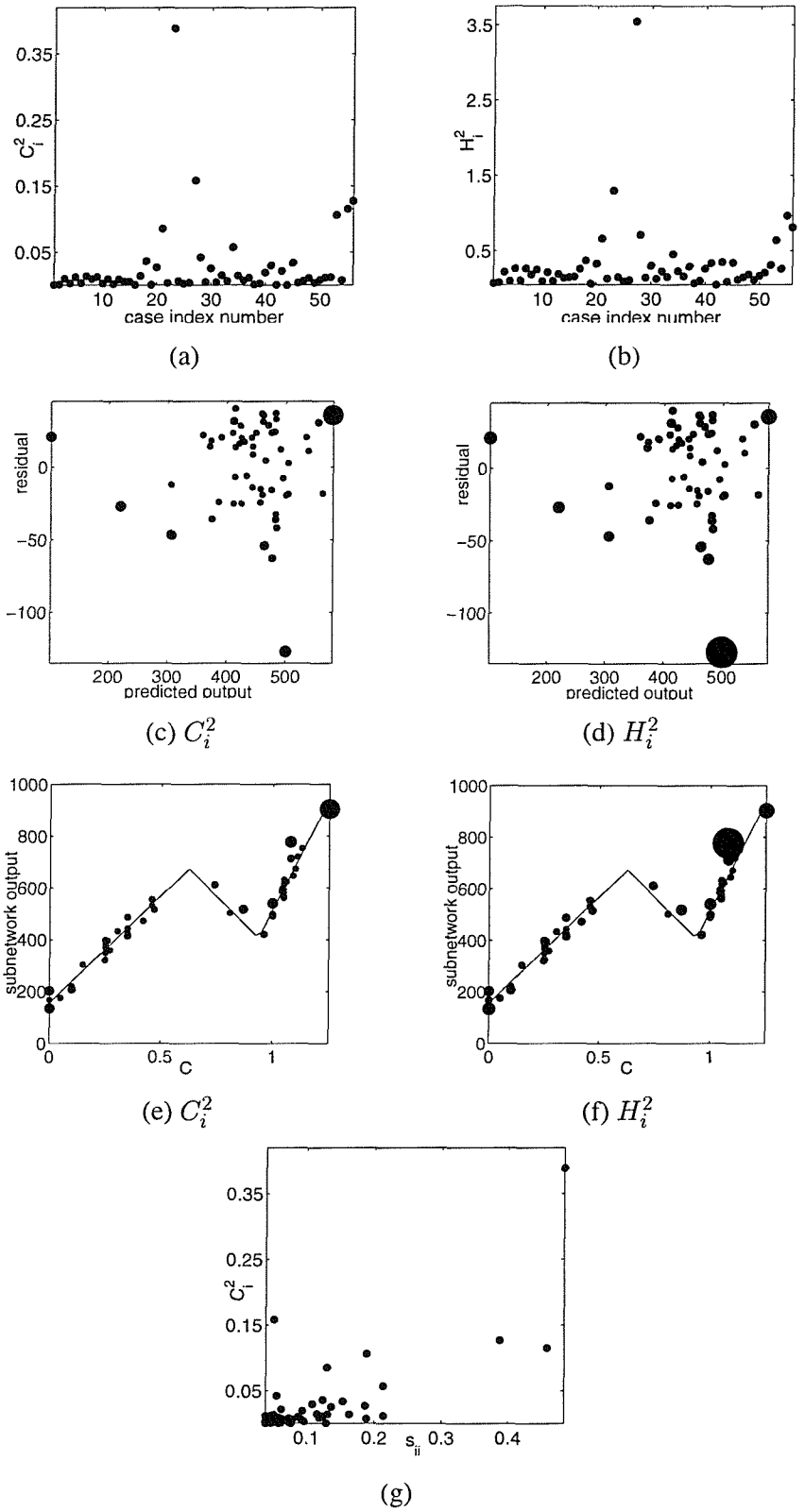


Figure 7.11: Influence plots obtained for the neurofuzzy model for uts, from using Hadi's and Cook's measures of influence: index influence plots (a)-(b), proportional influence plot (c)-(d), C subnetwork response and corresponding modified residuals (e)-(f), and s_{ii} versus C_i^2 (g).

7.7.2 Assessing stability of the model selection

In addition to the forward selection / backward elimination (FS/BE) model construction procedure employed in the ASMOD algorithm, a stepwise (SW) pass structure of the form:

pass
- univariate addition
- tensor product
- tensor split
- subnet deletion
- knot insertion
- knot deletion
- reduce order

whereby all model building and pruning refinements are considered at each iteration, was used to assess whether the same network structures would be inferred from using a different search strategy. The effect of initialising networks with univariate structures reflecting simple metallurgical understanding of the trends believed to characterise input-output dependencies was also assessed. In both cases, the network structures obtained were identical to those determined by the FS/BE construction, indicating that the data supported only the identification of simple dependencies.

As discussed in previous chapters, the use of SS measures allows all data to be used in training. However, it remains desirable to assess the performance of a technique on unseen data to obtain an estimate of the generalisation performance. The sample size, distribution and properties of the input data, outliers and/or a subset of the samples may unduly influence the model search. As such, it is desirable to assess whether small perturbations in the training set (e.g. resampling procedures, addition of noise, etc.) can cause large changes in the results (e.g. model structure, output estimates, parameter values, etc.).

When the sample size is very limited, reserving part of the data as a test set will inevitably remove informative samples from the training data. Furthermore, the generalisation error estimate inferred from a very small sample may be unreliable, exhibiting a large variance depending on the particular training-test split. However, by training $\mathcal{M}$ models on different (resampled) training-test sets, it may be possible to infer a more reliable estimate of the generalisation error. Thus, the sensitivity of the ASMOD algorithm to small changes in the training data was investigated by training multiple models on different training-test sets. The complete data set (56 data pairs) was resampled (randomly) 20 times, so that different training-test sets were generated: in each of these partitions, 90% of the data was used for training, while the remaining 10% (corresponding to 6 data pairs) was used to obtain the performance on unseen data pairs for each model $\mathcal{M}$, from which the average test error would provide an estimate of the generalisation performance of the neurofuzzy framework. These

empirical results will be discussed in section 7.9, as a similar approach was used to assess the generalisation capabilities of the SVM methods. The normalisation coefficients employed in the neurofuzzy input data normalisation were determined from the complete data set, thus ensuring that test samples were within the $[-1, 1]$ interval (the range on which the basis functions are defined).

The ANOVA terms determined by the ASMOD construction algorithm for each one of the 20 models identified from the resampled data sets, for all three tensile properties[11] are summarised in Tables 7.10 to 7.12, while the typical form of the subnetwork responses, corresponding to the most significant terms identified are shown in Figures 7.12 to 7.14.

As was the case upon training on the complete data set, the terms present in the $\sigma_{0.2}$ models are seen to comprise simple univariate linear approximations, differing only by the inclusion/omission of a carbon univariate contribution. The model structure comprising four univariate terms was determined 11 times out of the 20 model runs, and is consistent with the structure inferred from the complete data set.

From the models determined for uts, Table 7.11 shows that over the multiple runs all input variables are selected at least once. Generally model structures are seen to differ by the number of degrees of freedom used in the approximation (linear/piecewise linear) characterising the C dependency and the inclusion/omission of $\mathrm{PV}_{b/a}$ and Li. These results for uts show that the model structure determined with greatest consistency (12 times) is seen to differ from that attained from training on the complete data set. A comparison with the $\sigma_{0.2}$ resampling results shows the model structures as differing in terms of the inclusion of a Li contribution.

All models determined for $\%el.$ (Table 7.12) are also seen to be consistent with those determined from the complete data set, comprising univariate linear C and PV_c terms (Figure 7.14).

These results highlight to some extent the sensitivity of model selection processes, whereby small perturbations in the training data can result in different inferences.

Since the variance estimates obtained are seen to be rather high and the model structures relatively simple, it is useful to compare the approximation abilities of the neurofuzzy framework with those exhibited by the SVM methods presented in Chapter 6.

[11] In these tables ‡ indicates an identical model structure as that determined from the model inferred from the complete data set, while † indicates the model structure which over the multiple model runs is the most consistently inferred.

run	PV_c	$PV_{b/a}$	Mg	C
1	•	•	•	
2†,‡	•	•	•	•
3†,‡	•	•	•	•
4†,‡	•	•	•	•
5	•	•	•	
6†,‡	•	•	•	•
7	•	•	•	
8†,‡	•	•	•	•
9	•	•	•	
10†,‡	•	•	•	•
11	•	•	•	
12†,‡	•	•	•	•
13	•	•	•	
14	•	•	•	
15†,‡	•	•	•	•
16†,‡	•	•	•	•
17	•	•	•	
18†,‡	•	•	•	•
19†,‡	•	•	•	•
20	•	•	•	
Total	20	20	20	11

Table 7.10: Summary of the ANOVA terms determined for $\sigma_{0.2}$ by the ASMOD algorithm. † indicates the model structure identified more consistently, ‡ the model structure equivalent to that determined in section 7.7.

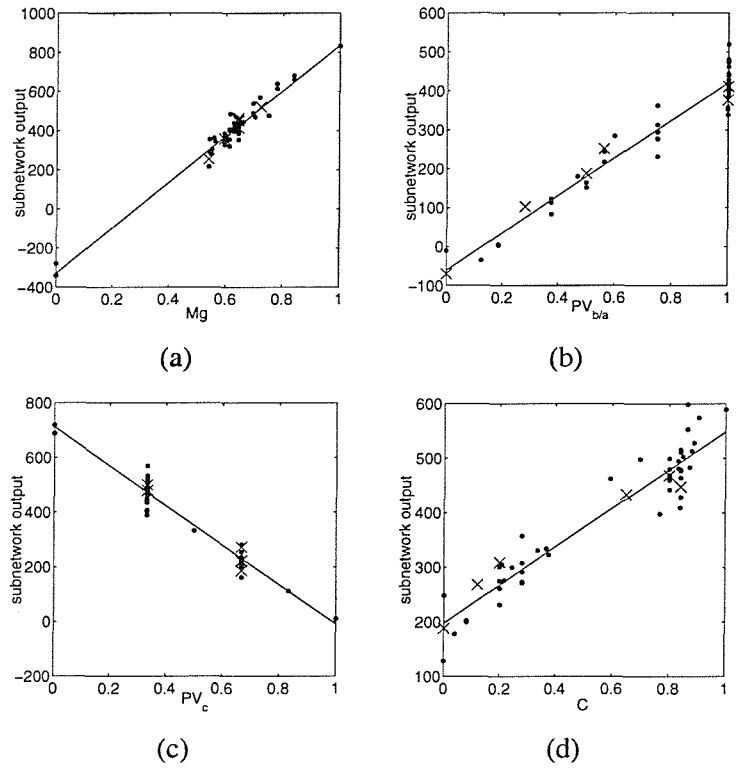


Figure 7.12: Subnetwork responses inferred for the set of ANOVA terms most consistently identified in the models determined by the neurofuzzy framework in modelling $\sigma_{0.2}$.

run	PV_c	$PV_{b/a}$	Li	Mg	C	O
1		•	•			
2†	•	•	•	•	•	
3‡	•			•	•	
4‡	•			•	•	
5†	•	•	•	•	•	
6‡	•			•	•	
7†	•	•	•	•	•	
8†	•	•	•	•	•	
9	•	•	•	•		
10‡	•			•	•	
11†	•	•	•	•	•	
12†	•	•	•	•	•	
13†	•	•	•	•	•	
14†	•	•	•	•	•	
15†	•	•	•	•	•	
16†	•	•	•	•	•	
17	•	•	•	•		•
18‡	•			•	•	
19†	•	•	•	•	•	
20†	•	•	•	•	•	
Total	19	15	15	19	17	1

Table 7.11: Summary of the ANOVA terms determined for uts by the ASMOD algorithm.

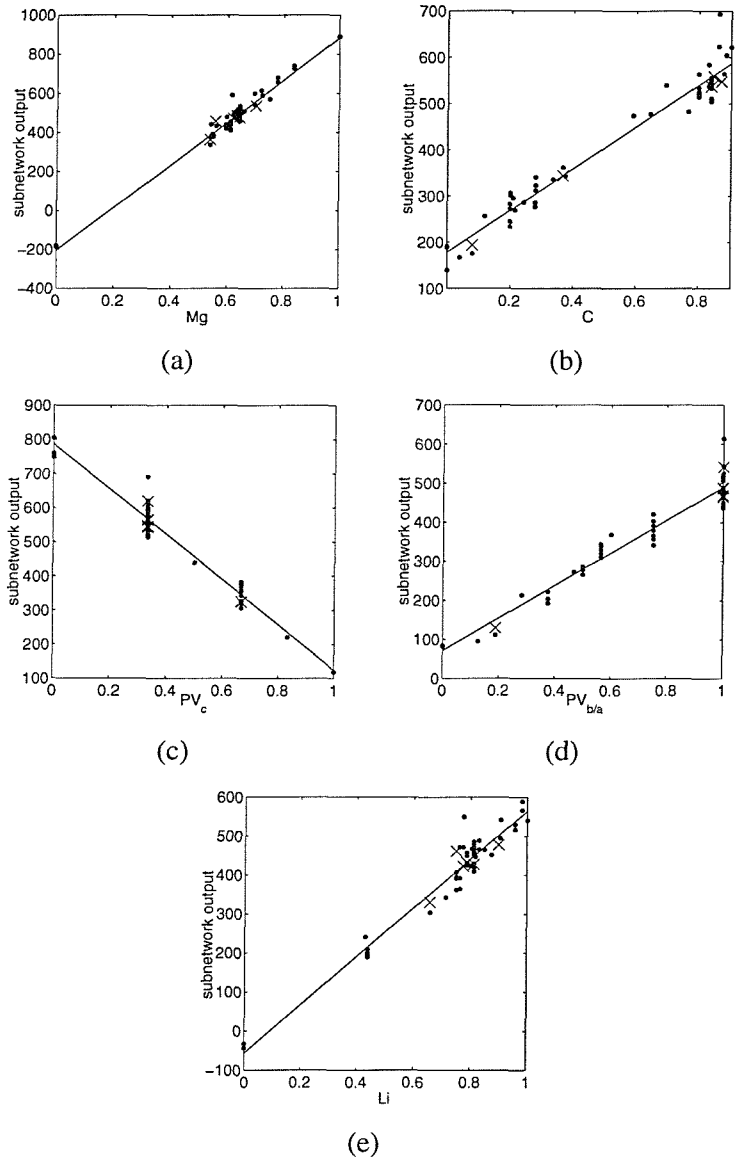
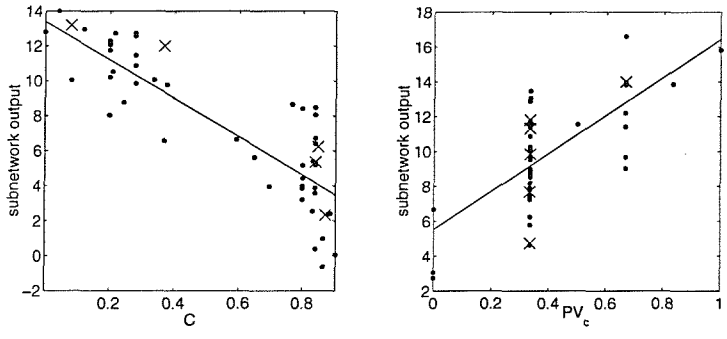


Figure 7.13: Subnetwork responses for the ANOVA terms most consistently determined in the models constructed by the neurofuzzy framework in modelling of uts.

run	PV_c	Li	C
1†,‡	•		•
2†,‡	•		•
3†,‡	•		•
4†,‡	•		•
5		•	•
6†,‡	•		•
7†,‡	•		•
8†,‡	•		•
9†,‡	•		•
10†,‡	•		•
11†,‡	•		•
12†,‡	•		•
13†,‡	•		•
14†,‡	•		•
15†,‡	•		•
16†,‡	•		•
17†,‡	•		•
18†,‡	•		•
19†,‡	•		•
20†,‡	•		•
Total	19	1	20

Table 7.12: Summary of the ANOVA terms determined for *%el.* by the ASMOD algorithm.



(a) (b)

Figure 7.14: Subnetwork responses for the ANOVA terms most consistently determined in the models constructed by the neurofuzzy framework in modelling *%el.*

7.8 Support vector regression

The SUPANOVA modelling framework described in Chapter 6 was used to determine models for the tensile properties using the same 20 training-test set partitions defined in section 7.7.2, from which an average estimate of the generalisation performance of the technique could be inferred. The modelling results were obtained by employing a quadratic loss function and an infinite ANOVA spline as the kernel function choice. Along with the empirical results attained by the SUPANOVA framework, variance estimates for a support vector regression corresponding to Stage I in the SUPANOVA framework, were also determined. The direct comparison of which allowed the sparseness of the ANOVA representations inferred by the subselection of ANOVA terms to be assessed.

The comparison between the empirical performances attained for the neurofuzzy, Stage I and SUPANOVA approaches for the three properties investigated is summarised in Table 7.13. As in tables presented in the previous section (which were seen to summarise the ANOVA terms determined by the neurofuzzy framework over the multiple runs), in Table 7.13, † indicates the model structure most consistently inferred (for both neurofuzzy and SUPANOVA results), whilst ‡ again indicates results corresponding to the neurofuzzy model structures consistent with those obtained from training on the complete data set. As the SUPANOVA results did not attain one consistent ANOVA representation in modelling $\sigma_{0.2}$, only the overall results obtained by the SVM methods are presented for this property.

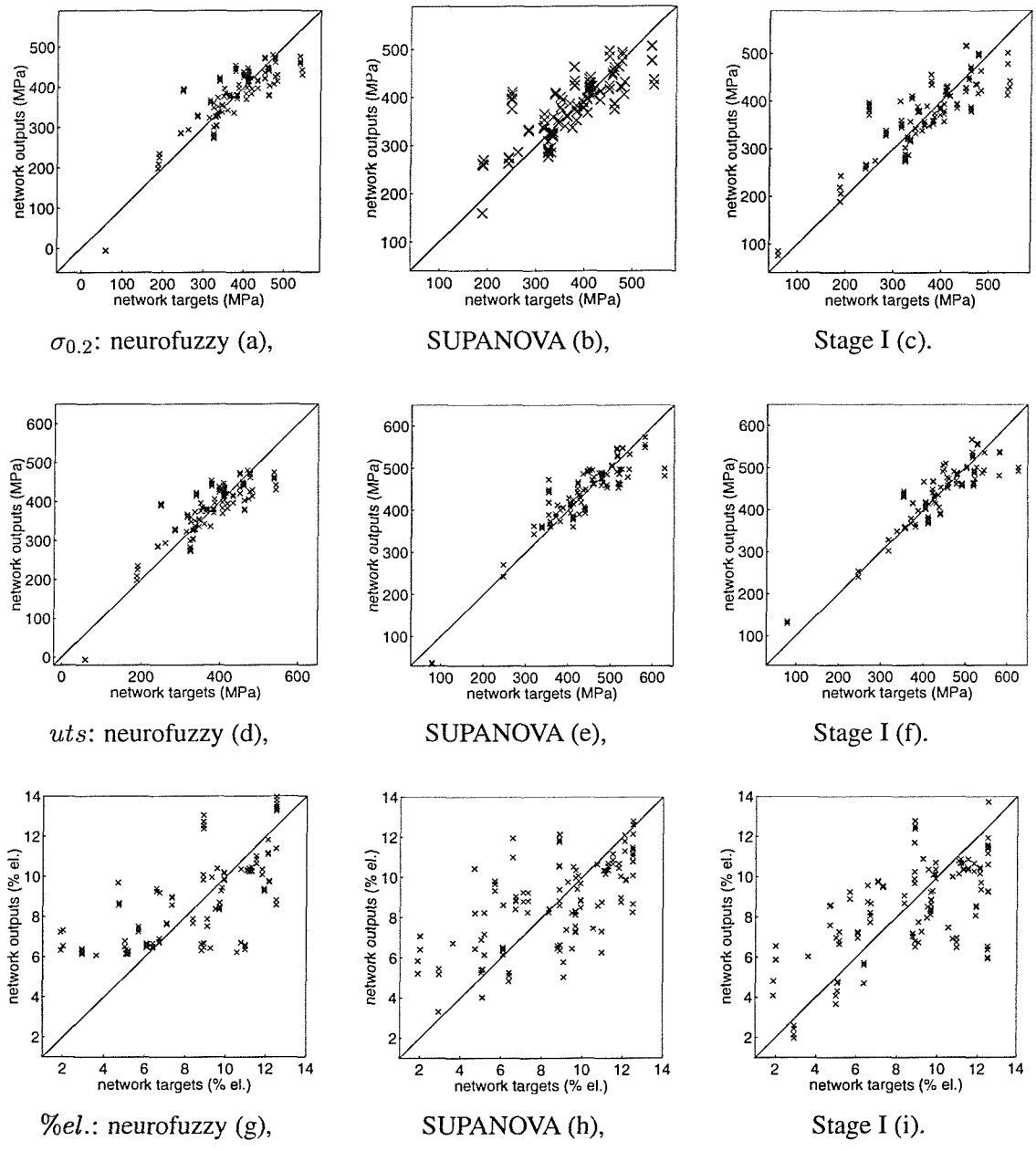
A comparison between the average training set variance estimates for the different modelling approaches show how the SVM methods exhibit better approximation capabilities on the training data, this being particularly evident in the Stage I results. Whilst the generalisation performance of the SVM methods for both $\sigma_{0.2}$ and $\%el.$ is generally comparable to those exhibited by the neurofuzzy framework (SUPANOVA marginally better, Stage I worst on the $\%el.$), the benefits of the kernel approximations are more evident in the uts results.

The predictions made by the different modelling approaches on the test data sets are graphically summarised in the scatterplots shown in Figure 7.15, from which the variance exhibited on individual test data pairs can be inferred.

The relatively poor test performance exhibited by both the neurofuzzy and SUPANOVA approaches in the $\sigma_{0.2}$ results over the complete 20 model runs are attributable to a small set of observations which were frequently present in a number of test set partitions (as can also be seen from inspecting Figure 7.15). This emphasises the difficulty in assessing the generalisation performance of a model from a very small data set.

A number of models were characterised by significantly larger values in the smoothing parameter C, and an inspection of the regression functions indicated an inadequate regularisation, resulting in significant overfitting of the training data.

For a subset of these under-regularised models, the general form of the approximation on C resembled that attained by the uts subnetworks in section 7.7. In general, there was a greater



$\sigma_{0.2}$: neurofuzzy (a), SUPANOVA (b), Stage I (c).

uts: neurofuzzy (d), SUPANOVA (e), Stage I (f).

$\%el.$: neurofuzzy (g), SUPANOVA (h), Stage I (i).

Figure 7.15: Predictions attained by the different modelling techniques on the test data for $\sigma_{0.2}$ (a)-(c), uts (d)-(f) and $\%el.$ (g)-(i).

variability in the estimation of the C (smoothing) parameter for both uts and $\%el.$ than for $\sigma_{0.2}$.

The ANOVA terms extracted by the sparse basis selection stage of the SUPANOVA framework for each one of the $\mathcal{M}$ models determined are summarised in Tables 7.14 to 7.16 for each tensile property. The general form of the approximations inferred for the most consistently selected terms are those shown in Figures 7.16 to 7.18.

Interpretation and validation of the SUPANOVA models, for which the contribution of the

	Frequency of selection			Training error $\mu\left(\hat{\sigma}_N^2\right)$			Test error $\mu\left(\hat{\sigma}_N^2\right)$		
	$\sigma_{0.2}$	uts	$\%el.$	$\sigma_{0.2}$	uts	$\%el.$	$\sigma_{0.2}$	uts	$\%el.$
Neurofuzzy	20	20	20	1339	1000	4.03	3000	1897	4.89
†	11	12	19	1341	947	4.06	1557	1311	4.36
‡	11	4	19	1341	921	4.06	1557	1716	4.36
SUPANOVA	20	20	20	962	813	3.02	2987	1499	4.72
†	–	10	11	–	816	3.37	–	968	4.84
Stage I	20	20	20	830	683	2.56	2906	1637	5.70
†	–	(10)	(11)	–	705	2.69	–	1123	5.17

Table 7.13: Summary of the mean training and test (generalisation) set error estimates inferred from the 20 resampled data sets in modelling $\sigma_{0.2}$, uts and $\%el.$ Comparison of the consistency (stability) exhibited by the neurofuzzy and SUPANOVA techniques in the determination of parsimonious models for the three properties investigated. The error estimates given are the mean error attained for the ANOVA representation most frequently selected (†). The table also presents empirical errors for the neurofuzzy models determined over the 20 resampled data sets whose structure was the same as that determined by training on the complete data set (‡). Error estimates for Stage I are also given, corresponding to the data sets from which the most consistent SUPANOVA models were obtained.

same variate is present in more than one term is not straightforward. The overall dependency must be understood by considering the superposition of all terms (e.g. the overall strengthening contribution of Li should be inferred by considering the effects arising from the univariate and any higher order terms present). Furthermore, interpretation of the bivariate terms should proceed with an understanding of how the input data is distributed. In regions where the data is sparse (e.g. high levels of PV_c and Li) the functions should be viewed sceptically as extrapolation.

Tables 7.14 and 7.15 show that over the 20 runs a considerable number of different terms (13 and 11 respectively) have been considered to account for strengthening effects. Univariate contributions of $PV_{b/a}$, Li and Mg are consistently selected, with the bivariate terms $PV_c \times$ Li and $PV_{b/a} \times$ Li present in a significant number of models. More generally, PV_c is combined with Li whilst $PV_{b/a}$ forms bivariate and trivariate terms with Li, C and O. It is seen that the higher order terms selected comprise Li and not Mg.

In modelling $\sigma_{0.2}$, $PV_{b/a}$ generally exhibits an almost linear strengthening effect, and both Mg and Li are seen to exhibit similar strengthening contributions. A few models exhibited a greater non-linearity in the C dependency, characterised by a smaller strengthening contribution for C concentrations where prior knowledge (see section 7.5) suggested a saturation of the grain refinement effect.

The presence of Li in the models identified is somewhat surprising, as metallurgical understanding indicates that Li additions to Al are generally recognised as having a very minor

solute strengthening effect (Noble et al. 1982; Sanders and Starke 1982; Starink et al. 1999) compared to Mg additions, as is also noted in section 7.2.3.

The $PV_c \times$ Li term suggests that more work is retained for higher Li contents and lower forging temperatures. In regions where most of the data is distributed, it may be concluded that there is an enhanced temperature sensitivity for Li-rich alloys, the reduction in strength being greater for higher Li contents. The absence of a dependency characterising PV_c and $\sigma_{0.2}$ at very low Li contents (Figure 7.16 (d)) corresponds to the region where data is lacking. The $PV_{b/a} \times$ Li term shown in Figure 7.16 (e) suggests an effect of Li on the deformation mode of the material: during hot deformation Li may influence how work is stored as it is known that solute elements (i.e. Li and Mg) result in a more planar slip mode (through lowering SFE).

These dependencies may affect the micro-fracturing process and grain size development (affecting any recovery process that may be present), but also may be related to a dispersoid strengthening effect in addition to solute strengthening contributions.

In the models determined for uts, there appears to be a more significant univariate C contribution compared with the $\sigma_{0.2}$ models. Generally, though the dependencies inferred for both $\sigma_{0.2}$ and uts are seen to be similar. The $PV_c \times PV_{b/a} \times$ Li term remains hard to validate as the $PV_{b/a} \times$ Li subdependency is seen to be inconsistent with that exhibited for $\sigma_{0.2}$. Its effect on uts should be viewed in light of the other terms present in the model.

In contrast to the strength models, where a considerable number of terms are seen to be considered over the 20 resampled data sets, models determined for $\%el.$ are seen to comprise very few ANOVA terms (1-2) and a smaller numer of terms considered (5) over the multiple runs. The $PV_{b/a} \times$ Li $\times$ C term is seen to be consistently selected, the general form of the approximation inferred being consistent with our understanding of the source of strengthening in these alloys. This term appears to reflect the significance of the powder processing in the attainment of adequate ductility levels.

run	bias	PV_c	$PV_{b/a}$	Li	Mg	C	PV_c × Li	Mg × Li	$PV_{b/a}$ × Li	$PV_{b/a}$ × C	C × O	$PV_{b/a}$ × C × O	C × Li	PV_c × $PV_{b/a}$ × Li
1	○		●	●	●		●	●						
2	○		●	●	●		●		●	●				
3	○		●	●	●		●			●				
4	○	●	●		●	●		●			●			
5	○		●	●	●		●	●	●					
6	○		●	●	●		●		●	●		●		
7	○				●		●		●	●				
8	○		●	●	●	●	●							
9	○	●	●	●	●				●					
10	○		●		●		●		●			●	●	
11	○		●	●	●		●							
12	○		●	●	●		●					●		
13	○	●	●	●	●		●		●					
14	○		●	●	●		●		●					
15	○		●	●	●		●					●		
16	○		●	●	●					●		●		●
17	○		●	●	●		●	●			●			
18	○		●	●	●	●	●							
19	○				●		●		●	●				
20	○	●	●	●	●				●					
Total	-	4	18	16	20	3	16	4	10	6	2	5	1	1

Table 7.14: Summary of the ANOVA terms determined for $\sigma_{0.2}$ by the sparse basis selection employed within the SUPANOVA framework.

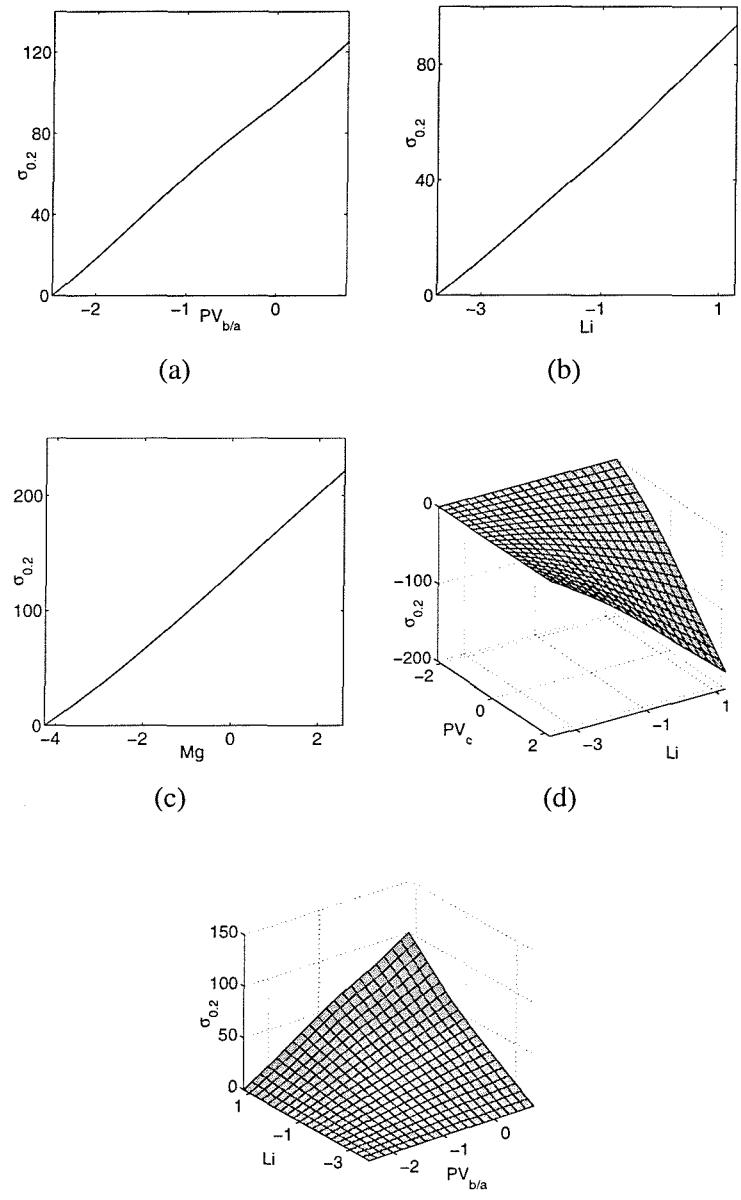
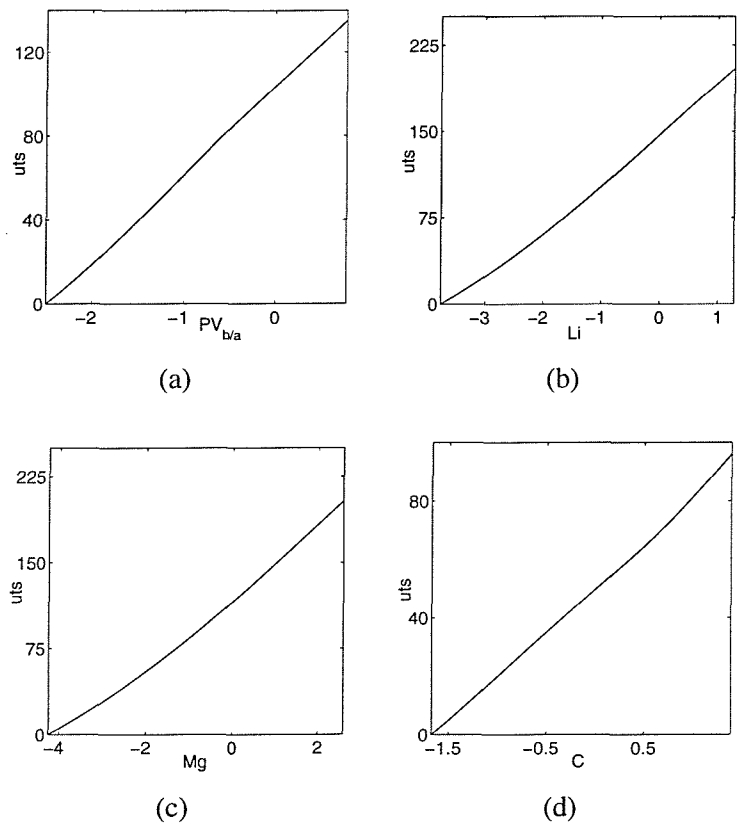


Figure 7.16: General form of the kernel approximations exhibited by the ANOVA terms most consistently identified in the basis selection stage of the SUPANOVA framework in the modelling of $\sigma_{0.2}$.

run	bias	PV_c	$PV_{b/a}$	Li	Mg	C	PV_c × $PV_{b/a}$ × Li	Mg × Li	$PV_{b/a}$ × Li	$PV_{b/a}$ × C × O	C × O	PV_c × $PV_{b/a}$ × Li × C
1	○	●	●	●	●	●						
2†	○		●	●	●	●	●					
3†	○		●	●	●	●	●					
4	○		●	●	●	●	●	●				
5	○		●	●	●	●		●				
6†	○		●	●	●	●	●					
7	○		●	●	●	●	●		●			
8†	○		●	●	●	●	●					
9	○	●	●	●	●					●		
10†	○		●	●	●	●	●					
11†	○		●	●	●	●	●					
12†	○		●	●	●	●	●					
13	○		●	●	●	●					●	●
14	○	●	●	●	●	●					●	
15	○		●	●	●	●						
16†	○		●	●	●	●	●					
17	○	●	●	●	●	●				●		●
18†	○		●	●	●	●	●					
19†	○		●	●	●	●	●					
20	○	●	●	●	●	●					●	
Total	-	5	20	20	20	19	12	2	1	2	3	2

Table 7.15: Summary of the ANOVA terms determined for uts by the sparse basis selection employed within the SUPANOVA framework.



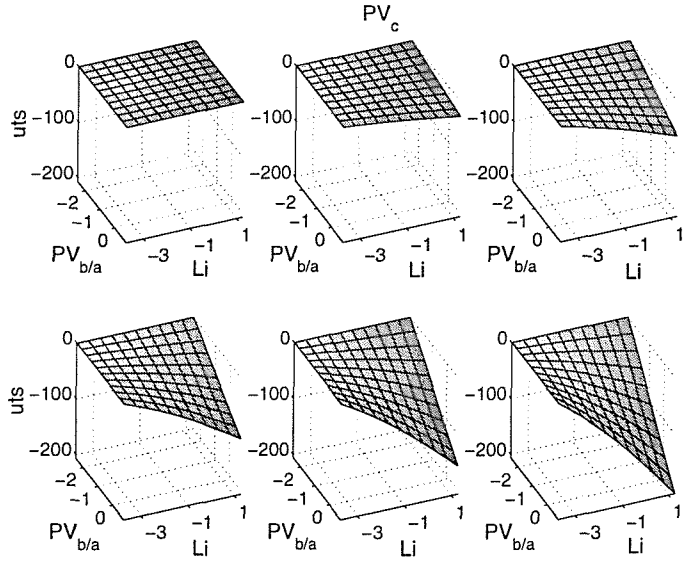
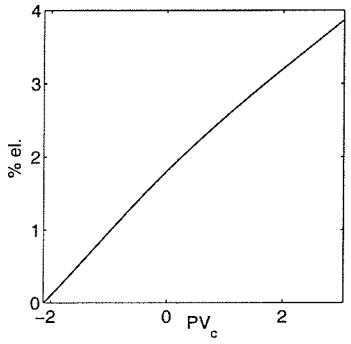


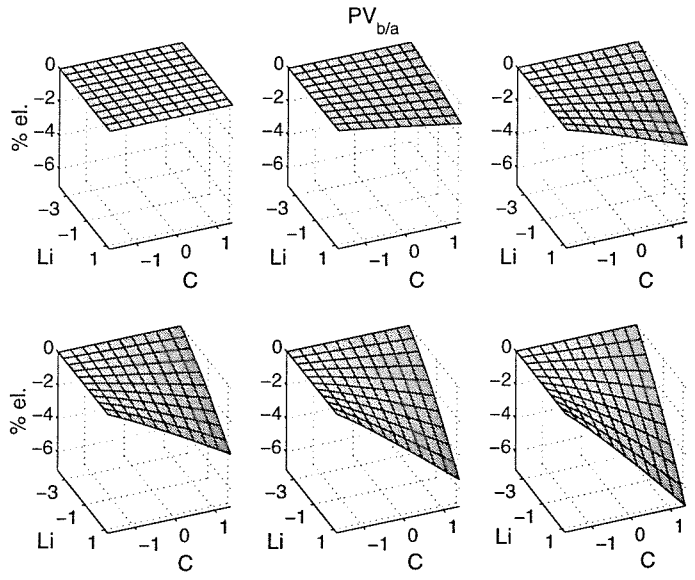
Figure 7.17: General form of the kernel approximations exhibited by the ANOVA terms most consistently identified in the basis selection stage of the SUPANOVA framework in the modelling of uts.

run	bias	PV_c	$PV_{b/a} \times$ C	$PV_{b/a} \times$ Li $\times$ C	Li $\times$ C	$PV_{b/a} \times$ C $\times$ O
1	○	•	•			
2	○	•		•		
3†	○			•		
4	○	•			•	
5†	○			•		
6	○	•		•		
7	○	•		•		
8†	○			•		
9	○					•
10†	○			•		
11†	○			•		
12	○	•				•
13†	○			•		
14†	○			•		
15†	○			•		
16†	○			•		
17	○		•			•
18†	○			•		
19	○	•		•		
20†	○			•		
Total	-	7	2	15	1	3

Table 7.16: Summary of the ANOVA terms determined for $\%el.$ by the sparse basis selection employed within the SUPANOVA framework.



(a)



(b)

Figure 7.18: General form of the kernel approximations exhibited by the ANOVA terms most consistently identified in the basis selection stage of the SUPANOVA framework in the modelling of $\%el$.

7.9 Discussion

In a comparison with the MLR analyses, the results obtained from the neurofuzzy and SUPANOVA modelling techniques achieve comparable approximations to the data though using a subset of the input variables and where considered appropriate, adapting the flexibility of the approximations in order to attain a better description of the dependencies present in the training data. Although the general form of the dependencies inferred by the adaptive modelling approaches were seen to be in general agreement, the SUPANOVA models were seen to include a number of higher order terms, suggesting the presence of interdependencies characterising the tensile properties and the processing conditions. In particular, a number of interactions between compositional elements and processing variables, suggested the presence of a number of strengthening effects in these alloys.

The larger number and higher order of the terms present in the SUPANOVA models, reflects the different implementations of the SRM principle used by the two modelling frameworks, showing how the SVM methods overcome the curse of dimensionality.

The instability of the model structures inferred and consequently the large number of terms considered in the SUPANOVA framework over the 20 resampled data sets is indicative of the considerable uncertainty in inferring a model from the data set, and hence the true representativeness of the present data of the properties investigated.

The nature of the dependencies were in agreement with the metallurgical understanding reviewed in section 7.2.3. Higher forging temperatures (PV_c) were seen to have an adverse effect on strength, approximated by simple linear terms in the neurofuzzy models, with PV_c generally being present as bivariate or trivariate terms in the SUPANOVA models.

The $\%el.$ dependencies may be interpreted in terms of the influence of microstructural refinement ($PV_{b/a}$, C) and work-hardening effects (PV_c) on ductility. In the SUPANOVA results for $\%el.$, $PV_{b/a}$ and C are systematically selected as higher order terms, whilst the ASMOD algorithm includes a univariate C dependency, $PV_{b/a}$ never being present.

For $\sigma_{0.2}$ both the neurofuzzy and SUPANOVA approaches identify a stronger univariate $PV_{b/a}$ contribution compared with a C dependency (the contribution of this latter being characterised in the SUPANOVA models in higher order terms).

In modelling uts, both techniques exhibit similar trends for Mg, PV_c and a general increasing trend for C. In the neurofuzzy results, training on the complete data was seen to result in a model structure omitting the contribution of both Li and $PV_{b/a}$, which however are seen to be present in the results obtained from determining multiple models (both neurofuzzy and SUPANOVA). The piecewise approximation on C may be held responsible for a smaller number of univariate additions, as the absence of $PV_{b/a}$ is concomitant with the presence of this term. Once again this raises issues as to whether the extra flexibility of the C subnetwork has unduly accounted for a proportion of the variance that could be achieved by the inclusion of $PV_{b/a}$. Additional degrees of freedom may have achieved a better overall approximation to the

data (MSE) but reflected a closer fit to a subset of the training data pairs (overfitting). Inferring from this that $PV_{b/a}$ is an insignificant variable is clearly misleading, and its omission is a consequence of the variable selection and complexity control approach used in the neurofuzzy framework.

Therefore, inferences based upon variable selection strategies will be highly dependent on the statistical properties of the training data set instance, as one variate may be a better explanatory variable in one data set, but not necessarily over all other possible data sets. As such, regularisation approaches to complexity control may be preferable to the greedy nature of maximum likelihood subset selection procedures.

The SUPANOVA models obtained for the strength models did not elucidate any further the dependencies between $PV_{b/a}$ and C as in the neurofuzzy results a degree of subset selection is present, particularly in the $\sigma_{0.2}$ models. In the SUPANOVA models determined for uts, both $PV_{b/a}$ and C are generally described by univariate and higher order terms.

The Mg dependencies can be understood in terms of an additive solid solution strengthening contribution, although a contribution arising from a dispersion strengthening effect through the formation of MgO is possible. As Mg was not combined with other processing variables, the strengthening effect of this alloying element was not further elucidated.

The inclusion of Mg in strength related models determined by the ASMOD algorithm requires some consideration. From Figures 7.4 and 7.5, the influence exerted in a regression of Mg by the baseline alloys is evident, and as such although in the final models the influence of the data pairs is not highly influential, the inclusion of a Mg subnetwork is nethertheless strongly influenced by these data pairs.

The omission of Li in the neurofuzzy models can also be understood in terms of complexity control and the one-step ahead iterative nature of the construction algorithms.

From an inspection of the univariate candidate refinements considered at a particular step in the model search (1^{st} in the uts, 2^{nd} in $\sigma_{0.2}$) the best refinement (corresponding to the inclusion of a univariate linear Mg subnetwork) gave only a marginally greater improvement in the model performance measures compared with an equivalent refinement performed by the inclusion of a Li subnetwork (identified as the second best refinement). Subsequent to the inclusion of the Mg subnetworks, the best constructive candidate refinements were seen to be achieved from the inclusion of other variables (PV_c and C) and not Li. This may be seen to be a consequence of maximum likelihood estimation and the greedy nature of iterative construction algorithms.

7.10 Conclusions

In this initial study, neurofuzzy networks and SVM methods have been used to determine models for three tensile properties of MA Al-Mg-Li materials, using compositional and alloy processing information.

Metallurgical understanding, conditioning and regression diagnostics have together allowed a greater understanding of the statistical properties of the data set, enabling a better interpretation and validation of the models identified by the adaptive approaches.

In the neurofuzzy framework, the SRM statistical significance measure prevented the development of over-parameterised models, limiting the capacity of the models to a complexity supported by the available data. Models identified in the neurofuzzy framework attained a comparable accuracy to the MLR results, though using a subset of the original set of input variables. In general, the SVM methods were seen to exhibit a better empirical performance, attained by use of the spline kernel approximation and the method by which the SVM approach addresses the curse of dimensionality. Overall however the output variance estimates on all three tensile properties were seen to be relatively high.

The simple nature of the structural relationships determined however basically questions the quality of the data set. Both in terms of the representativeness of the inputs, and whether the data was a sufficiently large sample of the process conditions to allow a useful model to be inferred.

Finally, the results have highlighted the instability of iterative model construction algorithms, whereby small perturbations of the data set can lead to the determination of different models. This also highlights the dilemma governing transparency and model complexity. The transparency attained by subset selection procedures may in certain processes be misleading. As several authors (e.g. see (Miller 1984; Derksen and Keselman 1992; Chatfield 1995)) note:

> *If you torture the data for long enough, in the end they will confess...what more brutal torture can there be than subset selection? The data will always confess, and the confession will usually be wrong.*

Chapter 8

Data Modelling of Structure-Properties of Experimental Trials in the Al-Zn-Mg-Cu Alloy System

8.1 Introduction

High strength 7xxx series alloys (Al-Zn-Mg-Cu) comprise an increasing volume of Al-alloys sold to the aerospace industry and due to increasing demands for property improvement, most research and development work of aluminium producers is directed towards this alloy system. Typically (in aerospace applications) these alloys have minimum target levels in three critical properties: yield strength, toughness and stress corrosion cracking (SCC) resistance. As SCC is problematic to measure on production-line timescales, the more easily obtainable electrical conductivity is used as a measure of SCC resistance. Of these three main properties, yield strength and conductivity are determined primarily by the precipitation processes that occur during commercial thermal treatments of the alloys. The third property, toughness, is a complex function of matrix flow characteristics, intermetallic particle populations (coarse primary constituents and dispersoids), grain structure and coarse heterogeneous precipitation (particularly on boundaries and dispersoids). The balance between these three main properties of 7xxx alloys is a precarious and sensitive one, with compositional and processing parameters having conflicting effects on the various properties. Modelling and optimisation of these properties is made problematic by the complexity of the relationships between primary process variables (composition, quenching rate, ageing time, ageing temperature, etc.) and target properties, particularly toughness and strength. To elucidate some of the relationships governing this balance of properties, a data set was formulated (as part of a commercial development programme),

and a series of experimental alloys were fabricated and tested by DERA on behalf of British Aluminium Plate (BAP).

The work included in this chapter presents the results obtained from the statistical investigation performed on this data (comprising a series of compositional and heat treatment trials) which broadly covers the commercial compositional range of Al-Zn-Mg-Cu alloys. Both the neurofuzzy framework and the SVM techniques have been used to determine models for yield strength and electrical conductivity from the experimental data. Simple statistical analyses have been performed on the data and MLR models obtained, allowing an assessment of the inferences obtained by the adaptive modelling approaches. In addition, in light of established understanding of the precipitation sequences characterising the 7xxx system, the investigation has been extended by assessing the effect of physical and heuristic based transformations of the input variables to develop more interpretable and representative feature sets. These transformations are expected to provide better characterisation of the main strengthening and physical characteristics of the alloys.

8.2 High-strength Al-Zn-Mg-Cu alloys

Based on the Al-Zn-Mg system, these alloys exhibit the greatest response to age hardening[1]. They respond to natural ageing, but unlike the 2xxx alloys, do not develop a stable naturally aged condition and so generally are not used in the T3 or T4 tempers. Pre-ageing deformation has little effect on the precipitation behaviour during the ageing treatment and generally there is no advantage upon stretching other than residual stress relief (this is due to the very low coherency strains of the GP zones formed in Al-Zn-Mg(-Cu) alloys). 7xxx alloys are generally employed in either a peak-aged T6 or overaged T7 condition. In this system, compositional changes do not alter the basic character of the hardening precipitates, but exert a subtle influence on the overall precipitated structure, e.g. the highest hardening levels and fastest precipitation rates are found in alloys with particular Zn:Mg ratios (Mondolfo 1971). Small amounts of Mn, Cr or Zr are typically added to control the microstructure during thermomechanical processing. In 7x75 Al-alloys $Al_{12}MgCr$ dispersoids aid in retaining the directional grain structure developed during processing of wrought products, preventing excessive growth of recrystallised grains which form during subsequent heat treatments. These incoherent dispersoids however increase the quench sensitivity of high-copper Al-Zn-Mg alloys since they act as nucleating agents for solute-rich precipitates during quenching. This process (forming the basis of the transformations derived in subsequent sections) eliminates some of the Cu and Mg from participating in the low-temperature ageing sequence, leading to lower strength levels (Polmear 1981; Starke 1989). Some alloys (e.g. 7010 and 7050) contain Zr for more

[1]The addition of Zn to Al-Mg produces an acceleration of the ageing process whereas the addition of Mg to Al-Zn alloys causes a retardation or an acceleration, but no appreciable change in the precipitation sequence (Mondolfo 1971).

efficient grain refinement and reduced quench sensitivity, also imparting improved strength and toughness. Due to their susceptibility to SCC, manufacturers have developed thermal treatments to attain a useful balance in strength and SCC resistance (e.g. T77, T79 and T74).

8.2.1 Precipitation in Al-Zn-Mg-Cu alloys

Whilst several precipitation sequences in 7xxx series alloys have been reported, e.g. see (Park and Ardell 1983), it is well established that the main sequence responsible for most of the age-hardening in 7xxx alloys is:

$$\text{ssss } \alpha \rightarrow \text{GP zones} \rightarrow \eta' \rightarrow \eta \tag{8.1}$$

where ssss α is the supersaturated solid solution of the Al-rich (α) phase, η' is thought to have a stoichiometry close to $Mg_4Zn_{11}Al$ (Park and Ardell 1988), whilst η is based on $MgZn_2$ (Mondolfo 1976; Park and Ardell 1988). It is known that in the peak aged condition most of the GP zones formed early on in the precipitation sequence are replaced by the η' phase, whilst the η phase (if at all present in the slightly overaged alloys investigated) will not be responsible for significant strengthening effects. Ageing alloys with high Zn:Mg ratios produces the transition precipitate η', the precursor to the equilibrium η phase ($MgZn_2$). For low Zn:Mg ratios, the T phase ($Al_2Zn_3Mg_3$) may form, though usually the amount of this phase is minor and has little effect on the precipitation process.

In concentrations below 1 wt.%, Cu does not appear to alter the basic precipitation mechanism and probably contributes a solid solution strengthening effect. With somewhat higher concentrations, copper does not produce new phases, but mostly goes to replace Zn and Mg in the precipitating compounds, thus participating in the precipitation process, decreasing the coherency of the precipitate when aged to peak strength. It remains uncertain whether or not Cu modifies the existing ageing process in Al-Zn-Mg alloys and/or introduces additional precipitates normally occurring in the Al-Cu-Mg system. The results obtained by Maloney et al. (2000) appear to confirm that Cu additions promote a rapid early hardening reaction when ageing at elevated temperatures, suggesting that Cu introduces clustering or precipitate processes which are additional to those observed in ternary Al-Zn-Mg alloys (i.e. accelerated nucleation of η and η' precipitates and induced precipitation of grain boundary zones and S phase). The addition of Cu increases the nucleation frequency of GP zones and of the η' phase without affecting the growth rate of the GP zones (Wagner and Shenoy 1991), thereby considerably increasing the strength of Al-Zn-Mg alloys, although increasing the quench sensitivity[2]. Overall, the effects of Cu additions are in increasing the ageing rate by increasing the degree of supersaturation and perhaps through nucleation of the S phase (Al_2CuMg) (the acceleration of the early stages of the transformation from GP zones to intermediate phase has been attributed

[2]In additions up to approximately 0.6 wt.% though Cu additions appreciably reduce the quench sensitivity of these alloys.

to the nucleation of this phase). Thus, the effects of Cu on the nucleation of the η' ($MgZn_2$, $(Al,Zn)_2Mg$) strengthening precipitate and any solution strengthening contribution, may be offset by the presence of coarse, incoherent S phase formation.

A large body of work on the thermodynamics, microstructure and microstructure-property relations of 7xxx series alloys exists (Strawbridge et al. 1948; Brown and Willey 1967; Mondolfo 1976; Liang et al. 1997; Sainfort et al. 1997; Warner et al. 1997). Based on this knowledge, modelling of the strength of ternary Al-Zn-Mg alloys on an analytical microstructure related basis (Deshamps and Brechet 1999) has been carried out, and modelling quaternary 7xxx series alloys on the basis of similar microstructure-property relations appears to be possible (Warner et al. 1997). Section 2.5 discussed recent developments and advances in thermodynamic and physical based modelling approaches. A review and case studies in modelling properties of Al-alloys using both adaptive modelling techniques and physical-based models is presented by Starink et al. (2000).

8.2.2 Main microstructure related strengthening and conductivity effects

From Chapter 2 chemical composition and processing are seen to control the microstructure and thus the physical, mechanical, and corrosion properties of heat-treatable Al-alloys. During homogenising and solution treatments some of the Cu and Mg present in the 7xxx series alloys will not dissolve in the Al-rich matrix, as some S phase will be stable at the solution treatment temperature. Hence, the Cu and Mg "tied-up" in the S phase will not cause any reduction in conductivity levels of the Al-rich phase (this phase being the only significant conductive pathway in the alloy). Additionally, these elements will not be available for subsequent precipitation hardening during ageing (precipitation hardening being the dominating strengthening mechanism). This means that the Cu and Mg present in S phase have become largely irrelevant in affecting both conductivity and strength levels. It must however be recognised that the S phase will have a detrimental effect on a third critical property, the toughness of the alloy.

The property variations as a function of copper and magnesium concentrations are summarised in Figure 8.1, see also Warner et al. (1997). The addition of Zn has in general the effect of increasing the maximum attainable strength with little effect on toughness at a given strength, though it decreases castability. The high-strength, Cu-containing alloys are quite susceptible to SCC in the under- and peak-aged tempers. In Chapter 2 it was seen that SCC susceptibility increases with increasing solid solution alloying content and that it is influenced not only by the sum of the alloying additions, but also by their ratios. To minimise this susceptibility to SCC, over-aged treatments are utilised at some sacrifice in tensile strength. Ageing practices have been developed to improve the corrosion resistance of 7xxx alloys containing more than 1% Cu, and are based on the reduction of selective corrosion at grain boundaries with increased overageing; the concomitant increase in electrical conductivity above a minimum value developed upon ageing, coincides with the onset of coarsening of the η' precipitates. Hepples (1987)

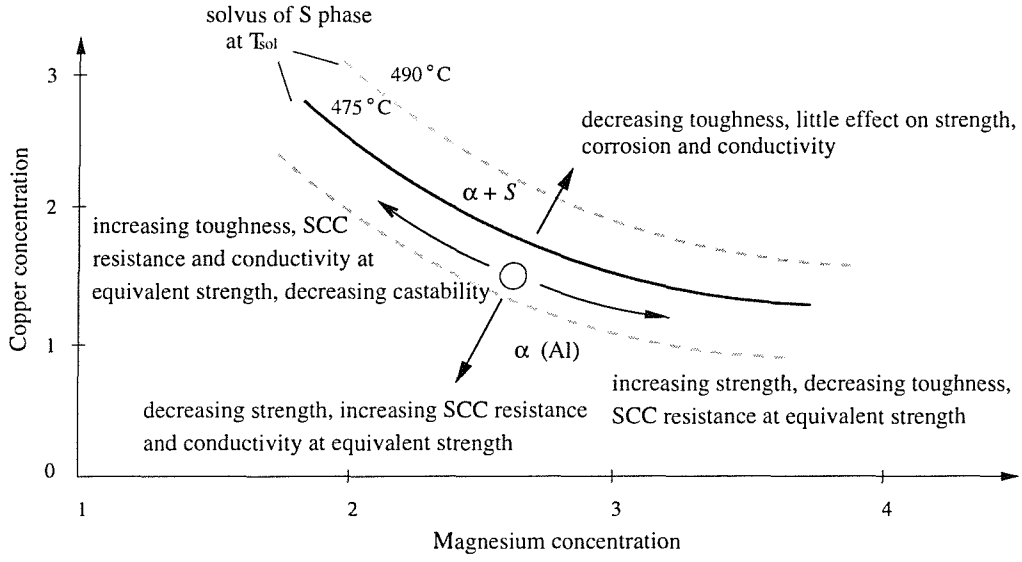


Figure 8.1: Schematic phase diagram and corresponding property variations for copper rich 7xxx alloys.

explained the reduced susceptibility to SCC of Al-alloys 7179 and 7475 upon extended ageing in terms of the Mg segregation to the grain boundaries. The increased susceptibility to SCC of low Zn:Mg alloys was thus attributed to the *free* Mg, i.e. that present in solution at grain boundaries (see section 2.3.4).

Improvements in SCC resistance have also been achieved through Cu additions (Starke 1989), increasing the Zn:Mg ratio (Al-alloy 7050) and through Zr additions, the latter reducing susceptibility caused by slow quenching. Increasing the Zn:Mg ratio above three has been considered to be detrimental to SCC resistance (Hatch 1984).

In section 2.3.4 it was discussed how a direct relationship between electrical conductivity and the SCC resistance of overaged 7xxx series Al-alloys appears to have been established (SCC resistance found to increase with increasing electrical conductivity). Electrical conductivity (the reciprocal of resistivity) of metals can be described by Matthiessen's rule[3]. Several

[3]According to Matthiessen's rule, the resistivity is the sum of a temperature and a composition dependent term:

$$\rho = \rho_P(T) + \rho_{res}(c_1, \ldots, c_n)$$

where $\rho_P(T)$ is the temperature dependent resistivity of the matrix, and $\rho_{res}(c_1, \ldots, c_n)$ is the contribution from the disturbances in the lattice structure from alloying elements with concentrations c_i. At low concentrations resistivity increases almost linearly with the amount of elements in solution. The resistivity contribution from precipitates is lower compared to elements in solid solution:

$$\rho = \rho_0 + \sum_i c_i' \rho_i + \sum_k c_k'' \rho_k \tag{8.2}$$

c_i' c_k'' are the concentrations (wt.%) of element i in solid solution and element k out of solution respectively, for $i, k = 1, \ldots, n$. The terms associated with the constants ρ_i, ρ_k yield the resistivity increase for each wt.% addition of element i in solid solution and element k out of solution respectively, e.g. see Olafsson et al. (1996).

physically-based models for the conductivity of 7xxx series alloys have recently been proposed. These interpret conductivity increases upon overageing to be due to the reduction in electron scattering resulting from coarsening of precipitates (Guyot and Cottignies 1996), reduction in the amount of dissolved atoms (Dorward 1999) resulting from the nucleation and growth of precipitates and finally due to the reduction in the amount of dissolved atoms in the PFZ adjacent to the grain boundaries (Starink et al. 2000).

8.3 The 7xxx experimental data set

The proprietary data set investigated comprised results obtained from heat treatment trials carried out by DERA, Farnborough, U.K. under contract from BAP, on a range of alloy compositions that broadly covered the high strength variants of the 7xxx series Al-alloys.

The main objective of this series of trials was to develop an understanding of how composition and ageing affected the strength/electrical conductivity balance around the 7010 composition (including compositional ranges of 7050 and 7075 Al-alloys). A total of 36 alloy/ageing time combinations were formulated, with selected compositional variations and ageing conditions (Pitcher 1998; Pitcher 1999).

To allow direct comparison, all alloys were identically processed: alloys were cast and processed to a thick plate form following industrial practices as closely as possible, receiving identical heat treatments - alloys were given the same solution heat treatment (475°C for 1 hour) and subsequently aged for varying times at 172°C with a heating rate of 20°C/h, similar to those used in commercial T7 tempers.

For each alloy, compositional levels (wt.%) of Zn, Mg, Cu, Zr, Fe and Si were quantified (comparison of the target compositions with those measured revealed that in general the intended compositions were attained) and the subsequent 0.2% proof stress ($\sigma_{0.2}$) and electrical conductivity (σ_{el}) were measured for each alloy variant. Transverse tensile test specimens were taken from the plates at the half plate thickness position. For each alloy variant, a number of tensile tests were repeated and conductivity levels monitored at different positions, from which mean property levels were estimated and used throughout the present analyses as the target outputs.

The wt.% concentrations of the alloying elements, $x_{Zn,w}$, $x_{Mg,w}$, $x_{Cu,w}$, $x_{Zr,w}$, $x_{Fe,w}$ and $x_{Si,w}$ along with the ageing time, t (hours), for each of the alloys was designated data set A. The mean and standard deviations for the inputs and output variables are summarised in Tables B.1 and B.2, from which it is seen that the main variables under investigation exhibit considerably larger standard deviations when compared to minor and impurity elements, the variability of the latter two being associated with uncontrollable processing conditions.

The pairwise input scatterplots shown in Figure B.1 do not identify any significant correlations existing between the input variables, while the conflicting requirements of attaining useful $\sigma_{0.2}$ levels whilst retaining satisfactory σ_{el} levels, shown in Figure B.4 are consistent

with the behaviour exhibited in Figure 2.1 (the r_{yy} correlation coefficient being -0.737). The simple correlation coefficients summarising the data distribution plots are shown below.

$$\mathbf{r}_{XX} = \begin{bmatrix} & \text{Zn} & \text{Mg} & \text{Cu} & \text{Zr} & \text{Fe} & \text{Si} & t \\ \text{Zn} & 1 & -0.332 & 0.118 & 0.118 & 0.417 & -0.333 & -0.211 \\ \text{Mg} & -0.332 & 1 & 0.038 & 0.366 & -0.028 & 0.158 & 0.572 \\ \text{Cu} & 0.118 & 0.038 & 1 & 0.112 & -0.221 & -0.205 & 0.001 \\ \text{Zr} & 0.118 & 0.366 & 0.112 & 1 & 0.669 & -0.635 & 0.199 \\ \text{Fe} & 0.417 & -0.028 & -0.221 & 0.669 & 1 & -0.534 & -0.023 \\ \text{Si} & -0.333 & 0.158 & -0.205 & -0.635 & -0.534 & 1 & 0.101 \\ t & -0.211 & 0.572 & 0.001 & 0.199 & -0.023 & 0.101 & 1 \end{bmatrix}$$

$$\mathbf{r}_{YX} = \begin{bmatrix} \text{Zn} & 0.354 \\ \text{Mg} & 0.125 \\ \text{Cu} & 0.365 \\ \text{Zr} & 0.174 \\ \text{Fe} & 0.052 \\ \text{Si} & -0.119 \\ t & -0.436 \end{bmatrix} (\sigma_{0.2}) \quad \mathbf{r}_{YX} = \begin{bmatrix} \text{Zn} & 0.204 \\ \text{Mg} & -0.410 \\ \text{Cu} & -0.081 \\ \text{Zr} & -0.109 \\ \text{Fe} & 0.123 \\ \text{Si} & -0.171 \\ t & 0.270 \end{bmatrix} (\sigma_{el})$$

From an inspection of the full residuals between inputs and output properties shown in Figures B.5 and B.6 it is difficult to identify clear structural relationships, the only suggested trend being that between Mg and σ_{el}. A limited number of observations suggest a dependency between output properties and ageing time, the general nature of which remains unclear particularly for $\sigma_{0.2}$, due to the large variation in properties attained at low ageing times for several differing compositional levels.

From the above analysis of the statistical properties and distributions of the data it is seen that although of limited size, the data set reflects its experimental design origins, as the distribution of the variables of interest were wide ranging, if somewhat sparse.

8.4 Data transformations

It may be argued that (non-linear) adaptive data-driven modelling techniques should be sufficiently flexible to automatically learn an appropriate input-output mapping without the necessity of any prior data processing. This may be the case when the inputs are both a representative set of features and the amount of data available is sufficient to allow reliable inferences to take place. However, the use of prior knowledge in the form of precipitation based understanding may not only augment the empirical performance of the modelling technique but may also enhance the interpretability of the dependencies inferred. As such, transformations of the input data based on physical/metallurgical understanding were considered.

The input transformations described in the following sections will be subsequently assessed in terms of whether they enhance the interpretability of the relationships and confidence in the estimates and improve the information (condition) of the data through use of prior system understanding to derive more descriptive input quantities.

8.4.1 Physical and precipitation sequence motivated transformations

In the first of the transformations considered the original compositional concentrations characterising the Al-Zn-Mg-Cu type alloys were transformed into variables that reflected key microstructural features, most notably the amounts of the main phases: the volume fraction of the main strengthening phase η' that forms in the alloys and the volume fraction of the main coarse intermetallic phase, the S phase[4]. Transforming the atomic fractions of Cu and Mg, x_{Cu} and x_{Mg}, into their corresponding fractions dissolvable in the Al-rich (α) phase, $x_{Cu,\alpha}$ and $x_{Mg,\alpha}$, formed the basis upon which further input features were derived.

If the stoichiometry of a phase is fixed, the solubility of an intermetallic phase can often be described by the regular solution model (Brown and Willey 1967; Starink and Gregson 1995; Starink and Gregson 1996). In this model, the solvus related to an intermetallic phase $M_mA_aB_bC_c$ (M_m is the main constituent of the alloy, and A_a, B_b, C_c are the alloying elements) is given by:

$$(c_A)^a (c_B)^b (c_C)^c = c_0 \exp\left[\frac{-\Delta H_{sol}}{k_B T}\right] \tag{8.3}$$

where ΔH_{sol} is the enthalpy of formation of one "molecule" of $M_mA_aB_bC_c$, k_B is Boltzmann's constant and c_0 is a constant. If appropriate values for ΔH_{sol}, c_0, a, b and c, for each phase can be derived from available solubility data, a phase diagram can be constructed. However, only for $T = 460°C$ are significant data on the solvi of all phases available (Strawbridge et al. 1948). For the S phase, the $\Delta H_{sol}(S)$ in ternary alloys has been determined previously (Starink and Gregson 1996), and by combining solvus data at 460°C (Strawbridge et al. 1948) with ΔH_{sol}, the S solvus as a function of the temperature can be estimated.

At the solution treatment temperature applied to the present data set, for the S phase, it is thus estimated:

$$(x_{Mg})(x_{Cu}) = 8.48 \times 10^5 \exp\left[\frac{-0.81eV}{k_B T}\right] \tag{8.4}$$

From this, the atomic fractions, $x_{Cu,\alpha}$ and $x_{Mg,\alpha}$, dissolvable in the α-rich phase can then be calculated by considering a dichotomy determined by the conditions described below.

If:

$$(x_{Mg})(x_{Cu}) \leq 8.48 \times 10^5 \exp\left[\frac{-0.81eV}{k_B T}\right] \tag{8.5}$$

[4]In the present investigation it is understood that the S phase is readily dissolved in the 7075 compositional range, but it is more difficult to dissolve in 7050 type Al-alloys.

then

$$\begin{cases} x_{Cu,\alpha} = x_{Cu} \\ x_{Mg,\alpha} = x_{Mg} \end{cases} \tag{8.6}$$

whereas if

$$(x_{Mg})(x_{Cu}) > 8.48 \times 10^5 \exp\left[\frac{-0.81eV}{k_B T}\right] \tag{8.7}$$

then

$$\begin{cases} x_{Cu,\alpha} = -\frac{1}{2}(x_{Mg} - x_{Cu}) + \frac{1}{2}\sqrt{(x_{Mg} - x_{Cu})^2 + 4 \times 8.48 \times 10^5 \exp\left[\frac{-0.81eV}{k_B T}\right]} \\ x_{Mg,\alpha} = x_{Mg} - (x_{Cu} - x_{Cu,\alpha}) \end{cases} \tag{8.8}$$

The atomic fraction of S phase, x_S, (i.e. the number of atoms present as S phase divided by the total number of atoms in the system) is given by:

$$x_S = 4(x_{Cu} - x_{Cu,\alpha}) \tag{8.9}$$

From section 8.2 it was seen that η' will be the main hardening phase, and by using the estimated composition of η' phase of $Mg_4Zn_{11}Al$, the maximum atomic fraction of η' that can form is given by:

$$x_{\eta'} = 16 \min\left(\frac{x_{Zn}}{11}, \frac{x_{Mg,\alpha}}{4}\right) \tag{8.10}$$

From the above, the main strengthening mechanisms involved are considered to be precipitation hardening due to η' and solution hardening of the α phase due to Zn, Cu and Mg. In the latter mechanism Mg is expected to have the strongest influence, as solution strengthening due to Mg is generally considered to be more important than due to Cu or Zn (Gomiero et al. 1992; Starink et al. 1999). Thus, the maximum level of precipitation hardening is expected to be determined by $x_{\eta'}$ while the maximum level of solution hardening is mainly determined by $x_{Mg,\alpha}$. The relative contributions of solution hardening and precipitation hardening will be dependent on ageing time as precipitation of alloying elements progresses, with precipitation hardening gradually replacing solution hardening as the main strengthening mechanism. For the present overaged alloys, significant η' precipitation will have occured in all alloys.

It is further known that the amount of Mg that is left in solution after complete formation of the main precipitates is a significant variable, influencing the properties (mainly σ_{el}) of overaged alloys (Anderson 1994). For all alloy variants comprising the data set, and indeed for all commercial 7xxx series alloys, $\frac{x_{Zn}}{11} < \frac{x_{Mg,\alpha}}{4}$, and hence the amount of η' will be controlled by the Zn content. Thus, the amount of excess Mg, $x_{Mg,xs}$, can be determined as:

$$x_{Mg,xs} = x_{Mg,\alpha} - \frac{4}{11}x_{Zn} \tag{8.11}$$

This excess Mg is expected to have a pronounced influence on conductivity, as the corresponding atoms will not be involved in the main precipitation sequence and will thus remain in the Al-rich phase.

In summary, it is considered likely that the main variables determining $\sigma_{0.2}$ are $x_{\eta'}$ and t, with secondary effects (Brown and Willey 1967) determined by $x_{Mg,\alpha}$ and $x_{Cu,\alpha}$, whereas for σ_{el}, $x_{Mg,xs}$ and t are considered the main variables, with further minor contributions from $x_{Cu,\alpha}$ and a very small influence due to x_{Zn} (or $x_{\eta'}$).

The set of features derived above, namely $x_{\eta'}$, $x_{Mg,\alpha}$, $x_{Cu,\alpha}$, $x_{Mg,xs}$ and x_S, together with the ageing time t, comprised the second data set investigated, designated the "microstructural" or data set B. The concentration of the minor and impurity alloying elements were omitted from this data set, as they were considered from a preliminary data analysis to be unlikely to influence the properties investigated.

8.4.2 Rule-of-thumb transformations

In technical publications, alloys are often categorised and discussed in terms of sums and ratios of the weight percentages of the main alloying elements. In the Al-Zn-Mg-Cu system, some of the quantities which have been considered include Zn:Mg, (Zn+Cu):Mg and Zn+Mg. These quantities can have physical meanings, for instance in section 8.2 it was seen that the Zn:Mg ratio will exert a considerable influence on the balance of the main precipitation sequences operating in the alloy system, with some literature defining particular dichotomies characterising the precise nature of the phase precipitated, e.g. see in particular (Mondolfo 1971)[5]. However, in terms of the solid state reactions, the relevance of adding weight percentages of atoms is generally unclear. Although, it can be seen that with the atomic weight of Zn being approximately 2.7 times that of Mg, adding Zn+Mg, in weight percentages, may be some measure of the amount of strengthening η' phase, provided that this phase has a broad range of stability around its central composition of $Mg_4Zn_{11}Al$.

Thus, certain quantities derived from sums and ratios of alloying elements participating in precipitation sequences can both provide a more interpretable and parsimonious description of the data. Hence, in order to fully investigate possible permutations of the input variables, the three quantities, $x_{Zn,w} : x_{Mg,w}$, $(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$[6] and $x_{Zn,w} + x_{Mg,w}$, have been considered, which, in addition to ageing time t constituted the third data set investigated, designated the "rule-of-thumb" or data set C[7].

[5]In this review article an extensive use of such quantities is employed to describe the physical properties of Al-Zn-Mg(-Cu) alloys. For example, the Zn:Mg ratio, considered to characterise the Zn bearing phase, in ratios greater than 2.5 is considered to promote $MgZn_2$ (η), while for lower levels is considered to give $Al_2Mg_3Zn_3$ (T). The definition of such dichotomies characterising precipitation in the alloy remains questionable.

[6]The motivation behind the (Zn + Cu) : Mg quantity originates from a comparatively similar behaviour of Cu with Zn in the ratio, as already seen previously, Cu may be dissolved in $MgZn_2$ or $Al_2Mg_3Zn_3$, as well as being present in solid solution and/or constituent particles (Mondolfo 1971).

[7]A further quantity which could have been considered was the Cu:Mg ratio, as in (Mondolfo 1971) it is seen

It should be understood that the two data sets derived above (B and C) are substantially different from both physical and statistical perspectives. The set of transformations giving rise to data set B are directly including system understanding to modify (augment) the statistical distributions of the original compositional levels, having the effect of "smearing" the data in input space (as may be inferred by comparing the data distributions of $x_{Mg,w}$, $x_{Cu,w}$ and $x_{Zn,w}$ with $x_{Mg,\alpha}$ (or $x_{Mg,xs}$), $x_{Cu,\alpha}$ (or x_S) and $x_{\eta'}$, respectively - see the following section) the degree of which being greater for some data pairs than for others.

In contrast, the variables comprising data set C are the result of a set of transformations which give rise to a set of features which may have benefits in interpretability, in terms of general rule-of-thumb understandings, but are not "pre-conditioned" from *a priori* physical based system understanding. For example, data set C (and consequently data set A) does not discriminate between the wt.% of the alloying additions (particularly Mg, Zn and Cu) present in solution or out of solution, these quantities known to have considerable differences in affecting σ_{el} as implied in section 8.2, see also (Hepples 1987).

Given a sufficiently representative and large data set, it is again acknowledged that data-driven adaptive modelling approaches should be capable of identifying and approximating non-linear input-output dependencies, without necessitating of the forms of pre-processing presented above; however, interpretability and knowledge representation requirements together with the limitations imposed by the small sample sizes, make both data sets B and C worthy of investigation.

As in Chapter 7, the input variables were transformed to have zero mean and unit variance. This simple pre-processing has already been seen to be beneficial as it removes non-essential sources of ill-conditioning, allowing a comparison of the magnitudes of the weights determined in a MLR analysis. Inferences were also obtained from a similar standardisation of the output quantities. For the neurofuzzy framework, the data was transformed to lie within ± 1, whereas the data was standardised to lie in the interval $[0, 1]$ for the SVM methods. Again, to preserve the commercial confidentiality of the data set, all plots involving input variables are presented in normalised forms, whilst final predictions and measured output values are presented in their original ranges.

8.4.3 Effects of the transformations

Whilst the transformations proposed in previous sections can generate more descriptive and interpretable features, more amenable to physical based understanding, a poor choice of transfor-

that for Cu>Mg, S phase can be present and furthermore, if Cu>>Mg, θ phase (Al_2Cu) can form. For the present alloy variants and indeed in 7xxx alloys, however this latter phase does not form. Thus, any effect of Cu in the system will be represented by the (Zn+Cu):Mg quantity.

mations can generate sources of ill-conditioning in a regression analysis, leading to misleading and/or nonsensical results, as discussed in previous chapters.

A simple inspection of the data distributions for the variables comprising data sets B and C will be useful in assessing the effect of the transformations compared to the original distributions. The application of conditioning diagnostic measures will be discussed in section 8.5, in the context of the MLR analysis, elucidating any sources of uncertainty.

Figures B.2 and B.3 show the input pairwise scatterplots for the two data sets, whilst the simple correlation coefficients between the variates are summarised below.

$$\mathbf{r}_{XX} = \begin{bmatrix} & \mathrm{x_{Cu,\alpha}} & \mathrm{x_{Mg,\alpha}} & \mathrm{x_S} & \mathrm{x_{\eta'}} & \mathrm{x_{Mg,xs}} & t \\ \mathrm{x_{Cu,\alpha}} & 1 & 0.033 & 0.531 & 0.207 & -0.015 & -0.011 \\ \mathrm{x_{Mg,\alpha}} & 0.033 & 1 & 0.085 & -0.308 & 0.977 & 0.570 \\ \mathrm{x_S} & 0.531 & 0.085 & 1 & 0.177 & 0.040 & 0.056 \\ \mathrm{x_{\eta'}} & 0.207 & -0.308 & 0.177 & 1 & -0.503 & -0.209 \\ \mathrm{x_{Mg,xs}} & -0.015 & 0.977 & 0.040 & -0.503 & 1 & 0.565 \\ t & -0.011 & 0.570 & 0.056 & -0.209 & 0.565 & 1 \end{bmatrix}$$

$$\mathbf{r}_{YX} = \begin{bmatrix} \mathrm{x_{Cu,\alpha}} & 0.370 \\ \mathrm{x_{Mg,\alpha}} & 0.138 \\ \mathrm{x_S} & 0.221 \\ \mathrm{x_{\eta'}} & 0.379 \\ \mathrm{x_{Mg,xs}} & 0.041 \\ t & -0.436 \end{bmatrix} (\sigma_{0.2}) \quad \mathbf{r}_{YX} = \begin{bmatrix} \mathrm{x_{Cu,\alpha}} & -0.091 \\ \mathrm{x_{Mg,\alpha}} & -0.418 \\ \mathrm{x_S} & 0.057 \\ \mathrm{x_{\eta'}} & 0.194 \\ \mathrm{x_{Mg,xs}} & -0.423 \\ t & 0.270 \end{bmatrix} (\sigma_{el})$$

$$\mathbf{r}_{XX} = \begin{bmatrix} & \mathrm{Zn:Mg} & \mathrm{(Zn+Cu):Mg} & \mathrm{Zn+Mg} & t \\ \mathrm{Zn:Mg} & 1 & 0.941 & 0.262 & -0.500 \\ \mathrm{(Zn+Cu):Mg} & 0.941 & 1 & 0.209 & -0.496 \\ \mathrm{Zn+Mg} & 0.262 & 0.209 & 1 & 0.125 \\ t & -0.500 & -0.496 & 0.125 & 1 \end{bmatrix}$$

$$\mathbf{r}_{YX} = \begin{bmatrix} \mathrm{Zn:Mg} & 0.086 \\ \mathrm{(Zn+Cu):Mg} & 0.166 \\ \mathrm{Zn+Mg} & 0.438 \\ t & -0.436 \end{bmatrix} (\sigma_{0.2}) \quad \mathbf{r}_{YX} = \begin{bmatrix} \mathrm{Zn:Mg} & 0.380 \\ \mathrm{(Zn+Cu):Mg} & 0.339 \\ \mathrm{Zn+Mg} & -0.035 \\ t & 0.270 \end{bmatrix} (\sigma_{el})$$

The correlations characterising input variables generated by the microstructural transformations are clear, particularly between $\mathrm{x_{Mg,xs}}$ and $\mathrm{x_{Mg,\alpha}}$, while dependencies characterising the other Mg-related features remain less evident.

The data distribution characterising x_S and several of the other quantities are seen to be affected by x_S, which for several alloys is descriptive of the situation where no S phase is considered to form.

The dependencies characterising the set of variables derived from the rule-of-thumb transformations are also clear and are a reflection of the compositional levels of the 7xxx system, particularly the dependency between $x_{Zn,w} : x_{Mg,w}$ and $(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$.

The full residuals for both data sets are shown in Figures B.7 to B.10, from which the effects of the microstructural transformations on the distribution of the samples can be clearly understood. Unsurprisingly, the trends exhibited for the Mg related features are retained. Inspection of the residuals for data set C reveal increasing trends between $x_{Zn,w} + x_{Mg,w}$ and $\sigma_{0.2}$, increasing trends between the Zn:Mg related variables and σ_{el}, while the addition of Zn to Mg has obscured the dependencies identified previously between the Mg content and σ_{el}[8].

8.5 Multiple linear regression

In order to assess the effect of the data transformations a MLR analysis was performed on all three data sets. Due to the small sample sizes and the different number of variables comprising the data sets, the results were compared in terms of the adjusted training MSE ($\hat{\sigma}^2_{df}$), for both unnormalised and normalised output vectors (the latter quantity indicating the output variance "explained" by the model) enabling an unbiased comparison between (training set) variance estimates obtained for the different models. The results obtained from the MLR analyses are summarised below in Tables 8.1 and 8.2. For completeness, variance estimates as determined from equation 3.4 are also given.

Data set	$\hat{\sigma}^2_N$	$\hat{\sigma}^2_{df}$	$\hat{\sigma}^2_{df,std}$	df	n_x
A	499.3	642	0.431	8/7	7
B	506.9	629	0.408	7/6	6
C	518.2	601	0.391	5/4	4

Table 8.1: Summary of the MLR models determined for the three data sets investigated for $\sigma_{0.2}$.

From these results, the models are seen to exhibit high training errors for $\sigma_{0.2}$, with no significant improvement obtained from data sets B and C (in all cases comparable to an effective standard deviation on the estimates of approximately $\pm$ 25 MPa). The models for σ_{el} are seen to account for a similar proportion of the variance with the $\sigma_{0.2}$ models (typically

[8]It should be understood that the interpretability of these univariate plots is partial. The loss of any input-output correlation induced by the transformations (e.g. data set B) should not necessarily be viewed negatively, as these transformations may mitigate the risk of discovering fortuitous dependencies among the variates. Furthermore, what may appear as univariate dependencies may be better described through multivariate and/or higher order dependencies.

Data set	$\hat{\sigma}^2_N$	$\hat{\sigma}^2_{\text{df}}$	$\hat{\sigma}^2_{\text{df,std}}$	df	n_x
A	2.70	3.47	0.51	8/7	7
B	2.63	3.27	0.48	7/6	6
C	3.03	3.52	0.52	5/4	4

Table 8.2: Summary of the MLR models determined for the three data sets investigated for σ_{el}.

50% compared with 60% for the former), and again training error variance estimates were unaffected by the transformations (giving an effective standard deviation on the estimates of typically $\pm$ 1.85 %IACS).

The set of standardised regression coefficients and associated measures of parameteric uncertainty for the properties investigated, determined from the three data sets are summarised in Tables 8.3, 8.6 and 8.9.

	$\sigma_{0.2}$			σ_{el}		
	w	$\hat{\sigma}_w$	τ	w	$\hat{\sigma}_w$	τ
$x_{\text{Zn,w}}$	0.434	0.139	3.13	0.038	0.151	0.25
$x_{\text{Mg,w}}$	0.596	0.172	3.47	−0.718	0.187	3.84
$x_{\text{Cu,w}}$	0.263	0.133	1.98	−0.048	0.145	0.33
$x_{\text{Zr,w}}$	0.173	0.235	0.74	−0.198	0.256	0.77
$x_{\text{Fe,w}}$	−0.134	0.202	0.66	0.124	0.220	0.56
$x_{\text{Si,w}}$	0.099	0.176	0.56	−0.191	0.192	0.99
t	−0.733	0.135	5.42	0.750	0.147	5.09

Table 8.3: Parametric inferences determined from the MLR models for $\sigma_{0.2}$ and σ_{el} from data set A.

From Table 8.3 it is seen that ageing time, $x_{\text{Mg,w}}$ and $x_{\text{Zn,w}}$ exhibit the largest weights and signal-to-noise levels in modelling $\sigma_{0.2}$, whereas for σ_{el}, ageing time and Mg are seen to be the variates with the most significant linear dependencies; a further comparison with the $\sigma_{0.2}$ model shows that for σ_{el}, $x_{\text{Zn,w}}$ and $x_{\text{Cu,w}}$ attain considerably lower weights, resulting in the low τ levels. Generally, the largest structural uncertainties (low τ's) are associated with $x_{\text{Zr,w}}$, $x_{\text{Fe,w}}$ and $x_{\text{Si,w}}$, the former two exhibiting somewhat larger parametric uncertainties (high $\hat{\sigma}_w$). From a variance-decomposition of $\mathbf{X}$ (performed on the normalised data), shown in Table 8.4, the uncertainty in the weights for $x_{\text{Zr,w}}$ and $x_{\text{Fe,w}}$ are to some degree reflected by the corresponding variance decomposition proportions, despite the lack of a significant collinearity problem (the simple correlation coefficient between $x_{\text{Fe,w}}$ and $x_{\text{Zr,w}}$ being 0.669, possibly a reflection of the small sample size), as indicated by the singular values given in Table 8.5. Thus, the large uncertainty in the weights exhibited for these variates are due to other sources of data weaknesses, primarily attributable to the lack of significant (linear) trends exhibited with the outputs.

Condition	Proportions of						
Index η	$\sigma^2_{w_{\mathrm{Zn}}}$	$\sigma^2_{w_{\mathrm{Mg}}}$	$\sigma^2_{w_{\mathrm{Cu}}}$	$\sigma^2_{w_{\mathrm{Zr}}}$	$\sigma^2_{w_{\mathrm{Fe}}}$	$\sigma^2_{w_{\mathrm{Si}}}$	$\sigma^2_{w_t}$
1	0.033	0.000	0.001	0.025	0.038	0.048	0.000
1.13	0.040	0.094	0.000	0.014	0.000	0.001	0.121
1.45	0.003	0.000	0.502	0.000	0.034	0.013	0.000
1.84	0.542	0.003	0.001	0.018	0.002	0.052	0.221
2.30	0.055	0.266	0.019	0.014	0.012	0.181	0.589
3.02	0.210	0.257	0.356	0.007	0.358	0.406	0.057
4.40	0.118	0.380	0.120	0.922	0.556	0.299	0.011

Table 8.4: Condition indices and associated variance-decompositon proportions for data set A.

Singular Values
$\mu_1 = 9.19$
$\mu_2 = 8.13$
$\mu_3 = 6.33$
$\mu_4 = 4.99$
$\mu_5 = 4.00$
$\mu_6 = 3.04$
$\mu_7 = 2.09$

Table 8.5: Singular values of $\mathbf{X}$ for data set A.

The weights determined for the $\sigma_{0.2}$ model from data set B, shown in Table 8.6, appear to be consistent with the analysis presented in section 8.4.1. Increasing ageing time will reduce the strength (as data is obtained mostly for the overaged regime) and among the compositional related variables $\mathrm{x}_{\eta'}$ and $\mathrm{x}_{\mathrm{Mg},\alpha}$ exhibit the largest weights. Consistent with the assessment in section 8.2 and unsurprisingly from the data distributions, x_{S} exhibits a near to negligible influence. However for σ_{el}, the results are less interpretable, particularly the sign and magnitude of the weights for the Mg related variates, which without a statistical interpretation are likely to give rise to erroneous conclusions.

From the uncertainty measures shown in Table 8.6, a MLR analysis has resulted in considerably high parametric uncertainties, suggesting some form of data weakness. As such, inferences based on the MLR model determined are questionable if not inappropriate. A comparison between the magnitude of the weights and the r_{xy} values reveals the nonsensical MLR inferences, suggesting the presence of an ill-conditioning problem to be affecting the MLR results. The definition of data set B, comprising all the quantities defined from the set of transformations used, has then resulted in the formulation of a linear (least-squares) model which is affected by an ill-conditioning problem.

An SVD of the design matrix for this data set (Table 8.8) and a subsequent variance

	$\sigma_{0.2}$			σ_{el}		
	w	$\hat{\sigma}_w$	τ	w	$\hat{\sigma}_w$	τ
$x_{Cu,\alpha}$	0.263	0.131	2.007	−0.152	0.14	1.07
$x_{Mg,\alpha}$	0.721	22.47	0.032	4.615	24.385	0.19
x_S	0.0001	0.136	$8e^{-4}$	0.160	0.147	1.09
$x_{\eta'}$	0.361	5.548	0.065	−1.238	6.021	0.21
$x_{Mg,xs}$	−0.064	24.747	$2e^{-3}$	−5.987	26.858	0.22
t	−0.733	0.132	5.56	0.749	0.143	5.24

Table 8.6: Parametric inferences determined form the MLR models for $\sigma_{0.2}$ and σ_{el} from data set B.

decomposition (Table 8.7), clearly identify the near-linear dependencies present among the variates: the collinearity detected between $x_{Mg,\alpha}$, $x_{\eta'}$ and $x_{Mg,xs}$, associated with a condition index $\eta_6 \simeq 511$, is responsible for the large uncertainty in the corresponding weights.

Condition	Proportions of					
Index η	$\sigma^2_{w_{Cu,\alpha}}$	$\sigma^2_{w_{Mg,\alpha}}$	$\sigma^2_{w_S}$	$\sigma^2_{w_{\eta'}}$	$\sigma^2_{w_{Mg,xs}}$	$\sigma^2_{w_t}$
1	0.000	0.000	0.000	0.000	0.000	0.050
1.28	0.185	0.000	0.175	0.000	0.000	0.003
1.89	0.054	0.000	0.031	0.000	0.000	0.232
2.28	0.017	0.000	0.095	0.000	0.000	0.634
2.40	0.710	0.000	0.603	0.000	0.000	0.080
511	0.035	1.000	0.097	0.999	1.000	0.000

Table 8.7: Condition indices and associated variance-decompositon proportions for data set B.

Singular Values
$\mu_1 = 9.63$
$\mu_2 = 7.55$
$\mu_3 = 5.11$
$\mu_4 = 4.23$
$\mu_5 = 4.02$
$\mu_6 = 0.02$

Table 8.8: Singular values of $\mathbf{X}$ for data set B.

Also, though not as damaging as this near-linear dependency, the variance-decomposition proportions reveal a certain amount of uncertainty associated with the weights for $x_{Cu,\alpha}$ and x_S. The inflated variances affecting these two variates is not clear (the r_{xx} being 0.531), though

the distribution of x_S values may be held responsible[9].

Finally, in the MLR analysis of data set C, Table 8.9 shows low τ levels associated with variates in the form Xx:Mg, resulting from higher parametric uncertainties than those exhibited by the other variables. Again, simple inferences based on the sign and magnitude of the weights may be misleading, e.g. the Xx:Mg quantities are seen to exhibit opposite trends. The difference in these trends may be associated with an effective Cu contribution.

	$\sigma_{0.2}$			σ_{el}		
	w	$\hat{\sigma}_w$	τ	w	$\hat{\sigma}_w$	τ
$x_{Zn,w} : x_{Mg,w}$	−1.206	0.322	3.74	0.930	0.371	2.51
$(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$	0.799	0.314	2.55	−0.103	0.361	0.29
$x_{Zn,w} + x_{Mg,w}$	0.677	0.116	5.86	−0.348	0.133	2.62
t	−0.726	0.128	5.66	0.726	0.148	4.92

Table 8.9: Parametric inferences determined from the MLR for $\sigma_{0.2}$ and σ_{el} from data set C.

It is necessary to elucidate to what extent the parametric estimates are affected by dependencies amongst the variates. A spectral decomposition and analysis was performed (Tables 8.11 and 8.10) revealing that despite the absence of a severe conditioning problem, some variate involvement is unsurprisingly present between $x_{Zn,w} : x_{Mg,w}$ and $(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$, as indicated also by the simple correlation coefficient (0.941).

Singular Values
$\mu_1 = 9.09$
$\mu_2 = 6.22$
$\mu_3 = 4.10$
$\mu_4 = 1.42$

Table 8.10: Singular values of **X** for data set C.

From the above analysis and interpretation of the MLR results, it should be clear that misleading conclusions may be drawn, if these are based on the assumption that all inputs are described by linear dependencies with the output properties, and that dependencies between the input variates are not understood. The near-linear dependencies identified are seen to inflate the parametric uncertainty in least-squares solutions. In order to obtain more reliable parametric inferences appropriate estimation methods, e.g. subset selection, regularisation methods should be considered even in simple linear analyses.

[9]In order to explain this dependency, variance-decomposition proportions were performed by the inclusion of a bias term in the normalised data set and on the original unnormalised data (also clearly including a bias term). The variate involvement was unchanged in the former decomposition whereas the variance for $x_{Cu,\alpha}$ was considerably reduced in the latter.

Condition	Proportions of			
Index η	$\sigma^2_{w_{\text{Zn:Mg}}}$	$\sigma^2_{w_{\text{(Zn+Cu):Mg}}}$	$\sigma^2_{w_{\text{Zn+Mg}}}$	$\sigma^2_{w_t}$
1	0.018	0.019	0.011	0.056
1.46	0.001	0.000	0.563	0.154
2.22	0.020	0.031	0.392	0.785
6.40	0.962	0.950	0.034	0.005

Table 8.11: Condition indices and associated variance-decompositon proportions for data set C.

In the following sections, the results obtained from the adaptive nature of both the B-spline approximations inferred by the neurofuzzy framework and those determined by the kernel methods used in the SVM technique will be used to highlight the limitations of MLR analyses, particularly in terms of inferencing and knowledge extraction.

8.6 Neurofuzzy data modelling

The neurofuzzy framework was used to determine models for $\sigma_{0.2}$ and σ_{el} from each data set, with the same criteria and methodologies as previously outlined and applied in Chapter 7.

A direct comparison between the approximation capabilities of neurofuzzy models and those obtained from equivalent MLR models for each data set enabled an assessment of the neurofuzzy approach, while the effectiveness of the transformation was validated in terms of typical model performance statistics, metallurgical understanding and initialisation of simple model structures.

The networks identified by the construction algorithms determined from empty model structures for data set A, B and C are shown in Figures 8.2 to 8.4, whilst the subnetwork responses for each model are shown in Figures 8.5 to 8.10.

The corresponding performance measures for these models are summarised in Tables 8.12 and 8.13. As in Chapter 7, training MSE ($\hat{\sigma}^2_N$) are presented together with unbiased variance estimates ($\hat{\sigma}^2_{\text{df}}$) for normalised and unnormalised output vectors.

Data set	$\hat{\sigma}^2_N$	$\hat{\sigma}^2_{\text{df}}$	LOOCV	$\hat{\sigma}^2_{\text{df,std}}$	SS	df	n_x
A	351	421.58	507.9	0.28	2394	6	4
B	454	526.68	590.6	0.35	2444	5	3
C	279	334.32	436.9	0.22	1898	6	4

Table 8.12: Summary and comparison of the neurofuzzy models for $\sigma_{0.2}$ identified from the three data sets considered.

Data set	$\hat{\sigma}^2_N$	$\hat{\sigma}^2_{\mathrm{df}}$	LOOCV	$\hat{\sigma}^2_{\mathrm{df,std}}$	SS	df	n_x
A	0.296	0.333	0.385	0.050	1.29	4	2
B	0.223	0.251	0.271	0.038	0.98	4	2
C	0.206	0.247	0.272	0.038	1.41	6	3

Table 8.13: Summary and comparison of the neurofuzzy models for σ_{el} identified from the three data sets considered.

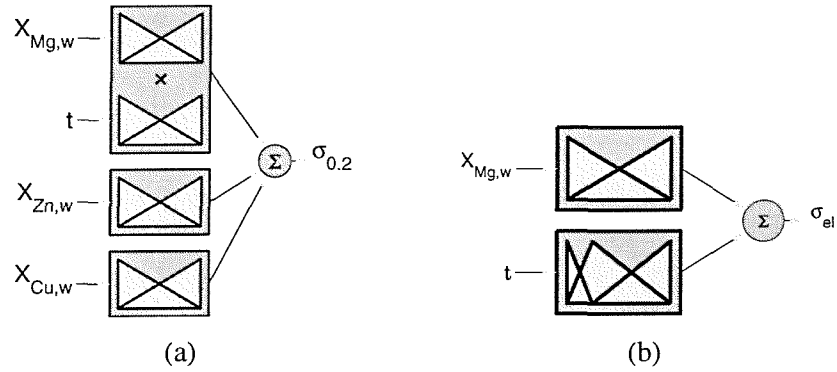


Figure 8.2: Network structures for the neurofuzzy models determined for $\sigma_{0.2}$ (a) and σ_{el} (b) from data set A.

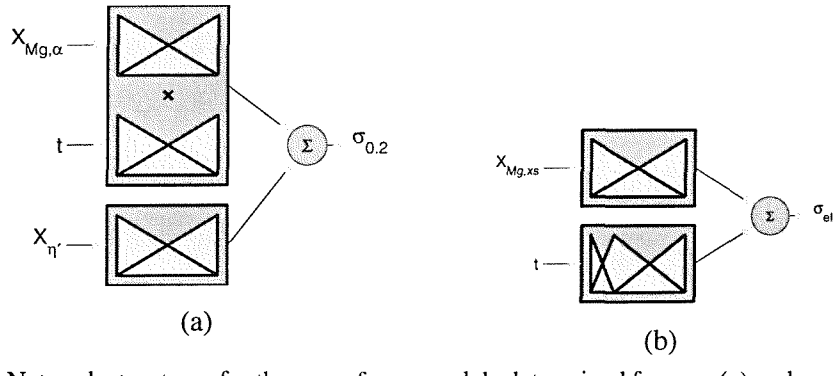


Figure 8.3: Network structures for the neurofuzzy models determined for $\sigma_{0.2}$ (a) and σ_{el} (b) from data set B.

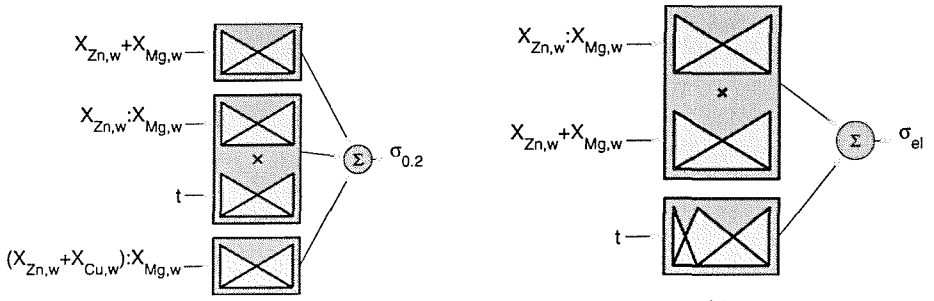


Figure 8.4: Network structures for the neurofuzzy models determined for $\sigma_{0.2}$ (a) and σ_{el} (b) from data set C.

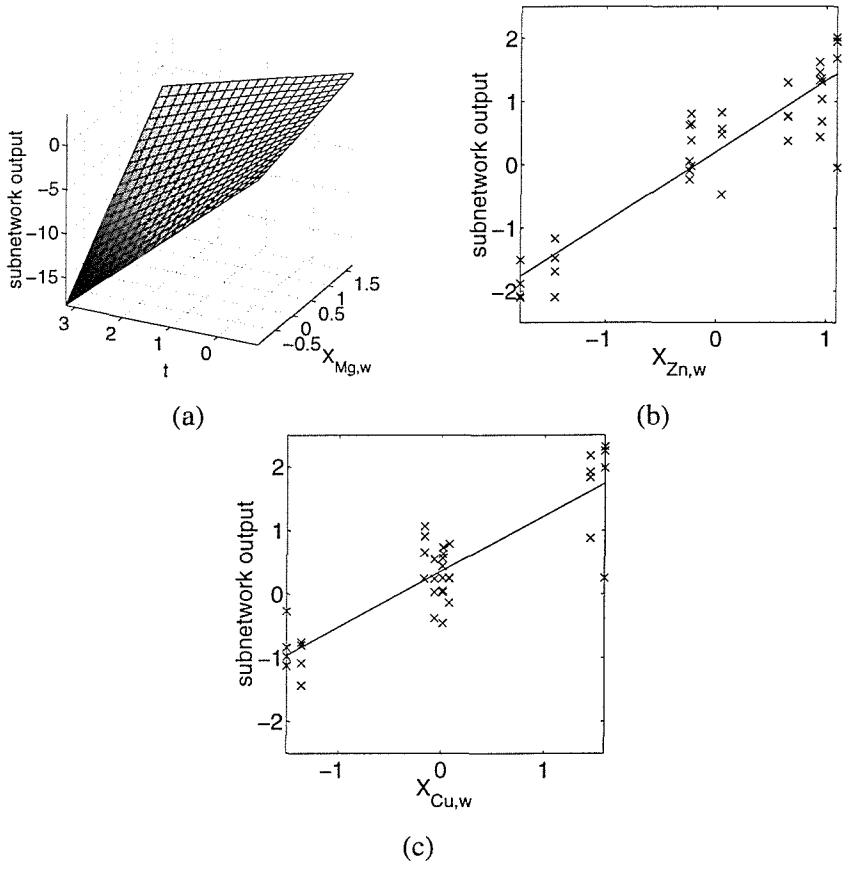
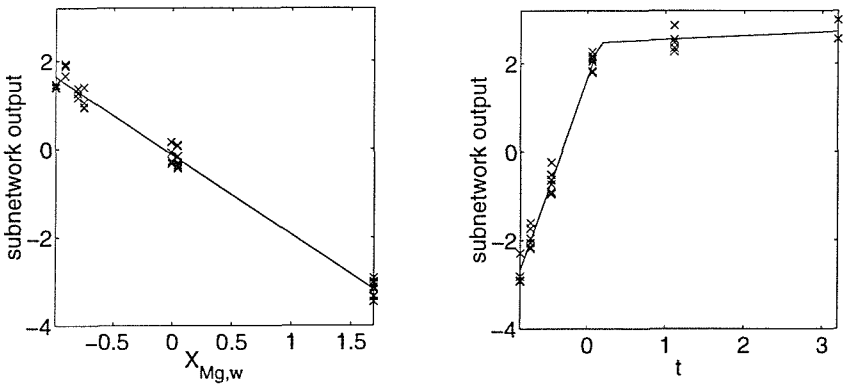
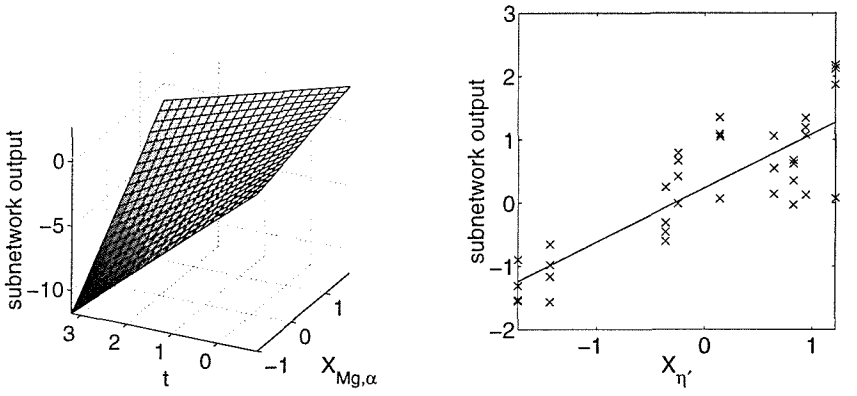


Figure 8.5: Subnetwork responses for the neurofuzzy model for $\sigma_{0.2}$ determined from data set A: $x_{Mg,w} \times t$ (a), $x_{Zn,w}$ (b) , $x_{Cu,w}$ (c).



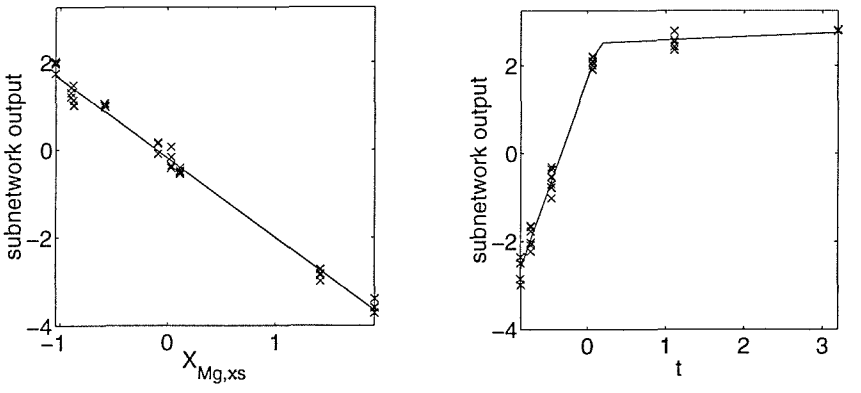
(a) (b)

Figure 8.6: Subnetwork responses for the neurofuzzy model identified for σ_{el} determined from data set A: $x_{Mg,w}$ (a), t (b).



(a) (b)

Figure 8.7: Subnetwork responses for the neurofuzzy model for $\sigma_{0.2}$ determined from data set B: $x_{Mg,\alpha} \times t$ (a), $x_{\eta'}$ (b).



(a) (b)

Figure 8.8: Subnetwork responses for the neurofuzzy model identified for σ_{el} determined from data set B: $x_{Mg,xs}$ (a), t (b).

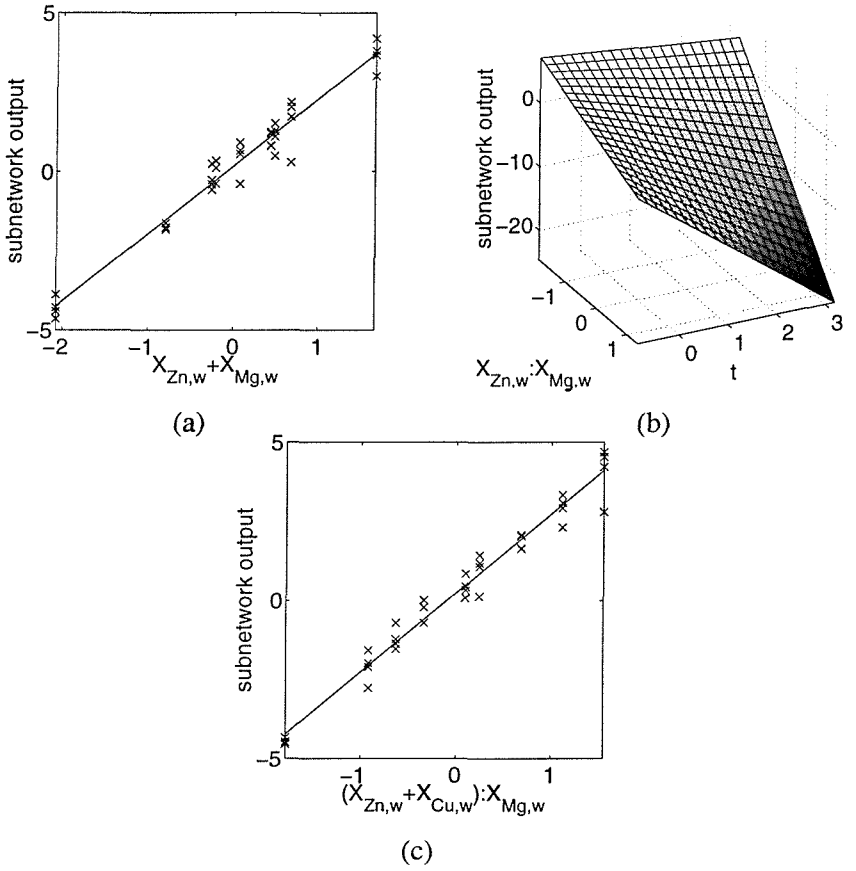


Figure 8.9: Subnetwork responses for the neurofuzzy model identified for $\sigma_{0.2}$ determined from data set C: $x_{Zn,w} + x_{Mg,w}$ (a), $(x_{Zn,w} : x_{Mg,w}) \times t$ (b) , $(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$ (c).

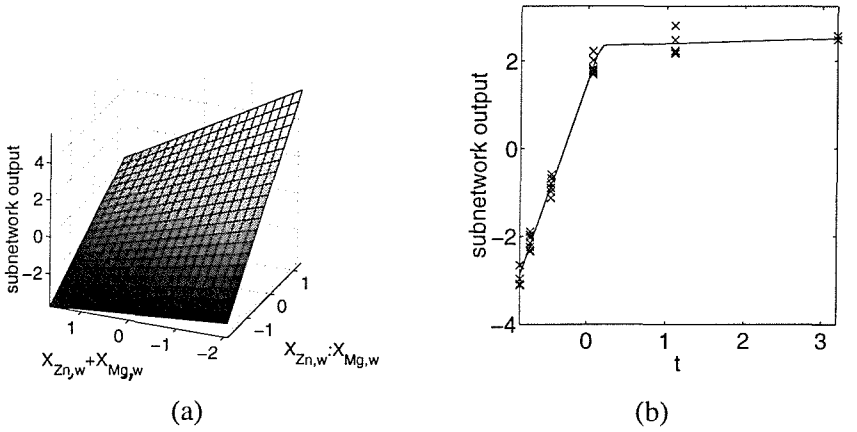


Figure 8.10: Subnetwork responses for the neurofuzzy model identified for σ_{el} determined from data set C: $(x_{Zn,w} + x_{Mg,w}) \times (x_{Zn,w} : x_{Mg,w})$ (a), t (b).

A comparison between the MLR and neurofuzzy results, indicates the latter attains a better approximation to the data for both $\sigma_{0.2}$ and σ_{el}, the improvement in the latter is particularly evident.

Figure 8.11 shows the network target versus network output scatterplots corresponding to the three data sets. As in Chapter 7, the uncertainty in the model output estimates (on the training data) are displayed with corresponding error bars characterising the weight uncertainty term only. A comparison between training set output estimates and the leave-one-out predictions typically shows limited variance, with only a small subset of model predictions exhibiting large discrepancies (more observable in the $\sigma_{0.2}$ models).

A comparison between the proportions of variance explained by the $\sigma_{0.2}$ models (ranging between approximately 65 and 78%), shows a marginal improvement in the approximation of the data has been achieved when compared to the MLR results, although generally employing a subset of the inputs. Whereas for σ_{el} the proportions explained are typically 96%, considerably improving on the levels attained by MLR, in all cases using a subset of the inputs.

In modelling $\sigma_{0.2}$ from data sets A and B a subset of the inputs are present in the final models identified, whereas for data set C, the model includes all inputs despite the correlations identified previously. In all cases an input variable related to Mg and t have been combined to form a tensor subnetwork, leading to the improvement attained in MSE levels over those achieved by the MLR analysis. The other terms present in the model exhibit simple linear approximations.

In the σ_{el} models, the ageing time contribution is described by a piecewise linear approximation, this extra flexibility allowing a considerable improvement in the variance estimates attained compared to the MLR results. The improvement exhibited upon comparing data sets A and B (which are seen to differ only in terms of the variable used to describe the Mg contribution), being significant.

The series of model refinements performed by the ASMOD algorithm in the construction of the models for $\sigma_{0.2}$ and σ_{el} on the three data sets are summarised in Tables B.3 to B.8. The similarity between the iterative model refinements is noticeable, particularly for data sets A and B. From these it is seen that the inclusion of a Cu related variable was attempted and although yielding an improved approximation to the data, the corresponding increase in the model's SS measure prevented the inclusion of this variable in the σ_{el} models determined. Finally, it is seen why the two univariate additive subnetworks, one for Mg and the other for t, were refined in favour of a tensor subnetwork of the from Mg $\times$ t.

The values of the r_{yx} coefficient between σ_{el}, $\mathrm{x_{Mg,\alpha}}$ and $\mathrm{x_{Mg,xs}}$, (-0.418 and -0.423 respectively), explain the inclusion of the latter in the model construction: the first refinement is seen to correspond to a univariate addition with a simple linear approximation and as such between these two Mg-related variates, the one with the highest correlation with the output will be selected.

From the refinement histories for the σ_{el} models, it is seen how the second refinement cor-

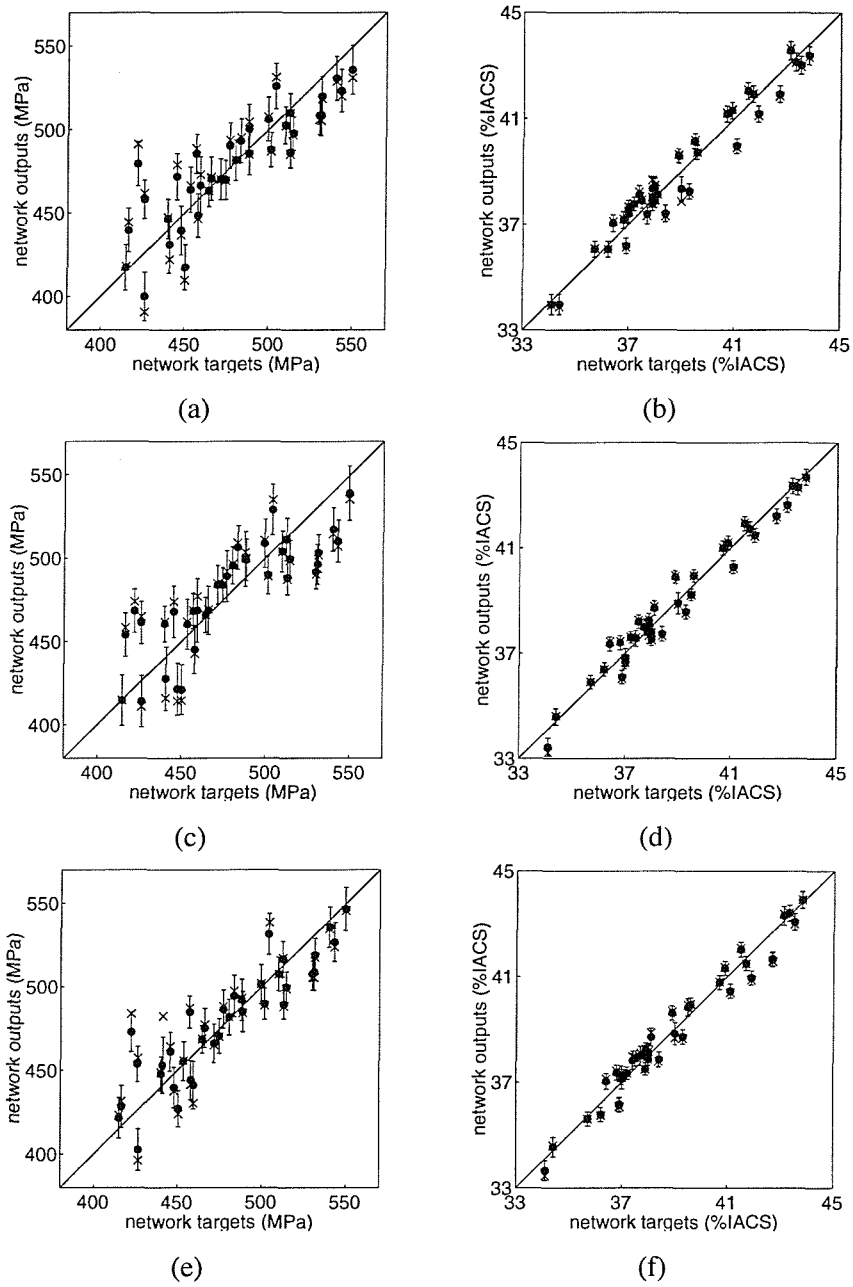
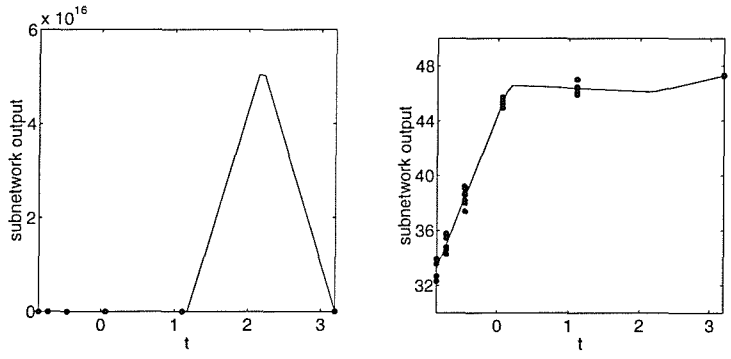


Figure 8.11: Prediction scatterplots for the neurofuzzy models for the output properties investigated for data sets A, B and C: $\sigma_{0.2}$ (a), (c) and (e) and σ_{el} (b), (d) and (f). Model estimates attained are shown for both the training data (•) and leave-one-out predictions (×), the former augmented with $\pm 1\sigma$ error bars.

responds to the univariate addition of ageing time with a highly flexible approximation (store 2), whereby the subnetwork uses three internal knots whose response is shown in Figure 8.12 (a). This is clearly seen to have resulted in a highly overfitted subnetwork structure and during the model pruning, two successive knot deletions are seen to remove the redundancy present in this subnetwork, the first of which modifies the approximation on ageing time to that shown in Figure 8.12 (b), the second yielding the final subnetwork response shown in previous figures. Although the pruning refinements have been seen to successfully remove undue degrees of freedom, the addition of ageing time with an unduly complex fit may have prevented other constructive refinements, resulting in the identification of a poor model for σ_{el}. A comparison with the results in sections 8.6.1 and 8.7 will identify whether this has resulted in a suboptimal model search.



(a) (b)

Figure 8.12: Subnetwork responses for ageing time corresponding to the 2^{nd} (a) and 3^{rd} (b) refinements performed in the model constructions for the σ_{el} models.

From an inspection of the subnetwork responses, it is seen that strength levels generally increase with increasing the Zn+Mg content, while for σ_{el} Figure 8.8 shows that decreasing the Zn+Mg content reduces the σ_{el} levels attainable.

The general effect of an increase in the $x_{Zn,w} : x_{Mg,w}$ quantity was to decrease $\sigma_{0.2}$ and increase σ_{el}. For both properties, the Zn:Mg dependency is further clarified by inspection of the tensor subnetwork response surface. From Figure 8.9 (b) it is suggested that alloy variants exhibiting higher $x_{Zn,w} : x_{Mg,w}$ ratios have the greatest susceptibility to a loss in strength with ageing. The general trend for decreasing $\sigma_{0.2}$ upon ageing is consistent with an increasing degree of overaging, i.e. coarsening of the η' phase upon extended ageing.

As in the MLR analysis, the nature of the approximation on the (Zn+Cu):Mg quantity may again be interpreted as the effective Cu contribution. The presence of both $(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$ and $x_{Zn,w} : x_{Mg,w}$ (despite a correlation value of 0.941 between these two quantities) is an indication that copper additions increase strength levels, whereas the omission of the Cu-related quantity in the σ_{el} model suggests Cu as having a lesser effect on electrical conductivity. These results are seen to be consistent with the alloy system and the

analysis performed by Pitcher (1998).

The results obtained show that, in general, the most parsimonious models were inferred from data set B, suggesting this data set comprises a subset of the variates having greater descriptive metallurgical features, enabling the determination of a more physically representative set of dependencies, more amenable to an explicit interpretation of microstructure-property relationships. The model construction procedures identified t and $x_{Mg,xs}$ as the most descriptive variables determining σ_{el}, whilst the selection of t, $x_{\eta'}$ and $x_{Mg,\alpha}$ for $\sigma_{0.2}$ are both in accordance with the metallurgical understanding described in section 8.4.1. One interesting result was the selection of the tensor product $x_{Mg,\alpha} \times t$. In terms of microstructure development this interaction identifies the amount of Mg initially dissolved in the Al-rich phase as having a significant influence on the rate of strength reduction, due to coarsening of the main phases, suggesting the $x_{Mg,\alpha} \times t$ dependency possibly being related to the interaction of Mg atoms with vacancies, this latter interaction significantly influencing diffusion rates, see e.g. (Starink and Zahra 1998). The Mg $\times$ t dependency is supported by the consistency with which input variables related to Mg have also been combined with t to form tensor product subnetworks in the other data sets. From inspection of the Mg $\times$ t subnetworks, there is a clear suggestion that the strength of the higher Mg alloys does not decrease as rapidly as in the case of low Mg alloys.

8.6.1 Model initialisation

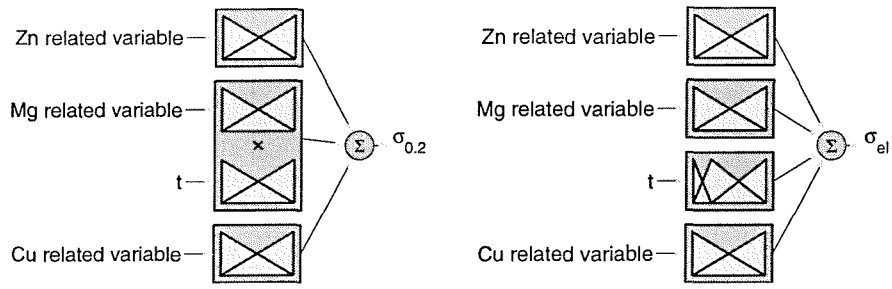
In Chapter 5 it was seen that one of the appealing features of the neurofuzzy networks considered in this work, was the ability to formulate a set of rules within a neural-type architecture. Therefore a simple initialisation of the network structures reflecting both metallurgical and empirical understanding of the dependencies characterising the present alloy system was investigated. The modelling task then corresponding to weight-rule confidence training.

For $\sigma_{0.2}$, models were constrained in the form shown in Figure 8.13 (a), reflecting the sources of strengthening believed to be present, introducing the Mg $\times$ t term identified by the previous model construction procedures.

A comparison of the network structures determined by the ASMOD algorithm shows that the models for σ_{el} are generally determined by an additive structure comprising a piecewise linear approximation of t and (depending on the data set) composition related inputs. Hence, a general model structure including all these terms is shown in Figure 8.13 (b).

Thus, constraining the models to the general forms shown in Figures 8.13 (a) and (b), reflected system understanding achieved both empirically and physically, and provided a further comparison of the effect of the data set transformations on modelling $\sigma_{0.2}$ and σ_{el}.

An assessment of the approximations attained for the different input subsets, by constrained network structures as described above, are shown in Tables 8.14 and 8.15. In addition to these sets, obtained from data sets A, B and C, a fourth set, D, was defined. This data set



(a) (b)

Figure 8.13: General form of the model structures suggested for the neurofuzzy networks for $\sigma_{0.2}$ (a) and σ_{el} (b).

being the same as that derived from data set B apart from taking $1/x_{Mg,\alpha}$ instead of $x_{Mg,\alpha}$ for the model for $\sigma_{0.2}$ model. This simple (non-linear) permutation was based on inspection of the model performances attained by inputs of the type $1/Mg$, and revealed this term to attain a better (linearised) approximation to the data[10]. As seen from Table 8.14 a tangible improvement is achieved by the simple linearisation yielded by the $1/x_{Mg,\alpha}$ term.

Data set	ANOVA terms	$\hat{\sigma}^2_{df}$	LOOCV
A	$t \times x_{Mg,w}, x_{Zn,w}, x_{Cu,w}$	421	507
B	$t \times x_{Mg,\alpha}, x_{\eta'}, x_{Cu,\alpha}$	437	528
C	$t \times x_{Zn,w} : x_{Mg,w}, x_{Zn,w} + x_{Mg,w}$, $(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$	334	436
D	$t \times 1/x_{Mg,\alpha}, x_{\eta'}, x_{Cu,\alpha}$	384	466

Table 8.14: Summary of the effects of the data transformations through defining network structures for $\sigma_{0.2}$. Note that for all models, $p = 8$ and $df = 6$.

Data set	ANOVA terms	$\hat{\sigma}^2_{df}$	LOOCV
A	$t, x_{Mg,w}, x_{Zn,w}, x_{Cu,w}$	0.251	0.291
B	$t, x_{Mg,xs}, x_{\eta'}, x_{Cu,\alpha}$	0.183	0.219
C	$t, x_{Zn,w} : x_{Mg,w}, x_{Zn,w} : x_{Mg,w}$, $(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$	0.359	0.452

Table 8.15: Summary of the effects of the data transformations through defining network structures for σ_{el}. Note that for all models, $p = 9$ and $df = 6$.

[10]This single modification to set B is thus seen to be based on the results attained by data set C.

8.6.2 Assessing stability of the model selection

As in Chapter 7, a stepwise (SW) model search was employed to determine models on all three data sets for both $\sigma_{0.2}$ and σ_{el} in order to assess the stability of the model structure inferred using a different iterative construction procedure. Identical model structures to those identified by the FS/BE construction were determined with the exception of models for σ_{el} for data sets A and C. For data set A, an additional univariate piecewise subnetwork for $x_{Fe,w}$ was included, while for data set C four univariate subnetworks were determined (of which ageing time and $(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$ were piecewise linear approximations). Although yielding improvements in the SS and both training set and LOOCV variance estimates, the inclusion of $x_{Fe,w}$ in preference to other refinements (from inspection of the model construction history) was marginal and the form of the approximation inferred problematic to validate. The piecewise approximation determined on $(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$ was promoted by removing redundancy early on in the model construction (on the t subnetwork), subsequently promoting the (piecewise) univariate addition of $(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$ as opposed to the tensor product $(x_{Zn,w} : x_{Mg,w}) \times (x_{Zn,w} + x_{Mg,w})$ refinement determined in the FS/BE model construction.

The sensitivity of the FS/BE construction algorithm to small perturbations in the data set was investigated by resampling the complete data set into multiple training-test sets (with a 90%-10% split). The generalisation performances attained were inferred from the average predictive performance obtained from the 20 models where, each of the model's test errors were estimated on four samples. These results subsequently allowed a comparison to be made with those obtained in section 8.7, from application of the SVM methods.

The ANOVA terms determined over each of the multiple model runs for both $\sigma_{0.2}$ and σ_{el} by the ASMOD algorithm are summarised in Tables 8.16 to 8.21[11]. The model structure determined with greater consistency (11 times) from data set A in modelling $\sigma_{0.2}$, as noted above, is seen to differ from that identified upon using the complete data set for training, which over these multiple runs is determined 7 times (Table 8.16). The trends for the former model structure are shown in Figure 8.14[12].

From Table 8.17 it is seen that the models determined by the ASMOD algorithm typically comprise $x_{\eta'}$, $x_{Mg,\alpha}$ and t univariate terms, with a number of models (8) exhibiting bivariate terms (mainly Mg $\times$ t) in preference to univariate terms. As for data set A, a number of models approximate the ageing time dependency with a piecewise approximation (5 out of the 11 univariate ageing time subnetworks). As in the case of training on the complete data set, the absence of a Cu contribution can be explained in light of the transformations performed

[11] As in Chapter 7, ‡ indicates an identical model structure as that determined from the model inferred from the complete data set, while † indicates the model structure which over the multiple model runs is the most consistently inferred.

[12] A few models (three) exhibited a univarite piecewise approximation to t.

in section 8.4. This can be understood in terms of both the properties of this data set and the smaller training set sizes, the latter is seen to further restrict the number of degrees of freedom allowed by the model complexity measure.

Models for $\sigma_{0.2}$ determined from data set C (Table 8.18), show that ageing time is never retained as a univariate term, being present as a tensor product (the two univariate contributions refined in favour of a bivariate term). The most consistent terms identified are seen to comprise the Zn+Mg univariate and the $(\mathrm{Zn} : \mathrm{Mg}) \times t$ bivariate terms. The model structure determined with greater consistency is then seen to differ from that determined from training on the complete data set by the absence of the $(\mathrm{Zn} + \mathrm{Cu}) : \mathrm{Mg}$ term.

A comparison between the models inferred for σ_{el} from data sets A and B show consistent results to those attained on training on the complete data set, with structures comprising a piecewise approximation for the ageing time dependency and a univariate term characterising that between σ_{el} with Mg. From data set B, a number of models are seen to include $\mathrm{x}_{\mathrm{Mg},\alpha}$ in place of $\mathrm{x}_{\mathrm{Mg,xs}}$, depending on the particular training-test split.

The models inferred from data set C (Table 8.21) are generally comprised by univariate contributions of Zn:Mg, Zn+Mg and t, the latter approximated by a piecewise linear term. The original model structure determined by training on all data is never identified from any of the multiple runs, the construction algorithms being unable to support (refine) a tensor product term between the Zn:Mg and Zn+Mg terms. As in previous analyses, the contribution of Cu was never included.

The error estimates inferred over the multiple model runs are summarised in Tables 8.22 and 8.23[13]. On all data sets, models for which $\sigma_{0.2}$ attained the same structure as those determined from training on the full data set exhibited lower training errors compared to the most consistently inferred network structures from the multiple runs. More significantly, models inferred from data sets B and C exhibited a lower validation error. However, due to both the small number of models inferred, and more significantly, the size of the data set considered, it is inappropriate to identify which of these two is a better model structure.

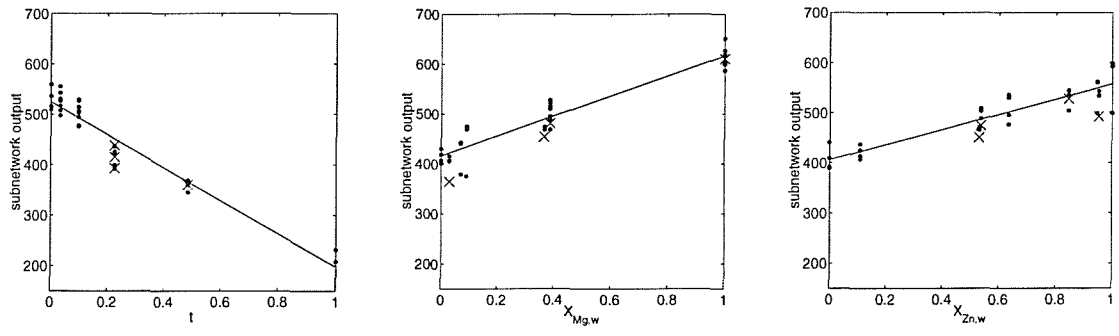
In terms of the different data sets considered, the results for $\sigma_{0.2}$ are consistent with those attained from training on all data, with data set C providing the best approximation abilities. A comparison between the test errors and the LOOCV estimates previously determined shows the latter as giving somewhat conservative estimates of the generalisation performance.

For σ_{el}, a comparison between validation error estimates shows that whilst data sets A and B attain similar accuracies, data set C exhibits poorer empirical performance.

[13]In the presentation of these results, the test errors refer to the mean of the error estimates on unseen data obtained over the 20 different training-test splits. In addition, the performance obtained by the most consistently identified model structures (i.e. ANOVA terms) are presented. These correspond to the results indicated by †. Neurofuzzy models attaining the same structure as that determined from using all the data for training correspond to the ‡ results.

run	Zn	Mg	Cu	t	Mg × t	Zn × t
1‡	•		•		•	
2†	•	•		•		
3†	•	•		•		
4	•				•	
5†	•	•		•		
6†	•	•		•		
7†	•	•		•		
8†	•	•		•		
9		•				•
10	•				•	
11‡	•		•		•	
12†	•	•		•		
13‡	•		•		•	
14†	•	•		•		
15†	•	•		•		
16			•		•	
17	•	•	•	•		
18†	•	•		•		
19	•				•	
20†	•	•		•		
Total	18	13	5	12	7	1

Table 8.16: Summary of the ANOVA terms determined for $\sigma_{0.2}$ by the neurofuzzy model construction algorithms from data set A for each of the 20 resampled data sets.

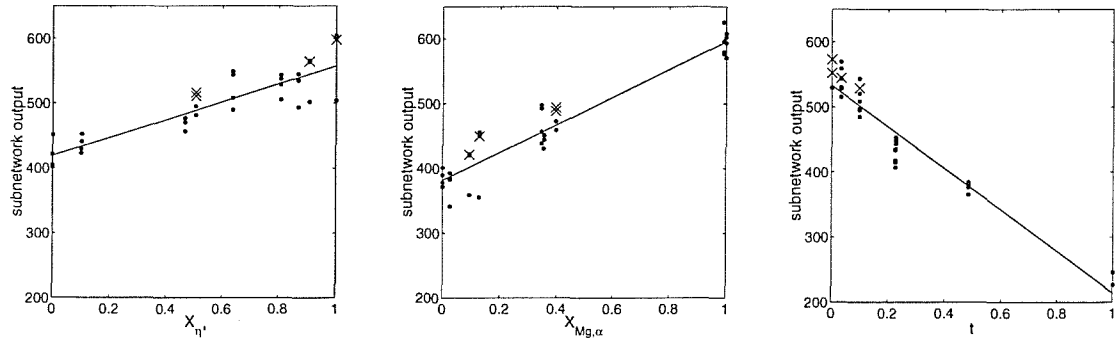


(a) (b) (c)

Figure 8.14: General form of the subnetwork responses for the ANOVA terms most consistently determined in the models constructed by the neurofuzzy framework from data set A in modelling $\sigma_{0.2}$.

run	$x_{Cu,\alpha}$	$x_{Mg,\alpha}$	$x_{\eta'}$	t	$x_{Mg,\alpha} \times t$	$x_{\eta'} \times t$	$x_{Mg,xs} \times t$
1	•		•		•		
2†		•	•	•			
3†		•	•	•			
4‡			•		•		
5†		•	•	•			
6†		•	•	•			
7†		•	•	•			
8†		•	•	•			
9		•				•	
10‡			•		•		
11‡			•		•		
12†		•	•	•			
13	•	•					•
14†		•	•	•			
15†		•	•	•			
16†		•	•	•			
17	•	•	•	•			
18†		•	•	•			
19‡			•		•		
20			•				•
Total	3	14	18	12	5	1	2

Table 8.17: Summary of the ANOVA terms determined for $\sigma_{0.2}$ by the neurofuzzy model construction algorithms from data set B for each of the 20 resampled data sets.

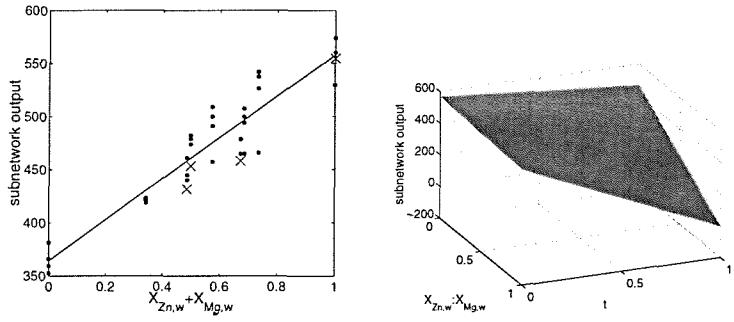


(a) (b) (c)

Figure 8.15: General form of the subnetwork responses for the ANOVA terms most consistently determined in the models constructed by the neurofuzzy framework from data set B in modelling $\sigma_{0.2}$.

run	Zn:Mg	(Zn+Cu):Mg	Zn+Mg	Zn : Mg × t	(Zn + Cu) : Mg × t
1‡		•	•	•	
2†			•	•	
3	•		•		•
4†			•	•	
5†			•	•	
6	•		•		•
7†			•	•	
8†			•	•	
9†			•	•	
10†			•	•	
11‡		•	•	•	
12	•		•		•
13‡		•	•	•	
14†			•	•	
15†			•	•	
16	•		•		•
17	•		•		•
18†			•	•	
19†			•	•	
20‡		•	•	•	
Total	5	4	20	15	5

Table 8.18: Summary of the ANOVA terms determined for $\sigma_{0.2}$ by the neurofuzzy model construction algorithms from data set C for each of the resampled data sets.

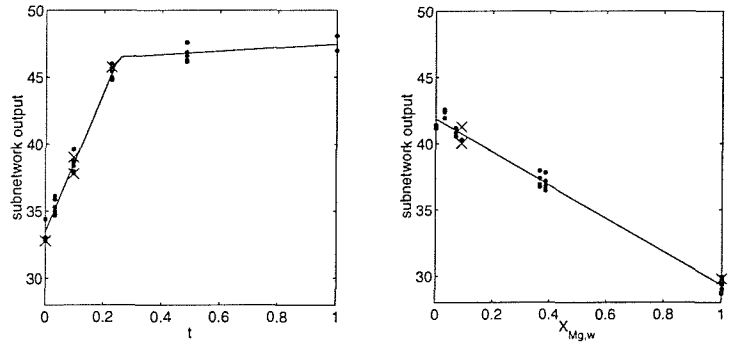


(a) (b)

Figure 8.16: General form of the subnetwork responses for the ANOVA terms most consistently determined in the models constructed by the neurofuzzy framework from data set C in modelling $\sigma_{0.2}$.

run	Zn	Mg	t
$1^{\dagger,\ddagger}$		•	•
$2^{\dagger,\ddagger}$		•	•
$3^{\dagger,\ddagger}$		•	•
$4^{\dagger,\ddagger}$		•	•
$5^{\dagger,\ddagger}$		•	•
$6^{\dagger,\ddagger}$		•	•
$7^{\dagger,\ddagger}$		•	•
$8^{\dagger,\ddagger}$		•	•
$9^{\dagger,\ddagger}$		•	•
10	•	•	•
$11^{\dagger,\ddagger}$		•	•
$12^{\dagger,\ddagger}$		•	•
$13^{\dagger,\ddagger}$		•	•
$14^{\dagger,\ddagger}$		•	•
$15^{\dagger,\ddagger}$		•	•
$16^{\dagger,\ddagger}$		•	•
$17^{\dagger,\ddagger}$		•	•
$18^{\dagger,\ddagger}$		•	•
$19^{\dagger,\ddagger}$		•	•
$20^{\dagger,\ddagger}$		•	•
Total	1	20	20

Table 8.19: Summary of the ANOVA terms determined for σ_{el} by the neurofuzzy model construction algorithms from data set A for each of the resampled data sets.



(a) (b)

Figure 8.17: General form of the subnetwork responses for the ANOVA terms most consistently determined in the models constructed by the neurofuzzy framework from data set A in modelling σ_{el}.

run	$x_{Mg,\alpha}$	$x_{Mg,xs}$	t
$1^{\dagger,\ddagger}$		•	•
2	•		•
$3^{\dagger,\ddagger}$		•	•
4	•		•
$5^{\dagger,\ddagger}$		•	•
$6^{\dagger,\ddagger}$		•	•
7	•		•
$8^{\dagger,\ddagger}$		•	•
9	•		•
$10^{\dagger,\ddagger}$		•	•
$11^{\dagger,\ddagger}$		•	•
$12^{\dagger,\ddagger}$		•	•
$13^{\dagger,\ddagger}$		•	•
$14^{\dagger,\ddagger}$		•	•
$15^{\dagger,\ddagger}$		•	•
16	•		•
17	•		•
$18^{\dagger,\ddagger}$		•	•
$19^{\dagger,\ddagger}$		•	•
20	•		•
Total	7	13	20

Table 8.20: Summary of the ANOVA terms determined for σ_{el} by the neurofuzzy model construction algorithms from data set B for each of the 20 resampled data sets.

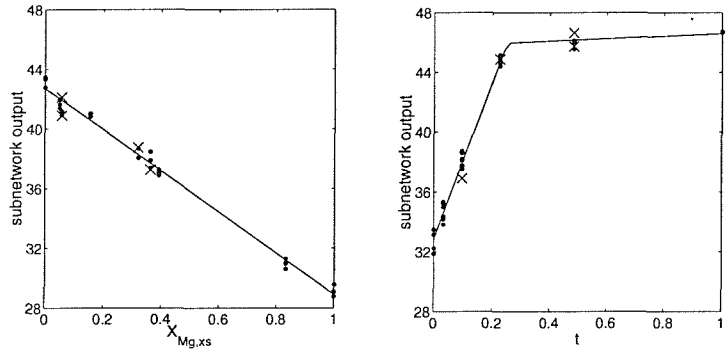
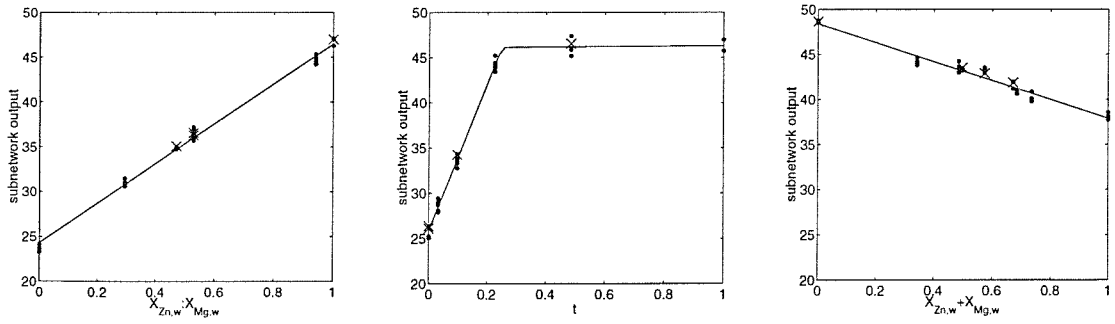


Figure 8.18: General form of the subnetwork responses for the ANOVA terms most consistently determined in the models constructed by the neurofuzzy framework from data set B in modelling σ_{el}.

run	Zn:Mg	Zn+Mg	t	Zn : Mg × t	Zn + Mg × t
1†	•	•	•		
2		•		•	
3†	•	•	•		
4†	•	•	•		
5†	•	•	•		
6		•		•	
7†	•	•	•		
8†	•	•	•		
9†	•	•	•		
10†	•	•	•		
11†	•	•	•		
12†	•	•	•		
13†	•	•	•		
14†	•	•	•		
15		•		•	
16†	•	•	•		
17		•		•	
18		•		•	
19†	•	•	•		
20	•	•			•
Total	15	20	14	5	1

Table 8.21: Summary of the ANOVA terms determined for σ_{el} by the neurofuzzy model construction algorithms from data set C for each of the 20 resampled data sets.



(a) (b) (c)

Figure 8.19: General form of the subnetwork responses for the ANOVA terms most consistently determined in the models constructed by the neurofuzzy framework from data set C in modelling σ_{el}.

8.7 Support vector regression

The approximation abilities of the SVM methods described in Chapter 6 were investigated on all three data sets in modelling both $\sigma_{0.2}$ and σ_{el}. The SUPANOVA approach was repeated on the same 20 training-test set partitions used in section 8.6, from which estimates of the generalisation performance could be determined.

As in Chapter 7, a quadratic loss function was used and an infinite ANOVA spline as the kernel choice to perform the non-linear mapping into feature space. As in section 7.8 the empirical performance attained by Stage I in the SUPANOVA framework was also determined.

The empirical results obtained for the neurofuzzy, Stage I, and SUPANOVA for $\sigma_{0.2}$ and σ_{el} are summarised in Tables 8.22 and 8.23[14].

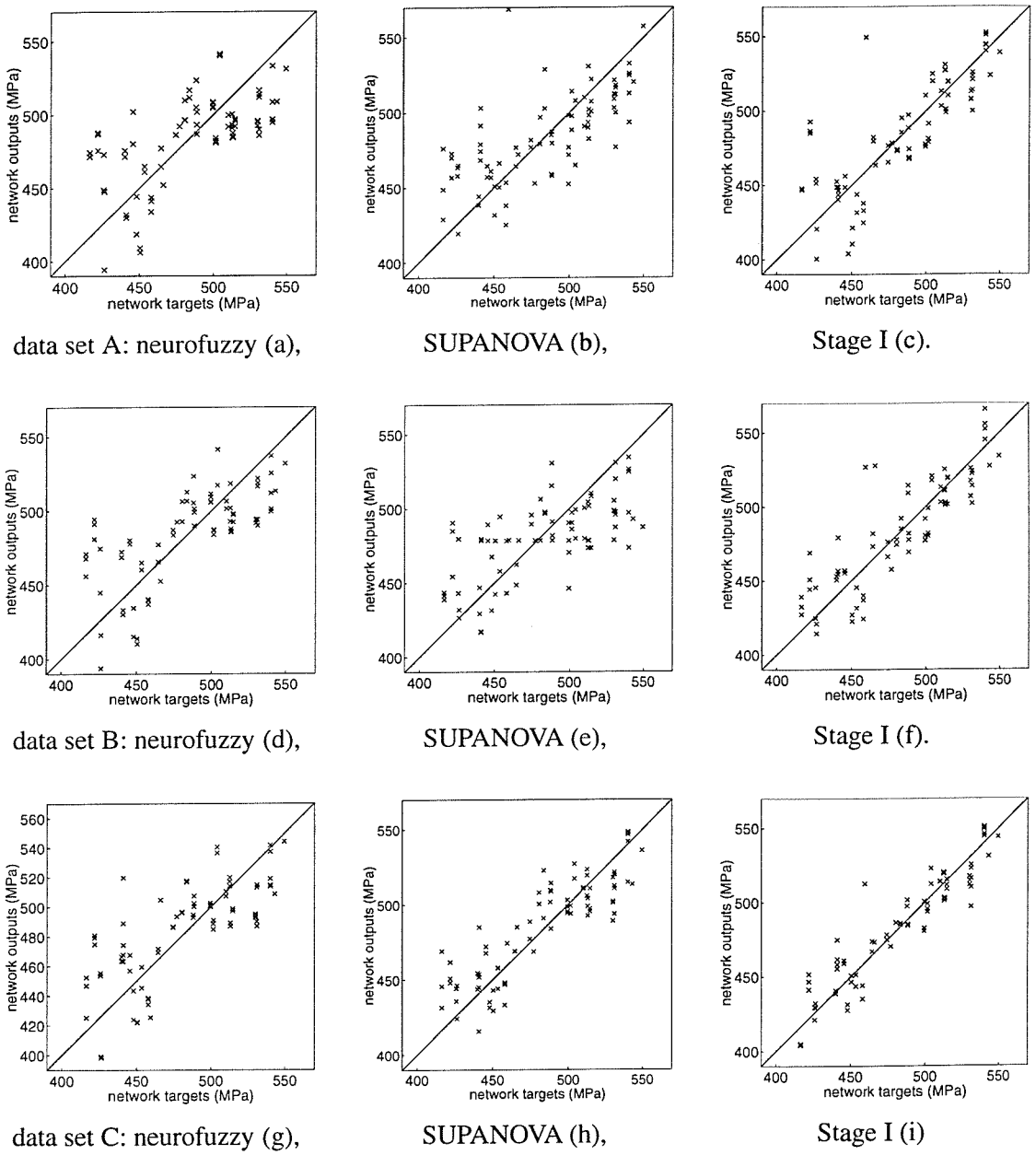
	Frequency of selection			Training error $\mu\left(\hat{\sigma}_N^2\right)$			Test error $\mu\left(\hat{\sigma}_N^2\right)$		
	A	B	C	A	B	C	A	B	C
Neurofuzzy	20	20	20	460	403	312	1159	860	709
†	11	11	11	540	444	373	1008	837	830
‡	3	4	4	273	435	236	1108	619	695
SUPANOVA	20	20	20	297	579	155	791	867	368
†	2	7	8	73	207	60	768	338	207
Stage I	20	20	20	108	55	49	614	510	195
†	(2)	(7)	(8)	94	71	36	1483	426	165

Table 8.22: Summary of the mean training and test (generalisation) set error estimates attained over the 20 resampled data sets in the modelling of $\sigma_{0.2}$ by the three techniques for all three data sets (A,B and C). The error estimates given are the mean error attained for the ANOVA basis most frequently selected. The table also presents empirical errors for the models determined over the 20 resampled data sets whose structure was the same as that determined by training on the complete data set. Error estimates corresponding to Stage I in the SUPANOVA models quoted are also given.

The test set predictions obtained for the different modelling approaches are graphically summarised in the target versus prediction scatterplots shown in Figures 8.20 and 8.21, from which the variance attained in predicting the same test cases can be observed.

In a comparison between the different modelling approaches, the results presented in Table 8.22 show how Stage I attains better generalisation abilities in modelling $\sigma_{0.2}$ over all three data sets. However, Table 8.23 shows that whilst attaining a similar training error on all three data sets, the performance of Stage I in terms of the validation/test variance estimates are seen to be inferior to the parsimonious modelling results. In general, the neurofuzzy models are seen to underperform, particularly in modelling $\sigma_{0.2}$.

[14]For the SVM results, the average performance obtained by the most consistently identified model structures (i.e. ANOVA terms) are presented by †.

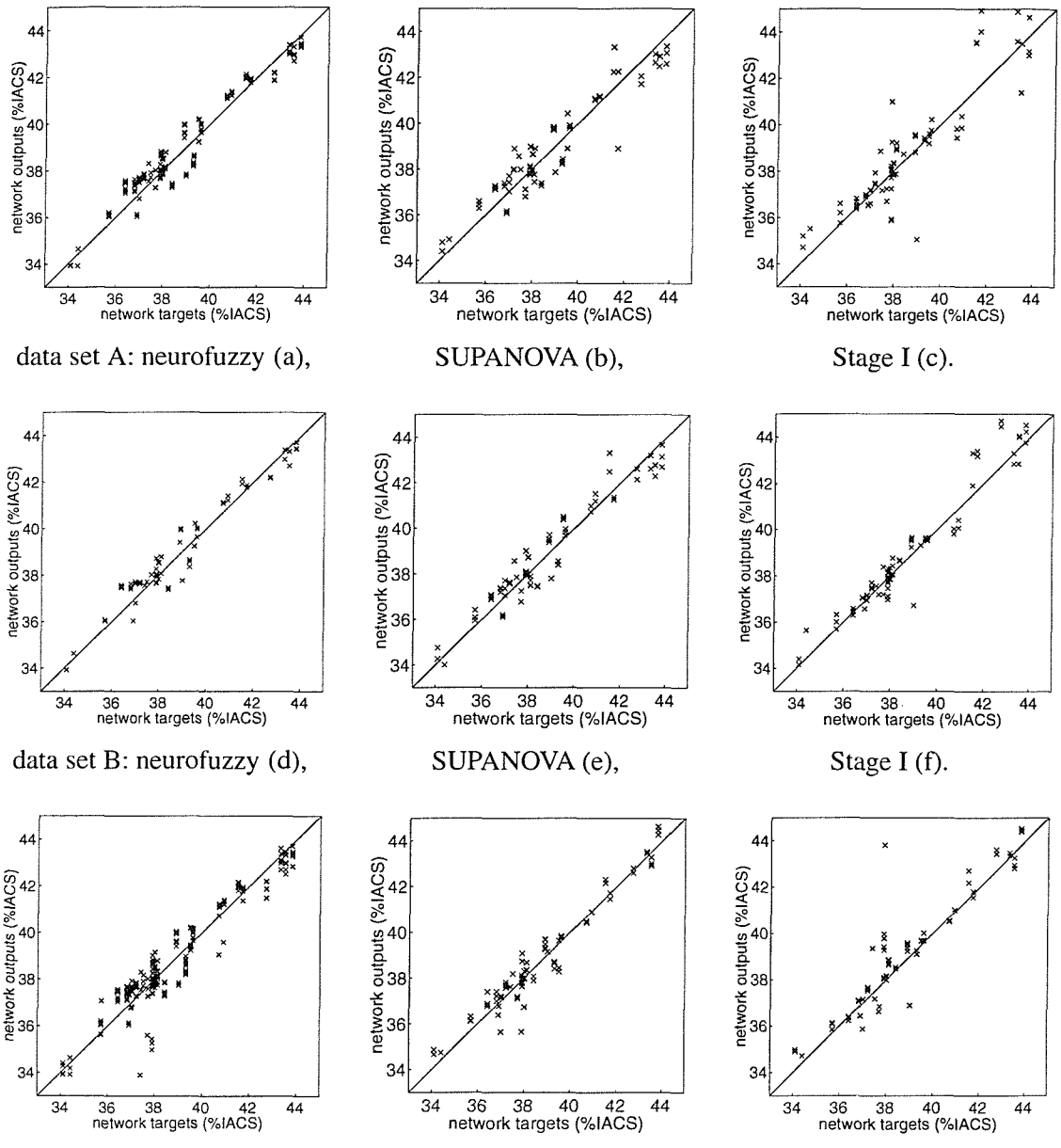


data set A: neurofuzzy (a), SUPANOVA (b), Stage I (c).

data set B: neurofuzzy (d), SUPANOVA (e), Stage I (f).

data set C: neurofuzzy (g), SUPANOVA (h), Stage I (i)

Figure 8.20: Predictions attained by the different modelling techniques on the test data for $\sigma_{0.2}$ from data set A: (a)-(c), data set B (d)-(f) and data set C (g)-(i).



data set A: neurofuzzy (a), SUPANOVA (b), Stage I (c).

data set B: neurofuzzy (d), SUPANOVA (e), Stage I (f).

data set C: neurofuzzy(g), SUPANOVA (h), Stage I (i).

Figure 8.21: Predictions attained by the different modelling techniques on the test data for σ_{el} from data set A: (a)-(c), data set B (d)-(f) and data set C (g)-(i).

	Frequency of selection			Training error $\mu\left(\hat{\sigma}_N^2\right)$			Test error $\mu\left(\hat{\sigma}_N^2\right)$		
	A	B	C	A	B	C	A	B	C
Neurofuzzy	20	20	20	0.290	0.227	0.58	0.37	0.35	1.44
†	19	13	14	0.296	0.211	0.295	0.349	0.362	0.429
‡	19	13	–	0.296	0.211	–	0.349	0.362	–
SUPANOVA	20	20	20	0.639	0.239	0.13	0.67	0.47	0.35
†	12	10	6	0.310	0.231	0.068	0.513	0.411	0.177
Stage I	20	20	20	0.087	0.04	0.06	1.27	0.42	0.88
†	(12)	(10)	(6)	0.088	0.023	0.023	1.25	0.594	0.106

Table 8.23: Summary of the mean training and test (generalisation) set error estimates attained over the 20 resampled data sets in the modelling of σ_{el} by the three techniques for all three data sets (A,B and C). Comparison of the consistency (stability) exhibited by the neurofuzzy and SUPANOVA techniques in the determination of parsimonious models for σ_{el} for all three data sets. The error estimates given are the mean error attained for the ANOVA basis most frequently selected. The table also presents empirical errors for the models determined over the 20 resampled data sets whose structure was the same as that determined by training on the complete data set. Error estimates corresponding to Stage I in the SUPANOVA models quoted are also given.

In general, the ANOVA representations determined by both SUPANOVA and neurofuzzy frameworks, exhibit a greater stability in the modelling results obtained for σ_{el}. Whereas there was seen to be considerable uncertainty (instability) in inferring ANOVA representations for $\sigma_{0.2}$, particularly from data set A. In the SUPANOVA framework, this can be seen to be a reflection of the size of the full ANOVA basis considered, from which the sparse subselection is performed. As both data sets A and B give rise to larger ANOVA basis, compared with data set C, the sensitivity in subselecting the same ANOVA components over the different training-test set partitions is understandably greater for the former two.

The general nature of the approximations determined for the most consistently selected ANOVA terms are shown in Figures 8.22 to 8.27. The smoothness of these approximations in certain data set instances differed by a considerable amount, resulting from the different degree of regularisation inferred from the cross-validation resampling procedure used for determining the smoothing parameter C. On certain data set instances the approximations inferred clearly resulted in a considerable overfitting of the training data.

Table 8.24 shows that for $\sigma_{0.2}$, the number of ANOVA terms selected varies from the presence of only a bias term to models having seven terms, with a considerable number of different terms considered (11) over the multiple runs. Consequently it is seen that there is a degree of uncertainty in the sparse representations, with no overall structure systematically inferred. Models are generally characterised by a number of univariate terms and by a bivariate term, the most consistently selected being the Zn $\times$ t dependency. The bivariate and trivariate

terms selected in a few instances (e.g. the Cu × Zr, Zn × Fe × Si) are difficult to validate.

In Figure 8.22, the selection of univariate terms $x_{Zn,w}$ and $x_{Mg,w}$ are both in agreement with the trends inferred in section 8.6. More problematic is the validation of the general form of the approximations characterising the $x_{Cu,w}$ and $x_{Zr,w}$ terms (the subset of models with a greater degree of regularisation than those shown in Figure 8.22 exhibited monotonically increasing trends for these two variates). As in section 8.6, the ageing time dependency is consistent with an overageing behaviour, the kernel approximation attaining a physically believable behaviour. The dependencies characterising $\sigma_{0.2}$ with $x_{Zn,w}$ and t are to be fully understood by considering both univariate and bivariate terms: alloys with a higher Zn content are more susceptible to a loss in strength arising from overageing.

It is seen that the Zn × t dependency inferred by the SUPANOVA framework is not present in the models identified by the ASMOD algorithm, where a dependency of the form Mg × t dependency was preferred[15].

Table 8.25 shows that in modelling $\sigma_{0.2}$ from data set B, the univariate terms $x_{\eta'}$, $x_{Mg,xs}$ and t terms are consistently selected, whereas compared to data set A, $x_{Cu,\alpha}$ is seldom present. As in section 8.6, this may be attributed to the effect of the input transformations, whereby other terms, particularly $x_{Mg,xs}$ and $x_{\eta'}$, are more parsimoniously accounting for the output variance. This consideration seems to be further supported by the fact that a comparison between the Zn and Mg-related approximations reveals a stronger non-linearity associated with the quantities present in data set B. Thus, an intrinsic contribution of the Cu levels may inherently be present.

In a further comparison with the other $\sigma_{0.2}$ results, data set A in particular reveals a considerably higher proportion of models (5) comprising only a bias term. This suggests that the dependencies present in data set B are reflected in the sparse subset selection performed in the SUPANOVA framework, whereby the solution corresponding to minimising the $\|\cdot\|_1$ norm of the c coefficients (see section 6.5.2) is seen to result in the subsequent selection of only the bias term.

A comparison in terms of the ANOVA terms determined by the two parsimonious modelling approaches shows that the SUPANOVA approach is seen to include the $x_{Mg,xs}$ rather than the $x_{Mg,\alpha}$ univariate term. The bivariate terms determined in the SUPANOVA framework differ from those determined by the ASMOD algorithm, the neurofuzzy models include bivariate terms with a greater consistency.

As in section 8.6.2, a term describing the dependency of $\sigma_{0.2}$ with ageing time and the balance of the precipitation sequence is (apart from one instance) always present in the models determined from data set C. The most consistently selected terms being the univariate approximations of t and Zn+Mg dependencies and the $((Zn + Cu) : Mg) \times t$ bivariate term, this latter selected in preference to $(Zn : Mg) \times t$, and seen to be consistently present in the

[15] An inspection of the model construction showed a consistency to select (from the SS scores) a tensor product (replacing the respective univariate subnetworks) between t and Mg rather than Zn.

higher order terms. The univariate trends are seen to be consistent with those inferred from the MLR analysis of section 8.5.

The bivariate term shown in Figure 8.24 (e) suggests that when the alloy is at its greatest strengthening potential it has the greatest susceptibility to a loss in strength with ageing. Where variates of the form Xx:Mg are present in the same model, the dependencies as shown in Figures 8.24 (a), (b) and (e) require some consideration. As in previous sections, Figure 8.24 (b) suggests an intrinsic strengthening effect arising from Cu additions. Then, Figures 8.24 (a) and (e) indicate a decreasing dependency of $\sigma_{0.2}$ with the effective quantities characterising the balance of the precipitation sequence.

From Figures 8.25 and 8.26 models inferred for σ_{el} from data sets A and B are seen to be characterised by similar univariate approximations determined on the Mg-related variates and t, with a bivariate term Mg $\times$ t present in several models. Both the SUPANOVA and neurofuzzy frameworks are seen to identify similar terms, with models differing in the Mg-related variate ($x_{Mg,xs}$ selected with greater consistency in the former, $x_{Mg,\alpha}$ always preferred in the latter). A comparison with section 8.6.2, shows how the complexity control approach used in the neurofuzzy framework prevents the inclusion of both univariate and higher order terms, which are present in the SUPANOVA models.

Figures 8.25 and 8.26 reveal a greater non-linearity in approximating the $x_{Mg,xs}$ contribution compared with $x_{Mg,w}$, and on $x_{\eta'}$ in comparison with $x_{Zn,w}$. This was seen to be consistent over the whole set of models determined from the different training-test partitions.

Similarly to the results obtained in modelling $\sigma_{0.2}$, the approximations attained from data set C show a systematic preference in characterising the σ_{el} dependency with ageing and precipitation balance with the (Zn + Cu) : Mg $\times$ t term. The trend in Figure 8.27 (d) may be attributed to an intrinsic effect of Cu on σ_{el} levels. Then from Figures 8.25 (a) and (e), the increasing trend exhibited by the quantities characterising the balance of the precipitates with σ_{el} is consistent with the results reported by Pitcher (1998). Figure 8.27 (e) suggests that at low ageing times there is little effect of the precipitate balance upon σ_{el} levels, but upon an increased overaged condition, a greater sensitivity of σ_{el} is exhibited.

From a comparison between the variance estimates presented in Tables 8.22 and 8.23 it is seen that the SUPANOVA models determined from data set C exhibit the best generalisation abilities for both properties. However, the results obtained by considering all 20 models determined for $\sigma_{0.2}$ are seen to be biased due to the presence of a number of models comprising only a bias term. Unsurprisingly, considerably better generalisation performance is inferred for the subset of models determined with greatest frequency.

run	bias	Zn	Mg	Cu	Zr	t	Zn × t	Mg × t	Cu × Zr	Zn × Cu × t	Zn × Fe × Si
1	○			●	●	●	●				
2	○	●		●	●	●				●	
3	○				●		●				
4	○	●		●	●	●	●	●			●
5	○	●		●	●	●	●				
6	○						●				
7	○		●	●	●	●	●				
8	○		●	●	●	●	●				
9	○	●					●				
10	○		●		●	●	●		●		
11	○	●	●	●	●	●	●	●			
12	○	●	●	●	●	●	●				
13†	○	●	●	●	●	●	●				●
14	○	●	●		●		●				
15	○	●	●		●		●				
16	○	●			●	●	●				
17†	○	●	●	●	●	●	●				●
18	○	●		●	●	●	●			●	
19	●										
20	○	●		●						●	
Total	1	13	9	12	16	13	17	2	1	3	3

Table 8.24: Summary of the ANOVA terms determined for $\sigma_{0.2}$ by the sparse basis selection employed within the SUPANOVA framework for each of the 20 resampled data sets obtained from data set A.

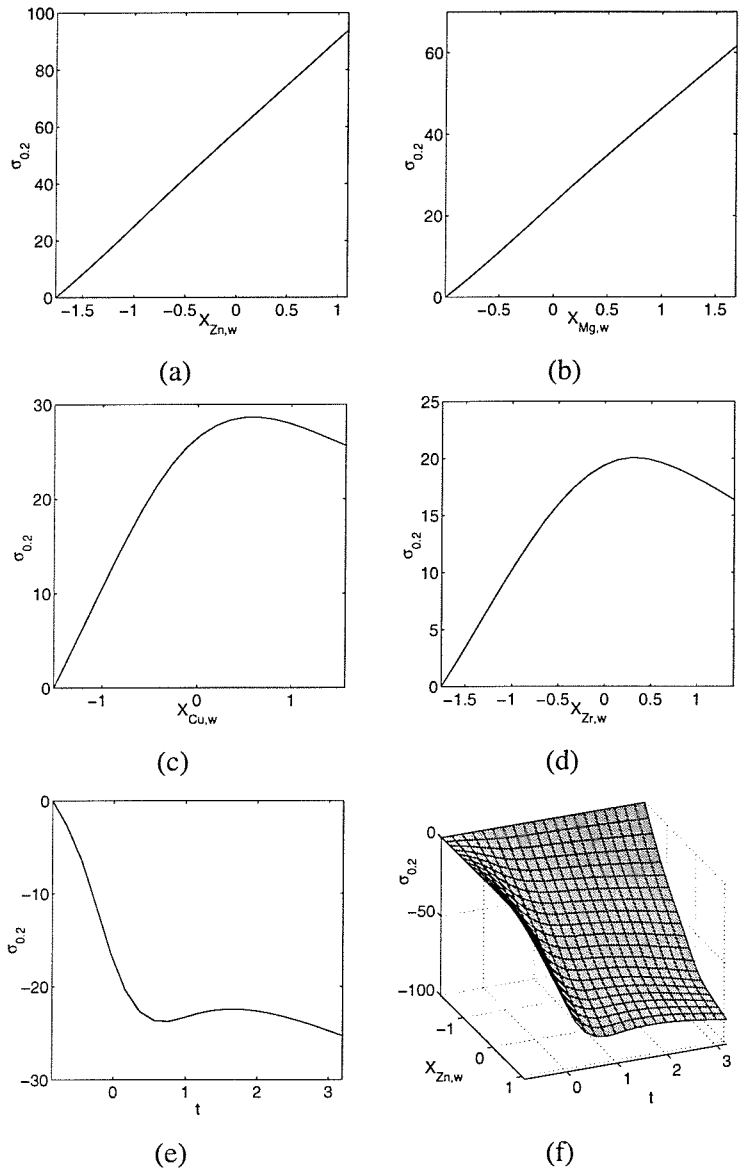
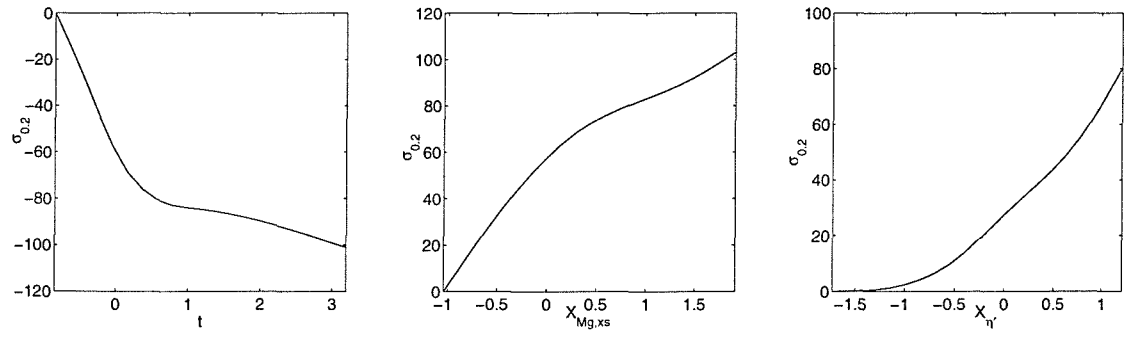


Figure 8.22: General form of the kernel approximations exhibited by the ANOVA terms most consistently identified in the basis selection stage of the SUPANOVA framework from data set A in modelling $\sigma_{0.2}$.

run	bias	$x_{Cu,\alpha}$	$x_{\eta'}$	$x_{Mg,xs}$	t	$x_{\eta'} \times t$	$x_{Cu,\alpha} \times x_{\eta'} \times t$	$x_{Mg,\alpha} \times x_{\eta'} \times t$
1	○				●			
2	○		●	●	●		●	
3	●							
4†	○		●	●	●			
5	●							
6	●							
7	○			●	●			
8†	○		●	●	●			
9†	○		●	●	●			
10	○		●	●	●	●		●
11	●							
12†	○		●	●	●			
13	○	●	●	●	●			
14†	○		●	●	●			
15	●							
16	○			●	●			
17	○	●	●	●	●			
18†	○		●	●	●			
19	○				●			
20†	○		●	●	●			
Total	5	2	11	13	15	1	1	1

Table 8.25: Summary of the ANOVA terms determined for $\sigma_{0.2}$ from data set B by the sparse basis selection employed within the SUPANOVA framework for each of the 20 resampled data sets.



(a) (b) (c)

Figure 8.23: General form of the kernel approximations exhibited by the ANOVA terms most consistently identified in the basis selection stage of the SUPANOVA framework from data set B in modelling $\sigma_{0.2}$.

run	bias	Zn:Mg	(Zn+Cu):Mg	Zn+Mg	t	(Zn + Cu) : Mg × t	Zn : Mg × t	Zn : Mg × (Zn + Cu) : Mg × t	(Zn + Cu) : Mg × Zn + Mg × t
1†	○	●	●	●	●	●			
2	○		●	●	●	●			
3	○			●	●	●			
4	○			●	●	●			
5	○			●	●	●			
6†	○	●	●	●	●	●			
7	○			●	●		●	●	
8†	○	●	●	●	●	●			
9	○	●	●	●	●	●			●
10†	○	●	●	●	●	●			
11	○		●	●	●	●			
12†	○	●	●	●	●	●			
13†	○	●	●	●	●	●			
14†	○	●	●	●	●	●			
15	○			●	●	●	●		
16	○	●		●	●	●			
17†	○	●	●	●	●	●			
18	○	●		●	●				
19	○			●			●		
20	○			●	●	●			
Total	-	11	11	20	19	17	3	1	1

Table 8.26: Summary of the ANOVA terms determined for $\sigma_{0.2}$ from data set C by the sparse basis selection employed within the SUPANOVA framework for each of the 20 resampled data sets.

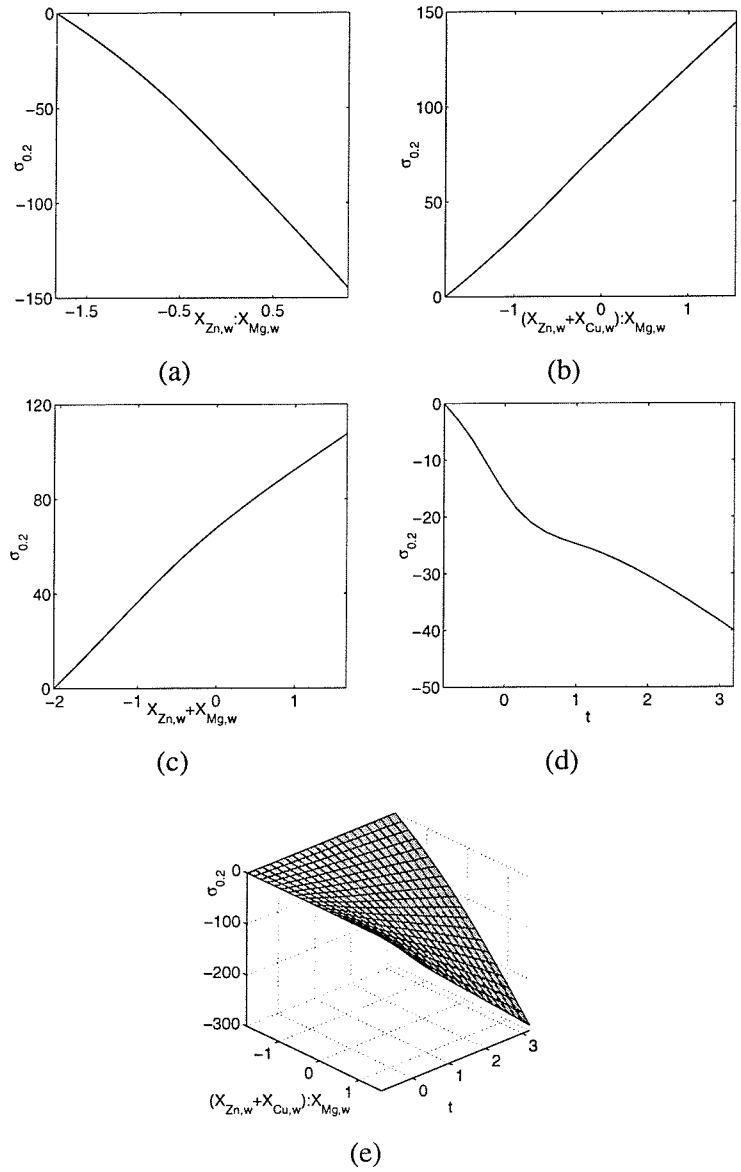
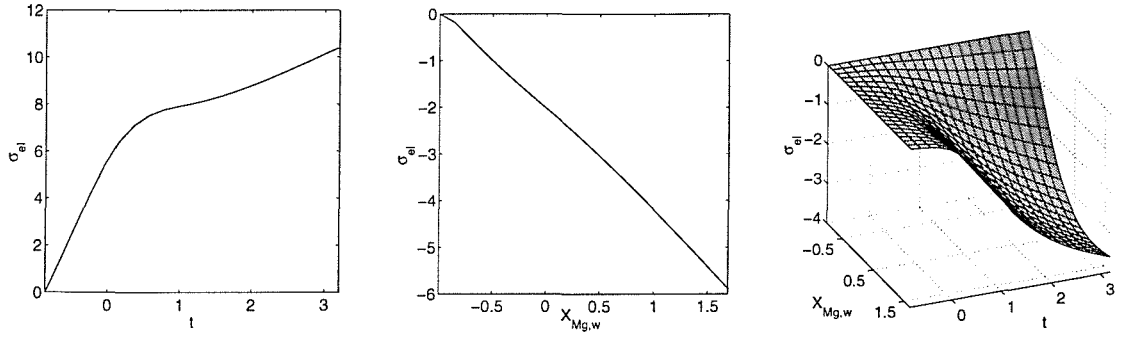


Figure 8.24: General form of the kernel approximations exhibited by the ANOVA terms most consistently identified in the basis selection stage of the SUPANOVA framework from data set C in modelling $\sigma_{0.2}$.

run	bias	Mg	t	Mg × Si	Mg × t
1	o	•	•	•	•
2†	o	•	•		•
3†	o	•	•		•
4†	o	•	•		•
5	o	•	•		
6	o	•	•		
7	o	•	•		
8†	o	•	•		•
9†	o	•	•		•
10†	o	•	•		•
11	•				
12†	o	•	•		•
13†	o	•	•		•
14†	o	•	•		•
15	o	•	•		
16†	o	•	•		•
17	o	•	•		
18	o	•	•		
19†	o	•	•		•
20†	o	•	•		•
Total	1	19	19	1	13

Table 8.27: Summary of the ANOVA terms determined for σ_{el} from data set A by the sparse basis selection employed within the SUPANOVA framework for each of the resampled data sets.

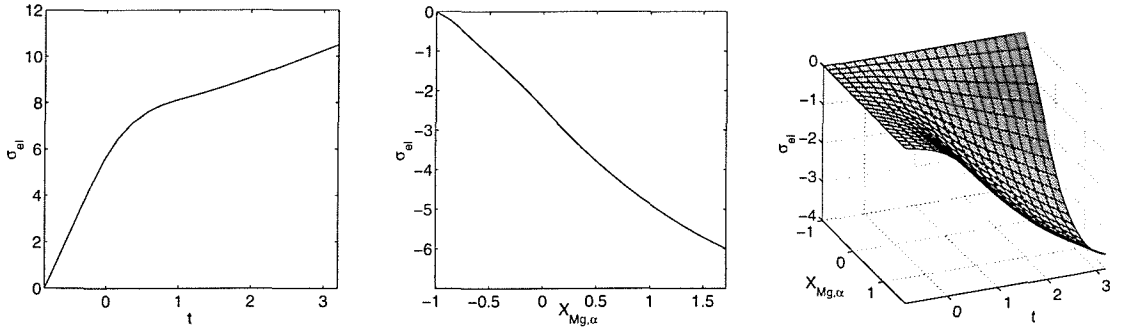


(a) (b) (c)

Figure 8.25: General form of the kernel approximations exhibited by the ANOVA terms most consistently identified in the basis selection stage of the SUPANOVA framework from data set A in modelling σ_{el}.

run	bias	$x_{Mg,\alpha}$	t	$x_{Mg,\alpha} \times t$
1	o	•	•	
2†	o	•	•	•
3†	o	•	•	•
4†	o	•	•	•
5	o	•	•	
6	o	•	•	
7	o	•	•	
8	o	•	•	
9†	o	•	•	•
10†	o	•	•	•
11	o	•	•	
12	o	•	•	
13†	o	•	•	•
14†	o	•	•	•
15	o	•	•	
16†	o	•	•	•
17	o	•	•	
18	o	•	•	
19†	o	•	•	•
20†	o	•	•	•
Total	-	20	20	10

Table 8.28: Summary of the ANOVA terms determined for σ_{el} from data set B by the sparse basis selection employed within the SUPANOVA framework for each of the 20 resampled data sets.



(a) (b) (c)

Figure 8.26: General form of the kernel approximations exhibited by the ANOVA terms most consistently identified in the basis selection stage of the SUPANOVA framework from data set B in modelling σ_{el}.

run	bias	Zn:Mg	Zn+Mg	t	(Zn + Cu) : Mg × t	Zn : Mg × (Zn + Cu) : Mg	(Zn + Cu) : Mg × Zn + Mg	Zn + Mg × t
1†	○	●	●	●	●			
2	○	●	●	●	●	●		
3†	○	●	●	●	●			
4†	○	●	●	●	●			
5	○	●			●			
6	○	●		●	●		●	●
7	○	●		●	●			
8	○	●		●	●			
9	○	●	●	●	●		●	
10†	○	●	●	●	●			
11	○	●		●	●			
12	○	●		●	●		●	
13	○	●		●	●		●	
14†	○	●	●	●	●			
15	○	●		●	●		●	●
16†	○	●	●	●	●			
17	○	●		●	●		●	
18	○	●	●	●	●		●	
19	○	●	●	●	●	●		
20	○	●	●	●	●		●	
Total	-	20	11	19	20	2	7	2

Table 8.29: Summary of the ANOVA terms determined for σ_{el} from data set C by the sparse basis selection employed within the SUPANOVA framework for each of the 20 resampled data sets.

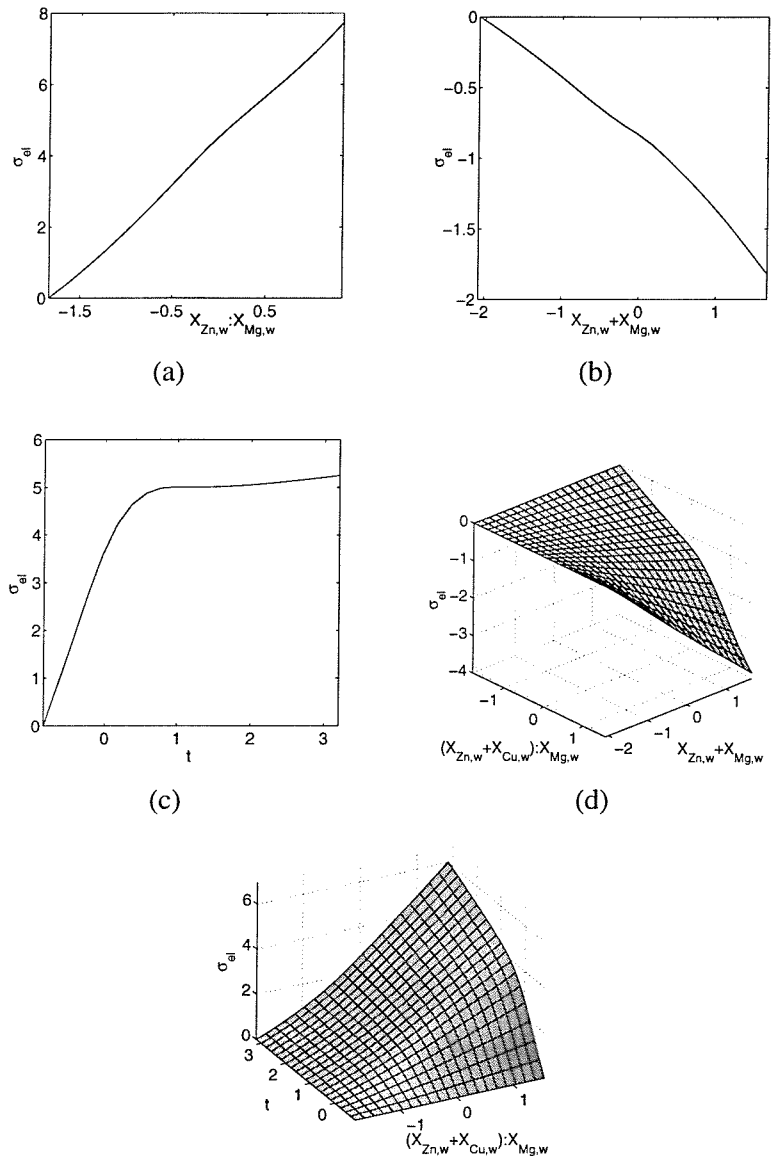


Figure 8.27: General form of the kernel approximations exhibited by the ANOVA terms most consistently identified in the basis selection stage of the SUPANOVA framework from data set C in modelling σ_{el}.

8.8 Discussion

For both $\sigma_{0.2}$ and σ_{el} (irrespective of data pre-processing) the neurofuzzy models consistently yielded improved approximation abilities compared to a MLR analysis. These improvements, particularly evident in the σ_{el} results, are attributable to a better functional representation than that attained by the simple MLR analysis.

In general, similar trends are extracted by the parsimonious techniques employed. Whilst the ANOVA spline kernels provided more flexible approximations and generally attained better approximation capabilities compared with the neurofuzzy models, a number of terms and non-linear approximations determined by the kernel approximations are difficult to validate (e.g. inclusion of Zr and the trends characterising Cu and t terms). On the other hand, the approximations inferred by the neurofuzzy construction algorithms attained rather simplistic approximations, generally limited to linear (in some cases piecewise linear) dependencies. The simple models determined in the neurofuzzy framework are not surprising, given the small size of the data sets and the model complexity control endorsed by the implementation of the SRM principle. This does however successfully constrain the inclusion of a large number of degrees of freedom and hence prevents undue flexibility which would have resulted in ill-conditioned basis functions and overfitting.

Regarding the model construction procedures used (FS/BE and SW), it remains problematic to infer which is the better strategy, since one approach may outperform the other, depending on the problem or training sample. More significantly, both approaches remain susceptible to local minima entrapment due to the iterative search procedure. In general though, the FS/BE procedure is preferable due to its coherent and structured search procedure. Stepwise searches, although performing a less structured and computationally less efficient search procedure, have the advantage of removing redundancy introduced early on in the model construction. The training data set sizes, particularly in the SVM models, where a validation set is necessary to determine the degree of regularisation, mean that it is unsurprising that the SUPANOVA results showed a higher sensitivity to the particular training-test splits, compared with the ASMOD construction algorithm. The number of variables present in the different data sets is seen to affect the stability of the ANOVA representations inferred, and can be understood in terms of the size of the full ANOVA basis considered. The variability in the regularisation coefficient inferred from cross-validation on the different training-test splits was reflected by a number of models significantly overfitting the training data. As the smoothing parameter C is determined by cross-validation and employing a gradient descent algorithm which may settle in a local minima, the variability in the regularisation parameter C is not surprising.

Overall, the results showed that there was a greater stability in the ANOVA representations inferred for σ_{el} compared with $\sigma_{0.2}$, particularly in training from data sets A and B. The instability in the subset selection of terms from data sets B and C can also be understood in light of the input dependencies present.

From sections 8.6.2 and 8.7 the amount of strengthening potential as characterised by the Zn + Mg term is always present (mainly as a univariate term) whereas a bivariate term of the form $(Xx : \mathrm{Mg}) \times t$ is better suited for describing the dependency of the balance of strengthening upon ageing time.

The approximations attained characterising the dependency of σ_{el} with t are comparable to those determined by Hepples (1987), while increasing the levels of alloying elements which may be retained in solid solution causes lower σ_{el} levels to be attained. In section 8.6 it was seen that $\sigma_{0.2}$ was characterised by a bivariate term (combining the contribution of t and that of an Mg-related variable), whilst for σ_{el}, a piecewise linear approximation for the t dependency was established. These results suggest that the former can be seen to represent the complex interaction of dissolved Mg, with vacancies bound to the Mg atoms slowing ageing in the present overaged alloys. Although the Mg $\times$ t term was less frequently selected in the results of sections 8.6.2 and 8.7, compared to the neurofuzzy models determined using all the available data, the dependency of ageing time and variates characterising the precipitation process supports the Mg $\times$ ageing time interaction.

Overall, the trends inferred from data sets A and C are consistent with the analysis performed by Pitcher (1998). The strengthening contributions of Mg and Zn being comparable to those previously reported, while the σ_{el} contribution attributed to Mg does not differ significantly from that concluded by our analysis of σ_{el}.

Comparing the performance of the different modelling approaches in terms of the data transformations and the modelling techniques used, a mixed picture emerges, and the results of modelling $\sigma_{0.2}$ and σ_{el} have to be discussed separately.

For σ_{el}, Tables 8.15 and 8.23 showed that in a comparison with the original data set, the use of transformed input data set B is clearly beneficial for both adaptive modelling techniques, suggesting that this data set enabled a greater and more explicit characterisation of the microstructural features influencing conductivity (the SVM methods however, attained the best performances from data set C).

The results obtained for $\sigma_{0.2}$ and summarised in Tables 8.14 and 8.22 show the results obtained from data set C as exhibiting better approximation abilities, suggesting that quantities derived from sums and ratios of the main variables involved in the precipitation dynamics in the alloy system considered comprise a descriptive set of features for the strengthening process.

Notwithstanding the better performance of data set C, there are still reasons for preferring models inferred from data set B. Firstly, from this data set a more parsimonious model was determined, enabling a better insight into the relevant physical quantites. Secondly, models determined from data set B are more likely to attain a better generalisation performance derived from the inclusion of *a priori* physical understanding, with models therefore being more reliable in extrapolating beyond the current compositional range[16]. This can be illustrated

[16]Here extrapolation is considered with respect to the original at.% (or wt.%) composition levels, i.e. no extrapolation beyond the limits defined by the physical based transformation will take place.

from an analysis of the dependency of yield strength upon Mg content, which would show that models obtained from data set C will, unrealistically, predict a single stage, monotonically increasing yield strength with increasing Mg content, whilst models obtained from data set B (in light of the transformations of section 8.4.1) will predict a (more realistic) levelling off of the strength increase obtainable through increasing the Mg content.

Although it can be argued that with sufficient data, particularly for alloys with $x_{Mg} >> 2$, an empirical modelling approach should be capable of describing this strengthening behaviour[17], the present data was insufficient to enable the reliable description of this physical behaviour, and so the use of prior knowledge can be seen to partly compensate for the sample size limitations. It is further noted that the empirical results obtained from data set D in part elucidate the source of the relatively weaker performance attained from data set B compared with data set C in modelling yield strength. This being mainly due to the subnetwork including a Mg related variable and t, i.e. related to the way in which Mg influences the kinetics of the reaction and so does not directly indicate that the simplified ideas concerning the strengthening mechanisms (section 8.4.1) are inaccurate. Nethertheless, the quantities present in data set C do offer some valuable description of the alloys in terms of quantities characterising the balance and amount of strengthening potential for the precipitation present in the alloy system. However, inferences based on models determined from these quantities remain more problematic, due to the dependencies present among the variates. The problem was seen to be reflected in the least-squares solutions of the parameters in the neurofuzzy models where the LOOCV estimates inferred from this data set were seen to be considerably inflated (particularly for $\sigma_{0.2}$) in a comparison to the training $\hat{\sigma}_N^2$.

The increasing dependency exhibited by σ_{el} with $x_{Zn,w} : x_{Mg,w}$ is not consistent with (Mondolfo 1971), where a higher Zn:Mg ratio is generally associated with worse SCC resistance.

8.9 Conclusions

In considering the overall implications of the modelling approaches attempted, the combination of data-driven approaches and transformation of the input data has elucidated the dependencies present in the 7xxx Al-alloy data set.

The use of adaptive data-driven modelling approaches in the description of underlying structures inherent in the data has been demonstrated as a useful approach in modelling where prior data set transformations are based on well-founded physically based relationships. In situations where more limited physical understanding exists, both neurofuzzy and SUPANO-

[17]It should be noted that alloys with $x_{Mg} >> 2$ are unsuited for commercial use due to the high volume fraction of undissolvable intermetallics (e.g. S phase) which cause unacceptably low toughness levels. Thus, commercial Al producers will in general not invest in producing high Mg containing alloys, thereby, preventing the generation of data sets with high Mg levels.

VA frameworks can offer a combination of pure empirical modelling and physically based modelling possibilities.

The transformations developed to derive data set B are seen to have the potential of facilitating the control of the Zn, Cu and Mg levels so that both deleterious intermetallics (e.g. S phase) and excess alloying elements can be eliminated (i.e. for a given amount of Zn, Mg and Cu, estimates of the expected excess of Mg and Cu can be determined). Thus, one or more of the alloying elements may be adjusted to maintain an alloy free of excess Mg or Cu[18].

From the inspection of the FS/BE model construction obtained in the initial results, general initialised models were defined for each property with a similar structure for each data set. This formed a basis for comparison of the effects on modelling performance for each data set transformation.

The present analysis can be used to draw out some of the microstructure-property issues for 7xxx alloys. Firstly, it is noted that compared to the original data set (A) the data set transformations, using some relatively simple information on the microstructure (B) yielded a considerable improvement in modelling performances attained for σ_{el} but no significant improvement in the modelling of $\sigma_{0.2}$. In retrospect this difference is not surprising as strength is the more complex property, more dependent on additional microstructural features that are more problematic to identify (e.g. grain size, precipitate size distribution, PFZs). In addition, tensile test results exhibit more noise, with higher experimental measurement error levels characterising the outputs. Overall, the analysis confirms the main expected structure-property relationships, e.g. the significance of the maximum amount of η', $x_{\eta'}$, determining strength. Of particular interest was the identification of an interaction present between ageing time and an Mg related variable, characterising the yield strength of the Al-Zn-Mg-Cu alloys, which enabled improved empirical modelling performances and possibly leading to improved understanding of the ageing dynamics in these alloys.

[18]Generally, an excess of Cu is preferable to an excess in Mg concentrations (Anderson 1994).

Chapter 9

Knowledge Discovery and Data Mining of 7xxx Series Al-Alloy Production Databases

9.1 Introduction

The analyses in Chapters 7 and 8 were seen to have been performed on relatively small data sets, both in terms of the number of input variables and training patterns available. Both experimental designs (Chapter 8), and the relatively less complex mechanical alloying process enabled relatively simple analyses to be carried out compared to those which are typically required in an analysis of large scale industial processes.

In Chapter 2 it was seen how the properties of heat-treatable wrought Al-alloys are obtained from a multi-stage fabrication process, which includes several thermal treatments and deformation processes. As a result, the fabrication of commercial wrought alloys may be determined by a large number of parameters and processing conditions, some of which may significantly influence the balance of physical, mechanical and structural properties attained. While several processing procedures may follow tight guidelines, others may exhibit a greater variability, determined by *ad hoc* procedures and practical processing requirements and constraints.

Together with this process and plant information, both quality control and physical/mechanical test results necessary to fulfil customer specifications such as minimum strength, toughness and electrical conductivity levels will also be stored in plant databases.

A simple statistical investigation performed using standard regression analysis tools will not be straightforward as certain fields in such databases may contain non-numeric information. Not only will the dimensionality and complexity of the system make simple statistical analysis problematic, but the large amount of data will necessitate some form of pre-processing prior to any such analyses so that sensible data sets which are amenable to a statistical investigation

can be defined. The analysis of this data can then proceed using appropriate exploratory data analysis techniques. Due to the dimensionality of the databases, knowledge discovery and data mining techniques will be required, in order to determine whether any useful knowledge or novel, unexpected patterns may be identified.

However, whether predictive process models can be reliably inferred from such "happenstance" data will be determined by the particular process and by the representativeness of the sampled data of the underlying physical processes, and as such, will be domain dependent. Although a parsimonious description of the data may be inferred, the development of a truly predictive model will be critically dependent on the data quality and the presence of any data deficiencies[1], as well as the properties and assumptions made by the modelling technique used.

This chapter presents the results obtained from a statistical analysis performed on 7xxx series alloy process data (specifically in terms of toughness performance), placing the overall data analysis and modelling performed within a KDD and data mining context. Compared to previous chapters, a considerable amount of the knowledge discovery process was in the data pre-processing and defining sensible data sets that were amenable to "mining". The adaptive approaches used in analyses performed in previous chapters have been employed in the modelling of process-property relationships from *raw* production data.

The processing stages and large number of fields present in the databases required a sound knowledge of both industrial practices and processing conditions, as well as a fundamental understanding of the physical metallurgy of high-strength Al-alloys. The nature of the process variables and the sampled ranges could then be understood prior to the analysis, subsequently using metallurgical understanding to validate the relationships determined and to understand whether any deficiencies in the modelling results could be explained in terms of data weaknesses.

The statistical investigation required a certain degree of interaction with the data set suppliers and proceeded with analyses being performed on a number of data levels, comprising different partitioning of the data and corresponding to particular sub-systems, all extracted from the original data. Simple visualisation of the data distributions and dependencies between processing conditions provided valuable insight into the statistical properties of the data, which were not possible to appreciate from inspection of simple summary statistics. Initially, the data mining capabilities of the neurofuzzy framework were assessed on the largest of these data sets, for which a large number of input variables were retained. Subsequently, the validitity of the data mining results were investigated by refining the analyses on the smaller data set partitions, on which both the neurofuzzy and SVM techniques were assessed. As in previous chapters results were also compared with simple linear regression analyses.

As well as describing general trends present in the data, in many processes it is of interest to detect the presence of interesting/novel patterns, such as peculiar (local) behaviours and

[1] Good generalisation will only be obtained if a sufficiently descriptive set of training samples and input features are *both* available.

the identification of a small set of data samples which do not comply with the rest of the data and the process model inferred (e.g. outliers). Hence, the influence measures defined in Chapter 5 formed an integral part of the analysis of the neurofuzzy results. Finally, Bayesian regularisation approaches were investigated on a subset of the data as an alternative approach to controlling excessive degrees of freedom present in neurofuzzy models.

9.2 The British Aluminium Plate data

As discussed in previous chapters, 7xxx series alloys exhibit the greatest response to age-hardening, their good combination of low density and high strength making them attractive materials for the aerospace industry. The processing and properties of these alloys have been discussed at length in Chapter 2, whilst section 8.2.1 provided a detailed treatment of the precipitation sequences present in Al-Zn-Mg-Cu alloys.

Production data for 7x75 type Al-alloys was extracted from several BAP process databases, each of which contained process data pertaining to particular stages in the production (e.g. casting, plate formation, etc.) and merged using customised data retrival routines. This data comprised a large amount of information on the production of wrought plate spanning a period of approximately 16 months[2], the amount of data determined by the number of plates produced.

Table 9.1 shows international standards for the compositional ranges of the major alloying elements, maximum levels of trace additions and maximum impurity levels for the different 7x75 Al-alloys considered[3]. From these compositional ranges it is seen that the 7175 and 7475

Al-alloy	Zn	Mg	Cu	Mn	Cr	Zr	Si	Fe	Ti	Ni
7075	5.1-6.1	2.1-2.9	1.2-2.0	0.30	0.18-0.28	0.20	0.40	0.50	0.20	-
7175	5.1-6.1	2.1-2.9	1.2-2.0	0.10	0.18-0.28	0.10	0.15	0.20	0.10	-
7475	5.2-6.2	1.9-2.6	1.2-1.9	0.06	0.18-0.25	0.06	0.10	0.12	0.06	-

Table 9.1: Composition ranges (wt.%) for the 7075, 7175 and 7475 Al-alloys.

alloys derive from the same base alloy (7075) but have modifications to certain compositional ranges and impurity levels. Compared to the 7075 alloy, 7175 has the same level of Cu, Mg and Zn, while Fe, Si, Mn and Ti are controlled to lower levels. Alloy 7475, the first Al-alloy developed to provide improved fracture toughness, is a leaner variant of the 7075 alloy, with

[2]Although significantly more off-line data is stored within other historical (archived) databases, it was considered that data amenable to a reliable statistical analysis was confined to the on-line databases, since old process specifications limited the integrity of any larger data sets.

[3]It should be recognised during subsequent analysis that BAP use significantly different (tighter) ranges of the major alloying elements (which is proprietary knowledge) as well as controlling trace elements and impurities to tighter limits than those specified in Table 9.1. Also, as will emerge below, BAP use a number of different alloy designations for alloys in the 7x75 compositional ranges.

lower limits for Fe and Si, lower maximum Cr content and modifications in the Cu, Mg and Zn concentrations. In all 7x75 alloys Cr is used as the main grain refining addition.

As discussed in section 8.2 the $Al_{12}MgCr$ dispersoid aids in retaining an unrecrystallised grain structure during processing of wrought products. Incoherent dispersoids such as $Al_{12}MgCr$ (and also $Al_{20}Cu_2Mn_3$) are responsible for an increased quench sensitivity of Cu-rich Al-Zn-Mg alloys as they act as nucleating agents for solute-rich precipitates during the quench.

The original data sizes were 2020×48 and 4475×48 for the 7175 and 7475 Al-alloys respectively[4]. Limitations in the database retrival system required fracture toughness data to be extracted separately from the reminder of the tensile data, which included percentage elongation, yield and tensile strengths. Thus, two separate data sets contained the mechanical properties of interest to alloy producers.

Table 9.2 exemplifies the different fields extracted from the BAP databases. The information retrieved is seen to comprise a description of the alloy fabrication in terms of processing conditions, which include the chemical composition of the melt, the as-cast ingot dimensions, heat-treatment type and conditions, initial and final temperatures of the hot rolling mills, final plate dimensions, post solutionising stretch, together with physical and mechanical properties and corresponding specimen test conditions. In addition, a number of other fields include information regarding the alloy type, various fabrication codes, fabrication dates, etc.

Toughness data were seen to have been determined for various combinations of test and specimen conditions: at the quarter (Q) and mid-thickness (C) positions in the plate, the 1/3 (.3W) and mid-width (.5W) positions and in three different test orientations (LT, TL and SL). Plate geometry and these test orientations are illustated below in Figure 9.1. A comparison of the records in terms of PLATE-NO identifies plates tested in more than one orientation[5]. As well as the international standard alloy designation, each plate alloy type was seen to have an additional BAP designation. In particular, records pertaining to a certain 7x75 alloy type are seen to have an ALLOY KG designation. While processing conditions with some exceptions, e.g. cast slab dimensions (see section 9.2.1), can be regarded as continuous variables, sample test information and alloy type comprise a set of categorical variables or labels.

As the definition of reasonable data sets on which to base statistical analysis requires both metallurgical and process understanding to ensure that valuable information has not been inadvertently discarded, processing conditions will be briefly discussed in terms of the set of explanatory variables available in the present analysis.

As discussed at length in Chapter 2, while generally the toughness of Al-alloys decreases as the strength level is increased by heat treatment, fracture toughness behaviour is complicated by the presence of a number of microstructural features. Although heat treatment conditions are well understood to influence the toughness behaviour, their direct effect is difficult to assess due

[4]Rows×columns. i.e. entries × fields.

[5]The number and test conditions will be according to the particular customer specifications.

Field	Description
PLATE-NO	Plate identifier (7 digit identifier)
PL-GAUGE	Slab gauge (mm)
PL-LENGTH	Slab length (mm)
PL-WIDTH	Slab width (mm)
PL-GAUGE-OFF	As rolled plate gauge (mm)
PL-LENGTH-OFF	As rolled plate length (mm)
PL-WIDTH-OFF	As rolled plate width (mm)
ALLOY	Alloy designation [7075, 7175, 7475]
CUSTCOND	Customer condition (temper designation) [T7351]
DIMNS(1)	Final plate gauge (mm)
DIMNS(2)	Final plate length(mm)
DIMNS(3)	Final plate width (mm)
CAST	Cast number identifier
LOT	BAP plate code [week + product code + plate number]
CONDY(1)	Minimum conductivity (σ_{el})
CONDY(2)	Maximum conductivity (σ_{el})
STRETCH	Post-solutionising stretch (%)
ENTRYDATE	Processing date [day, month, year]
DIRECTION*	Tensile test orientation [LT, TL, SL]
THICKPOSN	Thickness position of specimen [Q, C]
WIDTHPOSN	Width position of specimen [.3W, .5W]
02PS*	0.2% proof stress (MPa)
UTS*	ultimate tensile strength (MPa)
ELONG*	percentage elongation (%)
K_{Ic}	fracture toughness (MPa $\sqrt{m}$)
Cu	Concentration of Cu (wt.%)
Fe	Concentration of Fe (wt.%)
Mg	Concentration of Mg (wt.%)
Mn	Concentration of Mn (wt.%)
Si	Concentration of Si (wt.%)
Zn	Concentration of Zn (wt.%)
Ti	Concentration of Ti (wt.%)
Cr	Concentration of Cr (wt.%)
Ni	Concentration of Ni (wt.%)
Zr	Concentration of Zr (wt.%)
B	Concentration of B (wt.%)
SLGAUGE	Slab gauge (dm)
HRGAUGE	Hot rolled gauge (dm)
HEATTYPE	Heat treatment: homogenisation, solution or precipitation [H, ST, PT]
HEATTIME	Heat treatment time (mins.)
HTTEMPMIN(1,2)	Minimum heat treatment temperatures (°C)
HTTEMPMAX(1,2)	Maximum heat treatment temperatures (°C)
HEATCODE	BAP heat-treatment designation (alpha numeric code)
STEMP	Hot mill rolling start temperature (°C)
FTEMP	Hot mill rolling final temperature (°C)
ALLOY KG	BAP alloy type designation (two digit alpha code)

Table 9.2: List of the processing conditions, tensile properties and test information extracted from the 7x75 BAP production databases.

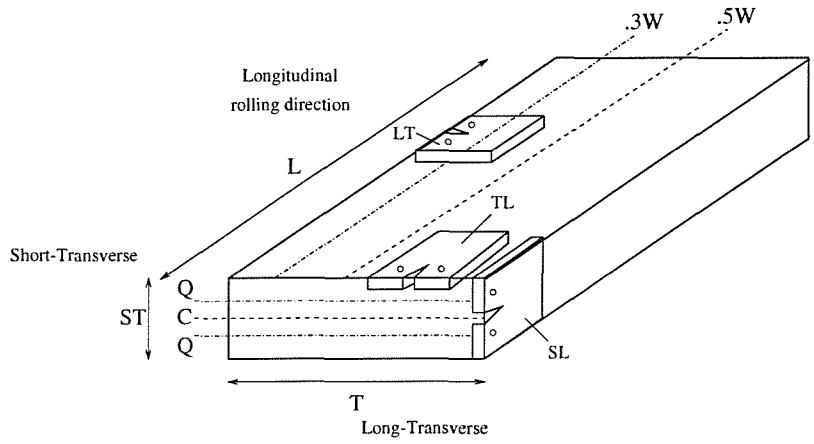


Figure 9.1: Schematic diagram of the plate geometry and fracture toughness test conditions.

to the interdependencies existing with the thermomechanical processing of these alloys. For example, the formation of coarse intermetallics (which may be subsequently fractured during primary processing or under service loading) provide crack nucleation sites and paths for low-energy crack propagation. Fracture toughness can therefore be improved by controlling the Fe and Si levels. The effect of reducing the presence of any soluble S phase through controlled homogenisation is likely to promote higher attainable toughness levels. Typically, lower quench rates will result in an increase in the proportion of intergranular fracture, decreasing the fracture toughness of the alloy. However, for quench sensitive plates, thermal gradients present within the material and macroscopic segregation from the cast may contribute to a through-thickness "W-shape" in properties, with the worst distribution of coarse secondary particles occurring at the quarter thickness position, as noted in section 2.3.2.

In industrial settings, quench procedures will be determined by the alloy system and the quench analysis performed. Although it is desirable to achieve the fastest possible quench rate, other factors, principally the control of residual stresses and flatness will limit the use of rapid quench rates. In thick section materials, where cooling rates will be limited by physical constraints, the scope for increasing toughness by optimisation of the quench rate is limited (the maximum attainable quench rates will be lower for thicker section sizes).

In order to compensate for the property degradation with increasing thickness, BAP use purer alloy variants (identifiable by the ALLOY KG designation) in the fabrication of larger gauge sizes. In thin sections on the other hand compositional ranges closer to those specified for 7075 can be used.

Toughness levels will typically be improved upon breaking down the as-cast structure of the alloy, achieved primarily through hot deformation processes. The final temperature (FTEMP, T_f hereafter) of the plates exiting the hot rolling mill can be expected to influence the

deformation substructure of the material and consequently the grain structure developed, as a higher temperature might be expected to limit the driving force and degree of recrystallisation during a subsequent solution treatment (recrystallisation entailing a loss in toughness levels). Typically, the degree of recrystallisation during the solutionising treatment will be greater in thin section sizes, resulting from the greater deformation occurring during processing.

Differences in the fracture toughness levels attained at the two different width positions may be associated with the casting conditions and understood in terms of microstructural profiles imparted by the different cooling rates and macrosegregation present within the material in the original cast.

9.2.1 Data representation and coding

From Table 9.2 it is seen that different slab and plate dimensions are measured at several processing steps (i.e. as-cast and scalped, the latter obtained prior to the hot deformation stage)[6]. While slab thickness (PL-GAUGE) and width (PL-WIDTH) correspond to the DC-cast dimensions[7], the final dimensions of the plate will be determined by the customer specifications.

For the purposes of defining a set of deformation variables that provide a physically meaningful representation of the TMP stage, the following strains associated with the hot rolling were defined and appended as extra fields to the original data. The strain in the LT direction (LT_ϵ) was defined as:

$$LT_\epsilon = \frac{\text{PL WIDTH OFF} - \text{PL WIDTH}}{\text{PL WIDTH}}$$

and that in the rolling direction (L_ϵ) as:

$$L_\epsilon = \frac{\text{PL LENGTH OFF} - \text{PL LENGTH}}{\text{PL LENGTH}}$$

from which the cross rolling strain (CRS) can be defined by:

$$CRS = \frac{LT_\epsilon}{L_\epsilon} \tag{9.1}$$

In addition, the overall gauge reduction (GR) was simply defined as:

$$GR = \frac{\text{PL GAUGE}}{\text{PL GAUGE OFF}} \tag{9.2}$$

The effect of increased deformation levels (a greater gauge reduction) on the fracture toughness levels developed can be interpreted together with the deformation temperature, in terms of texture developments and possible recrystallisation effects. It may be expected that the degree of cross rolling (in that it will promote a more isotropic in-plane distribution of microstructural features such as coarse intermetallics and grain structure) will have a

[6]This redundancy was unavoidable resulting from integration of the data extracted from the different databases.

[7]PL-GAUGE and PL-WIDTH dimensions revealing that only two cast moulds were used.

significant influence on toughness levels attained in the TL orientation.

The attributes in the fields determining the test conditions and alloy type were replaced with the numerical values shown in Table 9.3, encoding a natural ordering of these attributes where appropriate. In terms of the ALLOY KG labels shown in Table 9.3, the three designations represent changes in impurity levels (specifically Fe and Si) and primary alloying element levels (particularly Mg content). The sequence, LT→LE→LK represents increasing impurity levels, with a parallel reduction in major alloying elements. For any specific IADS alloy type (i.e. 7175 versus 7475), higher purity/leaner variants are used for thicker plates. As

	Fields and attributes									
	DIRECTION			THICKPOSN		WIDTHPOSN		ALLOY KG		
labels	SL	TL	LT	Q	C	.3W	.5W	LT	LE	LK
coding	−1	0	1	0	1	0.3	0.5	0	1	2

Table 9.3: Summary of the coding used to represent test information and alloy type fields as numerical entries.

will emerge from subsequent analyses, the majority of the 7175 data corresponded to the LT alloy type, whilst the 7475 data was almost entirely comprised by LK type alloys.

9.3 Data pre-processing

In the KDD process, and particularly in data mining, the quality and integrity of the data are seen to play a central role in the extraction of meaningful knowledge. In order to obtain a data matrix $\mathbf{D} = [\mathbf{X} : \mathbf{Y}]$ comprising the set of explanatory variates and output vector in a form from which sensible inferences and knowledge can be determined, the as-obtained data format required a considerable amount of pre-processing. For such purposes, a set of routines were written[8] with different pre-processing, cleaning and filtering operations allowing the desired data format to be obtained.

9.3.1 Data sifting

In addition to data selection (section 9.3.2), different forms of sifting were required in order to detect and correct a number of inconsistencies in the various data fields. Only after an initial understanding and inspection of the different fields were the different data problems revealed (e.g. entries with a gauge reduction value greater than one were clearly the result of an error in recording the plate thicknesses). The data sifting operations performed (detailed further below) can be summarised as:

[8]These procedures were written as a set of MATLAB scripts.

- data flattening,
- redundancy removal (e.g. averaging output values corresponding to records having identical input patterns),
- removal of unusual/anomalous batches,
- removal of zero entries,
- removal of missing values,
- removal of extreme outliers,
- removal of "physically unrealistic" records,
- de-duplication of records.

The BAP data covered plates that had been aged to the T7351 (overaged) condition by means of employing two different commercial precipitation heat treatments: a single or a double ageing step sequence (BAP 2000). These two ageing procedures were identifiable by the information contained in the HEATCODE, HEATTYPE and HEATTIME fields.

An analysis conducted by considering both these ageing procedures as a single "system" was considered inappropriate and hence the analysis was confined to the single ageing step treatment, which corresponded to the large majority of the data. Thus, in addition to removing the double step aged alloys, other non systematic heat treatments, such as re-solutionised treatments, extended ageing treatments (performed by re-heating alloys which attained inadequate conductivity levels) were also removed. As no K_{Ic} tests were performed below plate thicknesses of approximately 20mm, this yielded zero entries in the K_{Ic} and test condition fields, and as such these records were also removed for such thin materials.

The data set sizes after sifting the original BAP data and retaining only single step precipitation hardened alloys were 942×55 and 1958×55 for the 7475 and 7175 data sets, respectively. The expansion in the number of fields has arisen from the need to flatten the data (see section 4.2.3) by redefining the HEATTYPE and heat treatment variables (HEATTIME, HEATTEMPMIN(1,2), HEATTEMPMAX(1,2)) in terms of distinct fields for the homogenising (HT), solutionising (ST) and age-hardening (PT) heat-treatment processing conditions.

9.3.2 Data selection

An analysis conducted on a data set which retained all fields would have further assessed the data mining capabilities of the neurofuzzy framework in terms of computational efficiency, scaleability of the technique and assessing the effect of retaining a very large number of uninformative and redundant fields in terms of oversearching and overfitting phenomena. However, where process knowledge can be used to pre-screen the fields, identifying a subset of the inputs

considered to influence the physical behaviour of the system is more sensible compared to an analysis centred on simply assessing the data modelling technique and drawing inferences in the presence of a large number of irrelevant features. Thus, in this study input variables were pre-screened to discard what were considered the more uninformative and redundant variates.

As different processing conditions may characterise the two 7x75 alloys considered (although cast under the same conditions, different pre-heat treatments conditions are used) (BAP, 2000), the 7175 and 7475 data were investigated separately[9]. In the present analysis, tensile properties such as $\sigma_{0.2}$, uts and $\%el$ were not considered as input variables, as they have been in other neural network studies, e.g. (Fujii et al. 1996). These studies used $\sigma_{0.2}$ and uts as explanatory variates which were seen to exhibit strong (linear) dependencies with the corresponding output property (e.g. fatigue thresholds ΔK_{th}). Similarly, conductivity which previous analysis had indicated to be a useful indicator of the microstructural condition, providing information on the condition of age-hardened microstructures, was also not considered. Although these properties are undoubtedly linked to fracture toughness, they are not linked to the process control stage (being final properties themselves) and so were inappropriate in this study, where the effects of process control variables on K_{Ic} are considered. The objective being developing process control, which is not possible with a model that includes other processing outputs.

Table 9.4 shows the processing variables and test information used. This comprised the final plate thickness, the concentration of major alloying elements, grain refining elements, impurity elements, deformation/TMP conditions and solution heat treatment time, along with test condition and alloy type indicators. All heat-treatment temperatures were seen to be tightly

Plate gauge	%stretch	(DIRECTION)	(THICKPOSN)	(WIDTHPOSN)	Cu
Fe	Mg	(Mn)	Si	Zn	Ti
Cr	T_f	(ALLOY KG)	ST_t	GR	CRS

Table 9.4: Summary of the input variables retained in the analysis of K_{Ic} data. Variables in (·) are only present in certain data sets.

controlled and for the purposes of subsequent analyses were considered uninformative, as were the homogenising (H) and age-hardening (PT) thermal treatment times. Also, in all records Boron levels (wt.%) had identical values, hence this compositional element was also removed from the data set.

Unsurprisingly, initial and final hot-rolling temperatures (T_s and T_f) were seen to be somewhat correlated. The final rolling temperature, T_f and not the start temperature, was retained for subsequent analysis as the former was considered to be a more representative variable of the final hot rolled structure of the material.

[9] As there was a very limited overlap between the 7175 and 7475 test conditions, little benefit would have resulted from integrating the two data sets.

For the 7175 data set Mn was retained as the compositional range and variance of this element (section 9.4) was unexpectedly large. In the 7475 data only two plates were identified as the less purer variant (LE) of the 7475 alloy, the majority being of the LT type. In the 7175 data on the other hand a considerable number of LE alloys were present, corresponding to the thicker 7175 plates.

In terms of the labels in Table 9.3, data sets covering all variables (test information and processing data) were designated BAP1 and BAP4, addressing the 7175 and 7475 plates respectively. Subsequently, the 7x75 data were partitioned in terms of the thickness (THICKPOSN), width (WIDTHPOSN) positions, alloy type (ALLOY KG) and test directions (DIRECTION) attributes. These data sets, constituting all the available data that were suitable for subsequent modelling are summarised in Table 9.5. The 7175 plates tested in the mid-thickness position were seen to correspond to one BAP alloy variant, whilst those in the quarter position comprised two different BAP variants of the 7175 alloy. In Table 9.5, data set BAP1TLQ3W80 is a subset of BAP1TLQ3W, with the gauge range limited to 80mm for which only one alloy variant is present[10]. From these partitions the largest data sets are seen to correspond to plates tested in the mid-thickness position and in the LT/TL orientations. The considerable reduction in data set sizes compared to the original dimensions of the data

Alloy	DIRECTION	THICKPOSN	WIDTHPOSN	Plate gauge range (mm)	Dataset size	Number of inputs	Dataset designation
7175	all	all	all	25.57-102.06	302	16	BAP1
7175	TL	C	.3W	25.57-62.53	193	14	BAP1TLC3W
7175	TL	Q	.3W	64.42-102.06	109	14	BAP1TLQ3W
7175	TL	Q	.3W	64.42-77.56	70	14	BAP1TLQ3W80
7475	all	all	all	19.17-112.52	805	18	BAP4
7475	LT	C	.5W	19.17-112.52	296	13	BAP4LTC5W
7475	SL	C	.5W	64.81-112.52	107	13	BAPSLC5W
7475	TL	C	.5W	19.17-112.52	294	13	BAP4TLC5W

Table 9.5: Summary of the various data set partitions considered.

extracted from the databases was primarily due to considerable redundancy inherent in the database retrival routines and the presence of missing entries in several fields[11].

A first assessment of the representativeness of the data of the overall wrought production

[10]Partitioning the quarter thickness 7175 data in terms of either *Plate gauge* (<80mm, >80mm) or ALLOY KG (LK, LE) resulted in the same subset.

[11]In the majority of cases, missing values corresponded to compositional levels and time and temperatures of thermal and TMP processes. These missing values corresponded in the great majority of cases to the older records. As noted previously, missing output (and consequently test information) corresponded to alloys with section thicknesses below 19mm.

process indicates that certain processing stages are described in more detail than others: whilst heat treatments and TMP are described through a number of variables, information pertaining to the casting/solidification (e.g. rate of solidification) and quench variables (e.g. pump pressure, differential pressure, etc.) was lacking. Hence, any effect of these stages can only be conveyed indirectly through other inputs such as cast ingot dimensions and plate thickness.

The data sets summarised in Table 9.5 comprise the data sets on which the statistical analysis of subsequent sections will be performed. While in some cases, it may be desirable to obtain an overall representation of the system, in others, it may be more appropriate to perform separate inferences, for instance each one corresponding to a particular regime or system behaviour, particularly when the process is likely to be influenced by different factors with distinct effects.

9.3.3 Data normalisation

As in previous chapters the commercial sensitivity of the data means the processing conditions are presented as normalised values, while measured and predicted K_{Ic} values are presented in their original range of values. The MLR analysis was conducted on data normalised to have a zero mean and unit variance, while as in previous analyses, the adaptive techniques were performed on input data transformed to lie in the interval [+1, −1] in the neurofuzzy framework and in the [0,1] range for the SVM methods.

9.4 Data analysis

Although the data sets were of a high dimensionality, an inspection of the input and output distributions was nevertheless useful, revealing the presence of any strong dependencies and allowing greater understanding of the variance in both input and output quantities, and allowing better understanding of dependencies inferred in subsequent regression analyses. Inspection of the data distributions helped to determine whether any further pre-processing (e.g. further partitioning of the data set, outlier removal, etc.) was required.

Histograms for the distribution of K_{Ic} for the data set partitions considered are shown in Figure C.1, while Table 9.6 summarises the means and variances.

A distinct bimodal distribution in K_{Ic} is evident in Figure C.1 (b), which was attributable to the two alloy variants associated with this particular data set (two alloy variants arise in the 7475 data sets represented by Figure C.1 (e) and (f), however the vast majority of the results in these cases where for one alloy variant, hence a unimodal distribution is dominant). The distribution of K_{Ic} values for the various 7475 data sets showed that on average, higher toughness levels are associated with the LT direction, and lowest in the SL, with tests performed in the TL direction exhibiting intermediate toughness levels. As expected, the comparison of toughness levels for the different alloys showed that the purer 7475 plates attain

Data set	μ	$\hat{\sigma}^2$
BAP1	30.12	6.83
BAP1TLC3W	29.14	1.63
BAP1TLQ3W	31.87	11.27
BAP1TLQ3W80	29.86	3.46
BAP4	45.76	40.38
BAP4LTC5W	51.80	16.96
BAP4SLC5W	40.46	18.59
BAP4TLC5W	42.94	10.55

Table 9.6: Mean and variance of K_{Ic} for the different 7x75 data sets.

considerably higher toughnesses compared with the 7175 alloy. While a considerably lower variance in K_{Ic} levels was apparent for the 7175 alloys tested at the mid-thickness position compared to tests performed in the quarter position, the variance exhibited in the 7475 data sets was seen to have a high variance in all the different test conditions. The large variance in the data sets with larger plate thickness ranges (i.e. the 7475 data sets) does suggest toughness levels are dependent (either directly or indirectly) on the section thickness.

The full residuals for the complete 7175 and 7475 data sets (BAP1 and BAP4) for K_{Ic} levels and a selected number of the input variables are shown in Figures C.2 to C.10[12]. For the 7175 data (Figure C.2), thicker plates are seen to exhibit considerably higher toughness levels (linked to the higher purity alloy variant used in thicker material), whilst inspection of the residuals for both Fe and Si show clear dependencies with K_{Ic}, with purer alloys exhibiting higher toughness levels.

The influence of test specification on the data sets may be seen in the Fe plots in Figures C.4 and C.5, where the fact that mid-thickness tests (Figure C.4), are only performed for thinner plates, means that there is a reasonably unimodal Fe distribution, as only one alloy variant was used for the gauges in question. The higher gauge range represented by the data in Figure C.5 however spans two composition variants, with clearly bimodal composition characteristics, and corresponding toughness levels. It may be valuable to note that thicker plates/alloy variants received less cross rolling (see below).

The residuals shown in Figure C.3 corresponding to the BAP4 data show little structure, with only the residuals for Fe and Si showing slight dependencies with K_{Ic}. Overall, the inspection of the BAP1 and BAP4 residuals revealed only a limited number of first order dependencies between the inputs and K_{Ic}.

[12]To avoid overburdening the reader and presenting uninformative plots, the residual plots shown are limited to those for which dependencies are evident and those for which subsequent analyses considered as having some explanatory capabilities, e.g. elucidating the extent of extrapolation in poorly covered regions of the input space.

The presence of the two alloy variants in the BAP1TLQ3W data set may confound the interpretation of thickness effects with other variables (e.g. impurity levels) in terms of whether quench sensitivity or other effects were both present.

In the 7475 data, a comparison between Figures C.3 and C.7 reveals how the mid-thickness tests conducted in the LT orientation correspond to the plates exhibiting the higher toughness levels. For tests conducted in the SL direction, Figure C.9 suggests that increasing temperature at which the hot-rolling is performed imparts a positive effect on toughness.

In summary, the inspection of the residual plots revealed that whilst some general trends are consistent over different test conditions and alloy types, there appear to be a few differences in the dependencies characterising toughness levels exhibited in certain test conditions.

In order to infer whether there were any significant dependencies amongst the processing variables, pairwise scatterplots between the input variables were inspected and a number are shown in Figures C.11 to C.18. The bimodality in the 7175 data (due to the two alloy types) is evident from a number of the pairwise scatterplots. Unsurprisingly, final gauge thickness and the gauge reduction were seen to be correlated, exhibiting two different correlations corresponding to the two cast ingot gauges used. The dependency between the section thickness and gauge reduction, likely to result in problematic inferences in a MLR analysis, suggested a further possible partitioning of the data. However, this would have further limited the sample sizes. The inferences attained would reveal whether this dependency was problematic.

There appeared to be dependencies amongst the major alloying elements (Mg, Cu and Zn) and the impurities, with the degree of correlation between Fe and Si being greater in thicker plates. The thicker plates also exhibited greater control in Mn levels. Figure C.12 shows that in the 7475 data, dependencies between Cu, Zn and Mg are not as strong as for the 7175 data.

In a number of pairwise plots the distribution of the data between section thickness and solutionising time showed a clear lower limit in solutionising times which increased with plate thickness[13]. A number of plots also show that some thinner plates may exit the hot mill at lower temperatures.

The presence of two alloy types is clear in the pairwise scatterplots for the BAP1TLQ3W data set (Figure C.14), from which it may also be noted that CRS values appeared to be reduced for thicker plate gauges.

In general, the inspection of the data distributions revealed the presence of a number of outliers, in the X and Y spaces, the effect of which will be subsequently assessed.

While several other observations can be made from inspecting all the other plots shown, similar distributions and dependencies were generally present in the data sets. Overall, the distribution and ranges of the process conditions were characterised by varying degrees of control. Some variates which prior knowledge identified as exerting a considerable influence

[13]Confirmed by (BAP, 2000).

on toughness levels showed a high level of control and hence little structure with K_{Ic}. Although such control is desirable and the objective of industrial processing, the uninformative nature of these variates implies that the variance in K_{Ic} values in the present data sets is attributable to other effects.

9.5 Multiple linear regression

As in previous chapters, a MLR analysis was conducted prior to the application of the adaptive modelling approaches. In addition to fitting MLR models to all the data in each of the data sets

	Training errors			Test error		df
Data set	$\hat{\sigma}^2_N$	$\hat{\sigma}^2_{\text{df}}$	$\hat{\sigma}^2_{\text{std,df}}$	μ	σ^2	
BAP1TLC3W	0.74	0.80	0.49	0.94	0.14	14/15
BAP1TLQ3W	1.88	2.18	0.19	2.70	5.12	14/15
BAP1TLQ3W80	0.84	1.07	0.30	1.60	1.34	14/15
BAP4LTC5W	7.32	7.69	0.45	6.61	9.79	13/14
BAP4SLC5W	8.89	10.23	0.54	12.48	17.94	13/14
BAP4TLC5W	4.34	4.55	0.43	3.64	3.00	13/14

Table 9.7: Summary of the approximations attained from a MLR analysis conducted on the different 7175 and 7475 data sets.

in Table 9.5, an estimate of the generalisation performance of each MLR analysis was obtained from fitting multiple models and determining the average errors over 20 random (90-10%) training-test set partitions. This subsequently allowed the linear regression performance to be compared with the generalisation abilities obtained for the adaptive modelling approaches.

The performance attained by MLR models on the different data sets is summarised in Table 9.7, in terms of variance estimates obtained from an analysis performed on all the data ($\hat{\sigma}^2_N$, and the adjusted, $\hat{\sigma}^2_{\text{df}}$), for results obtained on normalised K_{Ic} data (zero mean, unit variance), and statistics (μ and σ^2 over the multiple test sets) summarising the generalisation performance inferred from multiple model fits.

As the model structure is prespecified and no model selection is performed, the high variance in the test errors exhibited on particular data sets reflects the different training-test set partitions. As such, these should be acknowledged when comparing the performance attained from the adaptive methods, whereby the complexity and structure of the models will be determined by training on 90% of the data and the test performance on the remaining 10%.

From the test error results, the linear regression models appear to attain a reasonable generalisation performance and accuracy on the training data on two of the three 7175 data sets, while for the 7475 data these are seen to give poor performance. The standardised variance estimates, $\hat{\sigma}^2_{\text{std,df}}$ reveal that typically only half of the output variance is accounted for ($\hat{\sigma}^2_{\text{std,df}}$

being the proportion of the output variance not explained by the model).

An indication as to whether the data is appropriately described by a combination of linear effects can be inferred by inspecting the parametric uncertainties and the confidence placed in the parameters in the MLR. Tables 9.8 and 9.9 summarise the standardised regression coefficients, associated parametric uncertainties and τ estimates for each of the input variables present in the model.

	BAP1TLC3W			BAP1TLQ3W			BAP1TLQ3W80		
	w	$\hat{\sigma}_w$	τ	w	$\hat{\sigma}_w$	τ	w	$\hat{\sigma}_w$	τ
Plate gauge	0.664	0.120	5.54	0.017	0.107	0.16	0.122	0.087	1.41
Stretch	−0.080	0.053	1.52	−0.043	0.044	0.97	−0.055	0.076	0.72
Cu	−0.156	0.083	1.88	0.070	0.067	1.05	0.091	0.120	0.76
Fe	−0.373	0.060	6.21	−0.462	0.106	4.36	−0.509	0.089	5.69
Mg	0.015	0.076	0.20	−0.233	0.191	1.22	−0.054	0.156	0.35
Mn	0.205	0.059	3.49	0.021	0.063	0.33	0.090	0.106	0.85
Si	−0.374	0.061	6.12	−0.357	0.096	3.71	−0.429	0.119	3.61
Zn	0.081	0.083	0.97	0.008	0.059	0.14	0.029	0.103	0.28
Ti	0.046	0.064	0.72	0.016	0.073	0.21	0.056	0.082	0.68
Cr	0.018	0.056	0.33	0.046	0.049	0.94	0.078	0.086	0.91
T_f	−0.103	0.056	1.85	−0.065	0.053	1.23	−0.109	0.095	1.15
ST_t	0.029	0.058	0.51	0.005	0.044	0.11	0.012	0.080	0.16
GR	−0.518	0.122	4.26	−0.039	0.051	0.76	−0.278	0.093	2.98
CRS	0.293	0.053	5.52	0.213	0.052	4.06	0.386	0.085	4.55

Table 9.8: Parametric inferences for 7175 K_{Ic} for each data set.

The largest weights and signal-to-noise parameters are generally obtained for Fe and Si in both the 7175 and 7475 data sets. A comparison over the different data sets shows that there is a certain degree of variability in both the weight magnitudes and parametric uncertainties for other variables (e.g. Mg, T_f, CRS) and that in several instances contrasting trends are inferred. For particular inputs, e.g. CRS, and different test directions metallurgical understanding may suggest this to be plausible whereas in others it is more problematic (e.g. Cu in the 7175 data sets). Where contradictory, trend inferences are associated with high parametric uncertainties, interpretation of individual effects may of course be misleading.

In a number of data sets, the highest parameteric uncertainties are seen to be associated with plate gauge and GR, suggesting that the correlations characterising these two variates may be held responsible for the lower τ's.

Overall, the small weight magnitudes and associated τ's show that the effect of a significant number of terms in the models are inadequately determined.

	BAP4LTC5W			BAP4SLC5W			BAP4TLC5W		
	w	$\hat{\sigma}_w$	τ	w	$\hat{\sigma}_w$	τ	w	$\hat{\sigma}_w$	τ
Plate gauge	0.348	0.060	5.76	0.255	0.105	2.42	0.306	0.059	5.15
Stretch	−0.012	0.041	0.30	0.058	0.082	0.70	−0.044	0.040	1.09
Cu	0.033	0.046	0.70	0.185	0.091	2.03	0.044	0.045	0.98
Fe	−0.487	0.048	10.16	−0.302	0.089	3.40	−0.525	0.047	11.16
Mg	−0.147	0.049	3.00	−0.157	0.108	1.45	−0.154	0.048	3.24
Si	−0.285	0.046	6.24	−0.108	0.089	1.21	−0.291	0.044	6.55
Zn	−0.059	0.051	1.16	−0.045	0.118	0.38	−0.013	0.049	0.27
Ti	0.015	0.043	0.34	−0.025	0.082	0.30	0.066	0.042	1.58
Cr	0.059	0.042	1.41	−0.093	0.083	1.12	0.051	0.041	1.25
T_f	−0.005	0.051	0.10	0.199	0.102	1.95	−0.064	0.050	1.26
ST_t	−0.067	0.047	1.43	−0.025	0.085	0.30	−0.163	0.046	3.53
GR	−0.101	0.063	1.59	0.009	0.100	0.09	−0.051	0.062	0.82
CRS	−0.230	0.043	5.31	−0.249	0.080	3.13	0.082	0.042	1.93

Table 9.9: Parametric inferences for 7475 K_{Ic} for each data set.

9.6 Neurofuzzy modelling

In Chapter 5 it was seen how neurofuzzy systems exhibit a number of useful knowledge representation properties, particularly attractive in data mining analyses, e.g. model initialisation, parsimonious system description and the ability to handle both continuous and discrete variables. Although previous chapters have highlighted the suboptimality of iterative model construction procedures, it should be acknowledged that the relatively simple approximations inferred in such analyses were a reflection of the relatively limited data sets, for which the SRM based statistical significance measure prevented the inclusion of a large number of degrees of freedom. In light of the relatively larger sizes of the BAP data sets and a number of linearities between the input variables and K_{Ic}, suggested by inspection of the data distributions (e.g. Fe, Si), it was of interest to assess whether the neurofuzzy model construction framework could adequately approximate these dependencies and discover any other relationships which were not evident from the residuals.

In the following sections, the analyses and inferences attained from the application of the neurofuzzy system identification and representation capabilities are discussed, enabling the merits of the approach and the various inferences drawn to be assessed.

9.6.1 Data mining

Following the same methodology employed in previous chapters, the ASMOD algorithm (using a FS/BE search pass structure) was used to determine network structures from the BAP1 and

BAP4 data sets. Models were determined using third order B-splines as well as (second order) piecewise linear basis functions, as the amount of data and the nature of the non-linearities exhibited in a number of the full residual plots suggested that higher order basis functions might attain better representation abilities compared to the linear piecewise approximation obtainable from second order B-splines.

In addition, simple network initialisations were considered for the BAP1 and BAP4 data sets to assess the influence of the test condition and ALLOY KG (categorical variables). These are shown in Figures 9.2 (a) and (b), whereby first order basis functions are used to encode the different test conditions and alloy types. Upon weight-rule confidence training (i.e. determining the network weights corresponding to these basis functions), the resulting biases associated with each of the attributes are illustrated in the subnetwork responses shown in Figures 9.3 and 9.4.

For the BAP1 network, these show that the data corresponding to the alloys tested in the quarter thickness position exhibit a higher mean in the K_{Ic} and that the LE alloy types on average attain considerably higher K_{Ic} levels than the LK alloys. The alloy type attribute is therefore understood to account for the toughness levels attained by the thicker alloys and thus explains why alloys tested in the mid-thickness position exhibit a lower K_{Ic} average than those tested in the quarter position (Table 9.6).

A similar bias for the thickness position is seen for the BAP4 data, whereby alloys tested in the quarter thickness position exhibit a marginally higher average K_{Ic}. In addition, the DIRECTION subnetwork confirms that SL data correspond to the lowest range in toughness values and the LT direction correspond to alloys with the highest toughness levels. Finally, the biases exhibited for the two width positions show alloys tested at the mid-width position to exhibit a higher mean K_{Ic}.

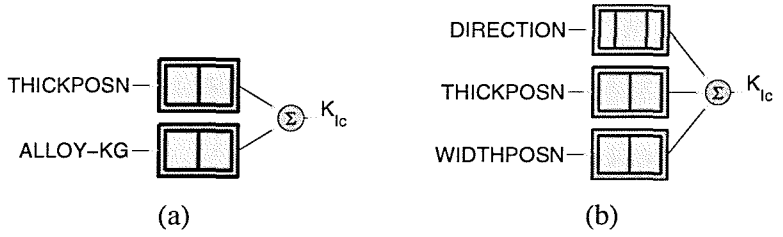
It should be recognised that interpretation of the initialised network structures *per se* is potentially misleading, as the biases determined for the alloy type and test attributes are seen to account for other sources of variance (e.g. purity levels, plate thickness dependencies, etc.). In this instance, it may be noted that the quarter thickness tests were all derived in the TL and LT directions, giving a data set dependent bias (as opposed to a "true" physical trend).

In terms of final network structures determined by initialising the ASMOD algorithm from both empty and initialised structures (as in Figure 9.2) gave identical model structures, whereas employing different order B-splines resulted in differences in the form of the approximations. Network structures determined for the BAP1 and BAP4 data sets using respectively third and second order basis functions are shown in Figures 9.5 (a) and (b). The corresponding prediction scatterplots are shown in Figures 9.8 (a) and (b).

The network structures inferred from the different initial conditions and the empirical performance attained are summarised in Table 9.10. These show that while a small improvement in the empirical performance in modelling the BAP4 data set can be obtain from using third order B-spline basis functions, similar approximations are inferred for the BAP1 data regardless of

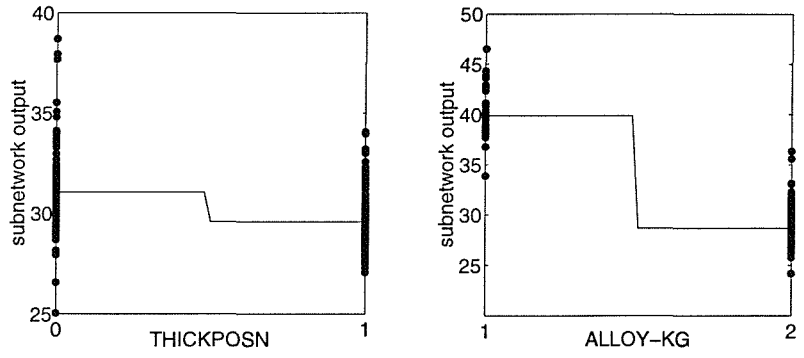
Data set	$\hat{\sigma}^2_{\text{Init},N}$	$\hat{\sigma}^2_{\text{df}}$	LOOCV.	B-spline order	ANOVA terms	df
BAP1	2.5	1.43	1.48	$1-2$	$\text{Fe}^{\text{pl}}, \text{CRS}^{\text{l}}, \text{Si}^{\text{l}}$	5
		1.42	1.49	$1-3$	Fe^{q}	5
BAP4	15.09	7.56	7.66	$1-2$	$\text{DIRECTION}^{\text{pl}}, \text{Fe}^{\text{l}}, \text{Si}^{\text{pl}}, \text{Plate gauge}^{\text{pl}}$	8
		7.39	7.48	$1-3$	$\text{DIRECTION}^{\text{c}}, \text{Fe}^{\text{l}}, \text{Si}^{\text{q}}, \text{Plate gauge}^{\text{q}}$	8

Table 9.10: Summary of the empirical results attained by the neurofuzzy models determined by the ASMOD algorithm for the 7175 and 7475 data sets, trained from either an initialised or an empty structure, using different order B-spline basis functions. In these results, the different order of the B-splines associated with the approximations inferred on the input variables is indicated by the superscripts in the ANOVA terms field (c, 1^{st} order, piecewise constant; l, 2^{nd} order linear approximation; pl, 2^{nd} order piecewise linear and q, 3^{rd} order approximation).



(a) (b)

Figure 9.2: Initialised model structures for the BAP1 (a) and BAP4 (b) data sets.



(a) (b)

Figure 9.3: Subnetwork response outputs corresponding to the initialised model structures shown in Figure 9.2 (a) upon weight-rule confidence training.

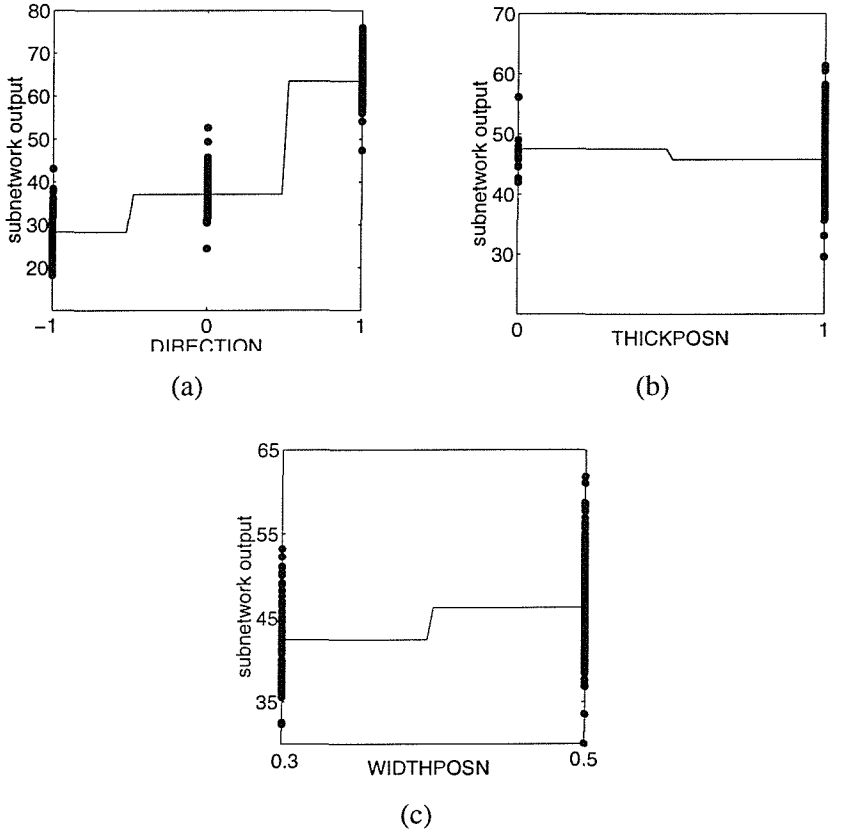


Figure 9.4: Subnetwork response outputs corresponding to the initialised model structures shown in Figure 9.2 (b) upon weight-rule confidence training.

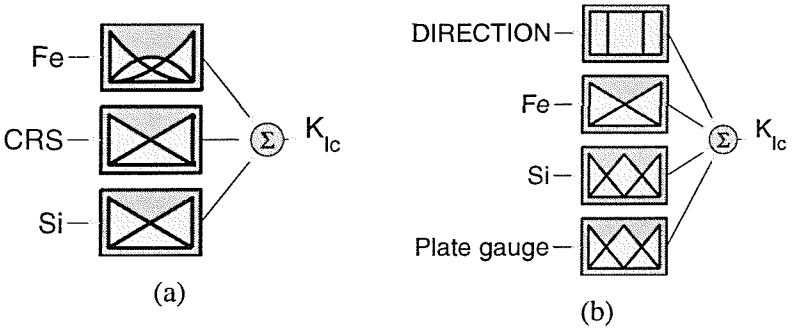


Figure 9.5: Network structures determined from initialised model structures by the ASMOD algorithm for (a) the BAP1 data set, employing univariate B-spline basis functions of orders 2-3, and (b) the BAP4 data set, employing 1-2 univariate basis functions.

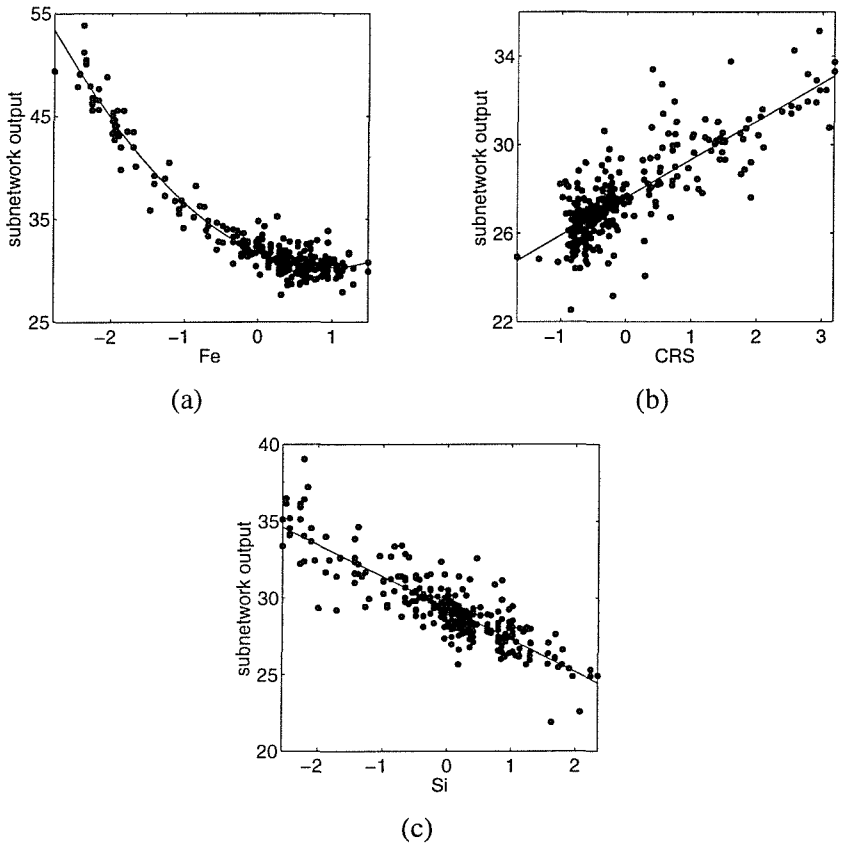


Figure 9.6: Subnetwork responses obtained from training on the BAP1 data set and allowing the ASMOD algorithm to employ 1-3 B-spline basis functions: Fe (a), CRS (b) and Si (c).

employing higher order B-splines. Subnetwork responses corresponding to models determined from employing third order basis functions are shown in Figures 9.6 and 9.7 for the BAP1 and BAP4 data sets respectively.

From these results it is seen that different model structures were determined for the BAP1 and BAP4 data sets, with the larger of the two (BAP4) exhibiting more degrees of freedom. Whilst training from empty and initialised model structures resulted in identical inferred networks, differences in the complexity of the approximating functions were found from employing second or third order basis functions in the candidate refinements.

The refinements performed and evolution of the model performance measures during the FS/BE construction stages are shown in Figure 9.9 for models trained from initialised structures and employing third order basis functions. These, together with the other results presented above show how in the BAP1 model evolution, both initialised dichotomies on THICKPOSN and ALLOY KG type have been removed, with the final network comprising three variables: Fe, Si and CRS. The model construction history shows how deleting the ALLOY KG and WIDTHPOSN subnetworks (at the 5^{th} and 6^{th} refinements) has negligible effect on the accuracy of the model, resulting from the contribution of Fe, Si and plate gauge subnetworks. Thus

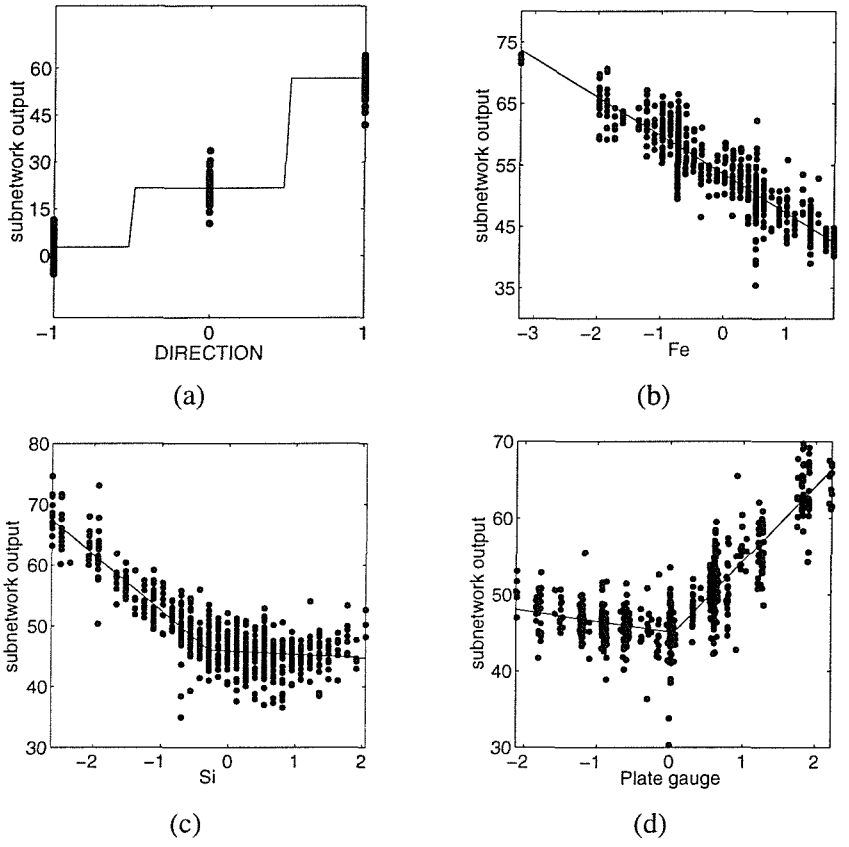


Figure 9.7: Subnetwork responses obtained from training on the BAP4 data set and allowing the ASMOD algorithm to employ 1-3 B-spline basis functions: DIRECTION (a), Fe (b), Si (c) and plate gauge (d).

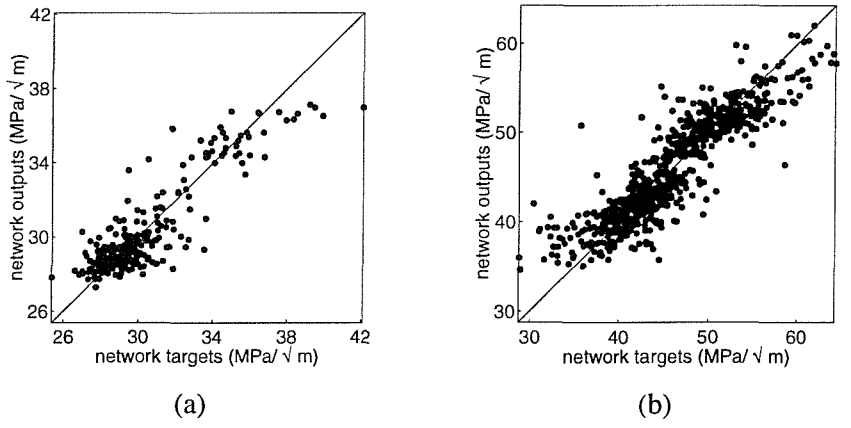
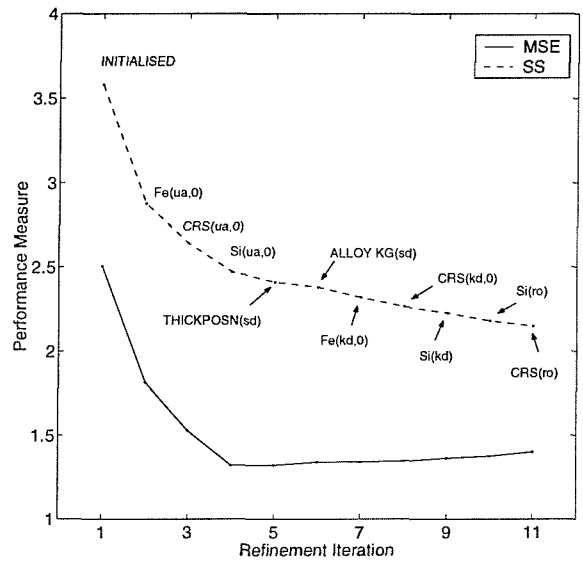
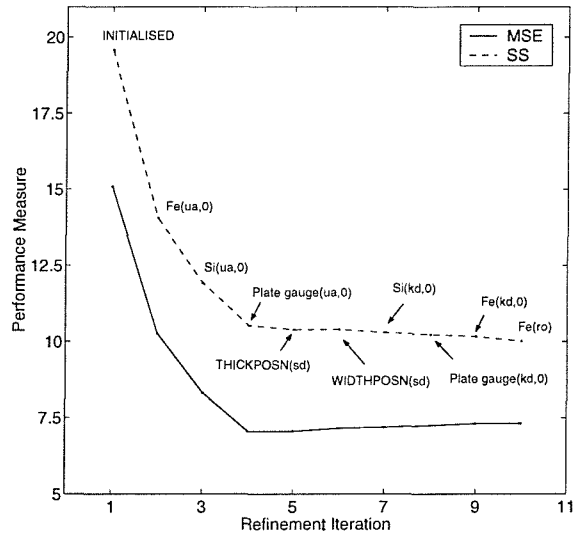


Figure 9.8: Prediction scatterplots obtained for the neurofuzzy models identified by the ASMOD algorithm upon training from the initialised network structures shown in Figure 9.2 for data sets BAP1 (a) and BAP4 (b), using 1-2 and 1-3 order B-splines.

it is seen that the model structure has been refined to comprise a set of explanatory variables, which more successfully explain the variance in K_{Ic} levels, compared to the approximations attained by the simple initialised model.



(a)



(b)

Figure 9.9: Evolution of the model perfomance measures during the iterative model search performed from the initialised structures for (a) BAP1 and (b) BAP4 data set respectively, upon initialising the model search to include third order basis functions.

The final model determined form the BAP4 data set shown in Figure 9.5 (b) is seen to retain the discrimination on test DIRECTION, whilst both THICKPOSN and WIDTHPOSN are removed during network refinement. The subnetwork response (and corresponding modi-

fied residuals), shown in Figure 9.7 is seen to maintain the $K_{Ic}(\mathrm{LT}) > K_{Ic}(\mathrm{TL}) > K_{Ic}(\mathrm{SL})$ ordering placed on DIRECTION. The contributions of Fe, Si, and plate gauge are seen to be included as either additive linear, piecewise linear and quadratic univariate subnetworks depending on the order of the basis functions considered during training. Figure 9.7 (d) shows how the piecewise approximation determined on plate gauge, initially decreases with a modest slope, whilst thicker alloys are seen to induce a different approximation, whereby higher plate gauges suggest improvements in toughness. This apparent recovery in toughness with increasing gauge for the thickest plates is perhaps surprising, and will be discussed further in relation to the various subsets of the BAP4 data later in the chapter.

In summary, it is seen that the final model structures include similar compositional dependencies (specifically in terms of Fe and Si) as well as simple relationships with dimensional/deformation inputs, as well as accounting for different test conditions where necessary. Although a limited number of inputs were included, the approximation obtained for the 7175 data are seen to be considerably better than those obtained for the 7475 data set, the latter data exhibiting very high training and LOOCV variance estimates.

9.6.2 Data modelling

Although attaining plausible structural dependencies, the accuracy of the models inferred above in the data mining approach was seen to be unable to account for a considerable proportion of the variance in toughness levels. This may be due to the formulation of either an inappropriate global model, whereby an additive system representation is unable to successfully account for differences in the behaviour exhibited in different test conditions, or the presence of a high noise component (missing features, inherent measurement scatter of K_{Ic} tests) limiting further significant improvements. It is also clear that the BAP1 and BAP4 data sets would be problematic to model in a transparent meaningful way (upon using a global, additive decomposition of the input space), given the presence of important interdependencies (such as gauge and alloy purity in the 7175), and due to the wide range of toughnesses and physical dependencies that may be involved with different test directions and locations. It was useful therefore to assess whether the fracture toughness of the 7x75 alloys in the different test conditions was characterised by different structural dependencies and noise levels. Hence models were determined for the different 7x75 data set partitions. These were again determined by using a FS/BE search, initialising the model construction algorithm from an empty model structure and again employing both second and third order B-splines.

The network structures and empirical performance attained are summarised in Table 9.11, where the ordering of the ANOVA terms reflects the order of inclusion of the inputs. Although the same input variables are seen to be present in the majority of the final network structures, both the order of inclusion of the inputs and the refinements attempted show the different model searches performed.

Data set	B-spline order	ANOVA terms	$\hat{\sigma}^2_{df}$	LOOCV	df
BAP1TLC3W	$1-2$	Fe^l, CRS^l, Si^l, Mn^l, Plate gauge^l, GR^l	0.81	0.87	7
	$1-3$	Fe^l, CRS^l	1.18	1.20	3
BAP1TLQ3W	$1-2$	$Fe^{pl}, CRS^l, Si^l, GR^{pl}$	1.60	1.79	7
	$1-3$	Fe^q, Si^l, CRS^l, GR^q	1.63	1.88	7
BAP1TLQ3W80	$1-2$	$Fe^l \times CRS^l, Si^l, GR^l$	0.98	1.09	6
	$1-3$	Fe^q, CRS^l, GR^l	1.26	2.17	6
BAP4LTC5W	$1-2$	Fe^l, Si^l, Plate gauge^{pl}	8.12	8.32	5
	$1-3$	Fe^l, Si^q, Plate gauge^q, CRS^l	7.09	7.33	7
BAP4SLC5W	$1-2$	Plate gauge^l, Fe^l, Si^c	10.64	11.56	5
	$1-3$	Plate gauge^l, Si^q, Fe^l	10.44	10.91	5
BAP4TLC5W	$1-2$	Fe^l, Si^l, Plate gauge^{pl}	4.60	4.70	5
	$1-3$	Fe^l, Si^q, Plate gauge^q	4.55	4.71	6

Table 9.11: Summary of the results determined by the neurofuzzy framework for each of the data sets.

The prediction scatterplots corresponding to the models attaining lower LOOCV errors are shown in Figure 9.10.

A comparison between the model structures inferred from employing second or third order basis functions shows greater discrepancies in the 7175 networks. As the implementation costs (i.e. in terms of the number of degrees of freedom) associated with third order basis functions are greater than that of similar piecewise linear approximations, they may be seen to prevent the inclusion of additional degrees of freedom during model building refinements, this would seem to have occurred in the small model determined from the BAP1TLC3W data set.

A comparison with the results obtained in section 9.6.1 shows that similar dependencies with the impurity elements (Fe, Si) and dimensional/deformation variables are determined. The network structures inferred all comprise univariate terms, with the exception of the Fe $\times$ CRS tensor product present in the model obtained using third order basis functions on the BAP1TLQ3W80 data set. The terms determined on the 7475 data sets are (with the exception of the addition of CRS in the BAP4LTC5W data sets, employing third order basis functions), seen to be the same as those determined on the BAP4 data, although in some instances differing in the form of the approximations, particularly in terms of plate gauge effects, which will be discussed later. Models determined on the 7175 data are seen to differ from the BAP1 network structure by the inclusion of GR subnetworks, and in the BAP1TLC3W data set (determined using second order B-splines), a linear Mn contribution. A number of subnetworks also exhibited different flexibilities, e.g. Fe being a linear effect in a number of models in Table 9.11 as opposed to piecewise linear or quadratic terms in Table 9.10 for the 7175 data sets.

A comparison with the training set variance estimates (in terms of the unbiased $\hat{\sigma}^2_{df}$)

obtained by the MLR analysis (Table 9.7) reveals that only the BAP1TLQ3W data set models determined from using either second and third order basis functions attain lower $\hat{\sigma}^2_{\text{df}}$. More generally, the 7175 results employing third order basis functions result in models attaining error variance estimates (both $\hat{\sigma}^2_{\text{df}}$ and LOOCV) marginally higher than models determined by using second order B-splines. The results obtained on the larger of the 7475 data sets however show that the use of higher order basis functions yield both lower training set $\hat{\sigma}^2_{\text{df}}$ and LOOCV estimates. Notwithstanding the underperformance and/or lack of significant improvement in empirical performance exhibited from the above in comparison with the MLR results in terms of $\hat{\sigma}^2_{\text{df}}$ levels, models determined by the ASMOD algorithm are seen to comprise more parsimonious descriptions of the data, enabling a number of non-linear approximations to be represented. The training set variance estimates, although allowing an assessment and comparison of the neurofuzzy models in terms of network parsimony and nature of the dependencies with the MLR results *per se*, cannot be used to infer the generalisation performance of the models as these should be determined on unseen data (section 9.6.3).

Model construction histories corresponding to networks determined from the BAP1TLC3W and BAP4TLC5W data sets are shown in Figures 9.11 and 9.12, respectively. These show how the inclusion of the input variates (e.g. T_f, Cu and Mg, ST_t in the 7175 and 7475 data set, respectively) cause an increase in the SS measure, whilst the inclusion of extra flexibility (i.e. knot insertion) is seen in some instances to improve (reducing) the SS measure (e.g. knot insertion on Si in the 7175 data set) but not sufficiently to successfully accept the refinement, as determined by the f_{tol}. Figure 9.11 (a) also shows how the evolution in the training set MSE ($\hat{\sigma}^2_N$) evolves in relation to the level obtained from the corresponding MLR model (shown by the ·--- line).

9.6.3 Assessing stability of the model selection

As in previous chapters, the stability of the representations inferred in the previous section were assessed by determining models on different training-test set partitions. This would also allow estimates of the generalisation performance of the neurofuzzy framework to be compared with those previously exhibited by the MLR analysis and with the SVM in section 9.7. Hence using the same 20 random training-test set partitions that were used in section 9.5 with the MLR, the ASMOD algorithm was used to determine multiple models for each of the 7x75 data sets.

Training was initialised from an empty model structure and models determined again employing either second and third order basis functions. The mean and variance of the training set MSE, LOOCV and test set error estimates inferred from the multiple model runs are shown in Table 9.12, while the frequency of selection of the different ANOVA terms for each data set, are summarised in Table 9.13.

Unsurprisingly, in most data sets, and particularly the smallest, the LOOCV estimates are optimistic estimates of the output variance compared with the generalisation error estimates

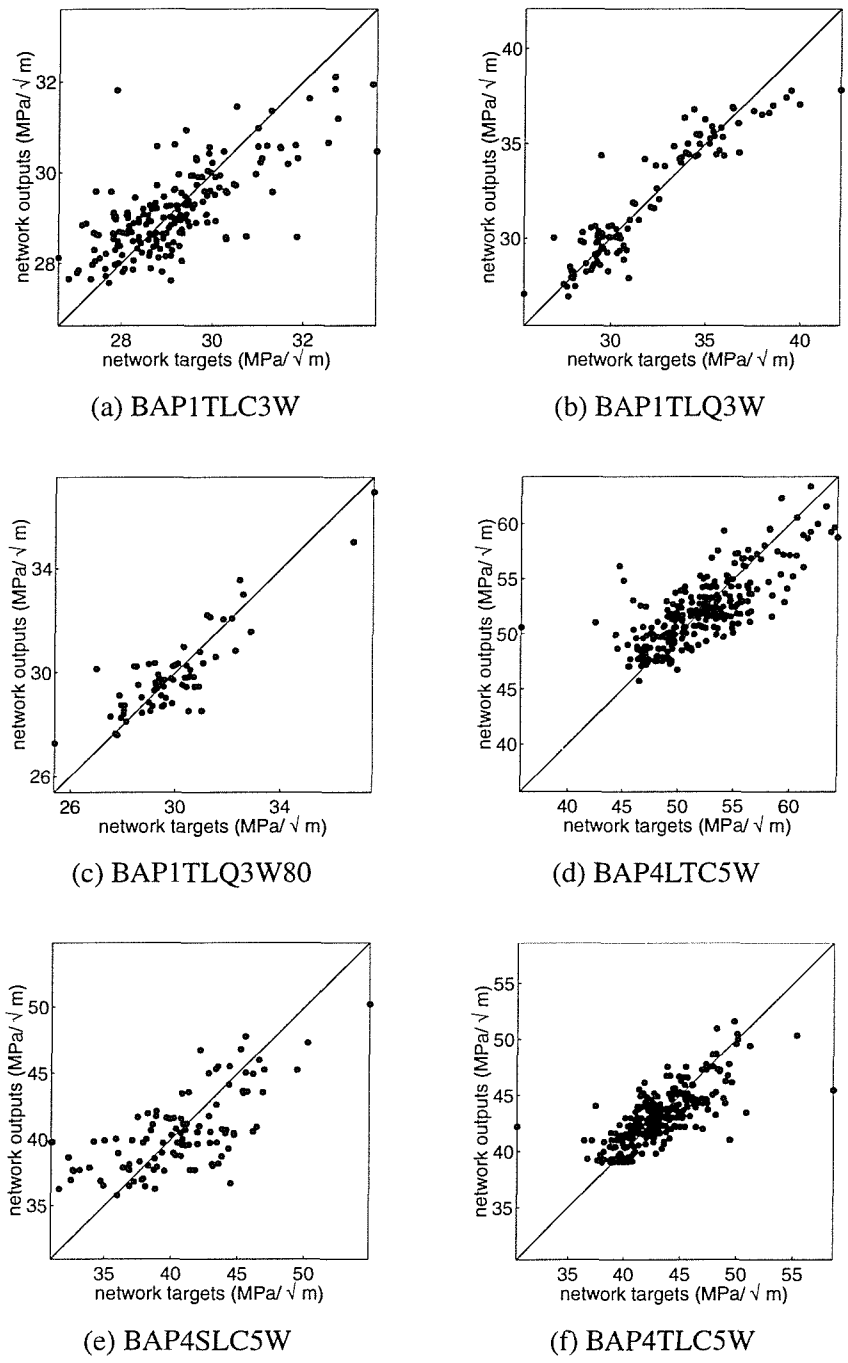
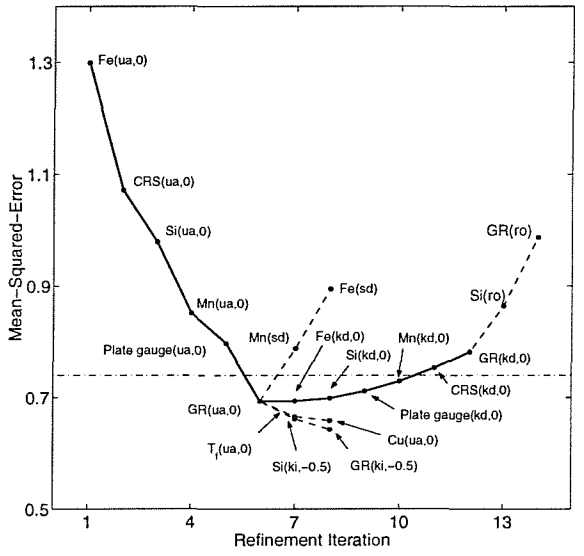
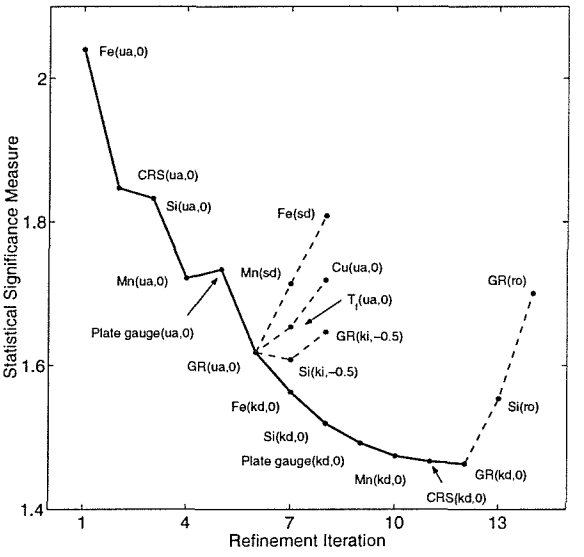


Figure 9.10: Prediction scatterplots corresponding to the neurofuzzy models having the lower LOOCV estimates, identified by the ASMOD algorithm from training on the different data sets, using different order B-spline basis functions: 1-2 for the 7175 data sets (a)-(c); 1-3 for the 7475 data sets (d)-(f).

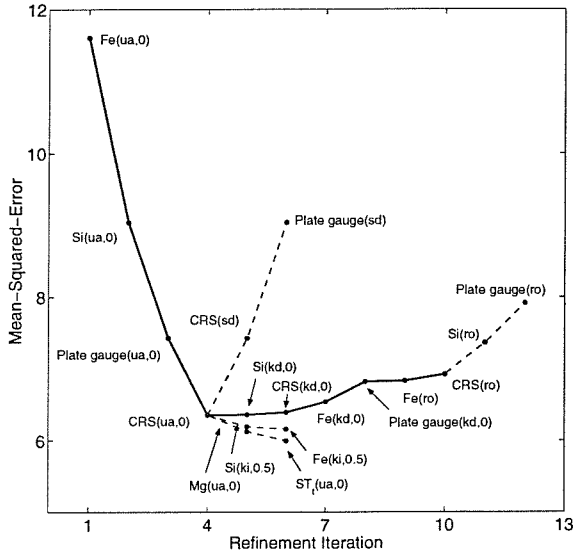


(a)

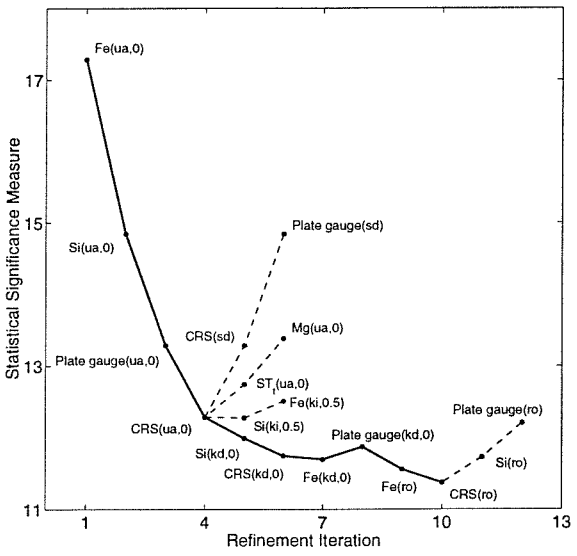


(b)

Figure 9.11: Evolution of the training set MSE (a) and SS measure (b) during the iterative model search performed in determining the network structure for the BAP1TLC3W data set, employing second order B-spline basis functions.



(a)



(b)

Figure 9.12: Evolution of the training set MSE (a) and SS measure (b) during the iterative model search performed in determining the network structure for the BAP4TLC5W data set, employing third order B-spline basis functions.

Data set	B-spline order	Training error μ	Training error σ^2	LOOCV μ	LOOCV σ^2	Test error μ	Test error σ^2
BAP1TLC3W	$1-2$	0.91	0.03	0.98	0.02	1.28	0.21
	$1-3$	1.18	0.01	1.22	0.01	1.43	0.29
BAP1TLQ3W	$1-2$	1.45	0.07	1.68	0.07	3.09	3.07
	$1-3$		0.04	1.80	0.06	2.74	3.13
BAP1TLQ3W80	$1-2$	0.90	0.01	1.22	0.13	1.79	1.38
	$1-3$	1.28	0.08			2.74	6.37
BAP4LTC5W	$1-2$	7.90	0.17	8.29	0.17	7.57	8.27
	$1-3$	7.17	0.11	7.60	0.11	6.49	6.32
BAP4SLC5W	$1-2$	10.56	1.25	11.53	1.06	15.64	30.37
	$1-3$	12.09	3.30	12.96	2.83	17.02	45.38
BAP4TLC5W	$1-2$	4.92	0.28	5.03	0.26	4.18	3.14
	$1-3$	4.80	0.22	5.03	0.19	4.04	2.95

Table 9.12: Summary of the empirical performance inferred from the multiple model runs obtained for the neurofuzzy framework.

inferred from unseen data in many cases[14]. A comparison with Table 9.7 shows that only the model determined for the BAP4LTC5W data set, using third order basis functions is seen to achieve a slight improvement over MLR in terms of the average generalisation (test set) error (μ). In a few instances (generally corresponding to the largest data sets) higher order basis functions were seen to attain a marginal improvement in approximation abilities. A comparison between the summary statistics shows the smallest data sets to exhibit the largest variances σ^2, particularly when using higher order basis functions.

The terms selected with the greatest consistency are in agreement with previous results: models typically comprised univariate subnetworks and the limited number of bivariate (tensor) terms comprised by a subnetwork in the form "deformation variable $\times$ impurity". A larger proportion of these are determined from models trained using lower order basis functions.

The general form of the subnetwork responses inferred for the most frequently selected terms are shown in Figures 9.13 to 9.18. A comparison between the subnetwork responses show that for a number of data sets, the contribution of particular terms is seen to differ from those attained from modelling the BAP1 and BAP4 data sets, implying that the dependencies between K_{Ic} and certain variates (e.g. GR, Fe in the 7175 data sets, CRS in the 7475) are dependent on the test conditions and the presence of different alloy types (7175).

[14]Although it should be understood that the LOOCV and generalisation performances were obtained on different data sets: the former from the training set used to determine the model structure, the latter from the 10% unseen data.

Data set	B – spline order	Plate gauge	CRS	GR	Fe	Si	Mn	Zn	Mg	T_f	ST_t	CRS × Fe	CRS × Si	GR × Fe	Fe × Si
BAP1TLC3W	1 – 2	6	20	6	20	14	14								
	1 – 3		15		20										
BAP1TLQ3W	1 – 2		18	14	16	14							2	4	
	1 – 3		20	18	18	16								2	
BAP1TLQ3W80	1 – 2		8	18	8	18		1		2		11		1	
	1 – 3		15	13	20										
BAP4LTC5W	1 – 2	20	3		20	20									
	1 – 3	20	19		20	20									
BAP4SLC5W	1 – 2	16	9		13	9			1	4					3
	1 – 3	13	1		7	12				6					
BAP4TLC5W	1 – 2	12			20	20					1				
	1 – 3	15			20	20									

Table 9.13: Frequency of selection of ANOVA terms over 20 model runs for the neurofuzzy models.

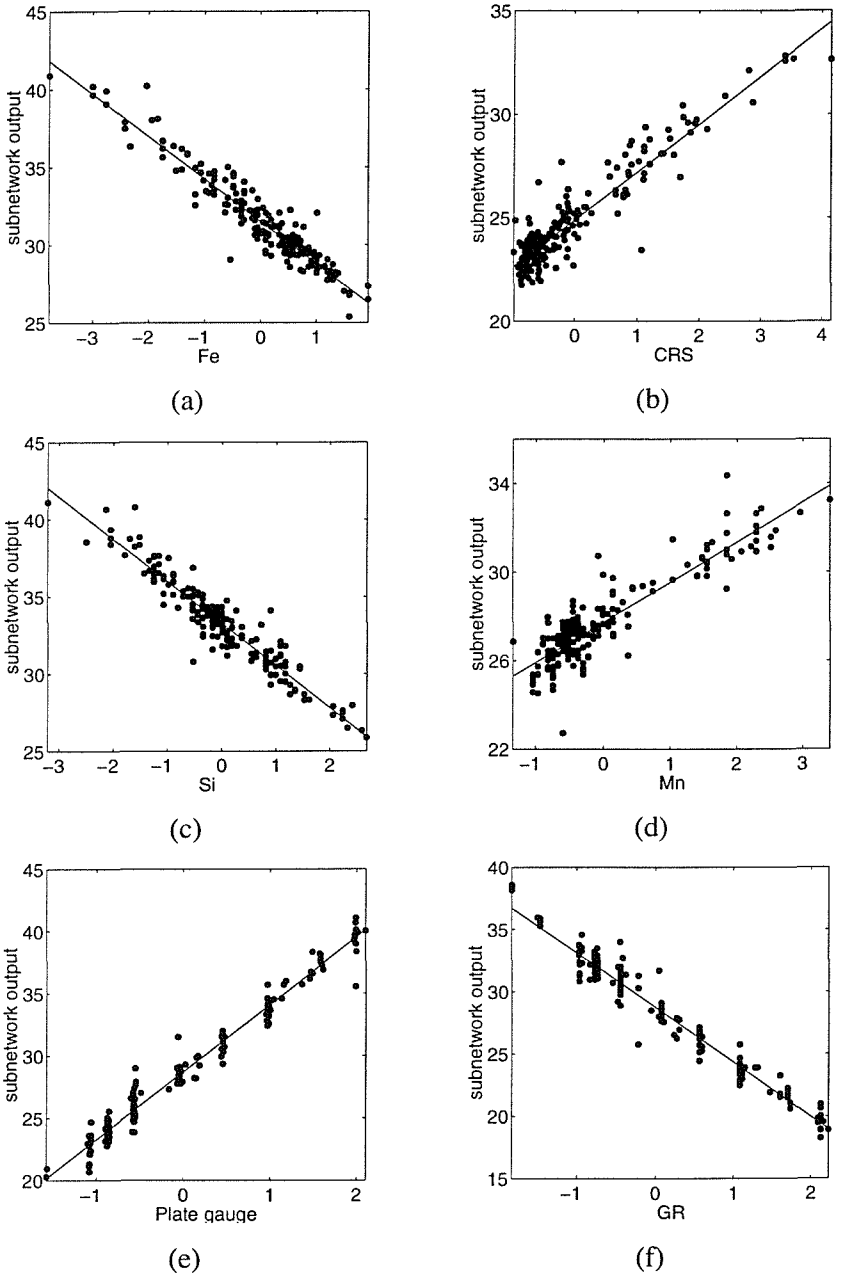


Figure 9.13: Typical subnetwork responses for the most frequently selected ANOVA terms determined by the ASMOD algorithm upon training on the BAP1TLC3W data set.

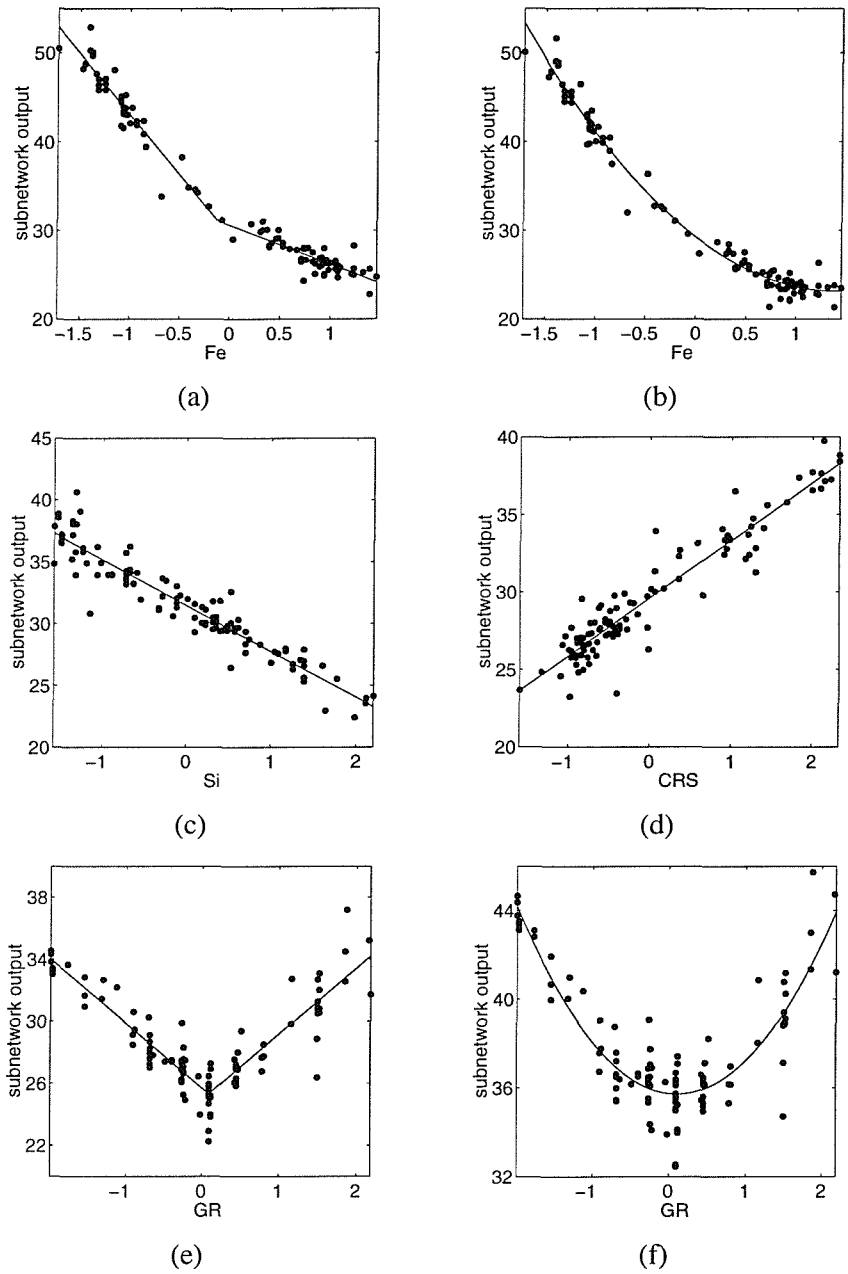


Figure 9.14: Typical subnetwork responses for the most frequently selected ANOVA terms determined by the ASMOD algorithm upon training on the BAP1TLQ3W data set.

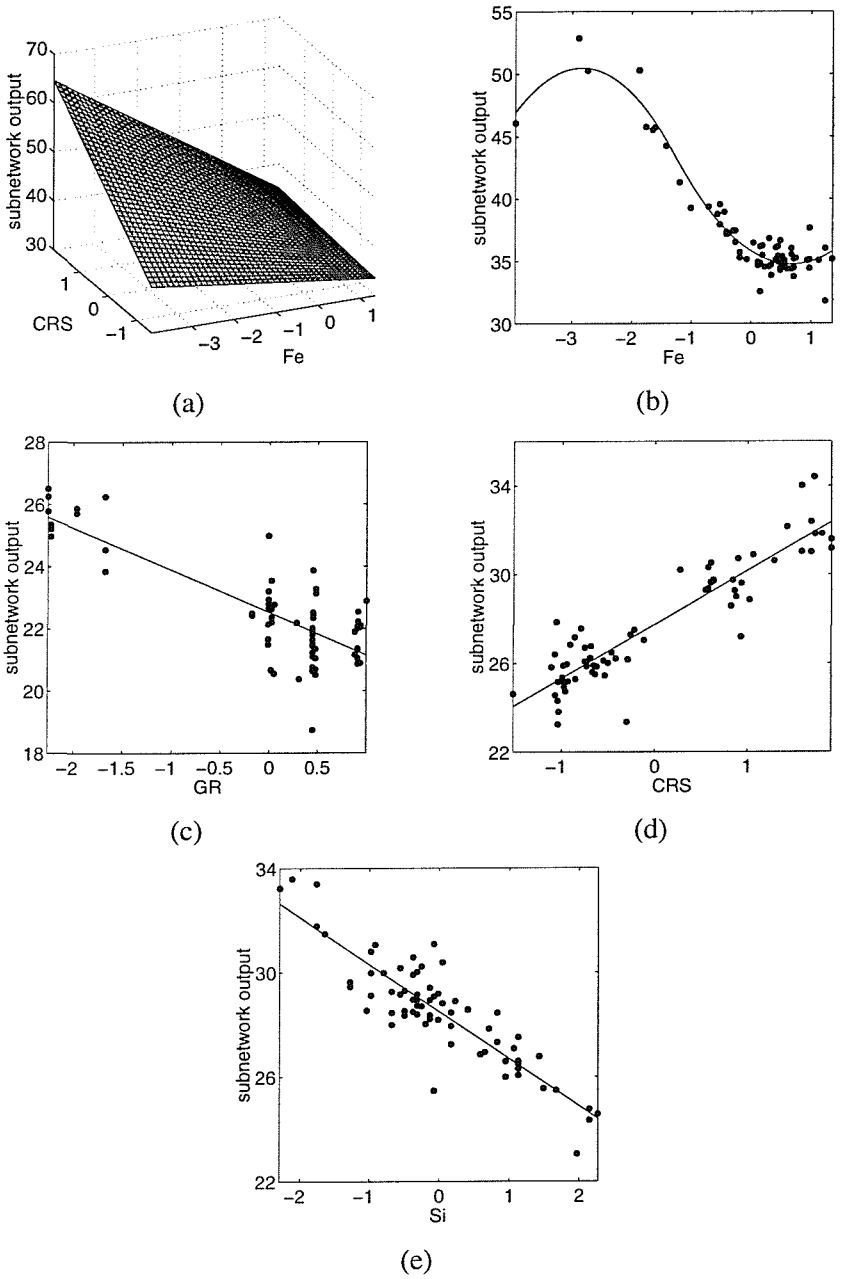


Figure 9.15: Typical subnetwork responses for the most frequently selected ANOVA terms determined by the ASMOD algorithm upon training on the BAP1TLQ3W80 data set.

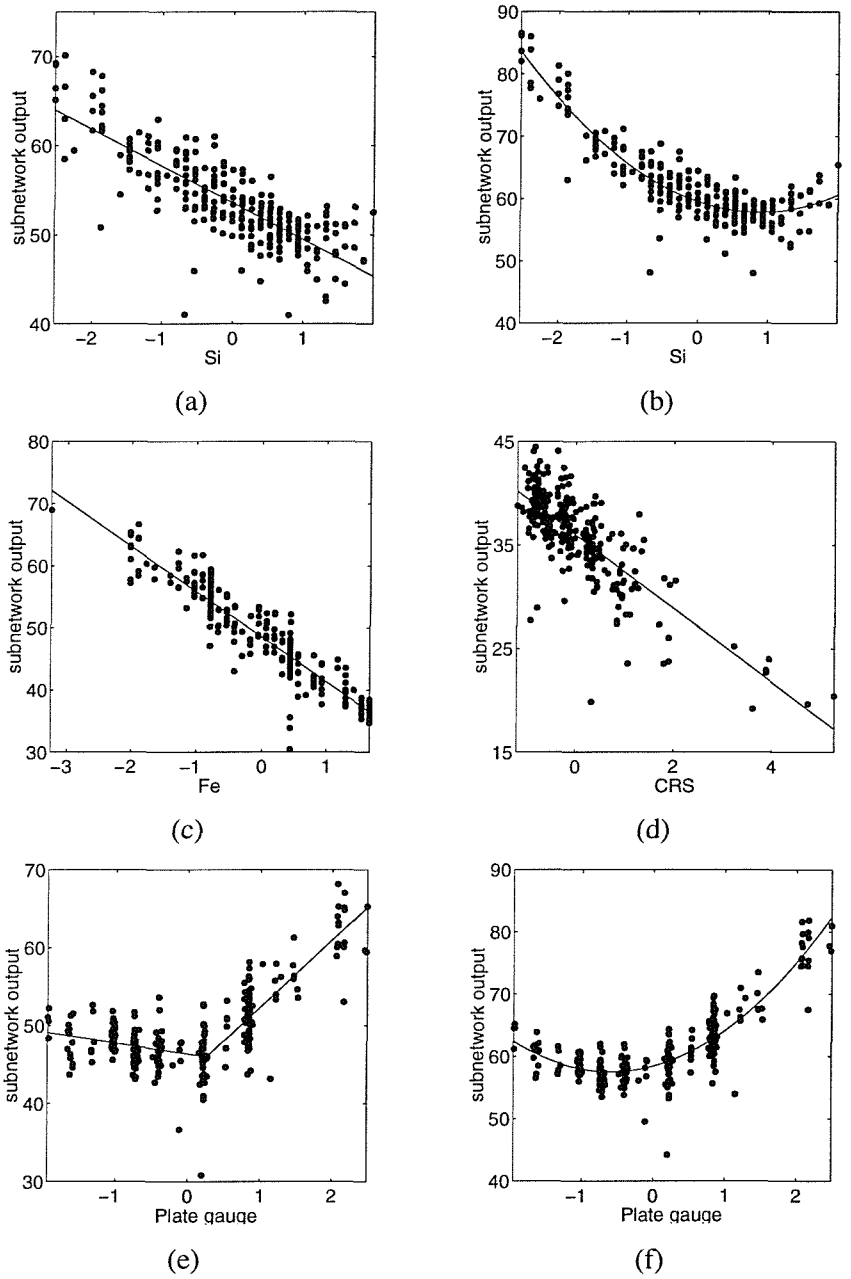


Figure 9.16: Typical subnetwork responses for the most frequently selected ANOVA terms determined by the ASMOD algorithm upon training on the BAP4LTC5W data set.

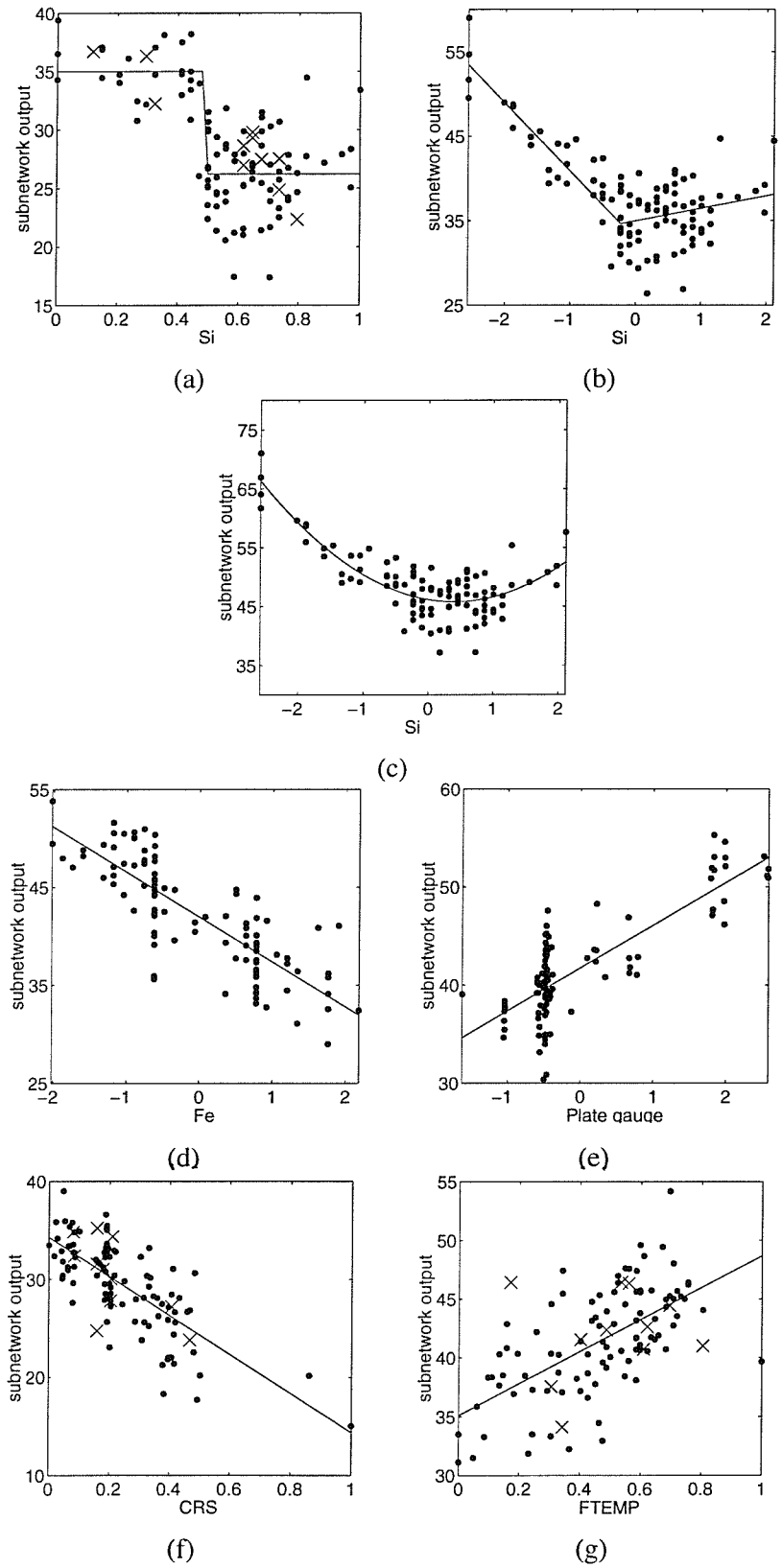


Figure 9.17: Typical subnetwork responses for the most frequently selected ANOVA terms determined by the ASMOD algorithm upon training on the BAP4SLC5W data set.

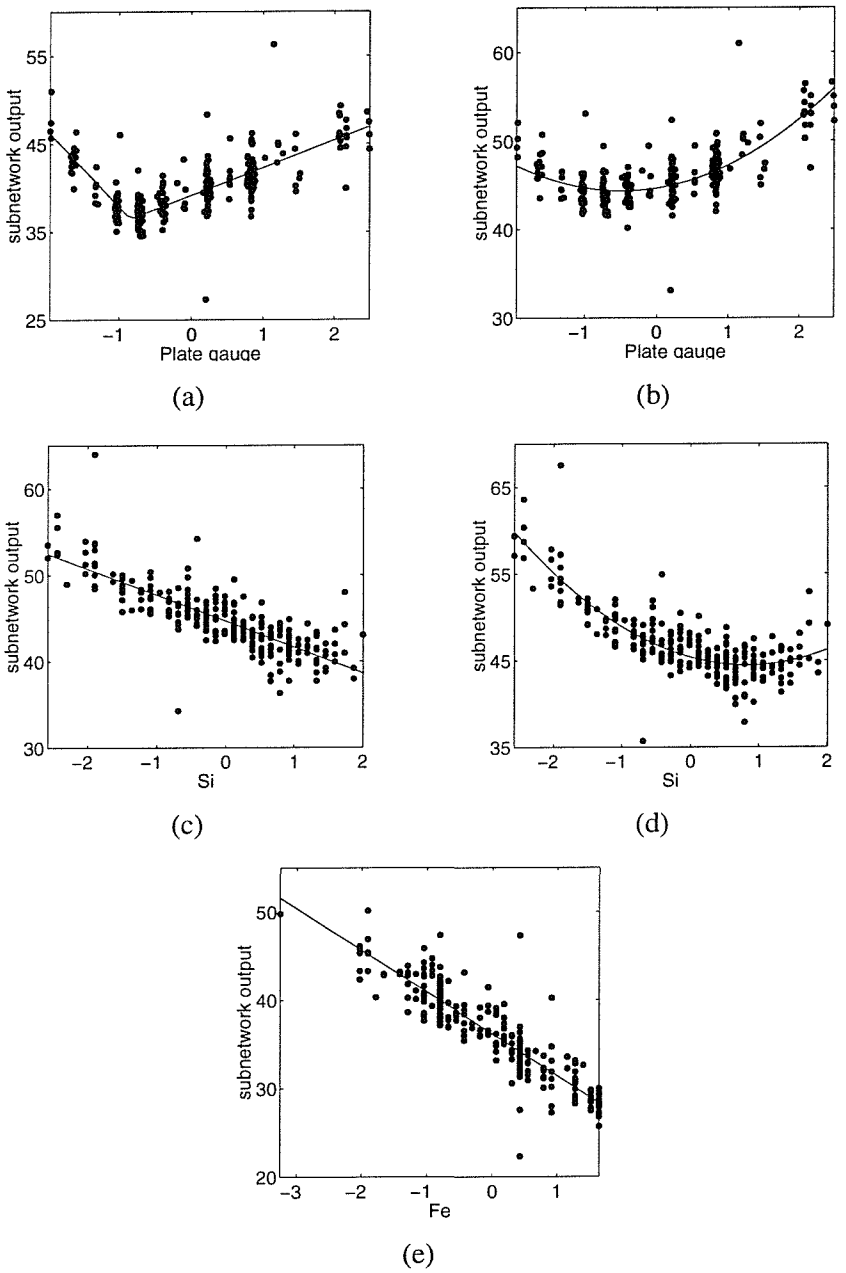


Figure 9.18: Typical subnetwork responses for the most frequently selected ANOVA terms determined by the ASMOD algorithm upon training on the BAP4TLC5W data set.

9.6.4 Influence analysis

From the various residual plots shown in previous sections, a number of outliers are seen to be present in the data sets. As such, it is important to assess whether these observations or other influential cases are seen to exert an undue leverage on the output estimates, especially as a number of (non-linear) approximations inferred were determined by a limited subset of training patterns.

For such purposes, the C_i^2 and H_i^2 measures defined in section 5.8 were used to infer whether particular data pairs attained a high influence. From the index and proportional influence plots[15] obtained for a subset of the models determined in section 9.6.2 for both C_i^2 and H_i^2 influence measures are shown in Figures 9.19 to 9.22. In several of the data sets a small subset (in some cases a single observation) of the training cases exhibit a high influence in comparison to the remainder of the training data. The proportional influence plots show that the majority of these influential cases are associated with large residuals, although a plot of the leverage values (diagonal elements of the **S** matrix) in terms of their corresponding C_i^2 values show that in in some cases (e.g. Figures 9.23 (a) and (b)), the largest leverages (s_{ii}) correspond to the largest C_i^2's.

These results imply that in the present data sets, influential cases are to a certain degree outliers in both the X and Y spaces and that the identification of the latter is generally less problematic than the former (although the conditional structure of the model can considerably affect results). To futher understand whether influential cases could be associated with particular sub-networks and possibly be induced by excessive degrees of freedom, the modified residuals were displayed reflecting their corresponding influence levels. For the most simple of cases, these confirmed the nature of the influence as can be seen in Figures 9.24 (a) and (b), where the influential case is seen to be a clear Y-space outlier. Figures 9.24 (c) and (d), highlight the problematic elucidation by simple residual plots of high leverage cases for non-trivial outliers.

The modified residuals shown in Figure 9.24 (e), corresponding to the Fe subnetwork for the model determined in section 9.6.2 for the BAP1TLQ3W80 data set, show a high influence associated with the low-Fe cases. The high influence associated with these observations suggests that the network weights associated with the respective basis functions are strongly influenced by these data pairs. It is interesting to note the one non-influential observation surrounded by the high influence cases. Although this univariate plot provides only a partial understanding of the distribution of the data and conditional model match, inspection of the other subnetwork responses (and corresponding modified residuals) suggest a masking effect from the various high influence points surrounding this one low influence case (see section 4.5).

[15] As in previous analyses, in the proportional influence plots, the relative magnitude of the influence of the observations is displayed in terms of the size of the dots, larger dots corresponding to increasingly higher influential data pairs.

In Chapter 7, it was seen that a small subset of the training patterns induced a piecewise linear fit but upon an influence analysis, were seen to be non influential cases. In contrast, the influence analysis on the BAP1TLQ3W80 data set, performed on a model exhibiting a third order (quadratic) Fe-subnetwork, identifies a significant influence associated with a set of observations where the data is sparse. These results somewhat illustrate the higher susceptibility to overfitting in regions of the input space where the data is poorly distributed upon using more complex approximating functions.

Although both influence measures (C_i^2 and H_i^2) are typically seen to agree in terms of identifying the same influential subsets, a comparison between the two shows a disagreement in the ranking of the observations in terms of their influence. Compared with Hadi's overall potential measure (which as discussed in Chapter 7) which attributes a high influence to both (Y-space) outliers and high leverage observations, Cook's distance measure attributes a higher influence to high leverage cases. This can be seen in the proportional influence plots shown in Figures 9.19 (c) and (d).

A further assessment as to whether particular training patterns exerted an undue influence in the model construction and hence in the final model structures determined can be obtained using a leave-one or m out strategy. A comparison with the results obtained from the above influence measures would then provide a further understanding of which influential cases result in the greatest influence in the model construction. Such exhaustive data analysis is however beyond the scope of the present thesis.

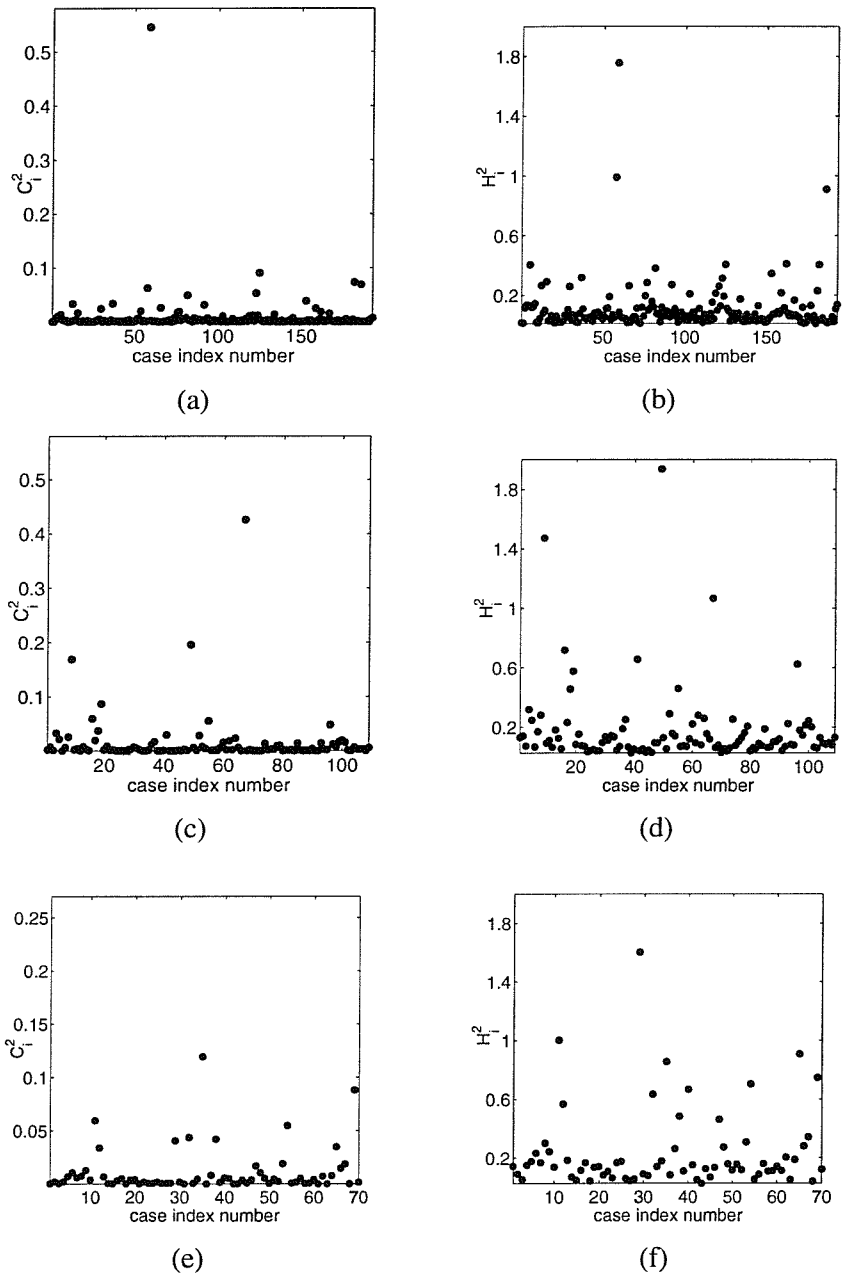


Figure 9.19: Index influence plots obtained for Cook's (C_i^2) and Hadi's (H_i^2) measures of influence corresponding to the neurofuzzy models (1-2 order B-splines) identified for the 7175 Al-alloy data sets: BAP1TLC3W (a)-(b), BAP1TLQ3W (c)-(d) and BAP1TLQ3W80 (e)-(f).

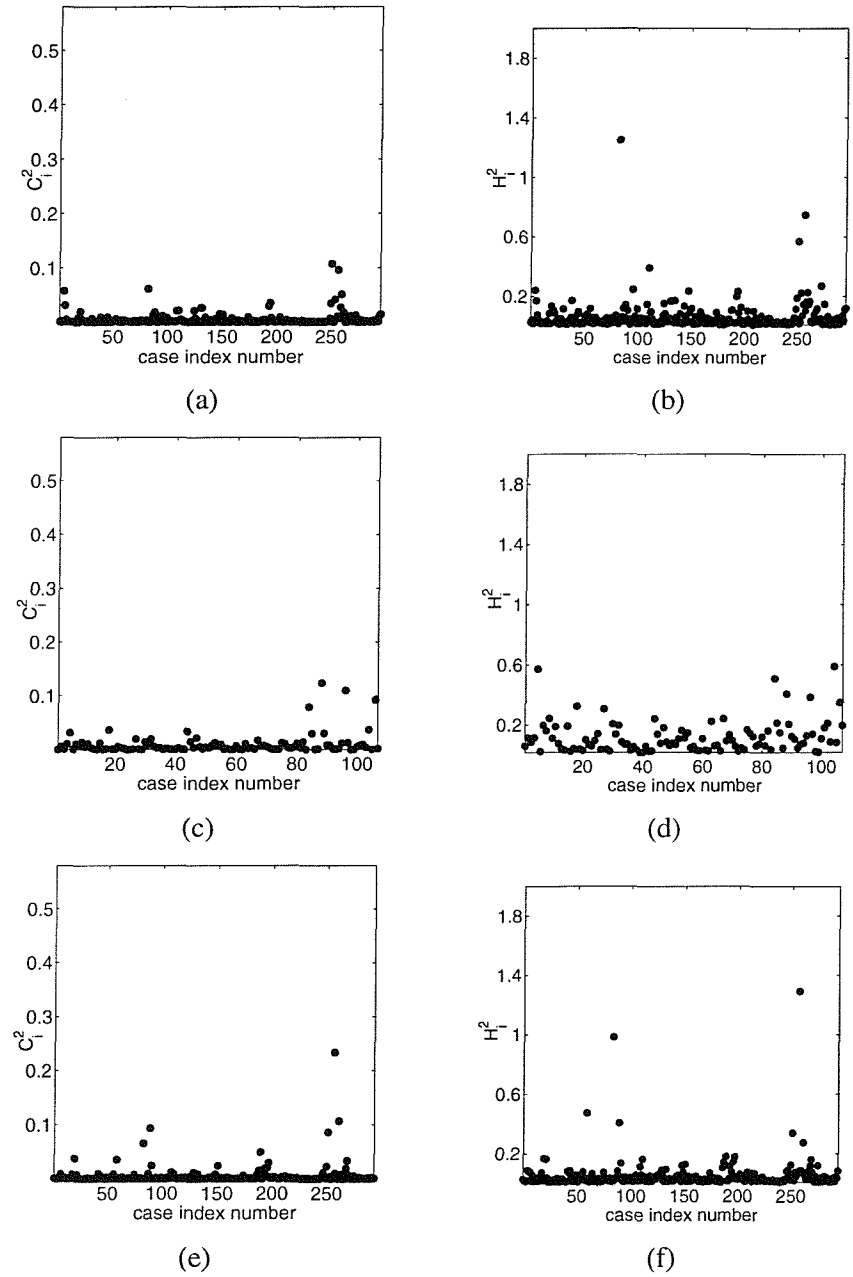


Figure 9.20: Index influence plots obtained for Cook's (C_i^2) and Hadi's (H_i^2) measures of influence corresponding to the neurofuzzy models (1-3 order B-splines) identified for the 7475 Al-alloy data sets: BAP4LTC5W (a)-(b), BAP4SLC5W (c)-(d) and BAP4TLC5W (e)-(f).

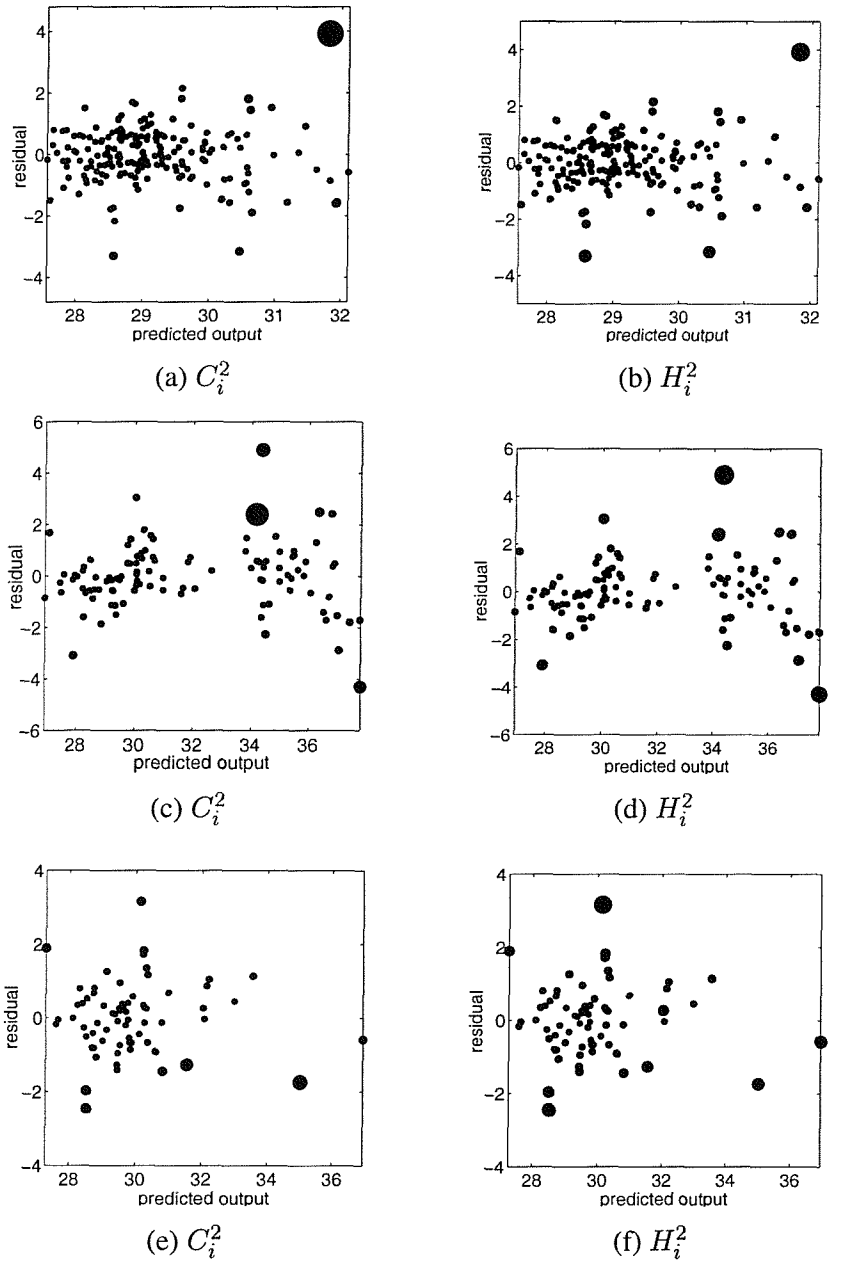


Figure 9.21: Proportional influence plots corresponding to the neurofuzzy models (1-3 order B-splines) identified by the ASMOD algorithm for the 7175 data sets: BAP1TLC3W (a)-(b), BAP1TLQ3W (c)-(d) and BAP1TLQ3W80 (e)-(f).

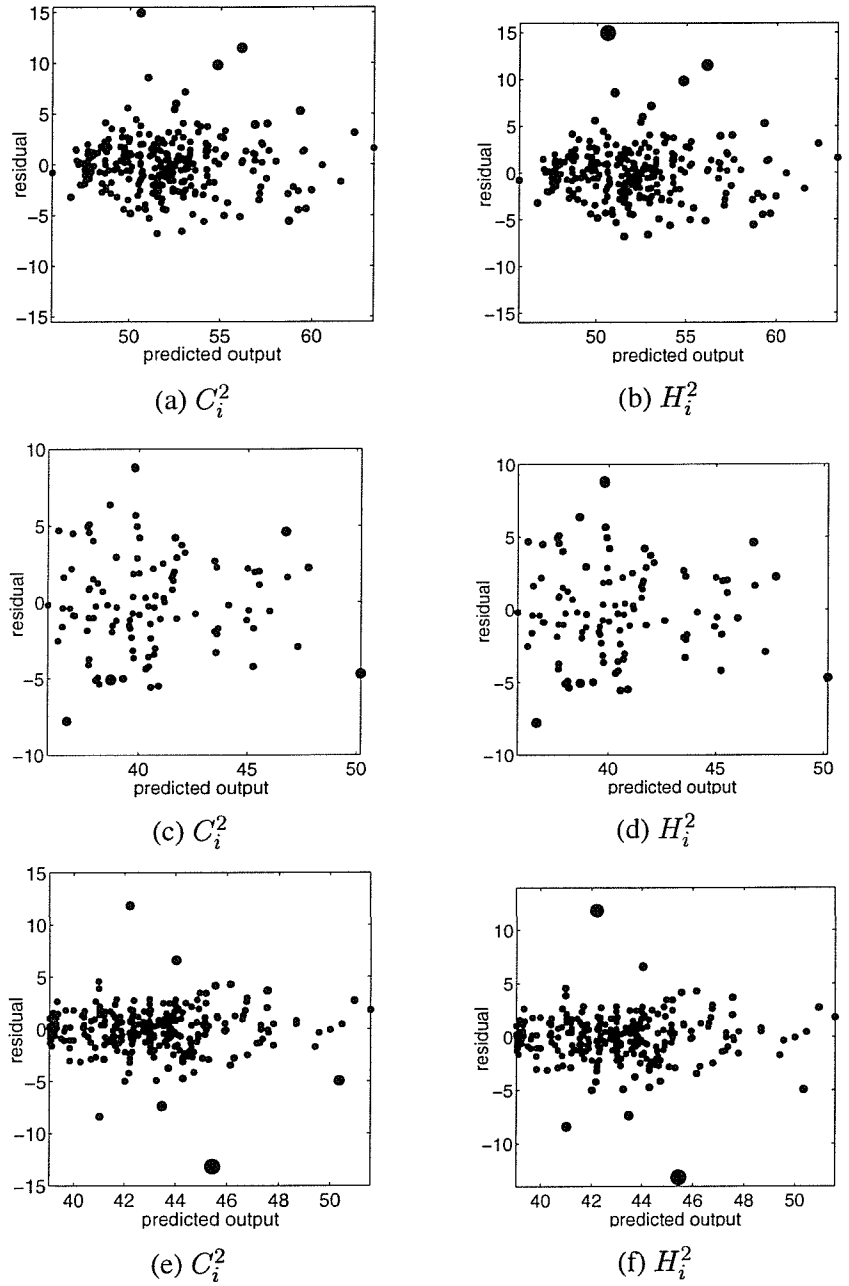


Figure 9.22: Proportional influence plots corresponding to the neurofuzzy models (1-3 order B-splines) identified by the ASMOD algorithm for the 7475 data sets: BAP4LTC5W (a)-(b), BAP4SLC5W (c)-(d) and BAP4TLC5W (e)-(f).

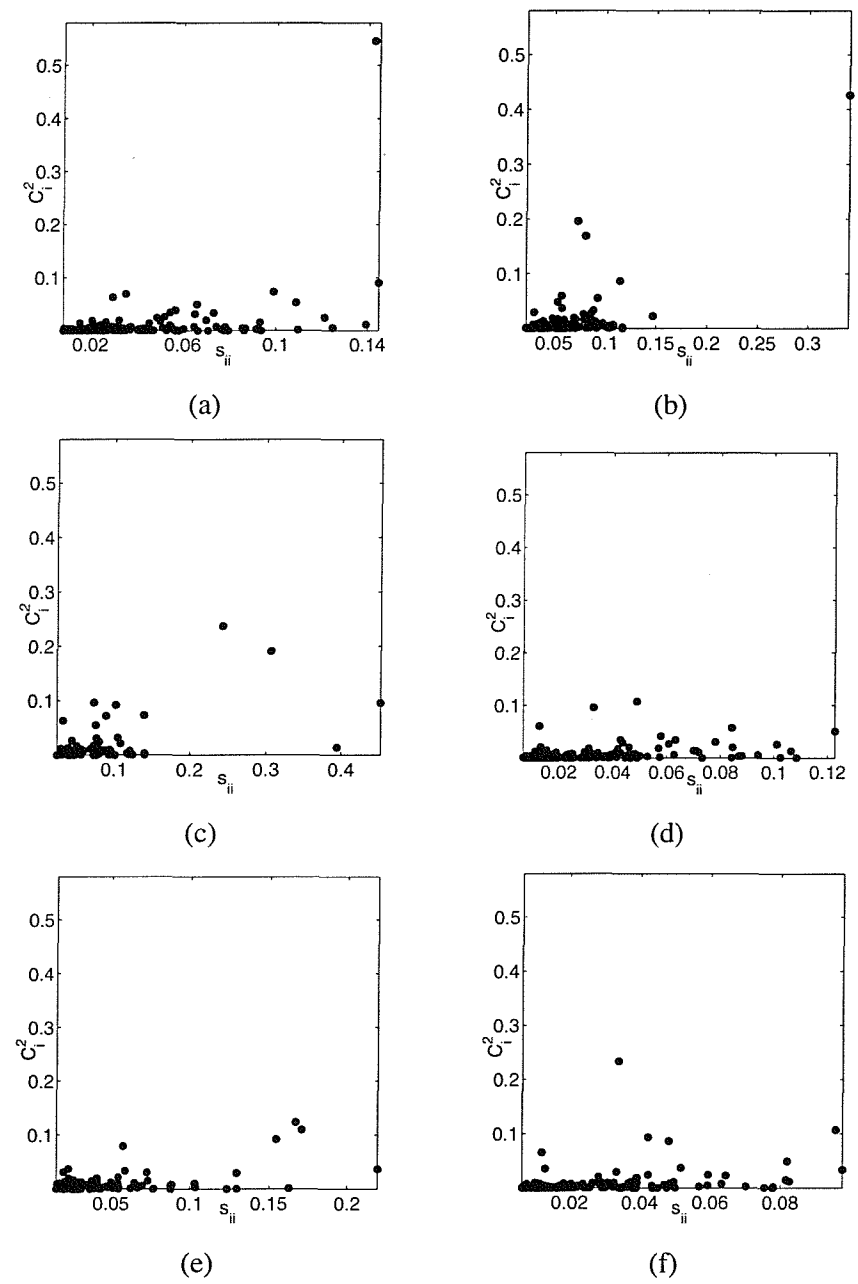
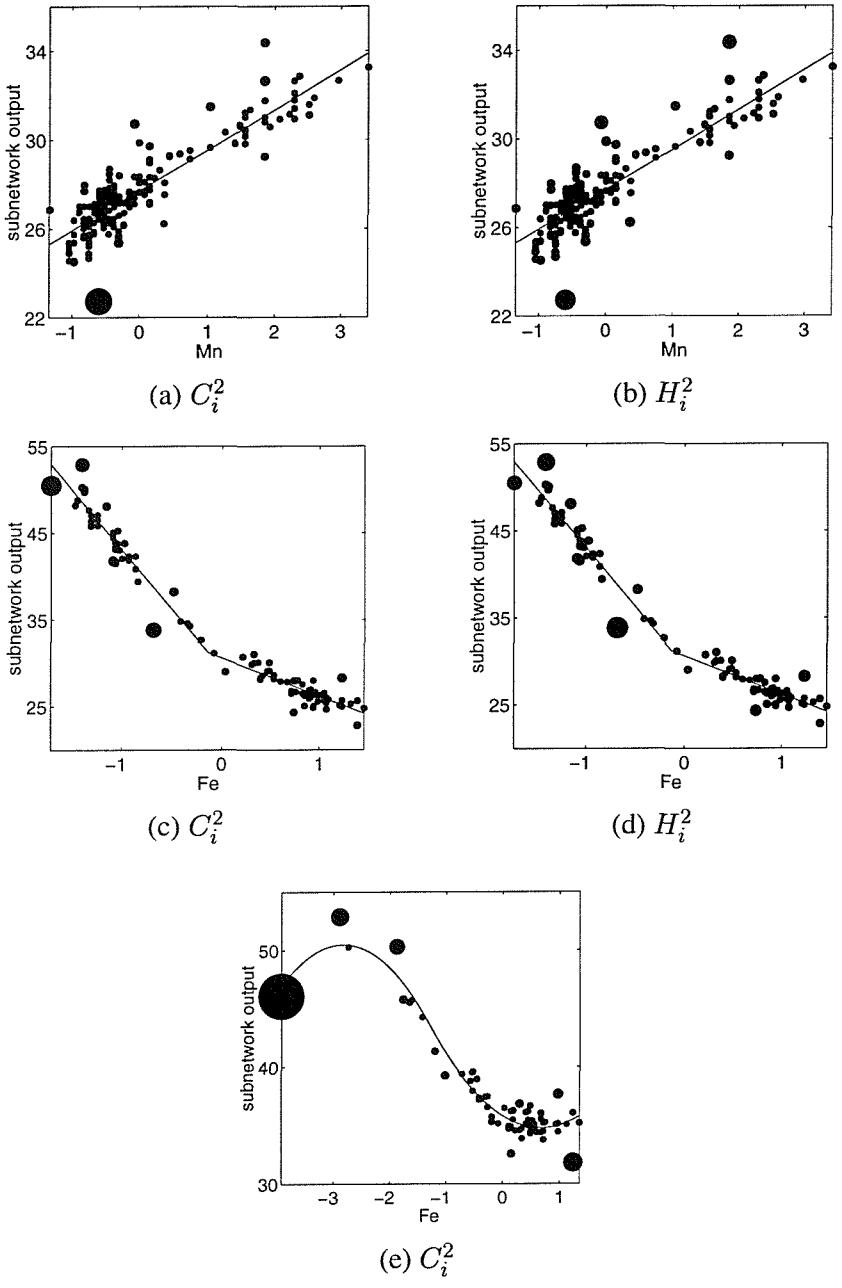


Figure 9.23: Leverage (s_{ii}) versus C_i^2 plots for the neurofuzzy models determined using second order (a)-(c) and third order (d)-(f) basis functions: (a) BAP1TLC3W, (b) BAP1TLQ3W, (c) BAP1TLQ3W80, (d) BAP4LTC5W, (e) BAP4SLC5W, (f) BAP4TLC5W.



(a) C_i^2 (b) H_i^2

(c) C_i^2 (d) H_i^2

(e) C_i^2

Figure 9.24: Subnetwork responses and corresponding partial residuals plotted in order to reflect the magnitude of the measures of influence used: (a)-(b) BAP1TLC3W and (c)-(d) BAP1TLQ3W, (e) BAP1TLQ3W80.

9.6.5 Regularised networks

Within the neurofuzzy model construction framework, redundancy and, more generally, the inclusion of excessive degrees of freedom is prevented by constraining the basis functions used to a low (second) order. More generally redundancy is controlled by use of the pruning refinements during the model construction. However, although the results obtained on the BAP data sets are in general seen to determine broadly plausible models employing third order basis functions, the FS/BE construction algorithms (and consequently the SS measure) in some instances were seen to unsuccessfully prevent models exhibiting an undue flexibility in poorly covered regions of the input space and as a result exhibiting a local overfitting behaviour on a subset of the training samples (e.g. Figure 9.15 (b), or 9.24 (e)). As discussed in Chapter 3, both maximum likelihood estimation and subset selection procedures may exhibit inferior generalisation capabilities compared with regularisation approaches.

The results obtained from the FS/BE model construction procedures show that only a subset of the inputs present are included in the final models for K_{Ic}, and although no subnetwork deletion refinements are performed, redundancy is controlled by a series of knot deletion and order reduction refinements. In certain instances however, pruning may result in rather simplistic model structures, which may exhibit a considerable bias component in the generalisation error. Parameters in linear unregularised models are optimised by minimising the quadratic cost function, J_N, corresponding to maximum likelihood estimation. This can reveal how a small subset of "greedy" variables may explain a higher variance in the output than they are truly accountable for, which will then compromise the generalisation performance of the model. In addition, a considerably biased representation of the system will be inferred. In view of these considerations and the control of redundancy in a maximum likelihood and subset selection framework, Bayesian regularisation was investigated (in a limited way) as an alternative approach to control excessive degrees of freedom in over-parameterised models.

As model parsimony was still desired, the present work proposed the use of model building refinements to determine a subset of inputs and an appropriate model structure which was then regularised. Modifying the model construction pass structure employed in previous analyses to include only model identification/building refinements led to the following two pass structure:

pass	
	univariate addition
	tensor product
	tensor split
pass	
	knot insertion

where the tensor split (model pruning) refinement was retained so that the construction procedures were given the flexibility to refine from a model structure including a tensor product

term.

A second order regulariser in the form discussed in section 5.5 was used, for which the values of the α and β hyperparameters were optimised using the re-estimation formulae given by equations 5.28 and 5.27, respectively. In order to allow different regularisers to be determined for each subnetwork, a local regularisation approach was implemented. In view of the similarities characterising the data sets and iterative model constructions determined in previous sections, regularisation was investigated for only three data sets, corresponding to those for which the model pruning refinements performed the largest number of refinements, and as such, corresponding to the models which retained the largest number of degrees of freedom and the greatest redundancy (as determined by the pruning methods). It should be however understood that these results do not necessarily correspond to the largest amount of regularisation performed nor the greatest difference in approximation abilities (e.g. training and test set performance), since different regularisation coefficients will be determined on the various different data sets.

To allow a comparison between the results obtained in sections 9.6.2 and 9.6.3, models determined from the FS + regularisation strategy were initially obtained for the BAP1TLC3W and BAP4LTC5W data sets using all data for training. This allowed the use of a second order regulariser to be investigated on networks comprising both second and third order basis functions. Subsequently, in order to determine whether the FS + regularisation entailed an improved generalisation performance compared to the FS/BE methodology, training and test set variance estimates were determined over multiple model runs (20) determined from the same training-test set partitions as were used in previous sections. In addition to the BAP1TLC3W and BAP4LTC5W data sets, the empirical performance on the BAP4TLC5W was also assessed.

The model structures determined for the BAP1TLC3W and BAP4LTC5W data sets from the two pass structure defined above are shown in Figures 9.25 and 9.26, while Figures 9.27 and 9.28 show the corresponding subnetwork responses.

A comparison with the networks identified by the FS/BE model construction approach shows a difference in terms of both the order and number of basis functions: in the 7175 model, the pruning refinements remove all six (internal) knots in the model, whilst for the 7475 model, in addition to removing one knot (present at the same position on the input axis) in each of the subnetworks, the B-spline order on the Fe and CRS axis were reduced to second order basis functions.

The form of the approximations obtained after the networks have been penalised with the second order regulariser are shown in Figures 9.29 and 9.30[16]. Table 9.14 summarises the values of the hyperparameters inferred and the training errors of the unregularised and regularised models obtained for the BAP4LTC5W data set. A comparison between the unregularised and regularised network responses shows how the different regularisers inferred

[16]The different offsets in these figures are not significant as they represent the individual bias for each subnetwork.

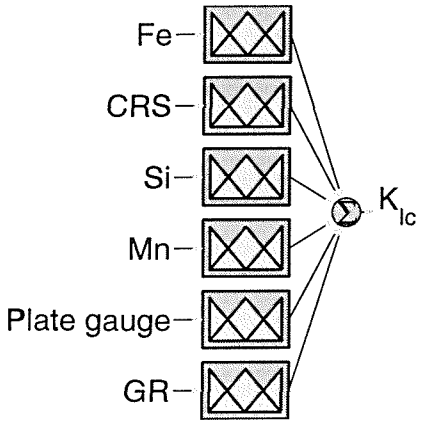


Figure 9.25: Neurofuzzy network structures determined from training on the BAP1TLC3W data set employing a forward selection constructive set of refinements.

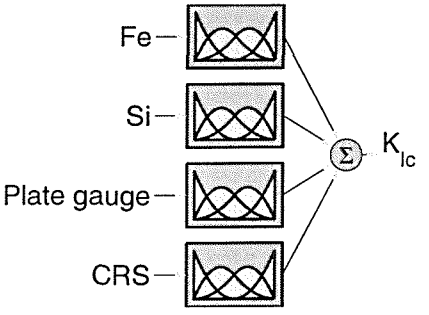


Figure 9.26: Neurofuzzy network structures determined from training on the BAP4LTC5W data set employing a forward selection constructive set of refinements.

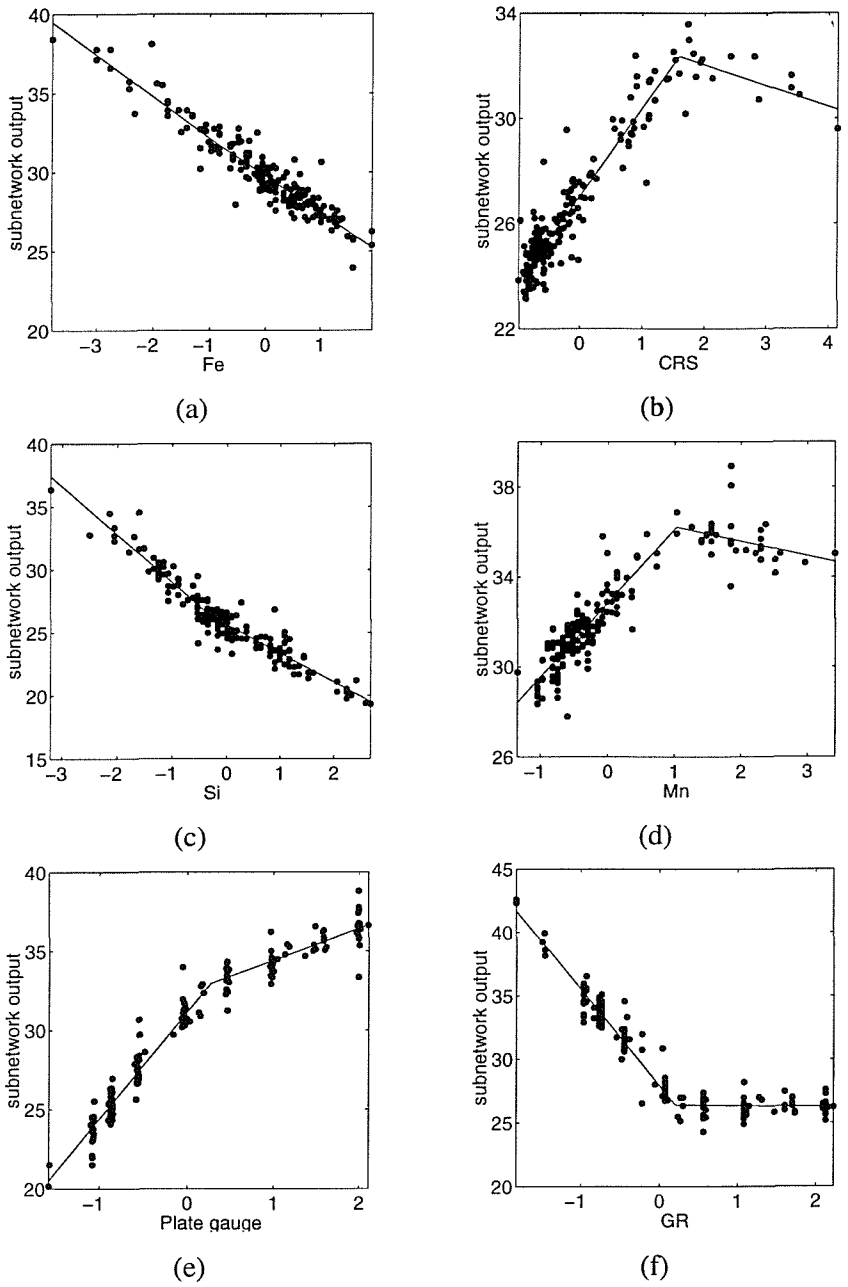


Figure 9.27: Subnetwork response outputs corresponding to the network structure shown in Figure 9.25

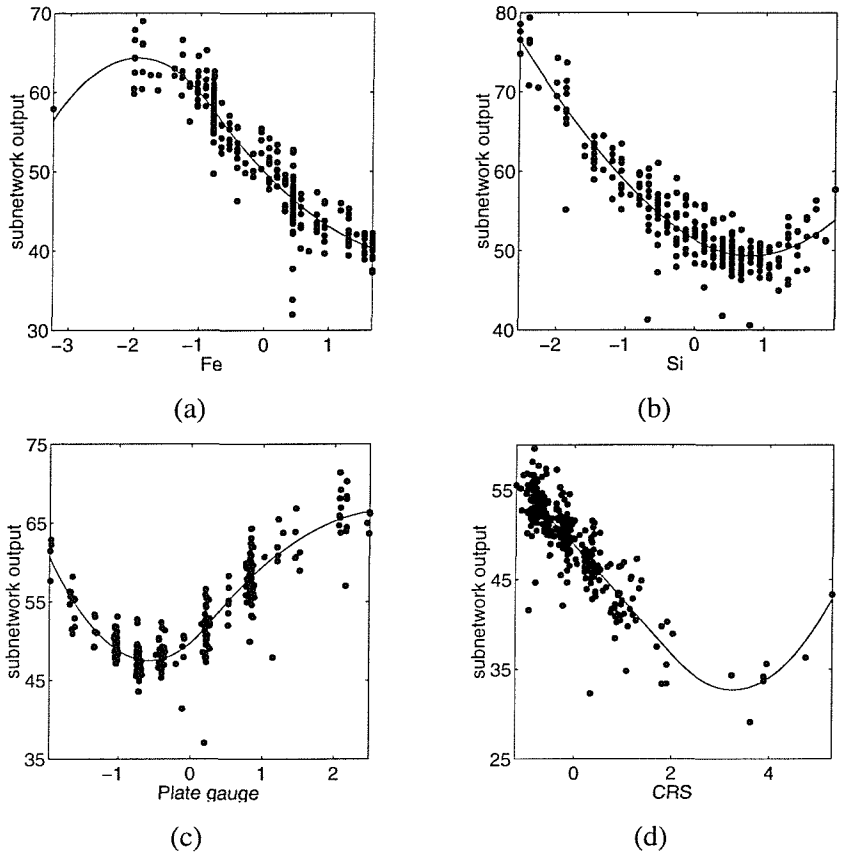


Figure 9.28: Subnetwork response outputs corresponding to the network structure shown in Figure 9.26.

subnetwork output curvatures, e.g. see the distinct change in slope at low Fe content in Figure 9.28 (a) for the unregularised model, as opposed to the same region of Figure 9.30 (a) for a regularised model.

For the BAP1TLC3W model, the same amount of regularisation was inferred for the different subnetworks, as can be seen from the (linearised) subnetwork responses.

β	α_{Fe}	α_{Si}	$\alpha_{\text{Plate gauge}}$	α_{CRS}	σ_N^2	$\sigma_{N,\text{reg.}}^2$
0.15	$1e^{-3}$	$2e^{-3}$	0.00	$2.8e^{-3}$	6.35	6.42

Table 9.14: Summary of the value of the hyperparameters and variance estimates determined for the BAP4LTC5W data set.

A summary of the results obtained from the multiple model runs is shown in Table 9.15, which may then be compared with Table 9.12. This shows that while an improvement in the generalisation performance on all data sets is attained, the greatest improvement is seen to be on the two 7475 data sets particularly when regularisation is employed to constrain the output curvature of third order B-spline subnetworks.

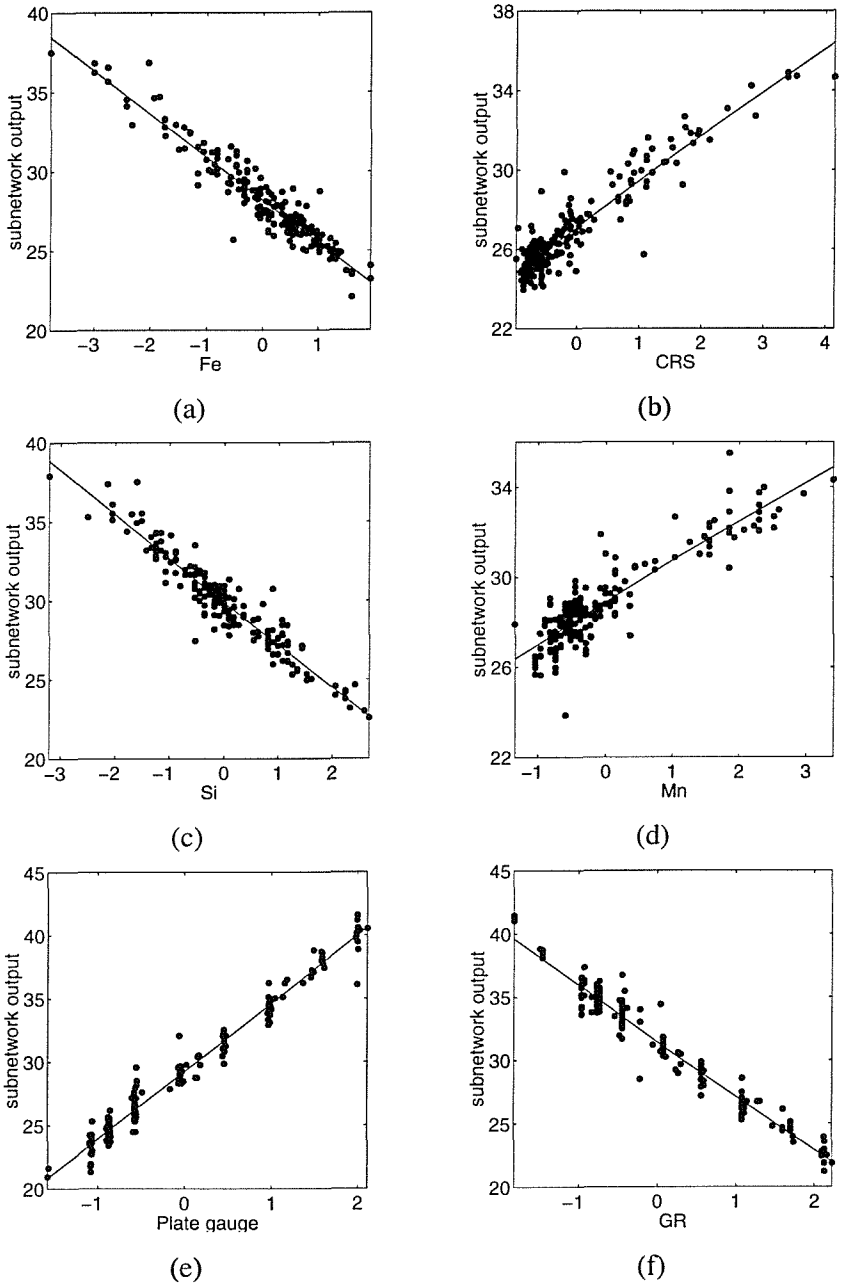


Figure 9.29: Subnetwork response outputs corresponding to the network structure shown in Figure 9.25 after second order regularisation.

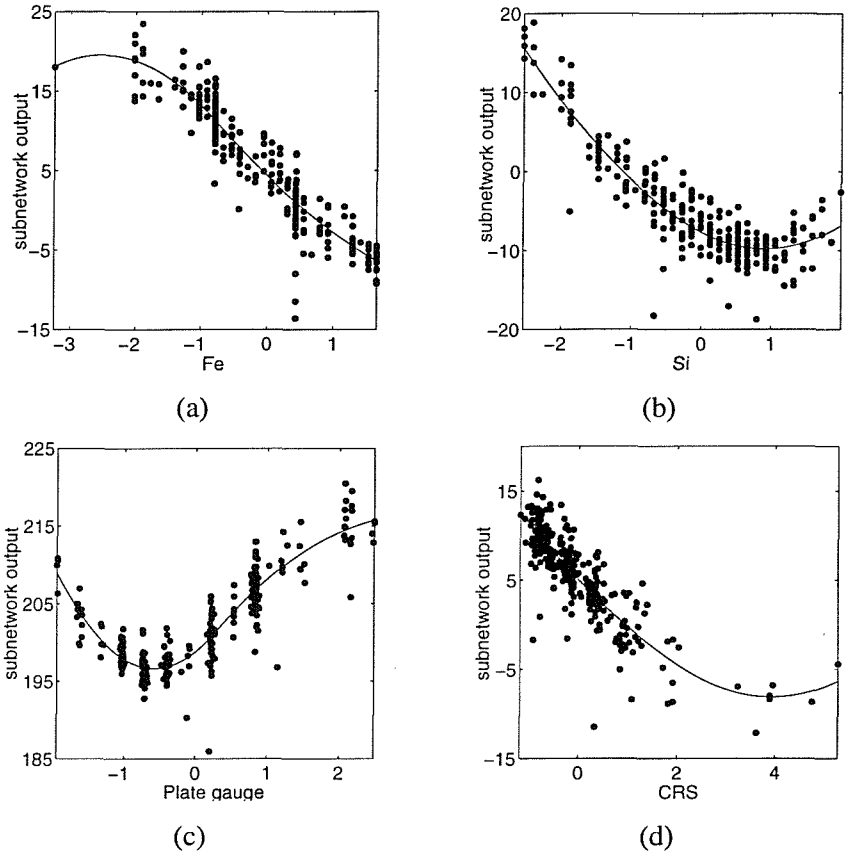


Figure 9.30: Subnetwork response outputs corresponding to the network structure shown in Figure 9.26.

Data set	B-Spline order	Training error μ	Test error μ
BAP1TLC3W	2	0.90	1.23
BAP4LTC5W	2	7.68	7.14
	3	6.59	5.78
BAP4TLC5W	2	4.81	4.07
	3	4.34	3.38

Table 9.15: Summary of the empirical results obtained for the FS + regularisation framework. The average errors were obtained in terms of the $\hat{\sigma}_N^2$ attained on each data set instance over the multiple model runs.

9.7 Support vector regression

In view of the lack in improved approximation abilities exhibited by the neurofuzzy modelling approach compared with the MLR analysis, it was of interest to assess whether the SVM methods could offer any improved performance.

Since the large number of variates would have resulted in a very large, high dimensional basis expansion in the first stage of the SUPANOVA framework, this was limited to include only interactions that could easily be visualised (i.e. up to trivariate terms). This also allowed the SVM models to be trained in more acceptable times. Thus, SUPANOVA models were determined by limiting the sparse selection to be performed on a full basis expansion comprising univariate, bivariate and trivariate terms.

The presence of a number of outliers identified in previous sections presented an opportunity to compare the performance of the SVM using different loss functions. As such, in addition to the quadratic loss function, used throughout this work, SVM were also trained using the Laplace loss function. Although the ϵ-insensitive zone did not introduce any sparseness to the solution ($\epsilon = 0$), in terms of the number of support vectors, the use of the Laplace loss provided an opportunity to assess whether any robustness was gained and reflected in the empirical results.

A Stage II sparse selection threshold (or tolerance level), S2tol $= 0.4$, as used in previous chapters, was seen in Chapter 8 to give rise to a number of SUPANOVA models comprising only of a bias term. In modelling the BAP data using this same threshold, particularly when larger ANOVA basis were considered (i.e. including bivariate and trivariate terms), a comparable if not larger number of models comprising only a bias term were produced.

Although the training set variance estimates attained by both the MLR and neurofuzzy results suggested a high output noise component, the dependencies inferred by the neurofuzzy model construction algorithms showed that a subset of the inputs exhibited structural relationships with K_{Ic}. Therefore SUPANOVA models which exhibited only a bias term on certain training sets were unsatisfactory, which will be discussed in more detail in section 9.9.

Adjusting the S2tol level, to some arbitrary value will determine the subset selection and hence the sparseness (in terms of the number of ANOVA components) of the final models. Thus the S2tol value can be seen to be a tunable parameter in the SUPANOVA framework, effectively determining the parsimony of the final models. To assess whether the number of models comprising only a bias term could be reduced by lowering the Stage II threshold, this was subsequently set to S2tol $= 0.01$. As in previous chapters an infinite ANOVA spline kernel was used to perform the non-linear mapping into feature space.

The SVM modelling approach was limited to the same 7x75 data set partitions defined in Table 9.5, and in order to allow a comparison with previous sections, the same 20 training-test set data splits used in the MLR and neurofuzzy analyses were used to infer the empirical performance of the SVM models. For brevity, the results have been presented in concise for-

mats, summarising the empirical performance inferred from the multiple model runs, whilst the nature of the dependencies determined are limited to those identified with the greatest stability.

As in previous chapters, the performance of the SUPANOVA approach was compared with that of a support vector regression comprising a full ANOVA spline kernel basis expansion, i.e. a model consisting of the product of piecewise cubic splines with the knots located at the training data points, corresponding to the Stage I model.

The empirical performance attained by Stage I models on the 7x75 data sets is summarised in Table 9.16, for models determined from using the quadratic (L_{quad}) and Laplace (L_ϵ) loss functions. From these results it is seen that the (average) training set variance estimates are considerably lower than those attained on unseen data, suggesting a significant degree of over-fitting. Unsurprisingly then, a comparison with the MLR results show how the simple linear models attain a better generalisation performance.

A further comparison in terms of the results attained from different loss functions shows how the models determined from using the L_ϵ loss are generally seen to exhibit higher average training error (μ), measured in terms of the quadratic loss[17]. While exhibiting similar variances (σ^2) with the L_{quad} results on the test sets, models determined from the L_ϵ loss function are seen to exhibit a considerably higher variance in the training errors.

	L_{quad}				L_ϵ			
	μ		σ^2		μ		σ^2	
Data set	Training	Test	Training	Test	Training	Test	Training	Test
BAP1TLC3W	0.22	1.87	$1.6e^{-3}$	0.56	0.20 (0.21)	1.05 (2.18)	$2.6e^{-3}$ (0.01)	0.03 (0.63)
BAP1TLQ3W	0.37	4.12	$8.2e^{-3}$	7.26	0.19 (0.46)	1.45 (4.15)	$7e^{-3}$ (0.05)	0.22 (7.59)
BAP1TLQ3W80	0.33	6.38	0.06	26.42	0.17 (0.55)	1.86 (6.45)	0.02 (0.46)	0.30 (23.50)
BAP4LTC5W	3.18	7.13	0.23	12.13	0.89 (4.00)	1.84 (7.73)	0.02 (1.14)	0.17 (14.32)
BAP4SLC5W	3.75	17.28	0.48	31.95	0.90 (3.66)	3.17 (17.35)	0.06 (1.09)	0.33 (56.40)
BAP4TLC5W	2.06	5.07	0.05	7.51	0.73 (2.84)	1.42 (5.06)	$8e^{-3}$ (0.30)	0.05 (8.64)

Table 9.16: Empirical performance inferred over multiple model runs corresponding to the Stage I models in the SUPANOVA framework determined from different training-test set partitions.

The training error statistics corresponding to the SUPANOVA models inferred from the

[17]In the following tables, summarising the empirical performance (training and test set error estimates) exhibited by the SUPANOVA and Stage I models, error estimates corresponding to models trained using the L_ϵ loss function are also measured in terms of the quadratic loss (L_{quad}). This allows a direct comparison between models trained using different loss functions.

multiple models, obtained using the two different Stage II sparse selection thresholds (0.4 and 0.01) and the L_{quad} loss function are summarised in Table 9.17.

Data set	Stage II tol.	Univariate basis		Bivariate basis		Trivariate basis	
		μ	σ^2	μ	σ^2	μ	σ^2
BAP1TLC3W	0.4	0.89	$1.4e^{-3}$	1.16	0.11	1.39	0.09
	0.01	0.74	$6.6e^{-3}$	0.71	0.02	1.12	0.09
BAP1TLQ3W	0.4	2.14	0.26	1.75	0.45	1.65	0.26
	0.01	1.04	0.12	1.24	0.20	1.25	0.18
BAP1TLQ3W80	0.4	1.45	0.16	1.50	0.50	2.58	0.93
	0.01	1.05	0.02	1.19	0.18	1.94	1.19
BAP4LTC5W	0.4	6.88	0.15	6.29	0.42	6.27	0.58
	0.01	6.24	0.17	6.08	0.24	6.10	0.44
BAP4SLC5W	0.4	10.10	2.53	9.24	2.22	10.40	6.17
	0.01	7.88	3.47	8.36	2.43	9.55	2.14
BAP4TLC5W	0.4	3.98	0.06	3.61	0.33	3.53	0.40
	0.01	3.54	0.14	3.05	0.08	3.24	0.35

Table 9.17: Training error variance estimates for the SUPANOVA models obtained from using a quadratic error function.

For the smaller of the data sets, a comparison of the average variance estimates (μ) shows that models determined from univariate basis expansions attain a better approximation to the training data when compared to models inferred from larger ANOVA basis (i.e bivariate and trivariate basis). This is primarily attributable to the presence of a number of models comprising only a bias term.

A comparison of these results in terms of the Stage II tolerance values used, shows that the use of the lower S2tol is reflected by a better approximation to the training data, resulting from a larger number of terms subselected and consequently a smaller number of models comprising only a bias component[18].

The generalisation performance of the SUPANOVA models are summarised in Tables 9.18 to 9.20 for the different loss functions and S2tol values used.

It is apparent that in instances for which the generalisation estimates exhibit a large variance, little confidence can be placed in the estimates, and comparisons between different training conditions should proceed acknowledging this high uncertainty.

With a few exceptions, it is seen that allowing a larger number of ANOVA terms to be included in the final models results in an improved generalisation performance. It remains difficult however to assess the generalisation performance exhibited by the SUPANOVA framework, since the presence of models comprising only bias components will bias empirical

[18]It should be noted that the regularisation coefficient will be the same, since C is determined in the first stage (Stage I) of the SUPANOVA framework.

Data set	Stage II tol. (0.4)				Stage II tol. (0.01)			
	L_{quad}		L_ϵ		L_{quad}		L_ϵ	
	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2
BAP1TLC3W	1.26	0.21	0.83 (1.33)	0.02 (0.39)	1.06	0.18	0.78 (1.07)	0.01 (0.17)
BAP1TLQ3W	3.44	3.25	1.32 (3.28)	0.09 (2.36)	3.22	4.80	1.16 (2.60)	0.11 (2.47)
BAP1TLQ3W80	2.64	4.58	1.11 (2.61)	0.12 (5.63)	1.68	1.55	0.99 (2.20)	0.13 (5.27)
BAP4LTC5W	6.71	6.54	1.80 (5.95)	0.08 (5.98)	5.63	6.77	1.72 (5.55)	0.06 (5.67)
BAP4SLC5W	14.10	33.71	3.06 (14.53)	0.36 (24.82)	14.71	21.76	3.11 (15.26)	0.34 (30.38)
BAP4TLC5W	3.26	3.26	1.24 (3.09)	0.04 (3.01)	3.20	3.33	1.21 (3.04)	0.04 (3.05)

Table 9.18: Empirical performance inferred over multiple model runs for the SUPANOVA framework determined from different training-test set partitions and constrained to form the ANOVA basis from univariate terms. Results are shown using two different sparse selection threshold values and loss functions.

Data set	Stage II tol. (0.4)				Stage II tol. (0.01)			
	L_{quad}		L_ϵ		L_{quad}		L_ϵ	
	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2
BAP1TLC3W	1.60	0.37	0.94 (1.66)	0.03 (0.63)	1.05	0.24	0.81 (1.23)	0.02 (0.40)
BAP1TLQ3W	4.21	3.65	1.38 (3.80)	0.17 (5.36)	3.47	2.78	1.26 (3.17)	0.13 (3.30)
BAP1TLQ3W80	4.37	20.95	1.12 (3.74)	0.48 (22.81)	3.08	16.46	1.18 (3.75)	0.39 (22.16)
BAP4LTC5W	7.09	7.47	1.89 (6.46)	0.10 (9.85)	6.78	9.91	1.84 (6.72)	0.09 (10.01)
BAP4SLC5W	14.25	38.24	3.13 (15.63)	0.49 (7.40)	13.65	25.19	3.07 (15.50)	0.39 (38.18)
BAP4TLC5W	3.53	1.49	1.35 (3.61)	0.05 (3.05)	3.33	1.46	1.25 (3.16)	0.05 (2.56)

Table 9.19: Empirical performance inferred over multiple model runs for the SUPANOVA framework determined from different training-test set partitions and constrained to form the ANOVA basis from univariate and bivariate terms. Results are shown using two different sparse selection threshold values and loss functions.

Data set	Stage II tol. (0.4)				Stage II tol. (0.01)			
	L_{quad}		L_ϵ		L_{quad}		L_ϵ	
	μ	σ^2	μ	σ^2	μ	σ^2	μ	σ^2
BAP1TLC3W	1.66	0.60	0.95 (1.70)	0.04 (0.70)	1.44	0.42	0.91 (1.58)	0.04 (0.69)
BAP1TLQ3W	3.98	4.60	1.43 (3.98)	0.20 (6.76)	3.69	2.42	1.38 (3.77)	0.14 (3.89)
BAP1TLQ3W80	4.79	19.67	1.45 (4.85)	0.30 (22.32)	4.52	20.90	1.40 (4.71)	0.38 (23.02)
BAP4LTC5W	6.66	12.25	1.81 (6.19)	0.09 (10.06)	6.95	10.41	1.76 (6.29)	0.13 (13.51)
BAP4SLC5W	14.90	39.40	3.11 (15.25)	0.45 (35.39)	13.77	39.24	2.97 (14.79)	0.51 (37.24)
BAP4TLC5W	3.32	2.31	1.32 (3.34)	0.03 (1.94)	3.37	1.64	1.23 (3.22)	0.04 (1.96)

Table 9.20: Empirical performance inferred over multiple model runs for the SUPANOVA framework determined from different training-test set partitions and constrained to form the ANOVA basis from univariate, bivariate and trivariate terms. Results are shown using two different sparse selection threshold values and loss functions.

comparisons, both with the MLR and neurofuzzy results but also between the SVM models determined from different basis expansions and loss functions considered. The effect of considering larger basis expansions and different Stage II thresholds on the number of ANOVA terms present in the final models is summarised in Tables 9.21 and 9.22.

The consistency with which particular input variables are selected in the SUPANOVA models and the degree of interaction with other variates, can be appreciated from Tables 9.23 to 9.25, summarising the number and type of ANOVA components inferred across the multiple model runs. These show that in results obtained from models determined from a univariate ANOVA expansion, contributions of Fe and Si are present in most of the 7175 and 7475 data sets.

Similarly to the results obtained in previous chapters, approximations inferred on the different training-test splits exhibit differing degrees of regularisation, with some terms over-regularised while others under-regularised. Again, this is seen to result from the use of a global smoothing parameter C, whereby in addition to being inferred at Stage I and hence from the complete set of inputs, some terms will be under whilst others over-regularised.

Data set	Loss function	Stage II tol	Basis	Number of ANOVA terms											
				Bias	1	2	3	4	5	6	7	8	9	10	11
BAP1TLC3W	L_{quad}	0.4	*uni*		1	1	12	5	1						
			biv	3	6	5	4	1	1						
			tri	11	3	1	4	1							
		0.01	*uni*					2	11	7					
			biv					7	3	4	5	1			
			tri	3	2	4	2	6	2	1					
	L_{ϵ}	0.4	*uni*	1	1	2	16								
			biv	3	9	5	3								
			tri	14	3	2	1								
		0.01	*uni*					8	11	1					
			biv	1		1	2	6	3	2	3	1	1		
			tri	5	1	8	3	1	2						
BAP1TLQ3W	L_{quad}	0.4	*uni*		5	10	2	3							
			biv				9	6	3	1		1			
			tri			3	9	5	3						
		0.01	*uni*			4	7	9							
			biv						5	4	4	5		1	1
			tri					3	9	4	1	3			
	L_{ϵ}	0.4	*uni*		5	7	7	1							
			biv				1	9	9	1					
			tri				11	8	1						
		0.01	*uni*			2	2	15	1						
			biv							4	8	3	3	2	
			tri				1	3	3	6	4	1	2		
BAP1TLQ3W80	L_{quad}	0.4	*uni*		8	7	4								
			biv	5	7	7	1								
			tri	11	5	4									
		0.01	*uni*				19		1						
			biv	2	1	1		15	1						
			tri	8	3		7		1						
	L_{ϵ}	0.4	*uni*	2	12	4	2								
			biv	6	2	12									
			tri	16	3	1									
		0.01	*uni*	2	2	6	4	1	1						
			biv	4	1		3	9	1		1	1			
			tri	11	1	4	3	1							

Table 9.21: Summary of the number of ANOVA terms determined across the multiple model runs for the SUPANOVA framework for the 7175 data sets from using different loss functions, Stage II thresholds and basis expansions.

Data set	Loss function	Stage II tol	Basis	Number of ANOVA terms: Bias	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
			uni				20												
	L_{quad}	0.4	*biv*				1	5	7	6	1								
			tri			9	4	1	5	1									
			uni				1	19											
		0.01	*biv*									3	4	1	6	5			
BAP4LTC5W			*tri*				6	2	1	5	4		2						
			uni				5	15											
	L_{ε}	0.4	*biv*				1	2	3	3	4	5	2						
			tri				4	4	4	1	3	2	2						
		0.01	*uni*					16	2	1	1								
			biv											1	4	2	3	8	2
			tri					1	2	1	1	2	1	1	1	2	6	1	
			uni		2	18													
	L_{quad}	0.4	*biv*		3	12	5												
			tri	1	5	13	1												
			uni			2	6	8	4										
		0.01	*biv*			8	6	1	5										
BAP4SLC5W			*tri*		3	15	2												
			uni		3	12	4	1											
	L_{ε}	0.4	*biv*		5	3	9	3											
			tri		1	2	15	2											
		0.01	*uni*			11	4	4	1										
			biv		5	1	3	8	1	1	1								
			tri		1	1	5	10		3									
			uni			1	19												
	L_{quad}	0.4	*biv*		3	1	3	7	4		2								
			tri		4	7	4	3	2										
			uni				7	9	4										
		0.01	*biv*							1	3	2	8	5	1				
BAP4TLC5W			*tri*			2	2	2	5	1	3	1	3	1					
			uni				9	7	4										
	L_{ε}	0.4	*biv*				1	2	7	5	4	1							
			tri			1	1	4	7	3	2	1							
		0.01	*uni*				2	4	10	4									
			biv									2	7	2	4	1	2	2	2
			tri							1	2	3	5	1	3	2	3		

Table 9.22: Summary of the number of ANOVA terms determined across the multiple model runs for the SUPANOVA framework for the 7475 data sets from using different loss functions, Stage II thresholds and basis expansions.

Data set	Loss function	Stage II tol	ANOVA terms													
			Bias	Plate gauge	Stretch	Cu	Fe	Mg	Mn	Si	Zn	Ti	Cr	T_f	ST_t	GR	CRS
BAP1TLC3W	L_{quad}	0.4		3			20		3	19							18
		0.01		19			20		19	20						7	20
	L_ε	0.4	1				17			18							18
		0.01		20			20		12	20	1						20
BAP1TLQ3W	L_{quad}	0.4					20			15						3	5
		0.01					20			17						9	19
	L_ε	0.4					20			12						3	9
		0.01					20	1		20						17	17
BAP1TLQ3W80	L_{quad}	0.4				1	20			5							12
		0.01				1	20			20				1			20
	L_ε	0.4	2				18			2							6
		0.01	2				18			10		1				2	16
BAP4LTC5W	L_{quad}	0.4		20			20			20							
		0.01		20			20			20							19
	L_ε	0.4		20			20			20							15
		0.01		20			20	2		20	1		4				20
BAP4SLC5W	L_{quad}	0.4		19			3							16			
		0.01		20			18			9				18			9
	L_ε	0.4		18			15							6			4
		0.01		20		2	18			1				7			7
BAP4TLC5W	L_{quad}	0.4		19			20			20							
		0.01		20		3	20			20					13	1	
	L_ε	0.4		20			20			20					10	5	
		0.01		20			20			20			3	2	18	13	

Table 9.23: Frequency of selection for the SUPANOVA univariate models determined over the multiple model runs.

Data set	Loss Function	Stage II tol	ANOVA terms: Bias	Plate gauge	Stretch	Cu	Fe	Mg	Mn	Si	Zn	Ti	Cr	T_f	ST_t	GR	CRS
BAP1TLC3W	L_{quad}	0.4	3	13(12,1)	7(·,7)		12(·,12)			5(·,5)				7(·,8)		5(·,5)	5(5,·)
		0.01		20(19,1)	15(·,15)		20(·,41)		13(13,·)	12(·,16)	2(·,2)		1(·,1)	16(·,19)		17(·,21)	20(20,·)
	L_ε	0.4	3	4(4,·)	15(·,15)		15(·,15)			8(·,8)			2(·,2)	2(·,2)		2(·,2)	1(1,·)
		0.01	1	17(15,3)	18(1,18)	3(·,3)	19(·,25)		5(5,·)	16(·,24)	11(1,11)	1(1,·)	5(·,5)	10(·,10)		7(·,9)	19(19,·)
BAP1TLQ3W	L_{quad}	0.4		20(20,2)		2(·,2)	20(·,40)			4(·,5)	1(·,1)		2(·,3)	18(·,18)		20(10,21)	4(4,·)
		0.01		20(20,6)		11(11,6)	20(·,43)			8(·,11)	3(2,2)		4(·,6)	19(9,19)		20(17,24)	20(20,·)
	L_ε	0.4		19(19,1)		2(·,2)	20(1,31)	2(·,2)		17(·,17)			15(·,22)	12(·,13)		20(10,28)	2(2,·)
		0.01		20(20,4)		10(7,6)	20(1,35)	10(·,10)		18(·,23)	2(1,1)		18(·,29)	19(·,25)	3(3,·)	20(14,36)	19(19,·)
BAP1TLQ3W80	L_{quad}	0.4	5				15(·,15)									15(1,16)	8(7,1)
		0.01	2	16(15,1)			18(·,19)			1(·,1)						18(16,19)	17(16,2)
	L_ε	0.4	6				13(·,13)									13(·,14)	13(12,1)
		0.01	4	12(12,·)			15(·,17)			3(·,3)	1(1,·)		1(·,1)		3(3,·)	15(13,18)	16(15,3)
BAP4LTC5W	L_{quad}	0.4		20(19,2)	19(5,24)	3(·,3)	20(·,38)	3(·,3)		20(·,30)		11(7,10)		6(2,7)	12(·,12)		2(·,2)
		0.01		20(19,22)	20(16,28)	12(·,13)	20(·,60)	16(·,17)		20(·,40)	18(·,18)	20(20,24)	12(12,·)	13(4,14)	17(·,17)	1(·,1)	13(·,13)
	L_ε	0.4		20(10,14)	14(11,15)		20(1,36)	17(·,24)		20(·,31)	2(·,3)	13(10,14)	1(1,1)	13(·,25)	2(·,2)	10(·,10)	19(·,19)
		0.01		20(17,25)	19(16,22)	13(2,15)	20(2,69)	20(·,31)		20(·,42)	11(1,17)	20(18,32)	20(20,3)	17(1,38)	18(1,18)	14(·,16)	20(·,21)
BAP4SLC5W	L_{quad}	0.4		19(16,3)			1(·,1)	5(·,5)						19(17,3)			5(·,5)
		0.01		20(18,6)	5(·,5)	1(·,1)	3(·,4)	10(·,10)				1(·,1)	1(·,1) 3(·,3)	20(19,11)			10(·,11)
	L_ε	0.4	5	15(7,8)		1(·,1)	6(·,6)	1(·,1)						15(8,8)			6(·,6)
		0.01	5	15(7,8)	4(3,1)	2(1,1)	8(·,9)	5(·,5)				5(3,3)		15(9,8)			10(·,10)
BAP4TLC5W	L_{quad}	0.4		15(·,17)	13(4,13)	9(·,9)	20(·,31)	10(·,10)		20(·,28)		8(4,8)		14(·,18)			
		0.01		20(·,41)	19(14,19)	20(·,21)	20(·,37)	20(·,20)		20(·,39)		20(20,13)	3(2,1)	20(·,46)		1(·,1)	20(20,·)
	L_ε	0.4		20(1,25)	18(12,18)	14(·,14)	20(2,29)	6(·,6)		20(·,36)		18(7,21)		14(·,31)			
		0.01		20(4,33)	19(18,19)	19(2,27)	20(2,35)	16(·,17)		20(·,43)		20(20,28)	14(14,3)	18(·,44)	3(·,3)		20(20,·)

Table 9.24: Frequency of selection for the SUPANOVA bivariate models determined over the multiple model runs. The numbers in the $(\cdot,\cdot)$ represent the total number of ANOVA terms containing the particular variable as either a univariate or bivariate term respectively.

Data set	Loss Function	Stage II tol	Bias	Plate gauge	Stretch	Cu	Fe	Mg	Mn	Si	Zn	Ti	Cr	T_f	ST_t	GR	CRS
			ANOVA terms														
BAP1TLC3W	L_{quad}	0.4	11	6(6, ·, ·)	6(·, ·, 6)		6(·, 1, 6)							5(·, ·, 5)		1(·, 1, ·)	8(8, ·, ·)
		0.01	3	15(15, ·, ·)	12(1, ·, 13)		12(·, 2, 13)		8(8, ·, ·)	2(·, ·, 2)				12(·, ·, 13)		2(·, 2, ·)	16(16, ·, ·)
	L_ε	0.4	14	2(2, ·, ·)	1(·, ·, 1)		1(·, ·, 1)			2(·, ·, 2)	2(·, ·, 2)		2(·, ·, 2)	1(·, ·, 1)			5(5, ·, ·)
		0.01	5	15(15, ·, ·)	6(·, ·, 7)		6(·, ·, 7)			3(·, ·, 4)	3(·, ·, 3)		3(·, ·, 3)	6(1, ·, 6)			14(14, ·, ·)
BAP1TLQ3W	L_{quad}	0.4		20(20, ·, ·)			20(·, 31, ·)							11(·, 11, ·)		20(8, 20, ·)	9(9, ·, ·)
		0.01		20(20, ·, ·)		13(13, ·, ·)	20(·, 34, ·)			1(·, 1, ·)	3(3, ·, ·)	2(2, ·, ·)	1(·, 1, ·)	14(2, 14, ·)		20(17, 20, ·)	20(20, ·, ·)
	L_ε	0.4		20(20, ·, ·)			20(·, 31, 2)	1(·, ·, 1)		1(·, 1, ·)			1(·, 1, ·)	13(·, 11, 2)		20(13, 20, 1)	3(3, ·, 1)
		0.01		20(20, ·, ·)		6(6, ·, ·)	20(·, 31, 4)	1(·, ·, 1)		1(·, 1, ·)	12(12, ·, ·)	3(3, ·, ·)	1(·, 1, ·)	15(·, 11, 4)	8(8, ·, ·)	20(19, 20, 3)	16(16, ·, ·)
BAP1TLQ3W80	L_{quad}	0.4	11				4(·, ·, 4)	1(·, ·, 1)						3(·, ·, 3)		4(·, ·, 4)	9(9, ·, ·)
		0.01	8	9(9, ·, ·)			9(·, ·, 11)	4(·, ·, 4)						7(·, ·, 7)		9(1, ·, 11)	12(12, ·, ·)
	L_ε	0.4	16				1(·, ·, 1)	1(·, ·, 1)								1(·, ·, 1)	4(4, ·, ·)
		0.01	11	1(1, ·, ·)			8(·, ·, 8)	3(·, ·, 3)						5(·, ·, 5)	4(4, ·, ·)	8(·, ·, 8)	9(9, ·, ·)
BAP4LTC5W	L_{quad}	0.4		20(20, ·, 1)	6(·, 5, 1)	2(·, ·, 2)	20(·, 21, 1)	5(·, 5, ·)		20(·, 29, ·)		6(1, 7, ·)		3(1, 2, 1)	1(·, 1, ·)		8(·, 8, ·)
		0.01		20(20, 2, 2)	7(1, 6, 1)	2(·, ·, 2)	20(·, 24, 1)	13(·, 17, ·)		20(·, 32, ·)	2(·, 2, ·)	18(18, 15, ·)	5(5, ·, ·)	8(4, 3, 2)	3(·, 3, ·)		10(·, 10, ·)
	L_ε	0.4		20(20, 1, 7)	7(·, 4, 4)	7(·, ·, 7)	20(·, 37, 3)	18(·, 27, 1)		20(·, 24, 6)	1(·, ·, 1)	10(·, 9, 2)		11(·, 12, 6)	7(·, 7, ·)	2(·, 2, ·)	20(2, 21, ·)
		0.01		20(20, 3, 19)	13(9, 4, 9)	14(2, ·, 18)	20(·, 47, 11)	20(·, 40, 3)		20(·, 26, 7)	7(2, 2, 3)	18(15, 18, 1)	18(18, ·, ·)	17(6, 23, 13)	13(·, 13, ·)	2(·, 2, ·)	20(2, 20, ·)
BAP4SLC5W	L_{quad}	0.4		16(16, ·, ·)				1(·, 1, ·)						17(17, ·, ·)			1(·, 1, ·)
		0.01		18(18, ·, ·)				2(·, 1, ·)						19(19, ·, ·)			2(·, 1, ·)
	L_ε	0.4		18(15, 3, ·)				2(·, 2, ·)						19(18, 3, ·)			2(·, 2, ·)
		0.01		19(16, 3, ·)	11(11, ·, ·)			3(·, 3, ·)				5(5, ·, ·)		19(18, 3, ·)			3(·, 3, ·)
BAP4TLC5W	L_{quad}	0.4		16(3, 2, 12)	2(·, 2, ·)	14(·, 2, 12)	20(·, 26, ·)	9(·, 11, ·)		20(·, 21, ·)		4(·, 4, ·)		12(·, 6, 12)			
		0.01		16(6, 4, 14)	4(1, 4, ·)	15(·, 4, 14)	20(·, 30, ·)	13(·, 18, ·)		20(·, 23, ·)		20(20, 7, ·)	5(5, ·, ·)	14(·, 11, 14)			18(18, ·, ·)
	L_ε	0.4		20(12, 4, 14)	6(·, 5, 1)	15(·, 4, 13)	20(·, 46, 4)	7(·, 7, ·)		20(·, 27, 6)		14(3, 11, 8)	1(·, ·, 1)	18(1, 18, 12)	1(·, 1, ·)		
		0.01		20(15, 5, 19)	13(9, 7, 2)	19(6, 5, 17)	20(·, 59, 4)	15(·, 15, 2)		20(·, 32, 6)		20(20, 20, 14)	13(12, 1, 1)	20(1, 24, 16)	3(·, 3, ·)		18(18, ·, ·)

Table 9.25: Frequency of selection for the SUPANOVA trivariate models determined over the multiple model runs. The numbers in the $(\cdot, \cdot, \cdot)$ represent the total number of ANOVA terms containing the particular variable as either a univariate, bivariate or trivariate term respectively.

9.8 Metallurgical considerations

In terms of the more physical interpretation of the models of the 7x75 data subsets, it is clear that Fe and Si are common significant input variables. In the neurofuzzy and SUPANOVA results, both are seen to reduce K_{Ic}, consistent with conventional metallurgical understanding of coarse intermetallic particle formation, and their influence on failure. As noted previously, evidence of overfitting was seen in some Fe univariate ANOVA terms (e.g. Figures 9.15 (b) and 9.24 (e)), but the general trends were consistent.

In terms of the categorical variable models applied to the BAP1 and BAP4 data sets in section 9.6.1 (as shown in Figures 9.3 and 9.4), it may be seen that the simple test direction effect in Figure 9.4 is consistent with the mechanical fibring of these plate materials. In terms of thickness position, the quarter position may be expected to exhibit lower toughness than the centreline (De Jong 1980). In terms of both the BAP1 and BAP4 data sets, a contradictory effect was seen. It should be noted however that the models represented in Figures 9.3 and 9.4 were specifically for the categorical variables (direction, thickness position, width position and alloy type). In these cases it must be noted that the categorical variables were "concealing" other important variables. In the case of the BAP1 data set, the quarter-thickness toughness results all came from the thicker, high purity alloy variant, whilst in the BAP4 data set, the quarter-thickness results all corresponded to the tougher LT and TL test directions.

It was seen that for the various TL orientation data sets for both alloys (as shown in Figures 9.13 to 9.15 and 9.18) cross-rolling (i.e. CRS), had a generally beneficial influence on toughness in the neurofuzzy models. This was closely confirmed by the SUPANOVA models (although the flexibility of the kernel functions used by the SUPANOVA modelling allowed some local non-linearities to appear). Such an effect may be considered consistent with a beneficial redistribution of coarse intermetallics and grain boundaries away from the nominal crack growth plane in this test direction. In keeping with this, increasing CRS was seen to reduce toughness in the LT orientation data set for the 7475 (Figure 9.16). The influence of such redistribution on SL toughness is not intuitively obvious, as the relevant microstructural features will be aligned in the nominal fracture plane by both straight and cross-rolling passes. A detrimental influence of cross-rolling on SL toughness is evident in Figure 9.17, which may merit further investigation.

Without further assessment, it is difficult to separate the effects of gauge and gauge reduction in this work, as the two are so clearly correlated. In the first instance, it should be noted that thicker material (increasing gauge or gauge reduction value) should have reduced fracture resistance due to more limited breakdown of the cast structure and slower quench rates on cooling. From Figures 9.13 and 9.15 for the neurofuzzy models obtained for the 7175 data, and Figures C.19 to C.26, for the SUPANOVA models it may be found that the two thickness related terms (i.e. gauge and gauge reduction value) exhibit a number of increasing and decreasing functions. The univariate neurofuzzy models may be considered in the first

instance as providing a simple assessment of these variable's effects.

In terms of the 7175 neurofuzzy models, Figure 9.15 (c) shows the expected simple trend of decreasing toughness with increasing thickness (in terms of gauge reduction in this diagram). Figure 9.13 (f) shows a similar trend in terms of gauge reduction, however there is an apparent compensatory influence acting via the gauge in Figure 9.13 (e). Figure 9.14 (e) or (f) show an initial decrease in toughness with increasing gauge reduction value, however a clear change occurs at high gauge reduction values (i.e. greater thickness). In the first instance it may be noted that the influence in Figure 9.14 was associated with a gauge range spanning two alloys variants, with a purer alloy being used for thicker plate, consistent with higher apparent toughness for the thicker material. This may then be consistent with the simple trend in Figure 9.15 (c) which was associated with just one alloy variant. In terms of the 7475 neurofuzzy models, all included plate gauge (and not gauge reduction), with Figures 9.16 and 9.18 showing a decreasing and then increasing trend in toughness with increasing gauge, whilst Figure 9.17 shows a simple linear increase in toughness with gauge.

Overall, it may be seen that for both 7175 and 7475 data sets, a number of decreasing and increasing trends in toughness were identified with varying gauge, that are not consistent with simple metallurgical understanding, even when alloy purity variations were considered. Given the complexities of the present data sets and associated processing routes, it is not possible to explain all these effects, and further work would be needed on this issue. In the first instance it may be noted that information on quenching practice and how this varies with gauge was not available for this work, and further investigation of this and other potentially significant information would be considered valuable.

It is clear that the SUPANOVA modelling suggested a variety of other potentially interesting functionalities. For example, it is interesting to note that SUPANOVA models (and indeed the neurofuzzy models) identified combined influences of deformation parameters (e.g. gauge reduction) with the intermetallic forming impurities (particularly Fe, see Figure C.20 (a)), which may then be related to the mechanical fibring process. Overall however it may be seen that the complexity and the relative indirectness of much of the data considered here (i.e. the inputs are directly from processing and do not include real microstructural factors such as grain size, dispersoid separation distances, etc.) it is not considered reasonable to assess all of the suggested functionalities in detail. The inherent value of transparent modelling approaches that allow such detailed model analysis may however be appreciated from the present work, although in models for which the contribution of individual variates appear in more than one term can be difficult to interpret.

9.9 Summary and conclusions

In this chapter adaptive modelling techniques have been used for the purposes of determining interpretable, parsimonious descriptions of how fracture toughness is influenced by processing

conditions of high-strength wrought Al-alloys. The statistical analysis of the industrial process investigated, was conducted on data obtained from commercial production databases. The process variables which comprised the data sets modelled included material compositional information, basic TMP conditions/practices and the testing variables.

The preliminary assessment of the integrity of the data performed in section 9.2, noted the lack of process information describing certain important processing stages (e.g. the cast, quenching). The subsequent definition of the data sets resulted in the omission of some processing variables that had been provided due to tight process control limiting the expected influence on material performance (e.g. the temperature associated with the precipitation treatment). The knowledge gained from these simple data assessments suggested that the highly desirable objective of developing a generally applicable predictive model was unrealistic, although the modelling of the present data was still considered to be of value for specific understanding and potential improvement of the process routes in question: at least within the input variable windows that were available.

Whilst on the majority of the data sets the initial MLR analyses provided high variance estimates, on the BAP1TLC3W data set in particular both the training and generalisation performance appeared comparable with accepted experimental K_{Ic} measurement scatter.

In light of the important missing information (e.g. quench data), it is perhaps not surprising to find that the generalisation performances attained by both the neurofuzzy and SVM methods were seen to be comparable, if not "inferior" to the MLR results. In view of the integrity and statistical properties of the input variables, data-driven modelling approaches were exposed to a number of data weaknesses which would naturally result in the identification of poor models, e.g. inferring non-causal interactions and dependencies, and more generally overfitting. Given the high interpretability of both adaptive methods used, it was however possible to validate the models obtained to some extent, distinguishing between the dependencies which constituted plausible interdependencies and those which were potentially suspect due to data deficiencies.

Notwithstanding such shortcoming, it was nevertheless useful to use adaptive modelling approaches for the purposes of attaining a parsimonious description of the various data sets defined, particularly since both the inspection of the data and the MLR results suggested the presence of a number of dependencies between K_{Ic} and a subset of the processing variables.

In terms of the model structures that were found, it was not surprising to find that the main variables included in the fracture toughness models were: 1. the impurity elements Fe and Si, which, as well as having a strong effect on toughness, may be expensive to control; 2. the final thickness section, which in the 7475 data sets in particular was seen to have a considerable range, and comprises an important factor in the quench; 3. the deformation variables resulting from the hot rolling (particulary cross rolling).

A comparison between the two adaptive approaches showed that the dependencies inferred by the SUPANOVA models were generally consistent with the simpler approximations attained

from the neurofuzzy framework. As in previous chapters, the approximations determined by the infinite ANOVA spline kernel functions enabled a greater flexibility to be attained compared to the practical limitations in the approximation capabilities obtainable in sparse data sets and high-dimensional problems by B-spline functions.

In light of the nature of the BAP data, the higher variance showed by the SUPANOVA framework both in terms of the ANOVA components subselected (particularly in the bivariate and trivariate basis expansions) and the smoothness (i.e. degree of regularisation) of the spline approximations, over the multiple training-test set partitions, is not surprising. The large basis expansion obtained from considering higher order ANOVA terms and the optimisation of the C smoothing parameter from a cross-validation procedure together can be held responsible for the considerable instability of the SUPANOVA framework. As such, a number of terms which are selected only in a few instances remain difficult to validate. In addition, the sparse subset selection performed in the SUPANOVA framework is seen to be sensitive to the size of the full ANOVA basis expansion. This has been seen to result in suboptimal inferences, whereby input variables which are subselected from univariate ANOVA basis expansions are absent in models inferred from higher order basis, and in the most severe conditions resulting in models comprising only a bias term.

Due to the more limited approximation capabilities of the neurofuzzy framework, both in terms of model structure and non-linear approximation capabilities, limited by the iterative search procedure and the implementation of the SRM principle, the neurofuzzy results showed a greater structural stability.

Although several of the outliers detected in section 9.6.4 were clear from the data distributions, the detection of non-trivial influential observations by a simple inspection of input and output scatterplots is generally more problematic in high-dimensional data sets. The influence measures were then seen to successfully detect the trivial, Y-space outliers, and a number of other influential cases, the latter in some cases arising from the conditional fit to the data. A full understanding of the nature of certain influential cases remained limited from simple residual plots. The results also highlight the deficiencies of the one-case influence measures, and that more appropriate techniques should be used to overcome masking effects.

The present analysis has illustrated the difficulty of extracting reliable knowledge from large observational databases of complex processes, showing how the modelling or data mining component in a statistical analysis of industrial databases is only part of the investigation. The different stages of the KDD process, from the definition of meaningful data sets amenable to valid statistical analyses, through to the modelling *per se* and the subsequent interpretation (validation) of results are seen to be re-visited several times, and that rarely will the first set of results fulfil all the modelling objectives.

In view of both the complexity of the process and the integrity of the process data, the development of a general predictive process model was seen to be an unrealistic objective, and instead the statistical analysis of the BAP data was conducted primarily within a KDD

context, with a greater emphasis on data description and knowledge extraction rather than the formulation of a model as a means to a predictive end.

In the neurofuzzy framework, the Bayesian regularisation employed was seen to successfully control excessive degrees of freedom present in a model, comparing favourably with model pruning refinements. Overall, the neurofuzzy framework has been seen to be a useful data mining tool allowing both sensitivity analyses and regression diagnostics to be performed.

Although not attaining improvements (as measured in terms of generalisation performance), compared with simple MLR analysis, both the neurofuzzy and SUPANOVA modelling results were seen to attain more reasonable descriptions of the data. This was achieved by means of adapting the structure and complexity of the models to account for non-linear dependencies between processing variables and fracture toughness, and suggested a number of interdependencies characterising different process variables (particularly in the SUPANOVA, as opposed to the neurofuzzy framework). Whilst the disappointing empirical results were understood in terms of the data weaknesses characterising the BAP data sets, a number of dependencies inferred by the adaptive methods were consistent with metallurgical understanding.

Chapter 10

Conclusions and Future Work

10.1 Process-property modelling

Modelling of processing-property relationships of complex alloying systems is of great potential benefit, however despite the recent advances that have been made in thermodynamic and thermomechanical processing, the development of reliable models which scale up to industrial processing environments remains elusive.

This thesis has investigated the performance of the neurofuzzy and SUPANOVA techniques in the modelling of physical and tensile properties of two different Al-alloy systems with the objectives of developing models exhibiting good predictive ability on unseen data whilst allowing interpretation of the processing-property relationships learnt.

In Chapter 7, models for the tensile properties ($\sigma_{0.2}$, uts and $\%el.$) of an Al-Mg-Li mechanically alloyed system have been determined from a small set of commercial trials using simple compositional information and processing variables, while in Chapters 8 and 9 models for both physical and tensile properties ($\sigma_{0.2}$, σ_{el} and K_{Ic}) were obtained from compositional information and thermal/thermomechanical variables, describing the processing-property relationships of high-strength Al-Zn-Mg-Cu (7xxx series) alloys. Whilst the data set investigated in Chapter 8 was the result of a series of experimental trials conducted in a carefully controlled processing environment, the analysis in Chapter 9 was performed on data obtained from large scale commercial production data.

The performance of the neurofuzzy and SUPANOVA modelling approaches have been assessed in terms of the approximation obtained on the training data sets, the predictive performance on test data and the form of the processing-property dependencies inferred. The empirical results were compared with those obtained by simple MLR analyses and a support vector ANOVA spline kernel regression, whilst the nature of the relationships were validated in terms of metallurgical understanding. In addition, the transparency of the neurofuzzy models

was used to initialise networks with simple structures accounting for processing/test conditions and prior system knowledge.

The properties of the Al-Mg-Li alloys were seen to be typically approximated by both adaptive methods with simple linear dependencies. In addition, the SUPANOVA results identified a number of interdependencies between the processing variables and a subset of the alloying elements. Overall, the results suggested that strengthening effects were determined by a number of mechanisms, in agreement with metallurgical understanding of the tensile properties of these alloys. These comprised solid solution, grain size and dispersoid effects. The empirical performances attained by the adaptive modelling approaches on training and test sets was seen to be relatively disappointing, with the former comparable to that attained by a MLR analysis.

The results from the Al-Mg-Zn-Cu experimental trials gave the most revealing relationships and improvements in empirical performances compared with the MLR analysis. The adaptive methods exhibited a better approximation capability in the modelling of $\sigma_{0.2}$, but the most significant improvement over the MLR models was obtained in the σ_{el} results, mainly determined by the non-linear approximations inferred from the precipitation hardening heat-treatment time. Limitations imposed by the small sample size were seen to bias the approximation capabilitites exhibited by the SUPANOVA modelling framework.

The interaction between the Magnesium content of the alloys and the ageing time, inferred from both adaptive methods was interpreted on a physical basis. The interaction of Mg atoms with vacancies is believed to influence the rate of ageing in these alloys. The benefits of rule-of-thumb and physical based data transformations were seen to improve the empirical performance of the adaptive methods, compared with those obtained on the original data set. The latter set of transformations providing the derivation of some novel quantities which were seen to enhance the characterisation of Al-Zn-Mg-Cu alloys with a set of microstructurally related features.

The approach adopted in the statistical analysis of the large 7x75 data sets showed how a significant amount of pre-processing was required in order to obtain valid statistical analyses and meaningful inferences. Although the empirical performances attained from the neurofuzzy and SVM approaches in the modelling of K_{Ic} did not generally exhibit improvements over MLR results (in some instances exhibiting worse performance), in terms of training set approximation errors and generalisation performance, the adaptive methods were seen to determine a number of non-linear approximations and interdependencies amongst the input variables that were validated/representative of both accepted behaviours (e.g. the role of Fe and Si) and particular processing conditions.

To a certain extent, the nature of the relationships inferred highlighted the conflicting requirements characterising the physical and mechanical properties; for example in the Al-Mg-Li system it was seen that forging conditions should be such that workability and strength levels attained are comparable with the required ductility levels. In the Al-Zn-Mg-Cu data set,

extensive overageing is seen to reduce $\sigma_{0.2}$ but increase σ_{el}.

The data analysis and results obtained in the various analyses indicate that for the present data sets, the much sought after optimisation and balance of properties were unrealistic objectives and beyond the scope of the data. As such, it is more appropriate to conclude by discussing issues such as data integrity and description. The following discussion leads to the suggestions made in section 10.3 for future areas of research, which directly address data quality issues.

It is perhaps indicative that the most successful modelling results in the present work have been attained in the statistical analysis performed on the experimental data set (Chapter 8): the carefully designed compositional/ageing time variants enabled a wider range of microstructural conditions to be attained, facilitating the development of more reliable descriptions of how variations of compositional and thermal treatments influence these properties.

The dependencies inferred with greater consistency (frequency) in the other investigations were also seen to be associated with the input variables exhibiting the greatest variance and range in values (e.g. Fe, Si, plate gauge, deformation variables), which may be effectively uncontrollable quantities or determined by production/fabrication requirements.

While the neurofuzzy models were generally seen to exhibit relatively simple dependencies, the approximations obtained by the SUPANOVA framework, in a number of instances exhibited pronounced non-linearities, in some cases clearly overfitting the training data due to a nonoptimal degree of regularisation inferred. Thus, a validation of both the general form of the structural relationships identified and the local nature of the approximations was needed in terms of the input ranges and the level of control which can realistically be obtained in an industrial processing environment.

Overall then, it emerged that the quality of the data and the need for experimental design are of utmost importance in inferring reliable dependencies between the quantities contained in the data. The results being a clear indication of the benefits obtained from carefully designed data sets and highlighting the problems associated in modelling observational data.

However, the statistical investigation and modelling of production (historical) data is often valuable as an assessment of the process route. The application of data mining methods may reveal some informative patterns, indicating suboptimal processing conditions or other useful knowledge which can be used to identify particular processing conditions responsible for the scatter observed in the properties attained in a number of materials. It may be argued that no experimentation is required in a statistical investigation of industrial processes since a large amount of historical data is available and from this it ought to be possible to extract information relating the changes in the material properties with the modification in the processing conditions which have occurred "naturally". However, it appears from this work that further planned experimentation will allow for a better appreciation of the underlying physical relationships.

The "hazards" of modelling observational data have emerged in the results obtained in Chapters 7 and 9. From which it was seen that important variables which *a priori* were considered to significantly influence the properties of the alloys investigated were limited by process

control, unrecorded variables can be held responsible for giving rise to nonsensical correlations and the sparsity of the data in certain regions of the input space leading to overfitting.

10.2 Neurofuzzy networks and SVM

From the observations made above and the more detailed results and discussions presented in respective chapters, both the success and limitations of the neurofuzzy and SUPANOVA frameworks can be put into perspective alongside the integrity of the data sets investigated. Suggestions for future topics of research, addressing the shortcomings and underperformances identified for both adaptive modelling techniques are made in section 10.3.

Although exhibiting fundamental differences and implementing in different ways the S-RM principle, both neurofuzzy and SUPANOVA techniques are seen to search for a similar ANOVA representation.

The results obtained in the application of the neurofuzzy framework showed the use of the SRM principle as a model complexity (statistical significance) measure was consistent with previous studies. It enabled parsimonious networks to be determined, limiting the degrees of freedom in a model to that which is supported by the data, although prone in some instances to identify inadequate model structures. The limitations imposed by the finite data set sizes, together with the high implementation costs (in terms of the number of degrees of freedom) in constructing tensor (bivariate) subnetworks was seen to limit the dependencies attained by the neurofuzzy models to typically univariate approximations or bivariate terms comprising a low number of basis functions defined on the input axis.

Model selection/performance criteria inherently based on the training set MSE, the number of parameters in the model and the training set sample size are used to select models whose parameters are determined through maximum likelihood estimation. The main criticism of this approach is that they remain prone to making incorrect choices, leading to the identification of nonsensical dependencies and overfitting regions of the input space where data is sparse. Consequently, the performance of iterative model construction algorithms will have the tendency to perform suboptimal refinements and hence become prone to local minima entrapment (in terms of the model performance measure). This will be accentuated in the presence of noisy data sets and interdependencies amongst the input features (Caruana and Freitag 1994; Vafaie and Imam 1994).

Furthermore, the bias resulting from the omission of inputs from using variable selection procedures, may be disconcertingly, interpreted in two contrasting ways: 1. a lack of structure with the output; 2. a dependency with the other explanatory variables present in the model, thus making the inclusion of an additional input redundant. In this respect, the inherent preference for simpler models which has been interpreted by Domingos (1999) as a misuse of Occam's razor, i.e. that greater simplicity does not necessarily lead to greater accuracy and generalisation

performance[1], is an inherent problem of parsimonious system modelling.

In the interpretation of the model structures obtained by the neurofuzzy models, it should be clear that the order of variable inclusion during model construction should not be used to infer the input variable importance, as the nature of certain dependencies may lead to misleading measures of relevance, whereby the notion of independent contribution to the output variance may have no meaning.

In several of the data sets investigated, it was seen that the ANOVA representations inferred by both adaptive modelling approaches exhibited a degree of instability. In both techniques this can be understood mainly in terms of the data set sizes being an insufficiently representative sample of the process, particularly for the observational data sets. In the SVM results this also resulted in the variance in the selection of an appropriate smoothing parameter.

Undoubtedly, both the neurofuzzy and SUPANOVA modelling approaches constitute a significant improvement over MLR analysis. In adapting their structure to identify the dependencies that are contained in the data, they allow a better description of the processing-property relationships.

Finally, in the analysis performed, assessing the generalisation performance exhibited by these techniques has been problematic, due to both the finite sample sizes and quality of the data sets.

10.3 Future research

Multivariate regression

In this work, modelling has been approached by obtaining a separate model for each output of interest. However, it may be valuable to obtain a model which describes the overall system, considering the interdependencies amongst the output quantities. Breiman and Friedman (1997) discuss predicting multivariate responses in a multiple linear regression framework and derive strategies which take advantage of correlations between the response variables to improve predictive accuracy compared to the general method of regressing each response variable on the common set of predictor variables. In a similar approach which Tibshirani and Hinton (1998) termed *coaching*, a set of variates ($\mathbf{z}$) (which for example may only be available during network training as they may be difficult to subsequently measure) is used to coach the inputs ($\mathbf{x}$) in estimating the response variables ($\mathbf{y}$). General classes of influence measures have been defined for multivariate linear models (Barrett and Ling 1992).

[1] Domingos (1999) notes that Occam's razor is generally interpreted in two ways: as favouring the simpler of two models with the same generalisation error because simplicity is a goal in itself, or as favoring the simpler of two models with the same training set error because this leads to lower generalisation error. Domingos finds the second interpretation to be provably and empirically false and argues that the first version is only a proxy for interpretability.

In their present form neurofuzzy networks and model construction procedures are not readily amenable to attaining multivariate responses, requiring the appropriate methods for dealing with multivariate responses, model selection and complexity control frameworks.

Regularised model construction

Chapter 9 has shown that model structures with a larger number of parameters can generalise reliably, if complexity is constrained through regularisation approaches. Here regularisation was used as a post-processing step once the structure of the model was determined by a forward selection procedure.

It may then be of interest to employ regularisation methods during the model construction (e.g. regularised forward selection), allowing more variables to be included in the model, whilst controlling redundancy by optimising the regularisation coefficients. The drawback of this procedure is that it imposes a significant computational overhead as the regularisation coefficients have to be re-estimated for each candidate refinement considered. Model construction will however remain susceptible to local minima. Notwithstanding this, it is considered a preferable approach to iterative model construction algorithms based on maximum likelihood estimation. It may offer a better (less biased) description of the process, whereby the inclusion of inputs which a greedy model construction algorithm considers uninformative, will cause the output variance to be averaged over two or more input variables, which may also result in improved generalisation performance.

Limiting the size and complexity of models will then require an appropriate model comparison framework. Two suggestions are made: one based on determining the Occam factors for the models, the other retaining the use of a SS measure but replacing p, the number of parameters, by p_{eff}, the *effective number of parameters* in the model.

Regularised model construction procedures have been previously proposed (Chen et al. 1996; Orr 1995b; Orr 1995a) whereby subset selection and zero order regularisation are combined in the construction of parsimonious regularised radial basis function networks.

Model averaging

Probably the main criticism of the approach adopted in this thesis is that conditioning on a single model ignores model uncertainty, resulting in underestimation of uncertainty when making inferences about quantities of interest (Raftery et al. 1993). Although this may be addressed through a complete Bayesian solution by means of averaging over all possible models when making inferences about quantities of interest, this approach remains impractical. Relatively simple methods have been proposed to improve the performance of unstable learners (Breiman 1996a) by means of averaging over a finite number of models, using a number of weighting schemes, e.g. see (Breiman 1996a; Tresp and Taniguchi 1995; Taniguchi and Tresp 1997). With such methods however the simple interpretable structures of the individual models are

lost.

Support vector methods

The SVM is a fast developing field with an increasing number of implementation and optimisation techniques being proposed. As such a considerable number of alternative implementation schemes to that used in this thesis are available.

The present results have shown that the use of cross validation procedures in the selection of the regularisation parameter was seen to be both computationally expensive and susceptible to inferring inadequate smoothing, alternative schemes for inferring the amount of regularisation are then of particular interest. In recently proposed Bayesian methods for SVM, MacKay's evidence framework can be used in the selection of the smoothing parameter and allow error bars on the predictions to be obtained (Kwok 1999; Sollich 2000).

Alloy characterisation (microstructural feature sets)

The present results have been either attained from the use of bulk alloy process data or microstructurally-related quantities derived from compositional information. Using a set of input features which allow a greater characterisation of the microstructural condition is likely to enhance the performance of adaptive modelling approaches. For instance, Li and Starink (2000) have used DSC (differential scanning calorimetry) data to achieve a greater understanding of the relation between composition, precipitation and the balance of strength and electrical conductivity in 7xxx Al-alloys, providing valuable information on optimal chemistry (Mg, Zn and Cu contents) and heat treatment. Further investigation of DSC results as additional microstructural features/inputs may be worthwhile.

Design of experiments

In Chapter 8, it was seen how a carefully chosen set of experiments have allowed a greater interpretation of the influence of compositional elements and ageing times to be understood.

Poorly designed experiments on the other hand (Chapter 7), not only waste resources, but may fail to provide unambiguous answers. The use of statistical techniques which can define optimal sampling of the input space is highly desirable. Design of experiments, e.g. see (Box et al. 1978; Box and Draper 1987; Atkinson and Donev 1992) is the application of geometric principles to statistical sampling which are useful in minimising the number of experiments necessary to obtain reliable statistical inferences (e.g. minimising the variance of estimated coefficients performed in a subsequent MLR analysis).

In a rare statistical approach in modelling the mechanical properties of (pressed and sintered powder metal) alloys, Mantini and Prucher (1993) showed that in a comparison between methods based on random data (i.e. stepwise regression) and experimental data, the latter enabled a more complete and accurate model to be inferred.

Appendix A

Data Modelling of Processing-Property Relationships in an Al-Mg-Li Powder Metallurgy Alloy System

A.1 Statisical properties of the data

x	μ_x	σ_x
PV_c (°C)	323.21	75.97
$PV_{b/a}$ (*)	121.54	48.84
Li (wt.%)	1.19	0.32
Mg (wt.%)	3.85	0.91
C (wt.%)	0.68	0.41
O (wt.%)	0.11	0.06

Table A.1: Mean and standard deviation for the input variables comprising the Al-Mg-Li data set. * proprietary processing variable (nominal units).

y	μ_y	σ_y
$\sigma_{0.2}$ (MPa)	376.50	88.88
uts (MPa)	440.31	84.38
$\%el.$ (%)	8.73	2.85

Table A.2: Mean and standard deviation for the output properties comprising the Al-Mg-Li data set.

A.2 Neurofuzzy model construction

Ref. no.	Inputs	Refinement	MSE	SS
1	Mg	ua(0)	4203.87	9417.65
2	C	ua(2)	1343.59	4857.32
3	PV_c	ua(0)	1014.31	4144.44
4a	$PV_{b/a}$	ua(0)	908.91	4217.67
5a	$PV_c \times$ C	tp(0,0)	1186.1	4287.99
4b	PV_c	sd	1343.59	4857.32
5b	C	sd	4203.87	9417.65
4c	C	ki(0.75)	959.79	4453.78
5c	C	ki(0.625)	858.67	4556.04
4	C	kd(-0.5)	1016.56	3675.04
5d	C	kd(0)	1411.67	4529.19
6d	C	kd(0.5)	1508.38	4300.38
5e	PV_c	ro	1348.75	4327.33
6e	C	ro	3004.39	8565.48

Table A.3: Summary of the iterative model search performed by the ASMOD algorithm in determining the uts model.

Ref. no.	Inputs	Refinement	MSE	SS
1	$PV_{b/a}$	ua(1)	4302.37	10891.5
2	Mg	ua(0)	2547.96	7264.19
3	PV_c	ua(0)	1634.28	5243.41
4	C	ua(0)	1338.27	4838.09
5	$PV_c \times PV_{b/a}$	tp(0,0)	1300.94	4703.15
6	$PV_c, PV_{b/a}$	ts(0,0)	1353.84	4343.65
7a	C	sd	1668.96	4758.17
8a	PV_c	sd	2612.61	6613.82
7b	Mg	ki(0)	1292.57	4672.86
8b	C	ki(0)	1274.05	5205.71
7c	C	ro	1668.96	4758.17
8c	PV_c	ro	2612.61	6613.82

Table A.4: Summary of the iterative model search performed by the ASMOD algorithm in determining the $\sigma_{0.2}$ model.

Ref. no.	Inputs	Refinement	MSE	SS
1	C	ua(0)	5.221	11.69
2	PV_c	ua(0)	4.068	10.30
3a	O	ua(0)	3.606	10.28
4a	C×O	tp(0,0)	3.160	10.14
3b	PV_c	sd	5.221	11.69
3c	C	ki(0)	3.902	11.13
4c	C	ki(0.5)	3.590	11.38
3d	O	ro	5.221	11.69
4d	C	ro	7.99	15.72

Table A.5: Summary of the iterative model search performed by the ASMOD algorithm in determining the $\%el.$ model.

A.3 Histogram plots

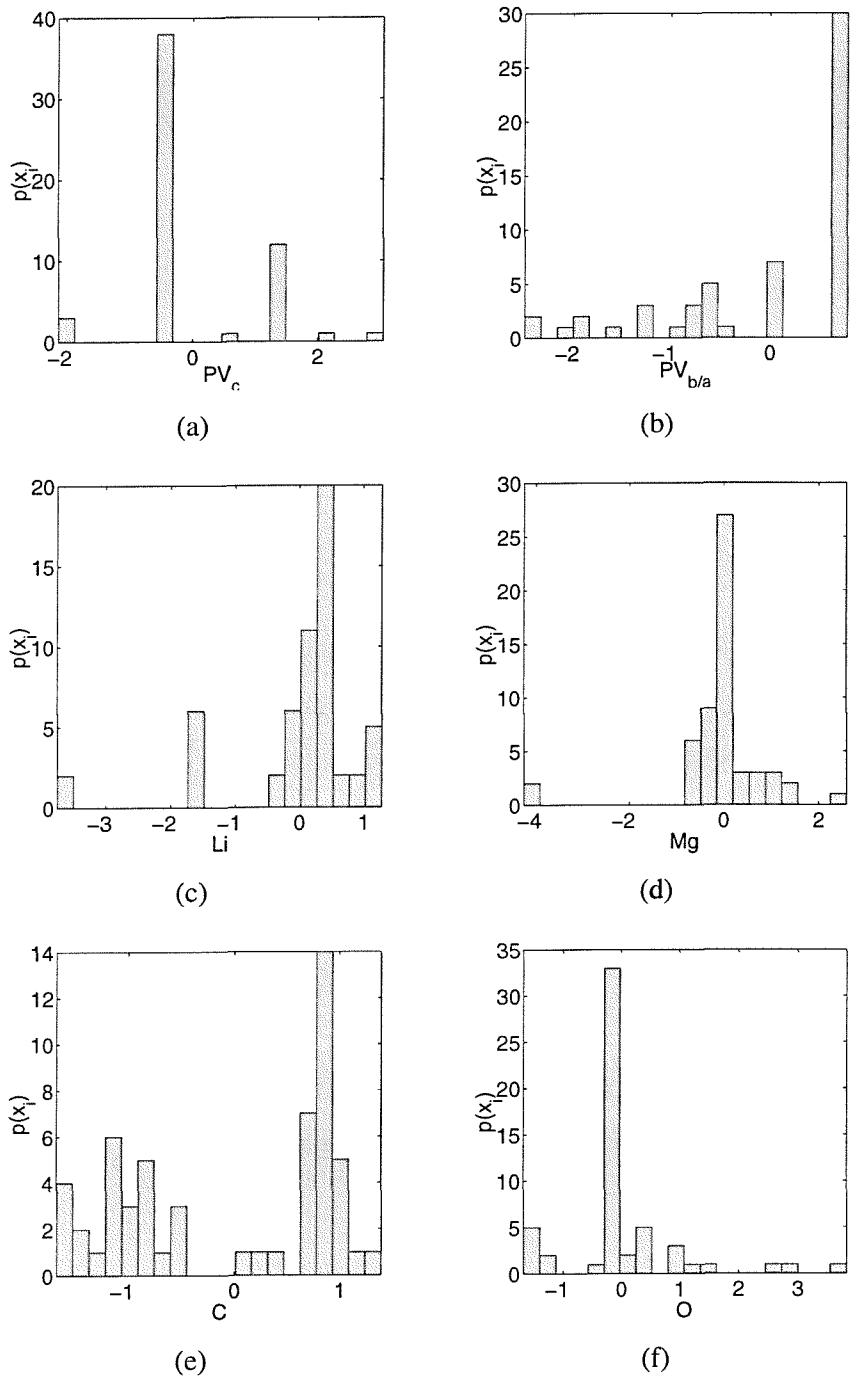


Figure A.1: Histogram plots for the input variables: PV_c (a), $PV_{b/a}$ (b), Li (c), Mg (d), C (e), O (f).

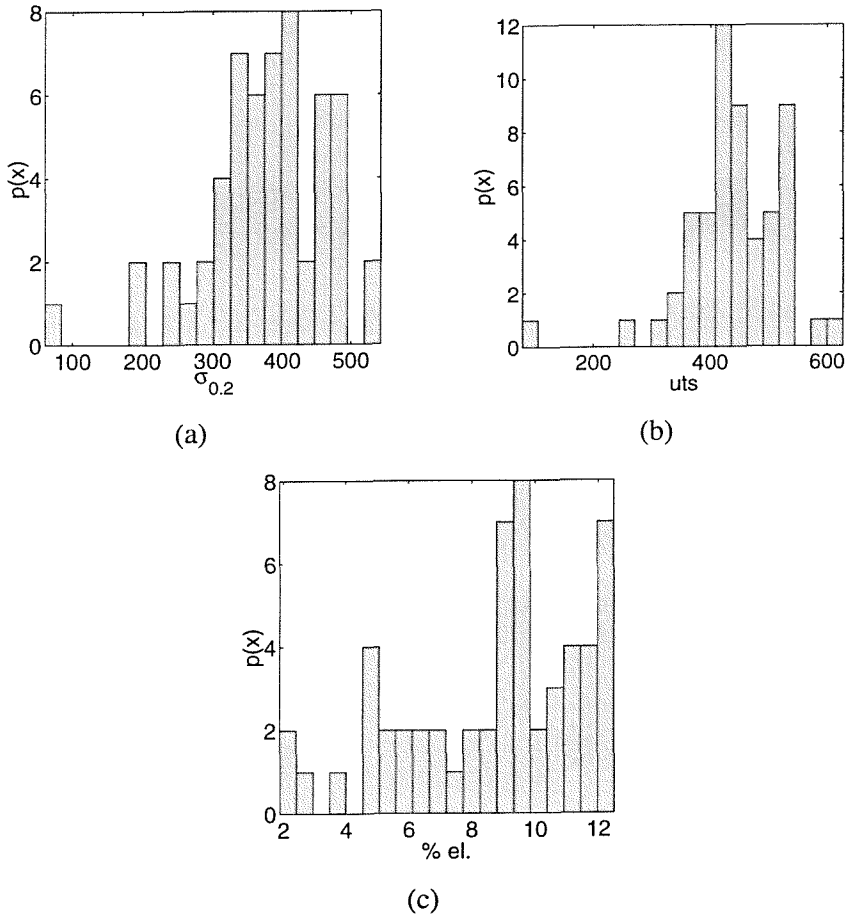


Figure A.2: Histogram plots for the tensile properties (outputs) investigated: $\sigma_{0.2}$ (a), uts (b) and $\%el.$ (c).

A.4 Pairwise scatter plots

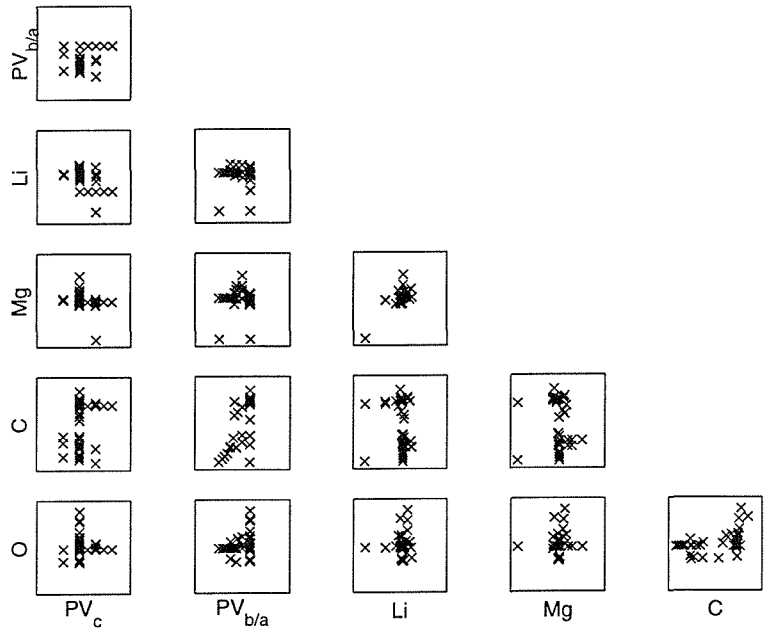
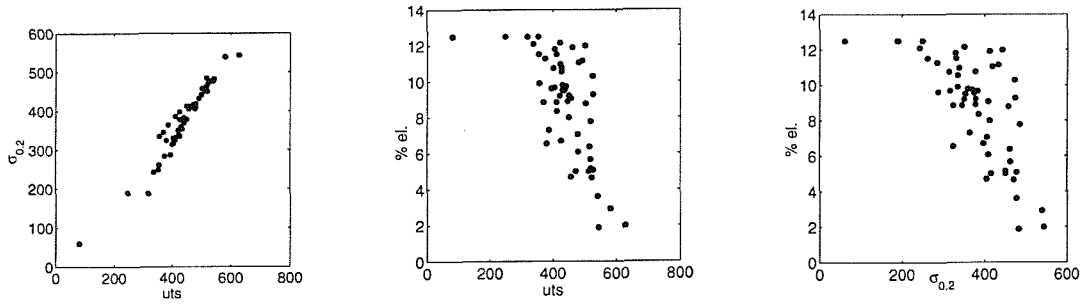


Figure A.3: Pairwise scatter plots between the input variables comprising the Al-Mg-Li data set.



(a) (b) (c)

Figure A.4: Pairwise scatter plots between the tensile properties (outputs) investigated.

A.5 Residual scatter plots

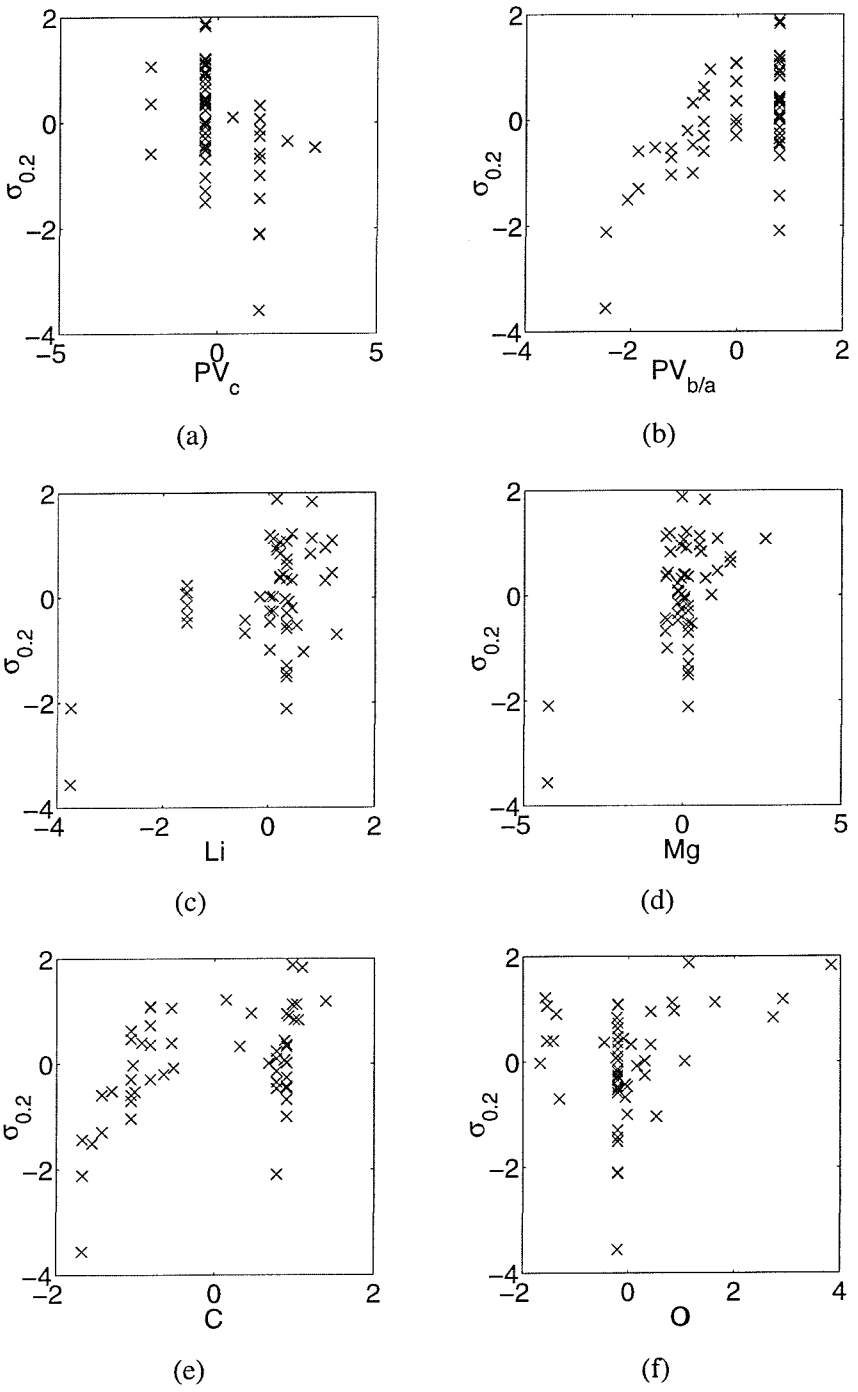


Figure A.5: Full residual plots between $\sigma_{0.2}$ and the input variables comprising the Al-Mg-Li data set: PV_c (a), $PV_{b/a}$ (b), Li (c), Mg (d), C (e), O (f).

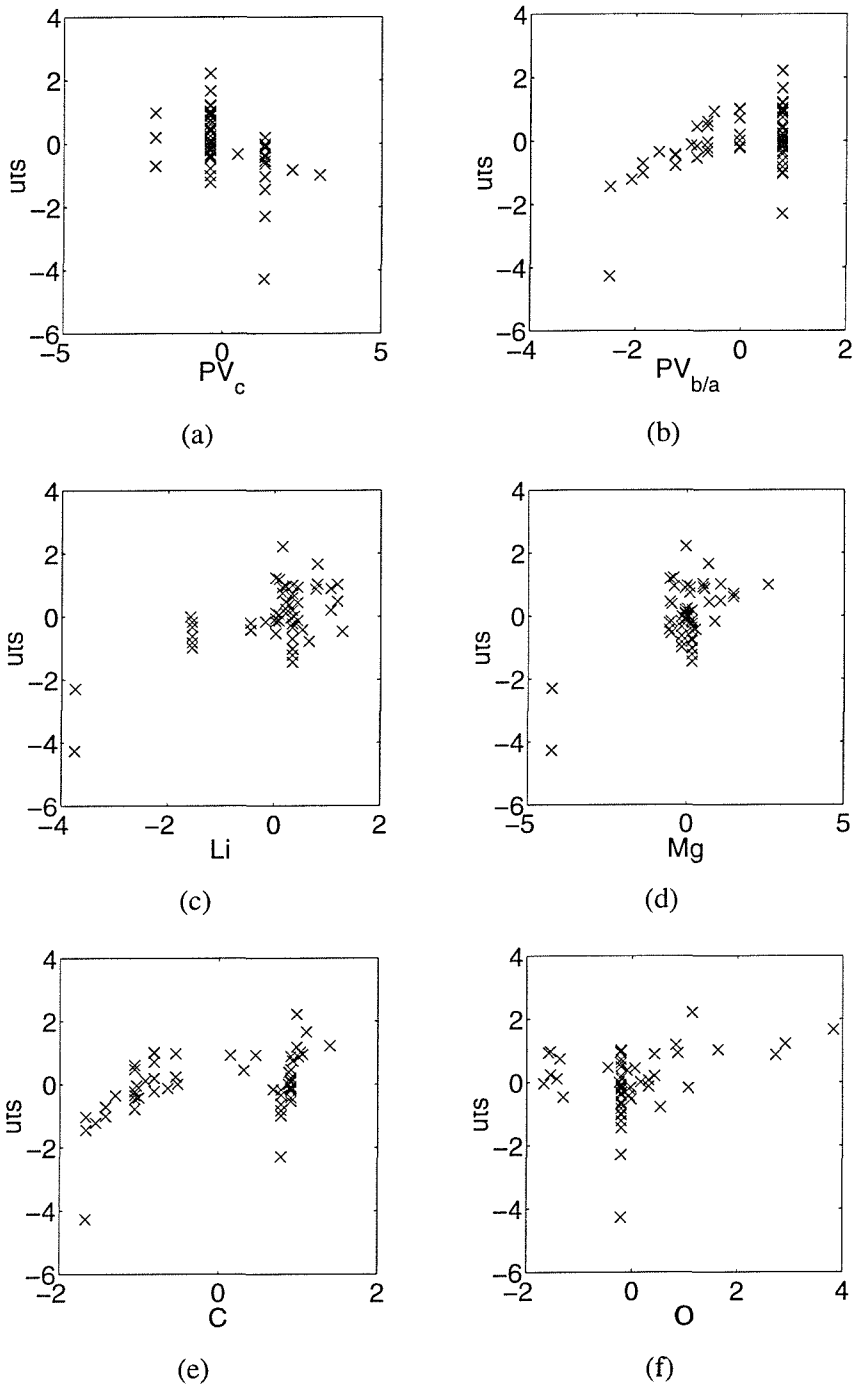


Figure A.6: Full residual plots between uts and the input variables comprising the Al-Mg-Li data set: PV_c (a), $PV_{b/a}$ (b), Li (c), Mg (d), C (e), O (f).

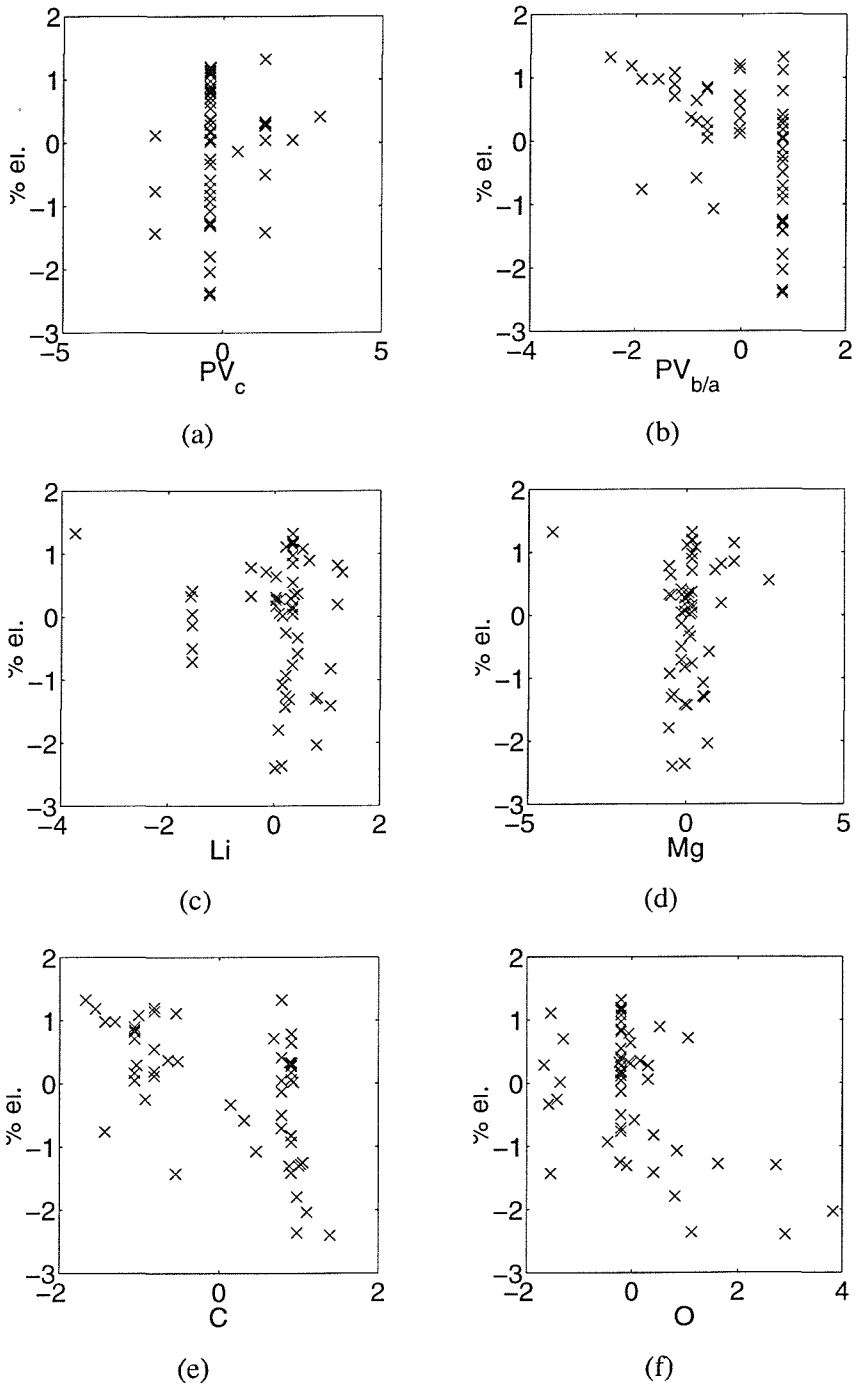


Figure A.7: Full residual plots between $\%el.$ and the input variables comprising the Al-Mg-Li data set: PV_c (a), $PV_{b/a}$ (b), Li (c), Mg (d), C (e), O (f).

Appendix B

Data Modelling of Structure-Properties of Experimental Trials in the Al-Zn-Mg-Cu Alloy System

B.1 Statistical summary of datasets

x	μ_x	σ_x
$x_{Zn,w}$ (wt.%)	6.09	0.648
$x_{Mg,w}$ (wt.%)	2.32	0.377
$x_{Cu,w}$ (wt.%)	1.87	0.492
$x_{Zr,w}$ (wt.%)	0.12	0.002
$x_{Fe,w}$ (wt.%)	0.06	0.009
$x_{Si,w}$ (wt.%)	0.02	0.001
t (hours)	7.50	7.663

Table B.1: Mean and standard deviations for the input variables comprising data set A.

y	μ_y	σ_y
$\sigma_{0.2}$ (MPa)	478.69	38.61
σ_{el} (%IACS)	38.92	2.56

Table B.2: Mean and standard deviations for the output properties of the Al-Zn-Mg-Cu data.

B.2 Neurofuzzy model construction

Ref. no.	Inputs	Refinement	MSE	SS
1	t	ua(0)	1174	3486
2	$x_{Mg,w}$	ua(0)	872	3128
3	$x_{Zn,w}$	ua(0)	637	2779
4	$t \times x_{Mg,w}$	tp(0,0)	471	2542
5	$x_{Cu,w}$	ua(0)	351	2395
6a	t, $x_{Mg,w}$	ts(0,0)	513	2764
6b	$x_{Cu,w}$	sd	471	2542
7b	$x_{Mg,w}$	sd	694	3031
6c	$x_{Zn,w}$	ki(0)	294	2634
7c	$x_{Zn,w}$	kd(0.5)	282	3539
7d	$x_{Cu,w}$	ro	471	2542
8d	$x_{Zn,w}$	ro	694	3031

Table B.3: Summary of the iterative model search performed by the ASMOD algorithm in determining the $\sigma_{0.2}$ model from data set A.

Ref. no.	Inputs	Refinement	MSE	SS
1	$x_{Mg,w}$	ua(0)	5.317	15.79
2	t	ua(2)	0.292	1.99
3a	$x_{Zn,w}$	ua(0)	0.225	2.02
4a	$x_{Cu,w}$	ua(0)	0.205	2.58
3b	t	sd	5.317	15.79
3c	$x_{Mg,w}$	ki(0)	0.292	2.61
4c	$x_{Mg,w}$	ki(-0.5)	0.291	3.64
3	t	kd(0.5)	0.292	1.57
4	t	kd(0)	0.296	1.29
5d	t	kd(-0.5)	2.907	10.43
5e	$x_{Mg,w}$	ro	3.74	13.42
6e	t	ro	6.22	18.47

Table B.4: Summary of the iterative model search performed by the ASMOD algorithm in determining the σ_{el} model from data set A.

Ref. no.	Inputs	Refinement	MSE	SS
1	t	ua(0)	1174	3486
2	$x_{Mg,\alpha}$	ua(0)	853	3061
3	$x_{\eta'}$	ua(0)	602	2627
4	$t \times x_{Mg,\alpha}$	tp(0,0)	454	2444
5a	$x_{Cu,\alpha}$	ua(0)	364	2483
6a	t , $x_{Mg,\alpha}$	ts(0,0)	507	2732
5b	$x_{\eta'}$	sd	691	3017
5c	$x_{\eta'}$	ki(0)	431	2936
6c	$x_{\eta'}$	ki(0.5)	414	3708
5d	$x_{\eta'}$	ro	691	3017
6d	$x_{Mg,\alpha}$	ro	1174	3486

Table B.5: Summary of the iterative model search performed by the ASMOD algorithm in determining the $\sigma_{0.2}$ model from data set B.

Ref. no.	Inputs	Refinement	MSE	SS
1	$x_{Mg,xs}$	ua(0)	5.248	15.58
2	t	ua(2)	0.219	1.49
3a	$x_{\eta'}$	ua(0)	0.174	1.56
4a	$x_{Cu,\alpha}$	ua(0)	0.147	1.84
3b	t	sd	5.248	15.58
3c	$x_{Mg,xs}$	ki(0)	0.218	1.95
4c	$x_{Mg,xs}$	ki(-0.5)	0.207	2.59
3	t	kd(0.5)	0.218	1.18
4	t	kd(0)	0.223	0.97
5d	t	kd(-0.5)	2.826	10.14
5e	$x_{Mg,xs}$	ro	3.741	13.42
6e	t	ro	6.220	18.47

Table B.6: Summary of the iterative model search performed by the ASMOD algorithm in determining the σ_{el} model from data set B.

Ref. no.	Inputs	Refinement	MSE	SS
1	$x_{Zn,w} + x_{Mg,w}$	ua(0)	1172	3481
2	t	ua(0)	818	2933
3	$x_{Zn,w} : x_{Mg,w}$	ua(0)	623	2720
4	$(x_{Zn,w} : x_{Mg,w}) \times t$	tp(0,0)	392	2114
5	$(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$	ua(0)	279	1899
6a	$((x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}) \times (x_{Zn,w} + x_{Mg,w})$	tp(0,0)	274	2456
7a	$(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$, $x_{Zn,w} + x_{Mg,w}$	ts(0,0)	279	1899
6b	$(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$	sd	392	2114
7b	$(x_{Zn,w} : x_{Mg,w}) \times t$	sd	1172	3481
6c	$(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$	ki(0)	250	2242
7c	$x_{Zn,w} + x_{Mg,w}$	ki(0)	237	2974
6d	$(x_{Zn,w} + x_{Cu,w}) : x_{Mg,w}$	ro	392	2114
7d	$x_{Zn,w} : x_{Mg,w}$	ro	818	2933

Table B.7: Summary of the iterative model search performed by the ASMOD algorithm in determining the $\sigma_{0.2}$ model from data set C.

Ref. no.	Inputs	Refinement	MSE	SS
1	$x_{Zn,w} : x_{Mg,w}$	ua(0)	5.470	16.24
2	t	ua(2)	1.034	7.05
3	$x_{Zn,w} + x_{Mg,w}$	ua(0)	0.302	2.70
4	$(x_{Zn,w} : x_{Mg,w}) \times (x_{Zn,w} + x_{Mg,w})$	tp(0,0)	0.205	2.57
5a	$x_{Zn,w} : x_{Mg,w}$, $x_{Zn,w} + x_{Mg,w}$	ts(0,0)	0.302	2.70
6a	$(x_{Zn,w} : x_{Mg,w}) \times (x_{Zn,w} + x_{Mg,w})$	sd	3.733	20.12
5b	t	ki(-0.75)	0.205	4.06
6b	t	kd(0.875)	0.195	8.38
5	t	kd(0.5)	0.205	1.84
6	t	kd(-0.5)	0.206	1.40
7c	$x_{Zn,w} + x_{Mg,w}$	ro	1.035	4.52
8c	$x_{Zn,w} : x_{Mg,w}$	ro	3.74	13.42

Table B.8: Summary of the iterative model search performed by the ASMOD algorithm in determining the σ_{el} model from data set C.

B.3 Pairwise scatterplots

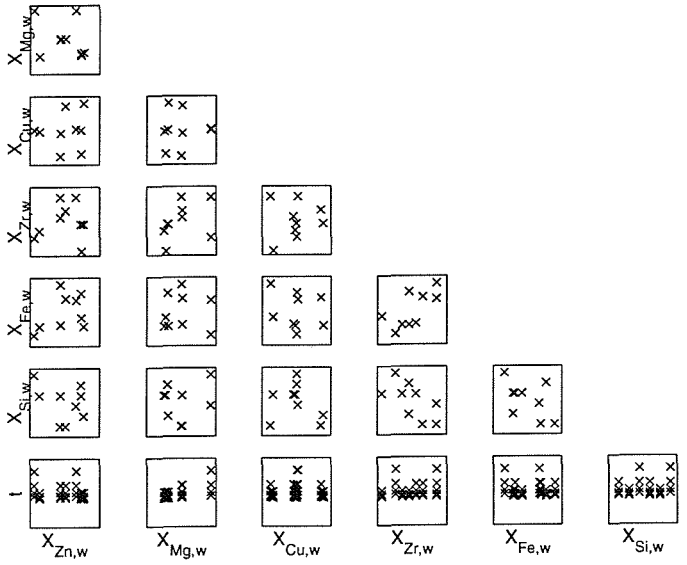


Figure B.1: Pairwise scatter plots between the input variables comprising data set A.

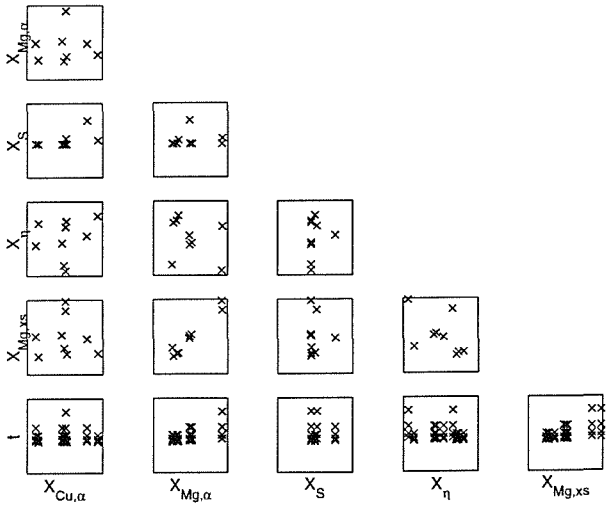


Figure B.2: Pairwise scatter plots between the input variables comprising data set B.

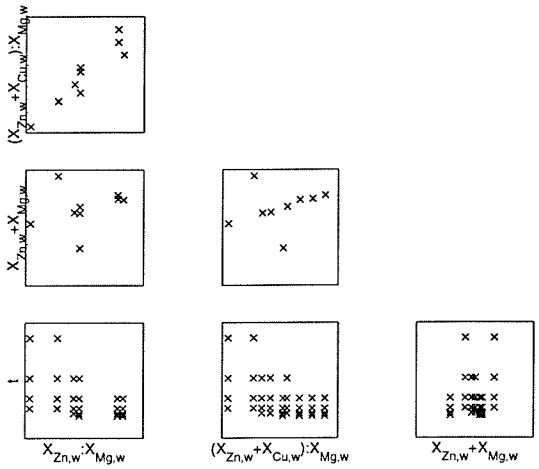


Figure B.3: Pairwise scatter plots between the input variables comprising data set C.

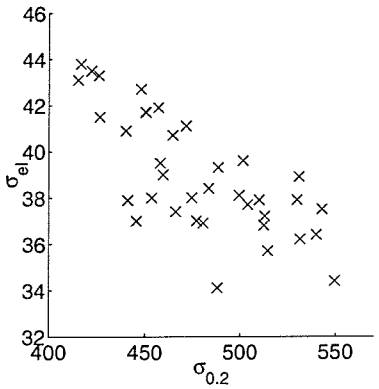


Figure B.4: Scatter plot between $\sigma_{0.2}$ and σ_{el}.

B.4 Residual scatter plots

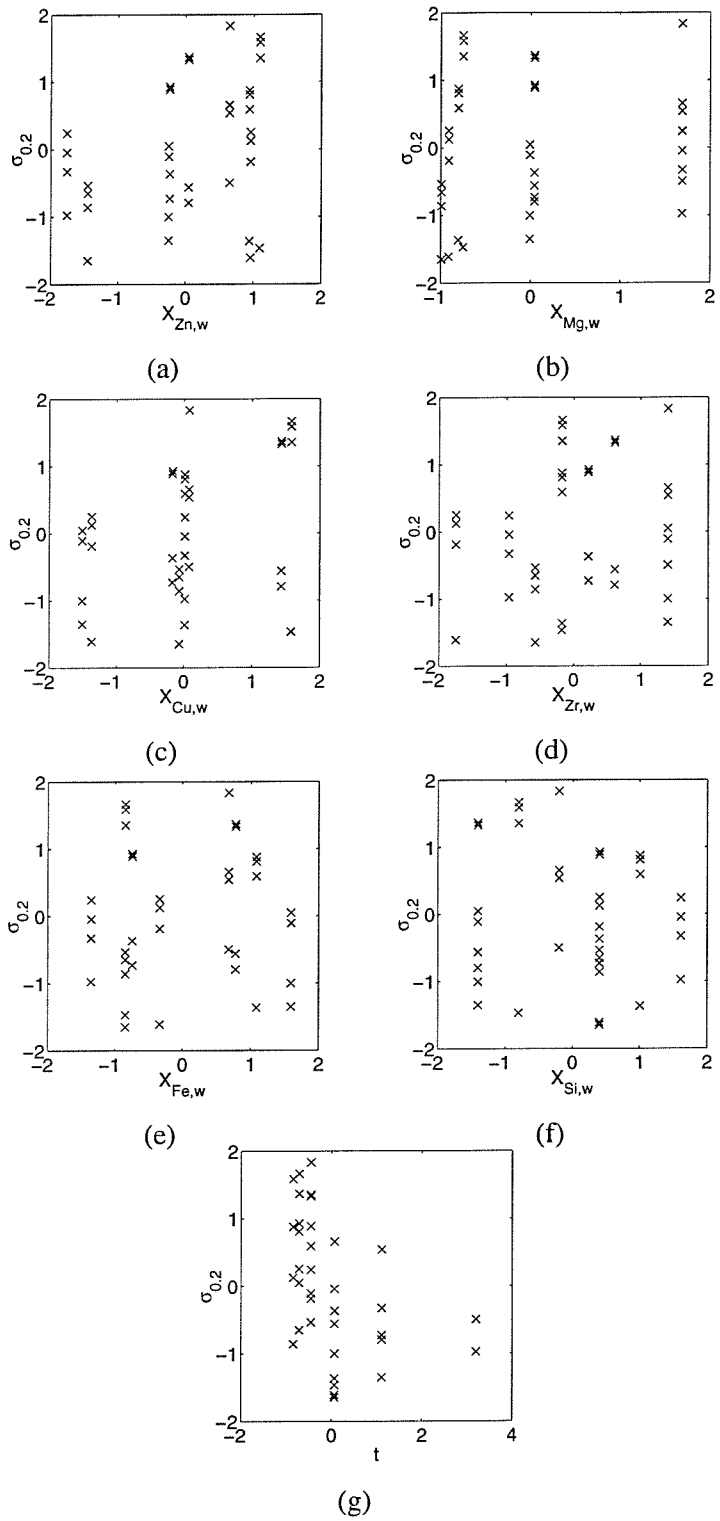


Figure B.5: Full residual plots between $\sigma_{0.2}$ and the input variables comprising data set A: $x_{Zn,w}$ (a), $x_{Mg,w}$ (b), $x_{Cu,w}$ (c), $x_{Zr,w}$ (d), $x_{Fe,w}$ (e), $x_{Si,w}$ (f), ageing time, t (g).

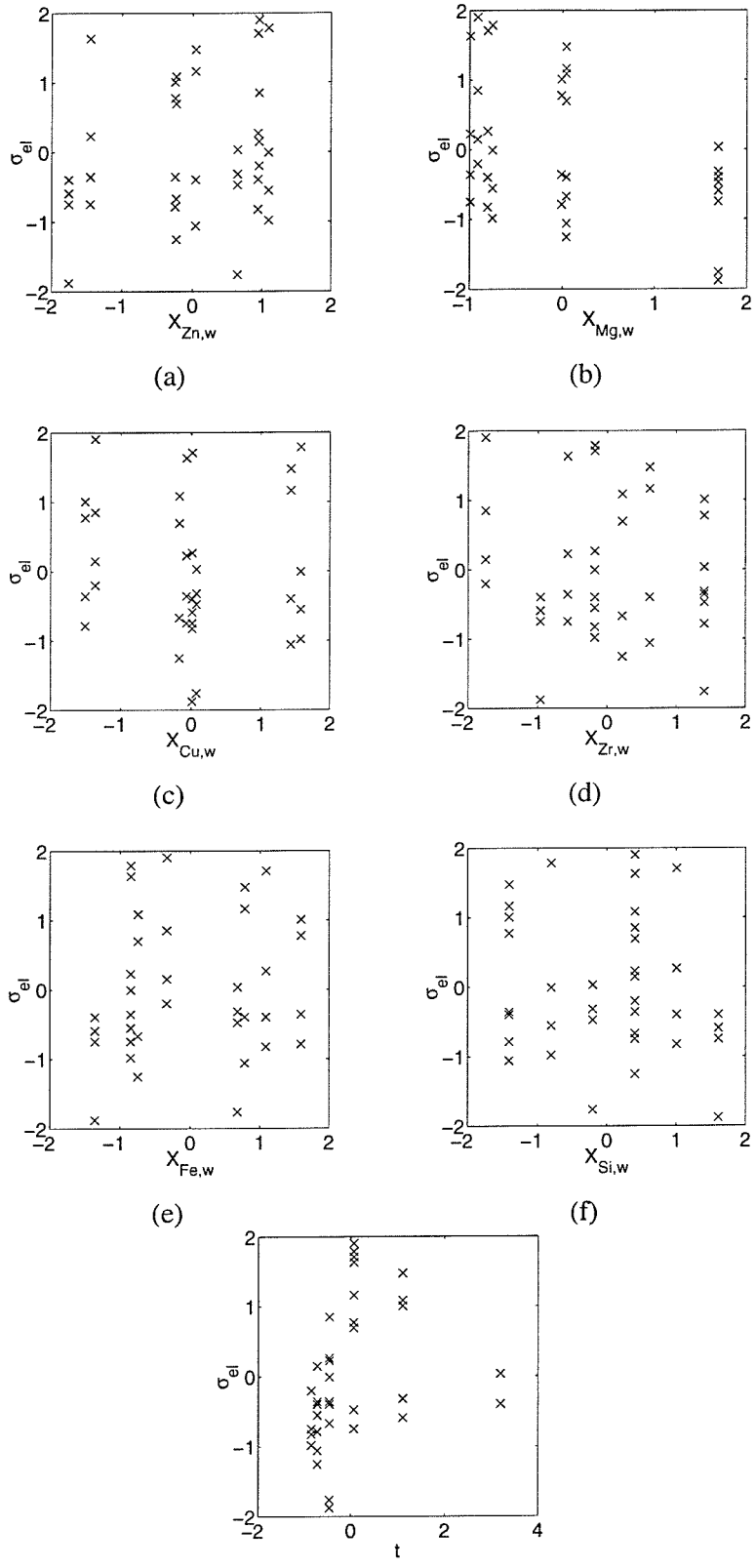


Figure B.6: Full residual plots between σ_{el} and the input variables comprising data set A: $x_{Zn,w}$ (a), $x_{Mg,w}$ (b), $x_{Cu,w}$ (c), $x_{Zr,w}$ (d), $x_{Fe,w}$ (e), $x_{Si,w}$ (f), ageing time, t (g).

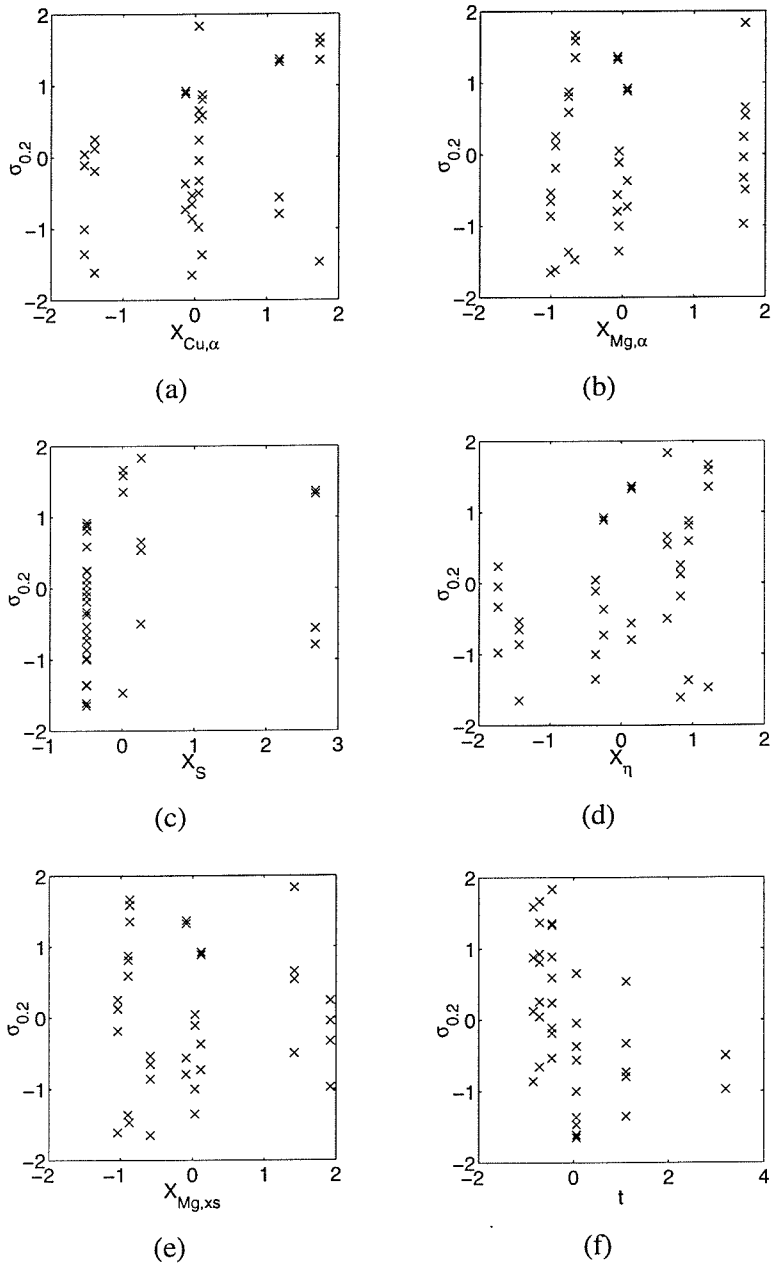


Figure B.7: Full residual plots between $\sigma_{0.2}$ and the input variables comprising data set B: $x_{Cu,\alpha}$ (a), $x_{Mg,\alpha}$ (b), x_S (c) ,$x_{\eta'}$ (d), $x_{Mg,xs}$ (e), ageing time, t (f).

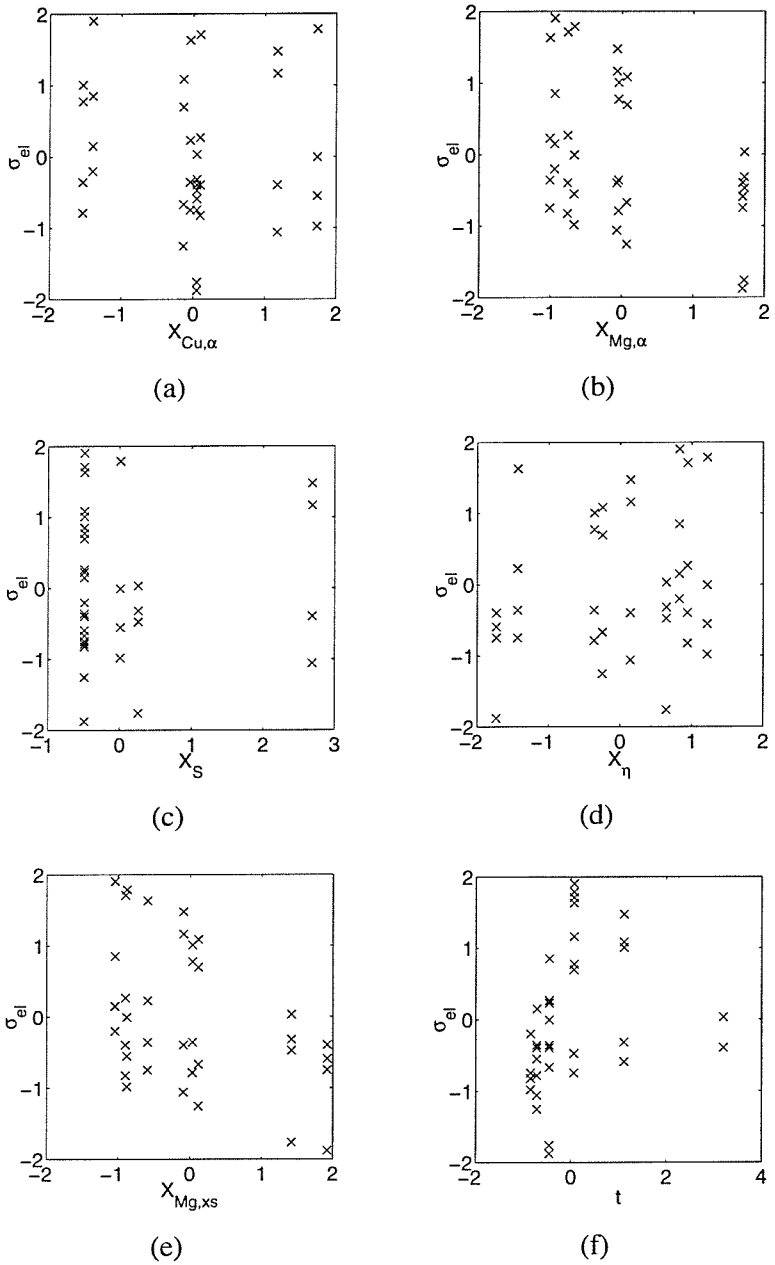


Figure B.8: Full residual plots between σ_{el} and the input variables comprising data set B: $x_{Cu,\alpha}$ (a), $x_{Mg,\alpha}$ (b), x_S (c) ,$x_{\eta'}$ (d), $x_{Mg,xs}$ (e), ageing time, t (f).

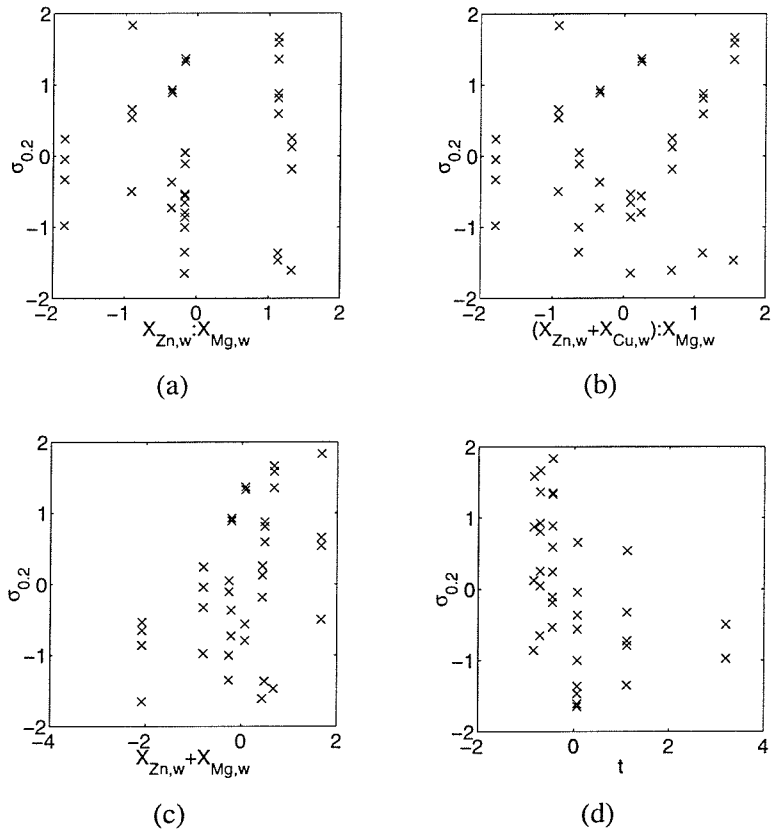


Figure B.9: Full residual plots between $\sigma_{0.2}$ and the input variables comprising data set C: $x_{Zn,w}$: $x_{Mg,w}$ (a), $(x_{Zn,w} + x_{Cu,w})$: $x_{Mg,w}$ (b), $x_{Zn,w} + x_{Cu,w}$ (c), ageing time, t (d).

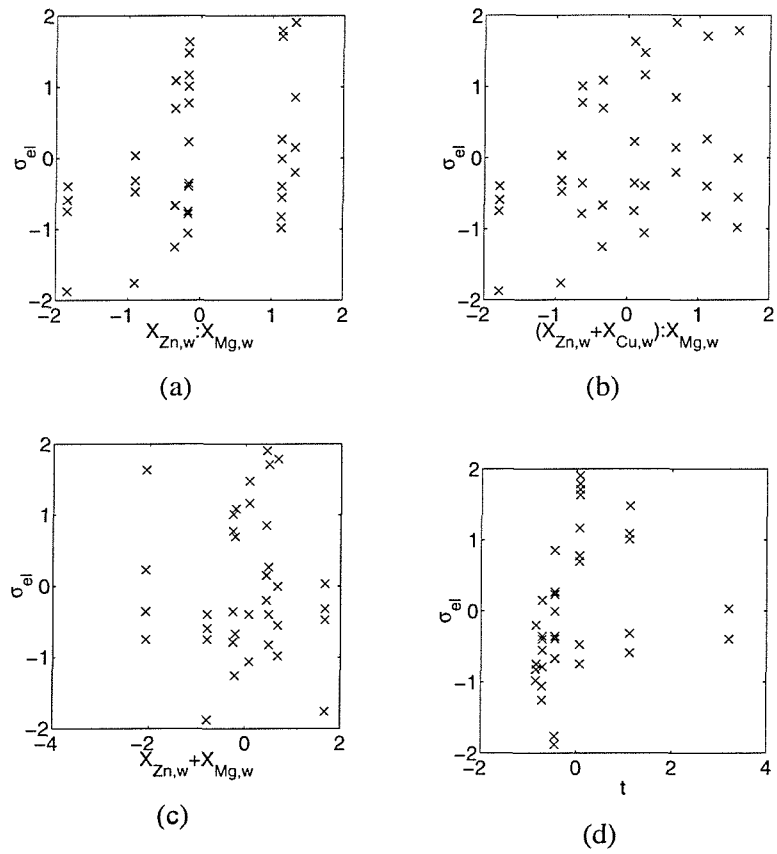
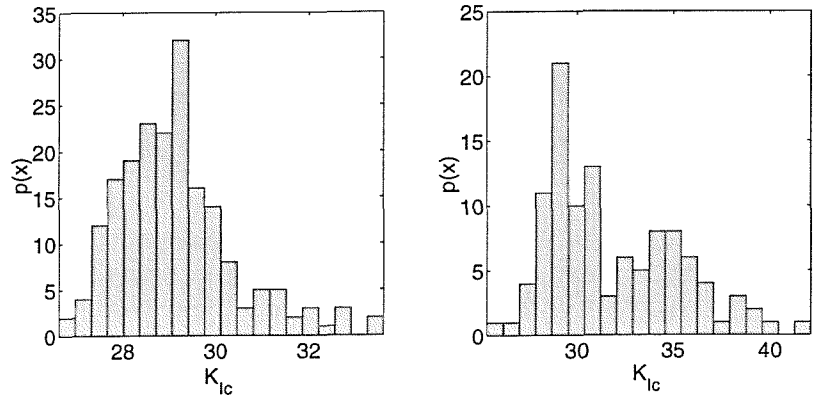


Figure B.10: Full residual plots between σ_{el} and the input variables comprising data set C: $x_{Zn,w}$: $x_{Mg,w}$ (a), $(x_{Zn,w} + x_{Cu,w})$: $x_{Mg,w}$ (b), $x_{Zn,w} + x_{Cu,w}$ (c), ageing time, t (d).

Appendix C

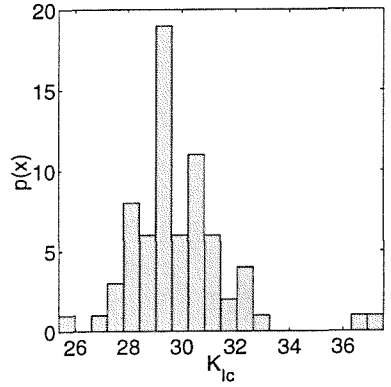
Knowledge Discovery and Data Mining of 7xxx Series Al-Alloy Production Databases

C.1 Histogram plots

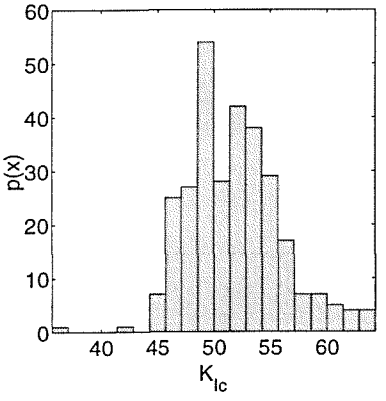


BAP1TLC3W (a)

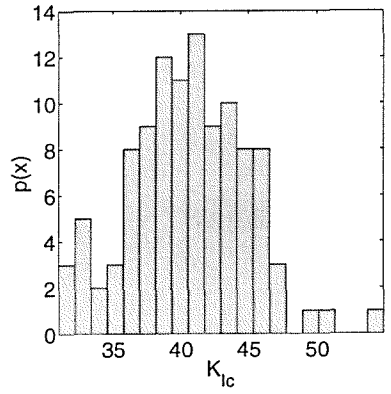
BAP1TLQ3W (b)



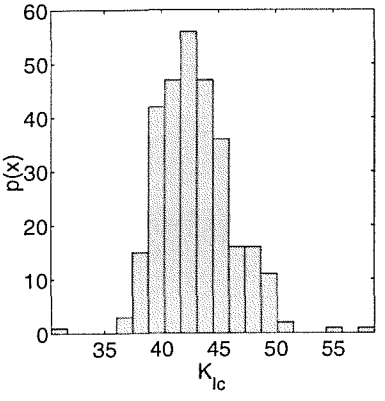
BAP1TLQ3W2580 (c)



BAP4LTC5W (d)



BAP4SLC5W (e)



BAP4TLC5W (f)

Figure C.1: Histogram plots for K_{Ic} for the 7xxx series data sets considered.

C.2 Residual scatter plots

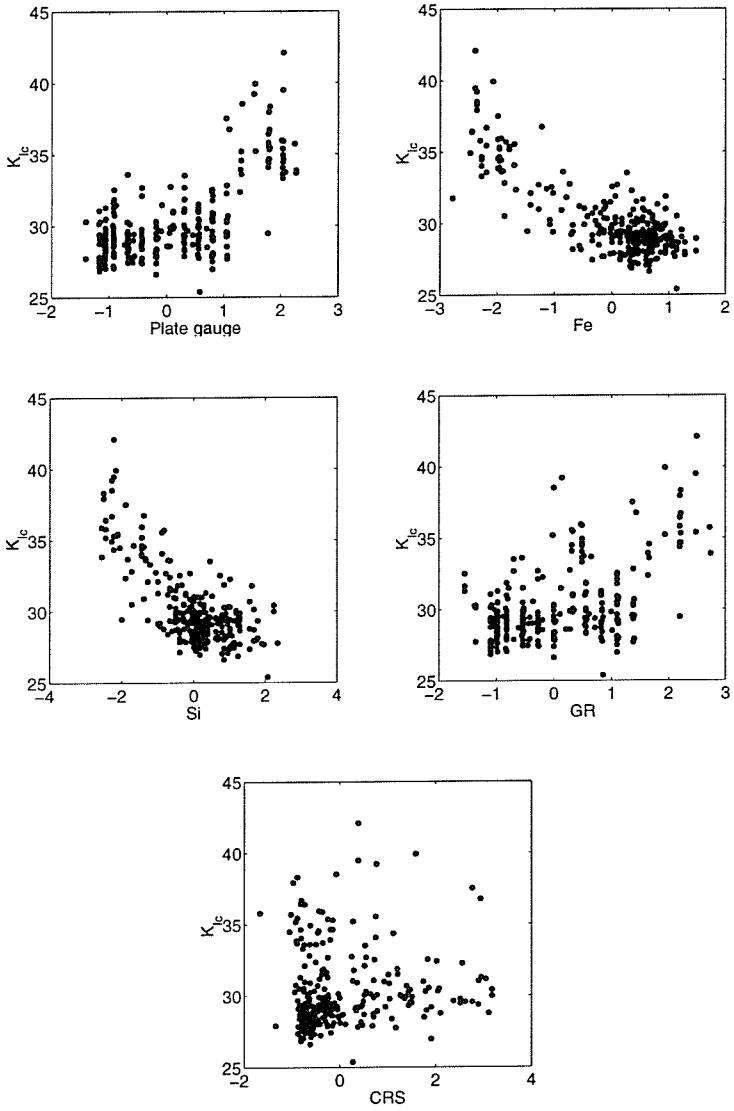


Figure C.2: Full residual scatterplots between K_{Ic} and a selected number of processing variables for the BAP1 data set.

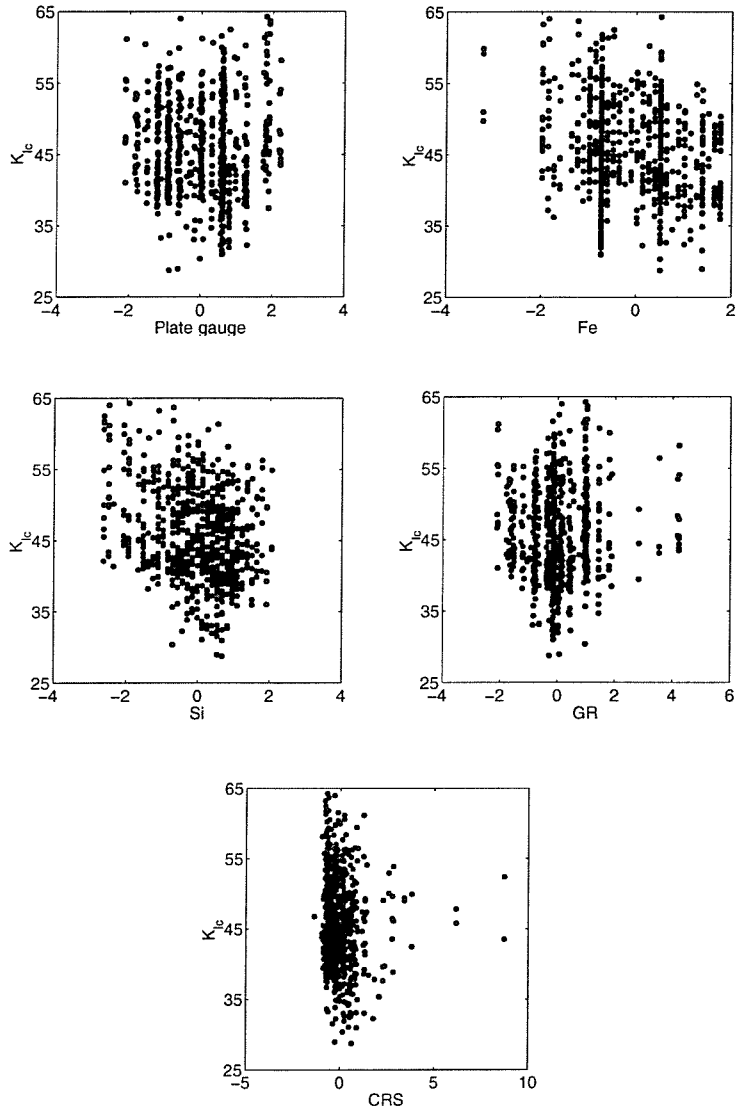


Figure C.3: Full residual scatterplots between K_{Ic} and a selected number of processing variables for the BAP4 data set.

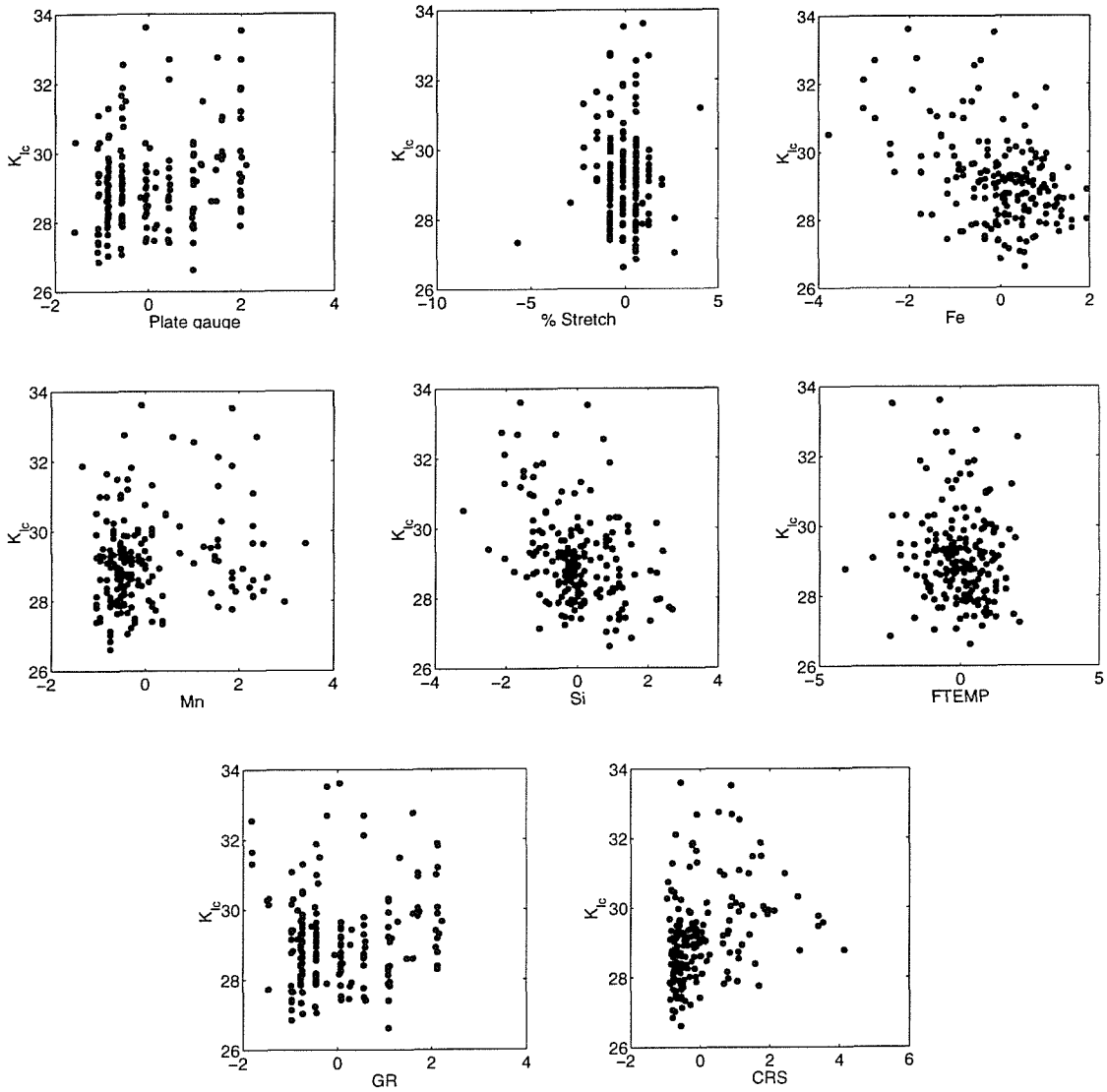


Figure C.4: Full residual scatterplots between K_{Ic} and a selected number of processing variables for the BAP1LTC3W data set.

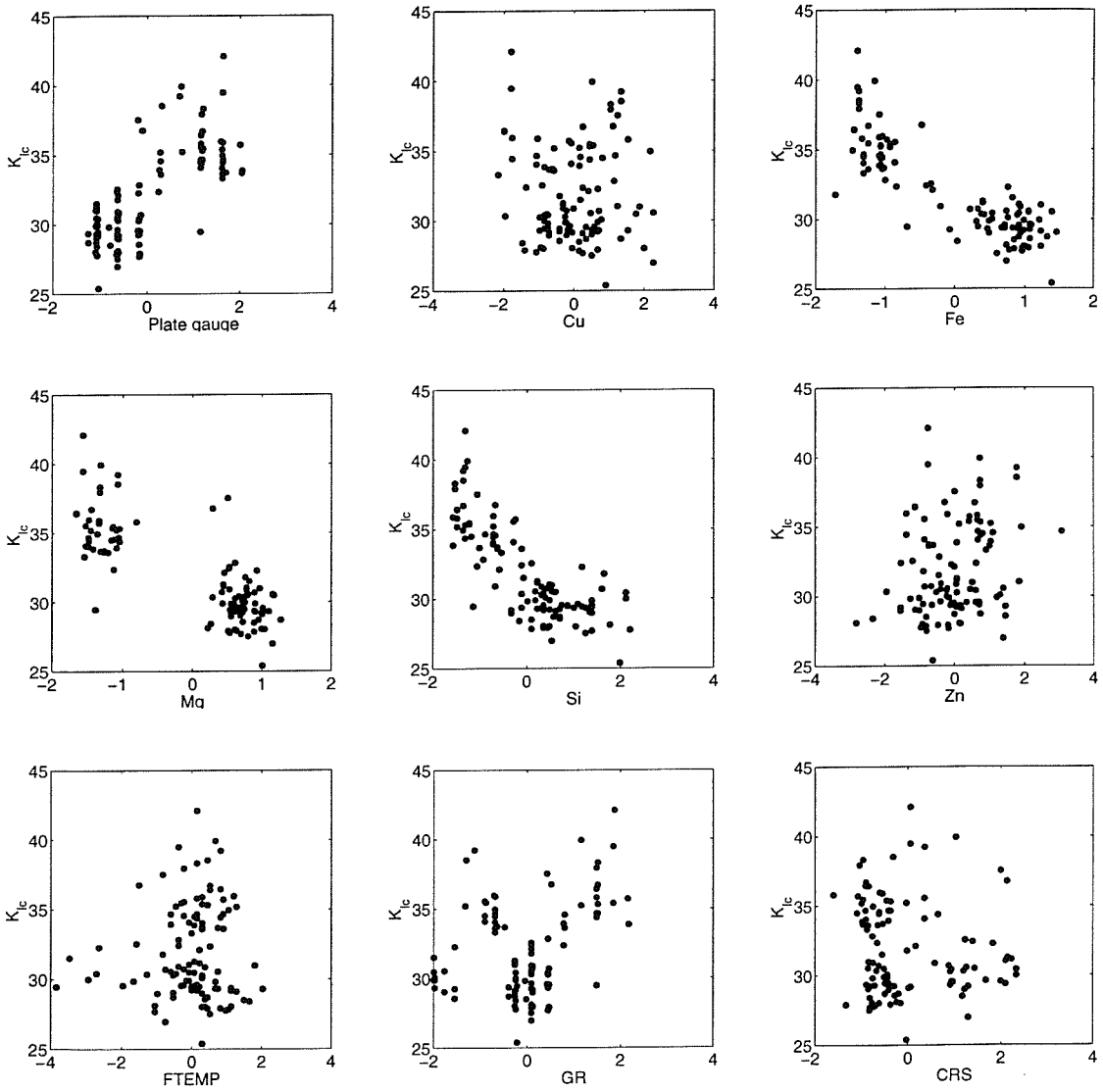


Figure C.5: Input scatterplots for a selected number of processing variables for the BAP1TLQ3W data set.

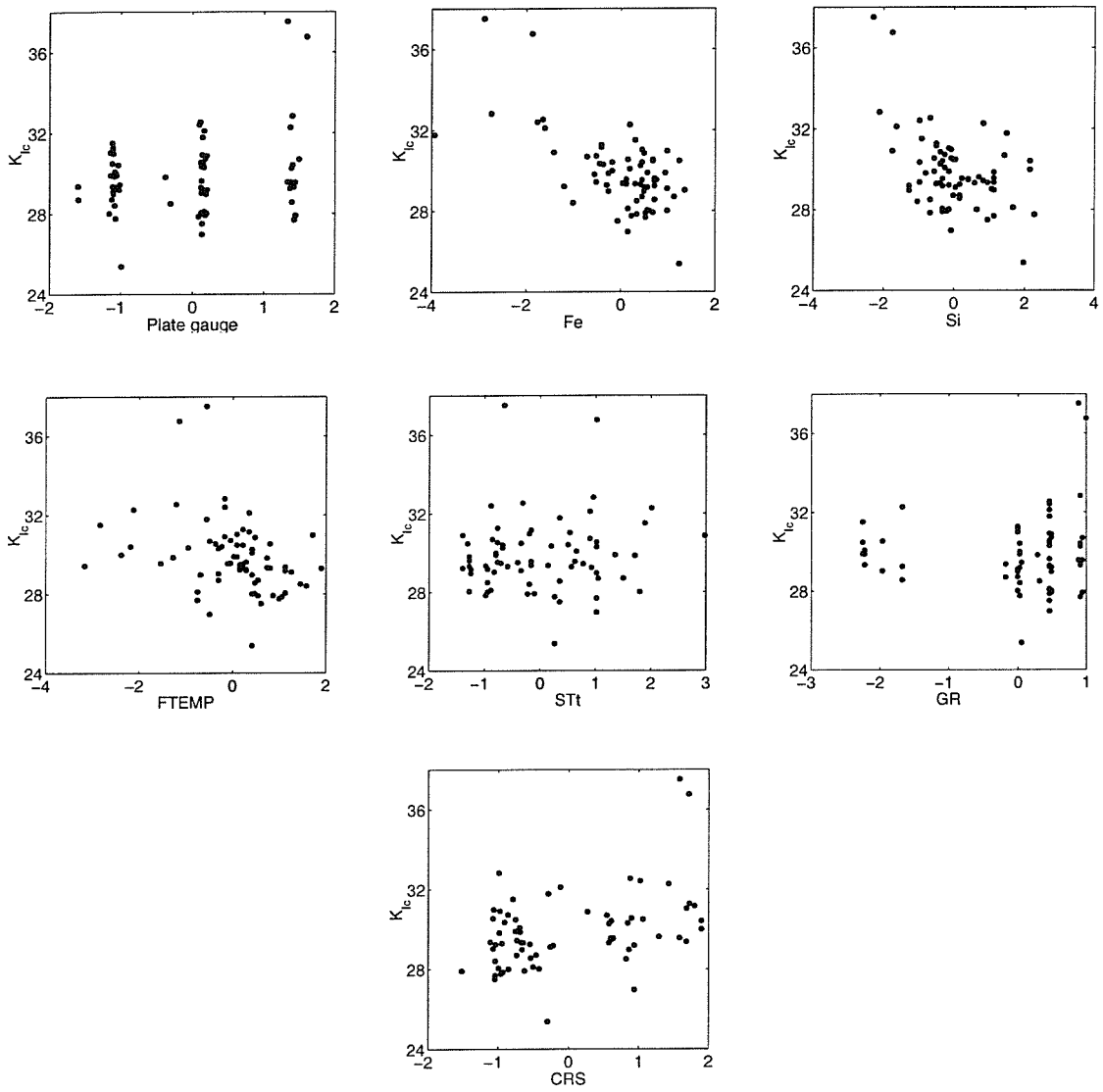


Figure C.6: Input scatterplots for a selected number of processing variables for the BAP1TLQ3W80 data set.

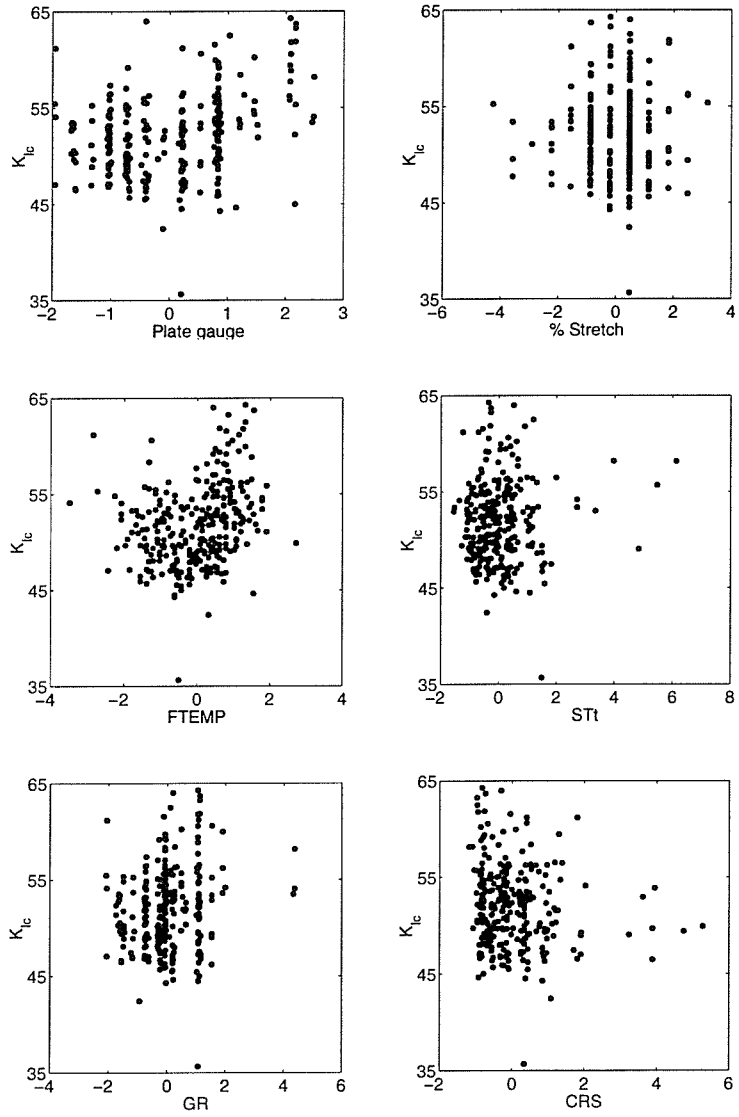


Figure C.7: Full residual scatterplots between K_{Ic} and the processing conditions for the BAP4LTC5W data set.

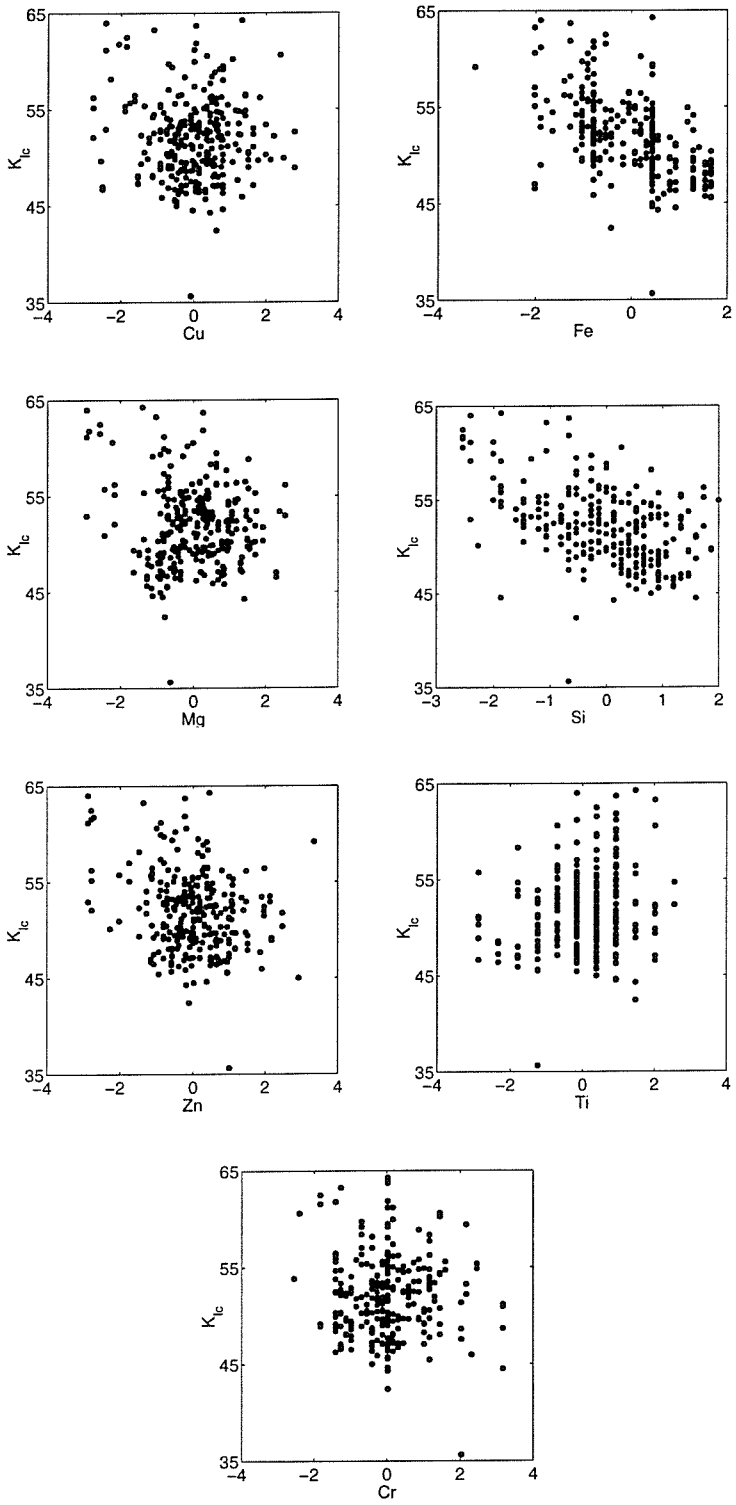


Figure C.8: Full residual scatterplots between K_{Ic} and compositional elements in the BAP4LTC5W data set.

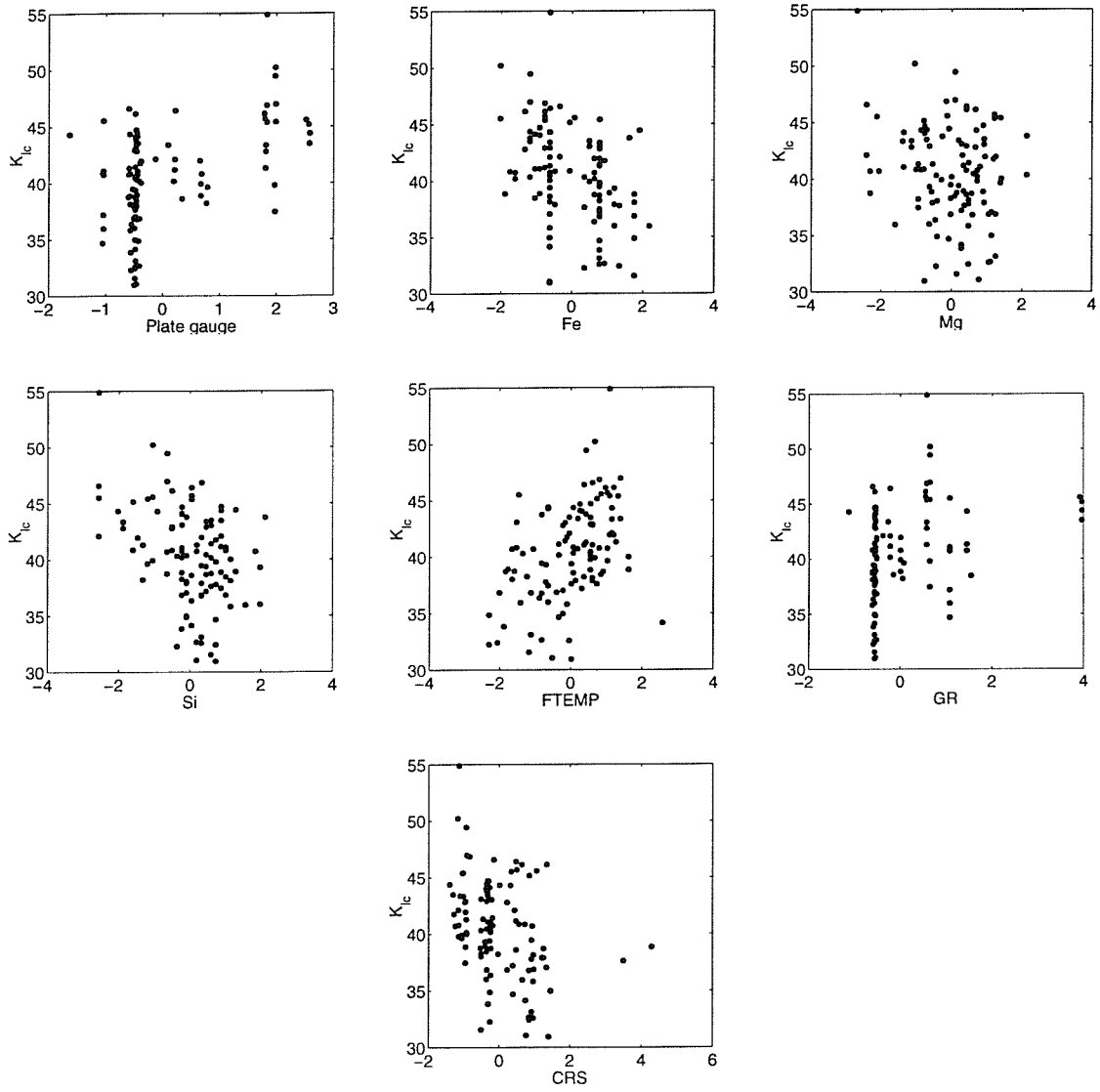


Figure C.9: Full residual scatterplots between K_{Ic} and a selected number of processing variables for the BAP4SLC5W data set.

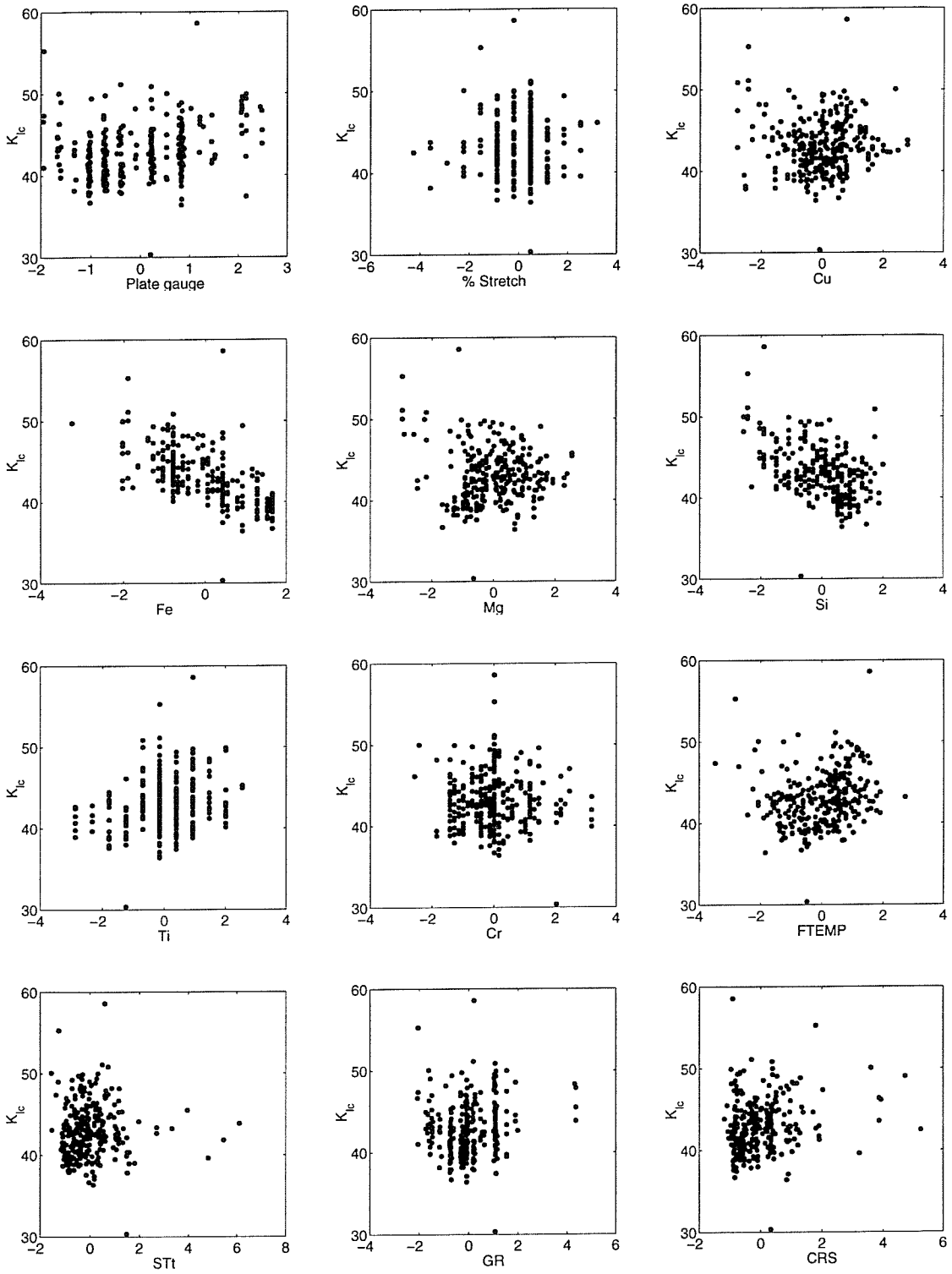


Figure C.10: Full residual scatterplots between K_{Ic} and a selected number of processing variables for the BAP4TLC5W data set.

C.3 Pairwise scatter plots

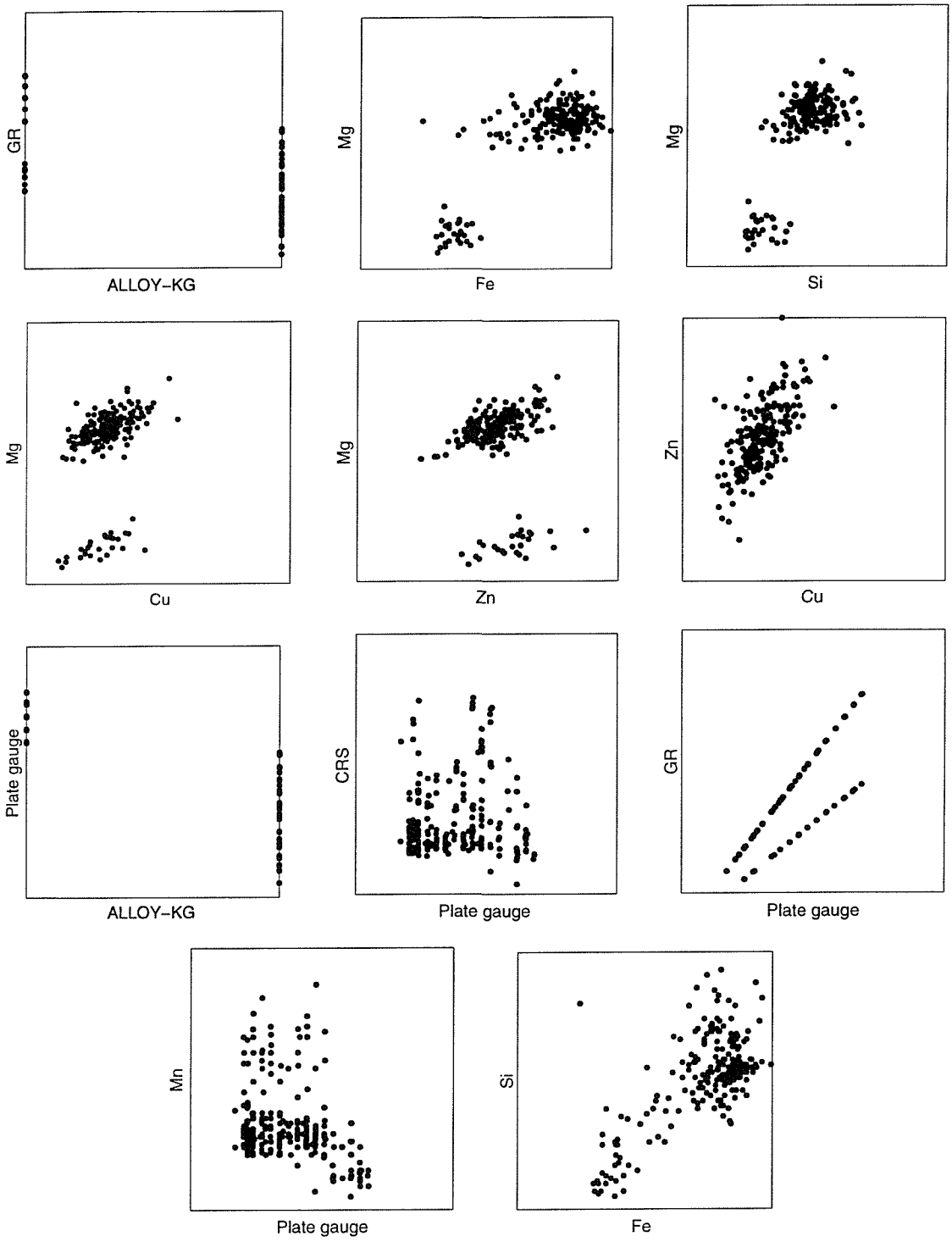


Figure C.11: Input scatterplots for a selected number of processing variables for the BAP1 data set.

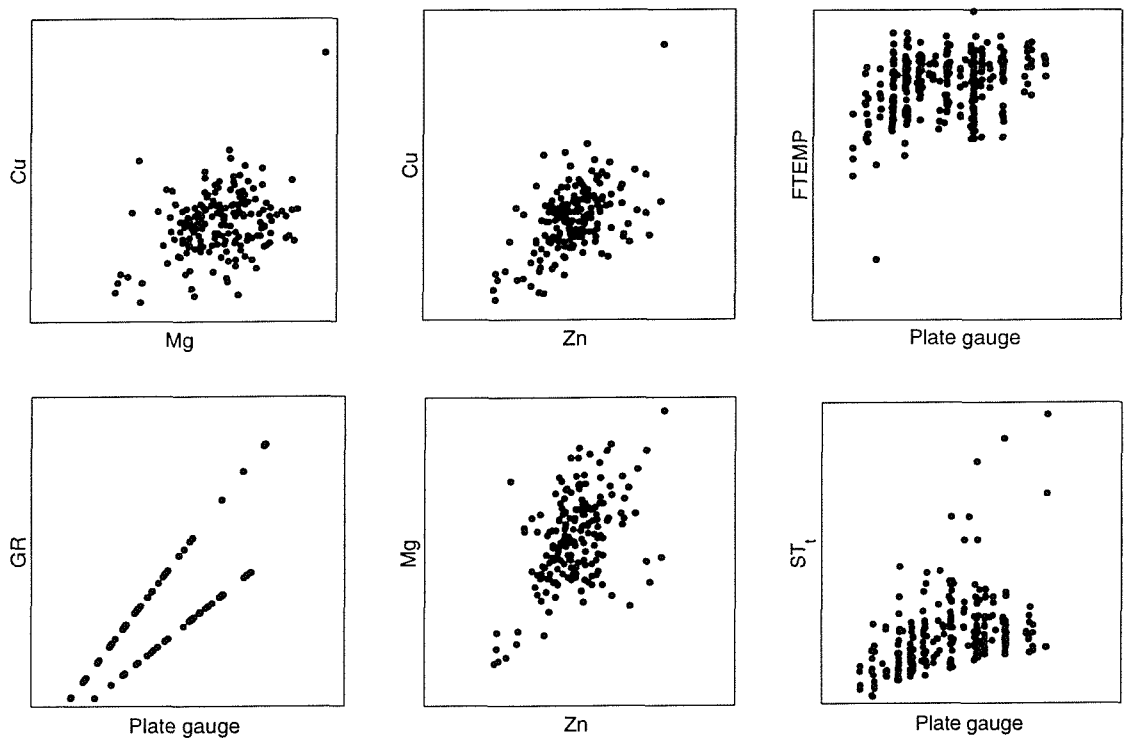


Figure C.12: Input scatterplots for a selected number of processing variables for the BAP4 data set.

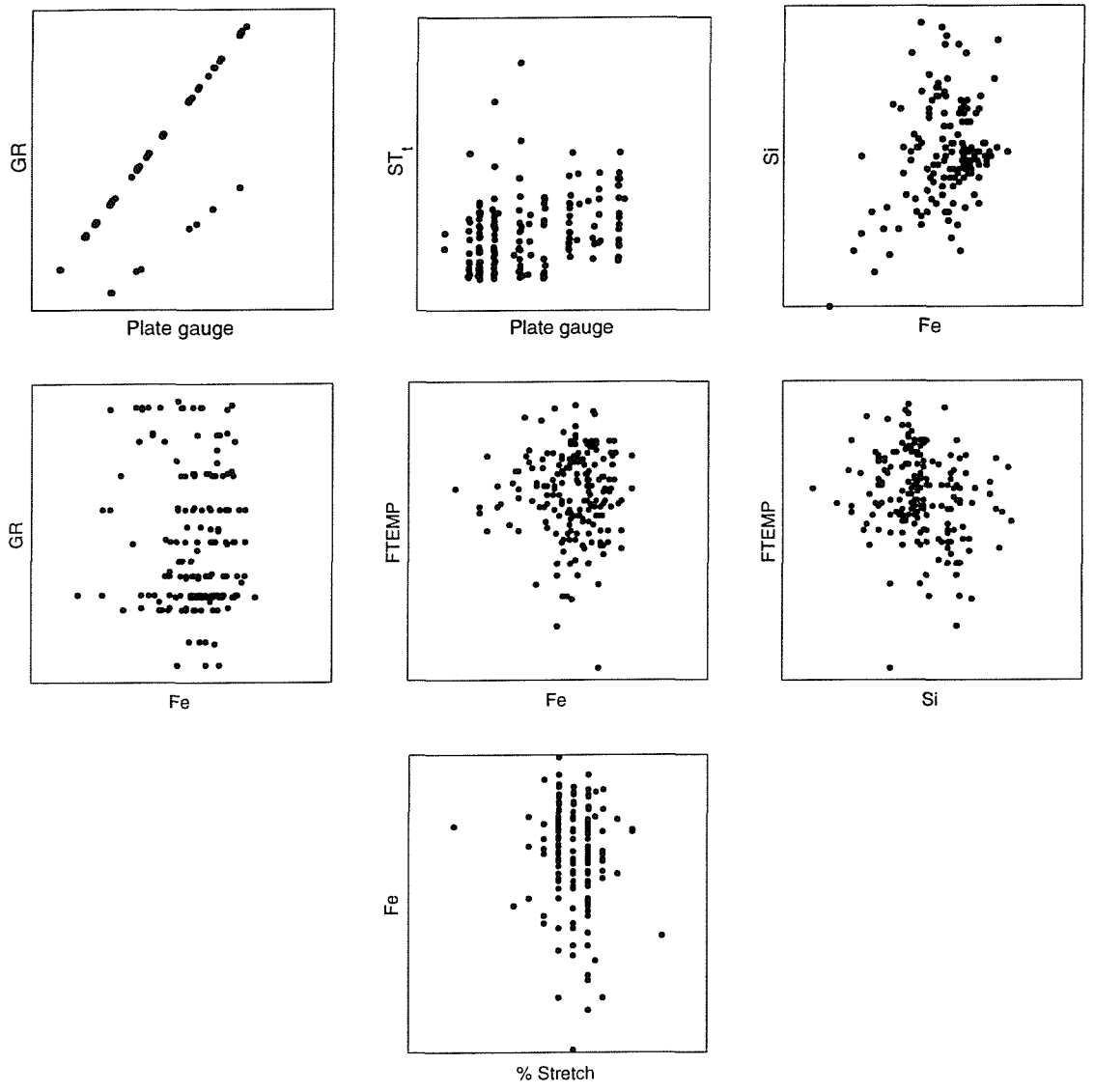


Figure C.13: Input scatterplots for a selected number of processing variables for the BAP1TLC3W data set.

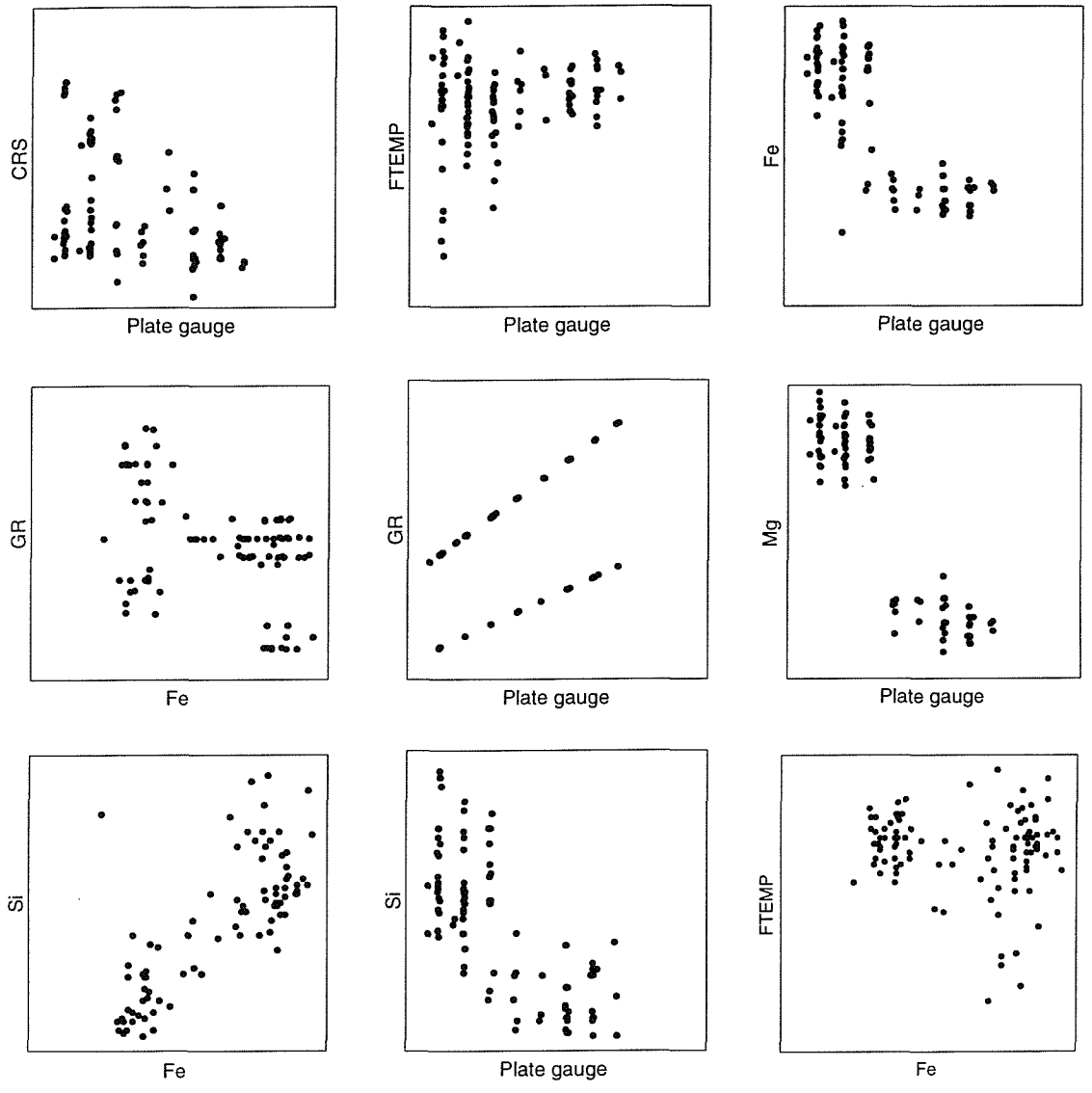


Figure C.14: Input scatterplots for a selected number of processing variables for the BAP1TLQ3W data set.

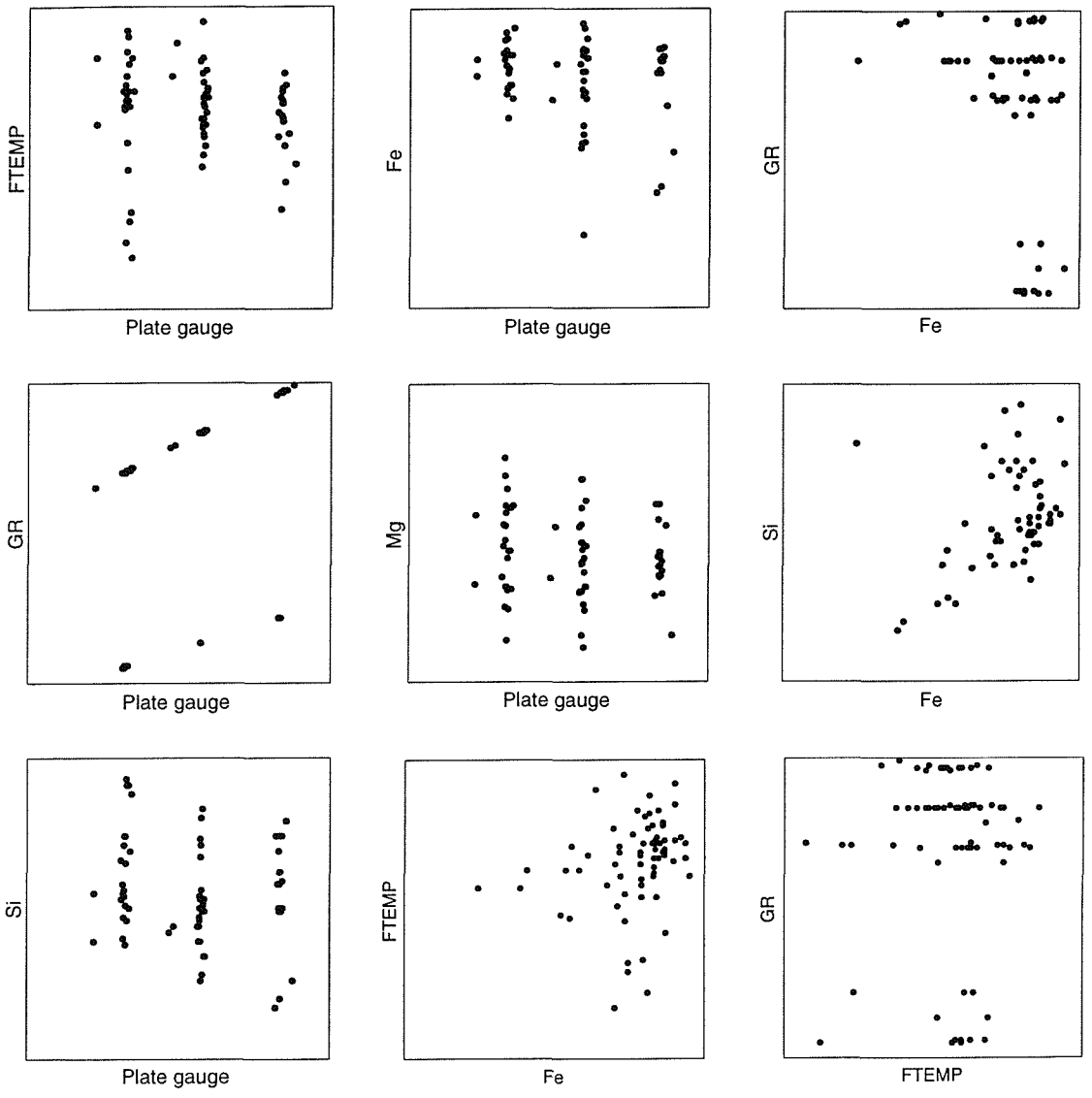


Figure C.15: Full residual scatterplots between K_{Ic} and a selected number of processing variables for the BAP1TLQ3W80 data set.

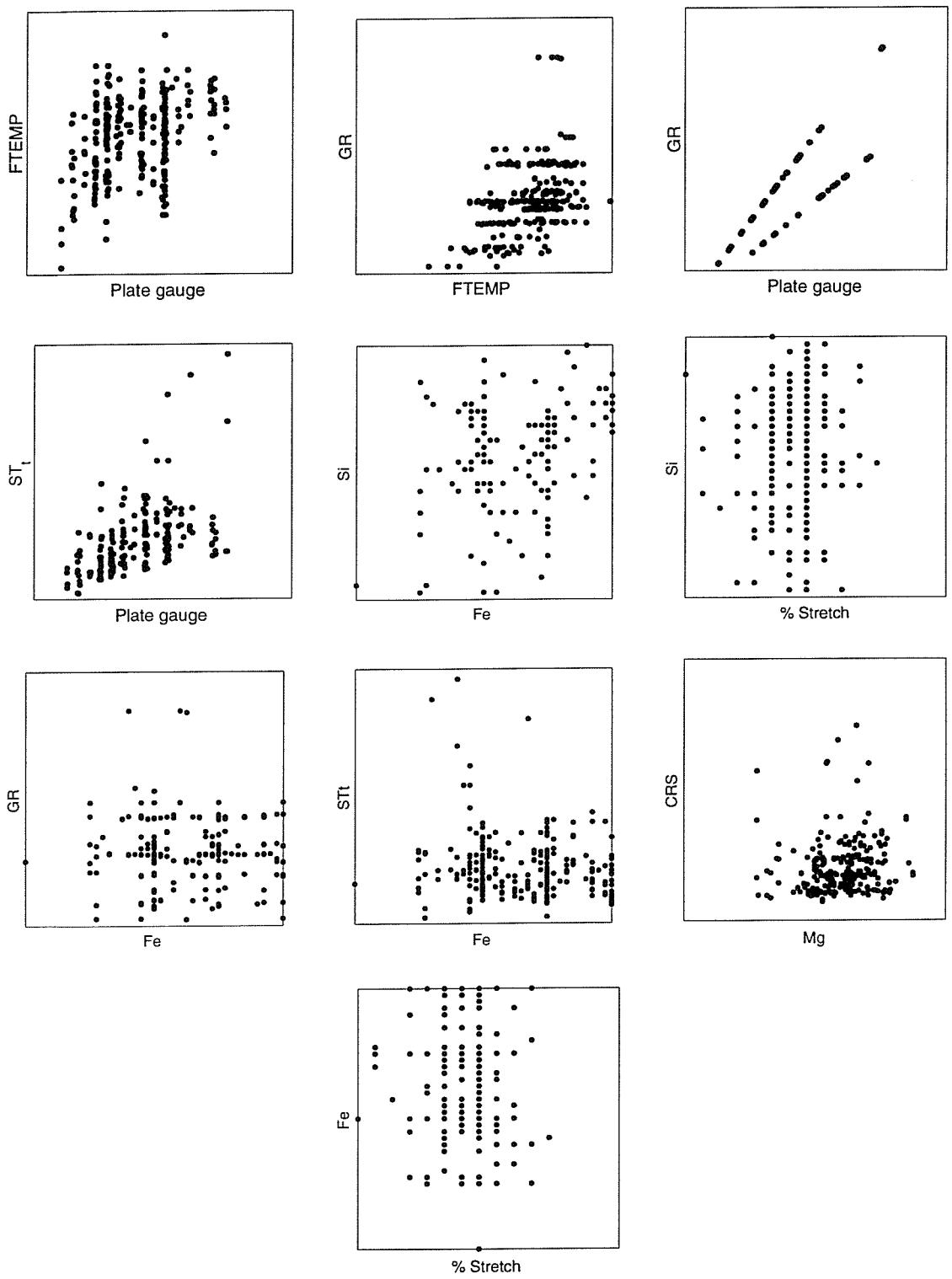


Figure C.16: Input scatterplots for a selected number of processing variables for the BAP4LTC5W data set.

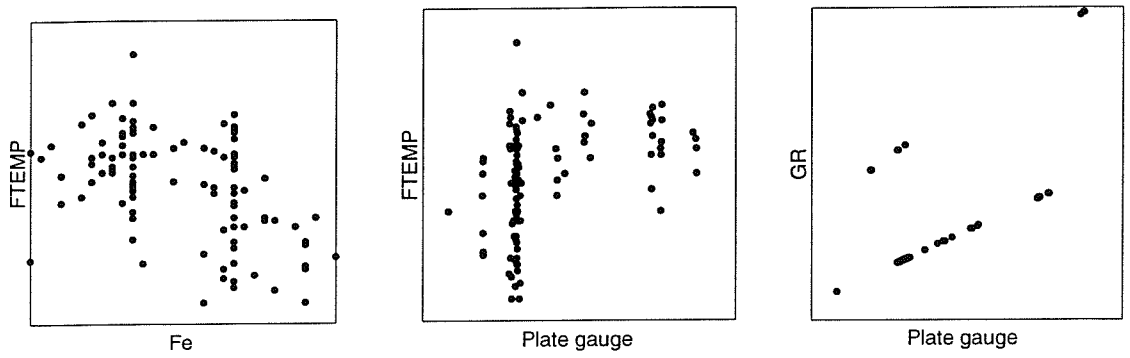


Figure C.17: Input scatterplots for a selected number of processing variables for the BAP4SLC5W data set.

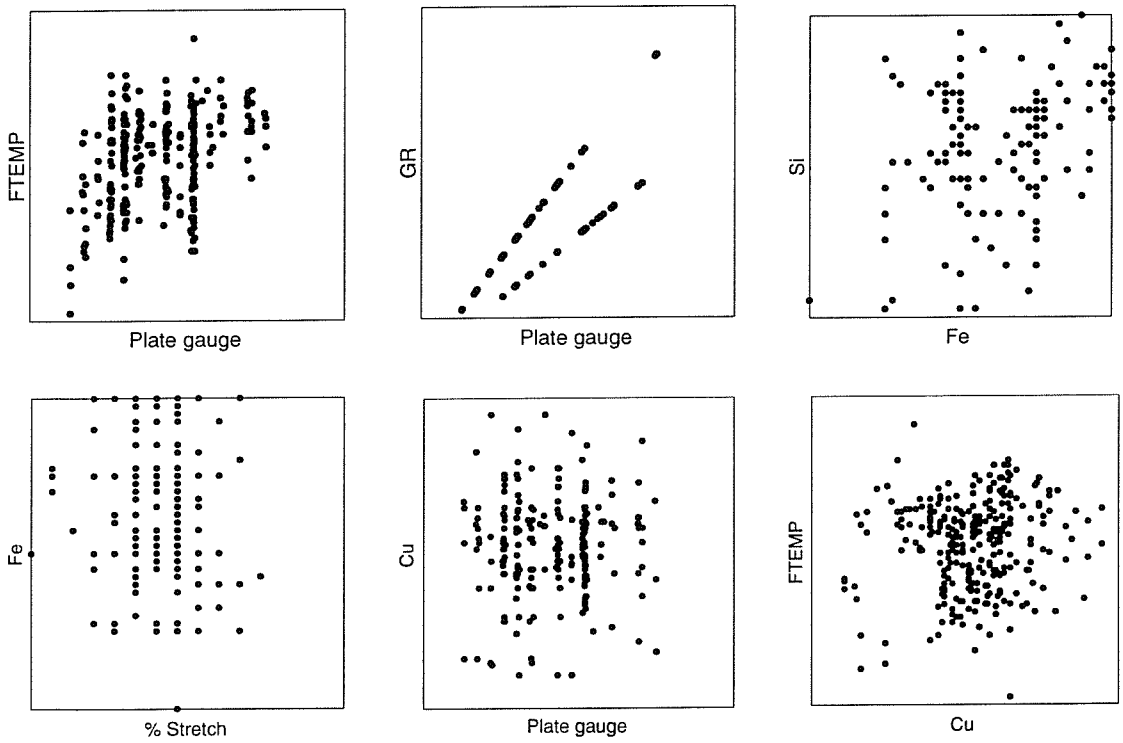


Figure C.18: Input scatterplots for a selected number of processing variables for the BAP4TLC5W data set.

C.4 SUPANOVA terms

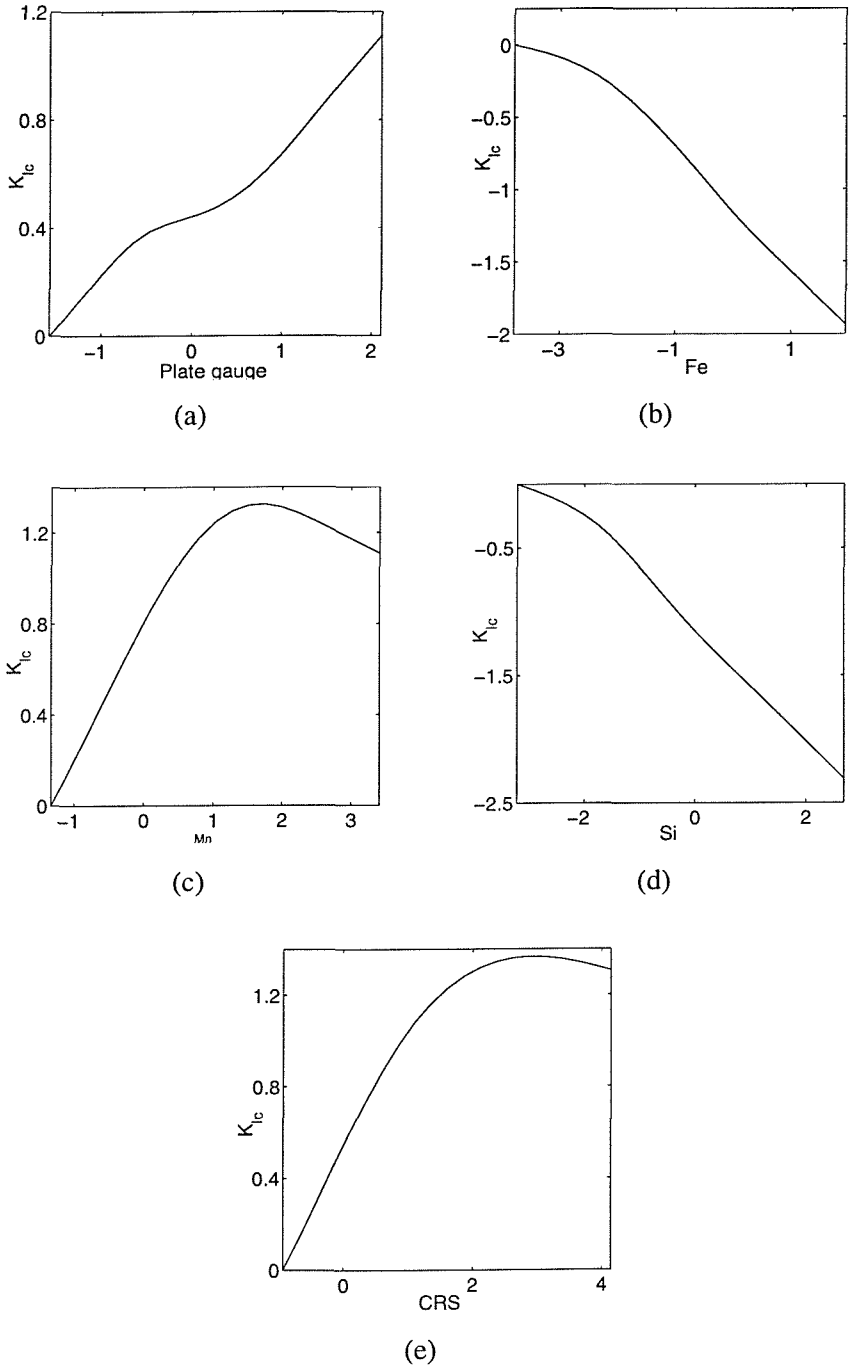


Figure C.19: General form of the univariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP1TLC3W data set.

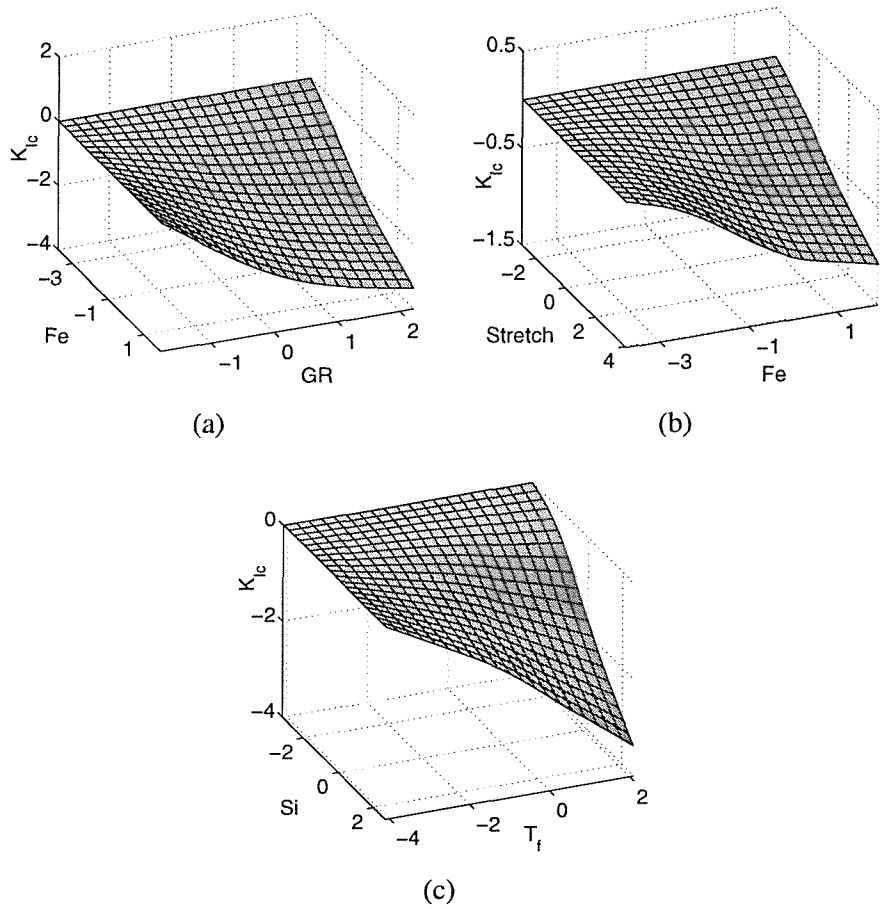


Figure C.20: General form of the bivariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP1TLC3W data set.

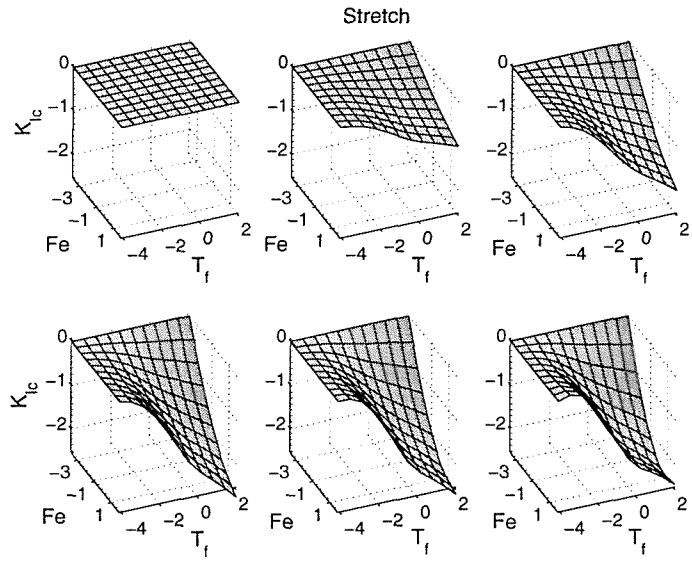


Figure C.21: General form of the trivariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP1TLC3W data set.

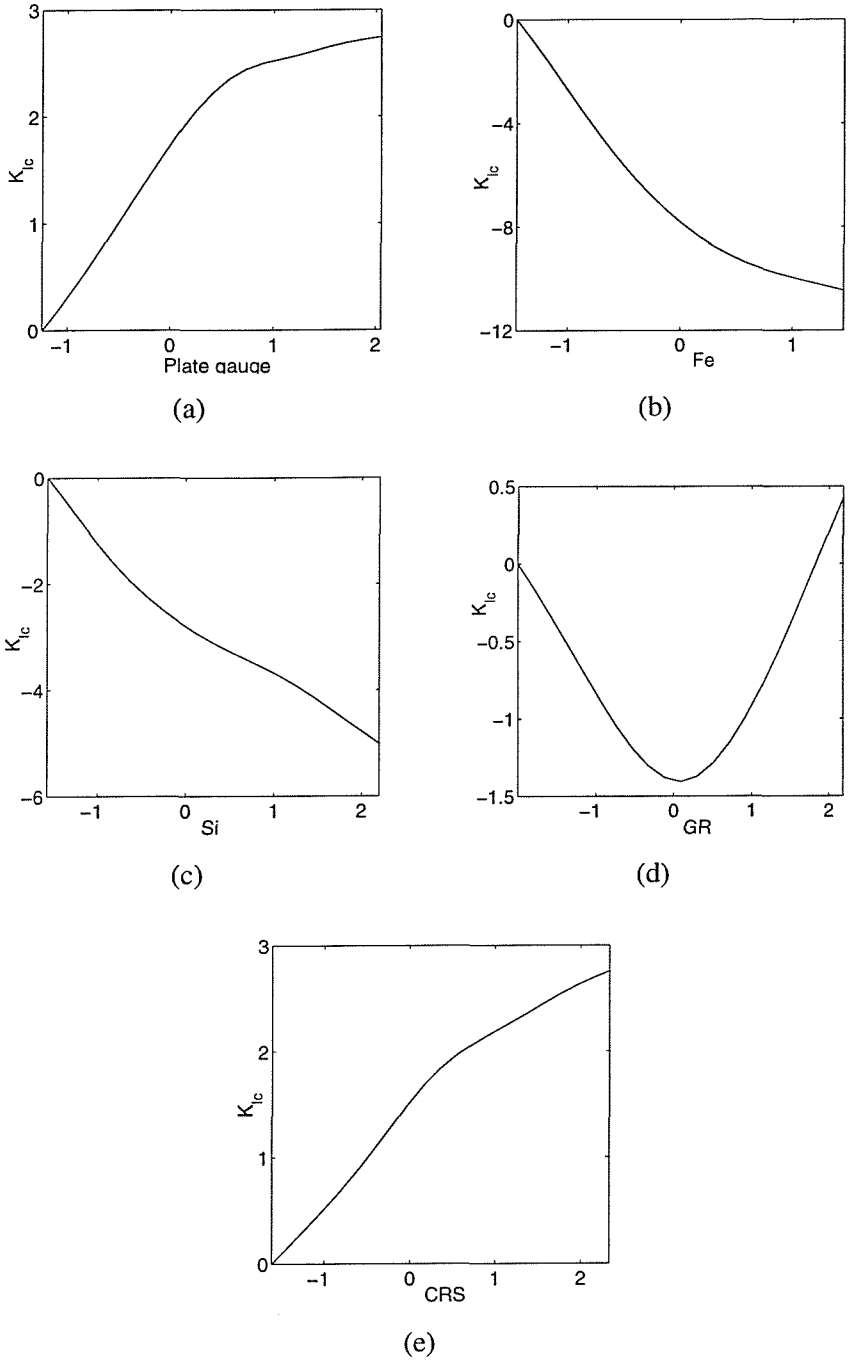


Figure C.22: General form of the univariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP1TLQ3W data set.

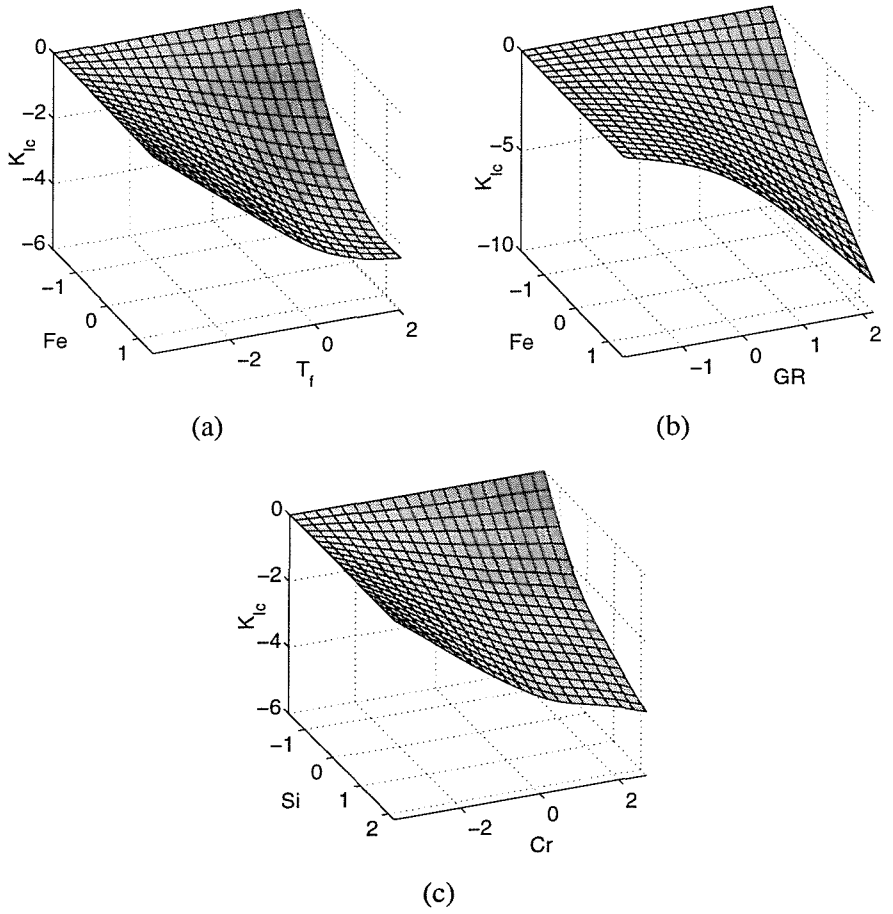


Figure C.23: General form of the bivariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP1TLQ3W data set.

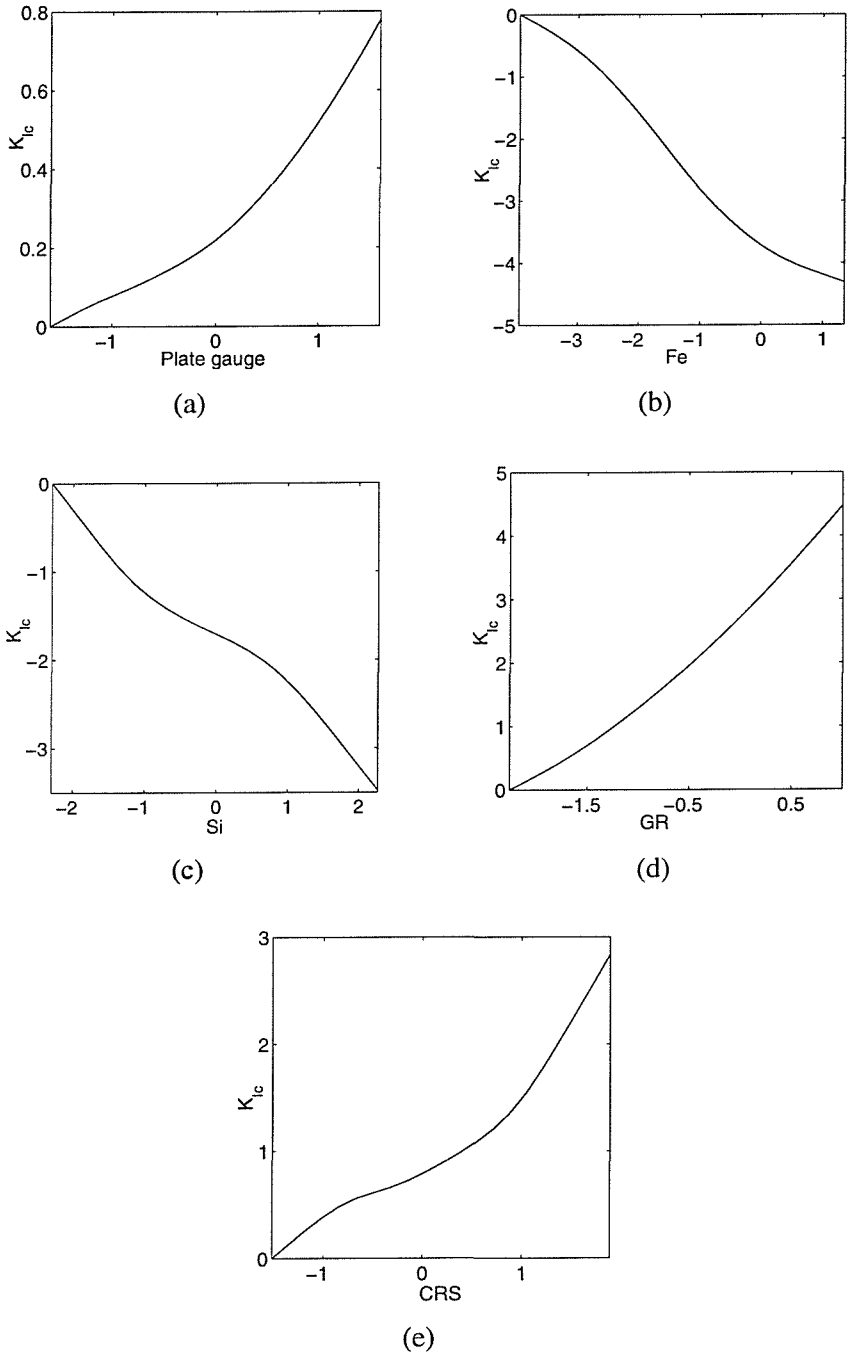


Figure C.24: General form of the univariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP1TLQ3W80 data set.

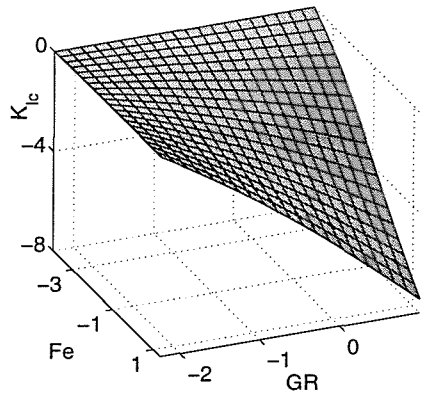


Figure C.25: General form of the bivariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP1TLQ3W80 data set.

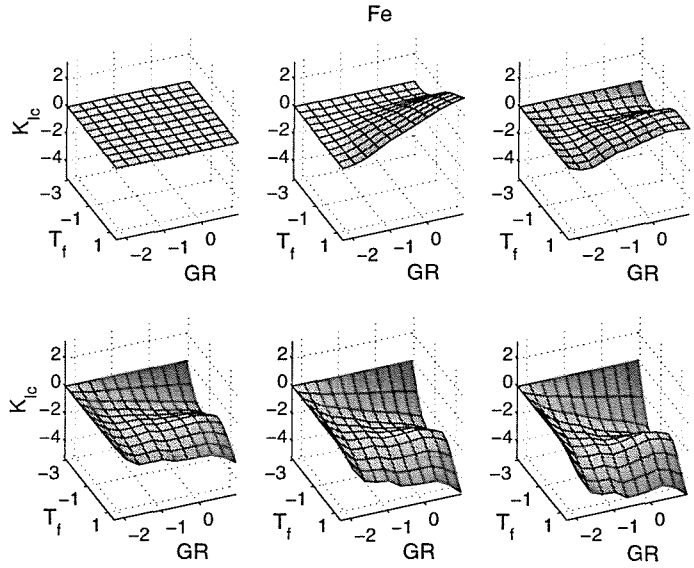


Figure C.26: General form of the trivariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP1TLQ3W80 data set.

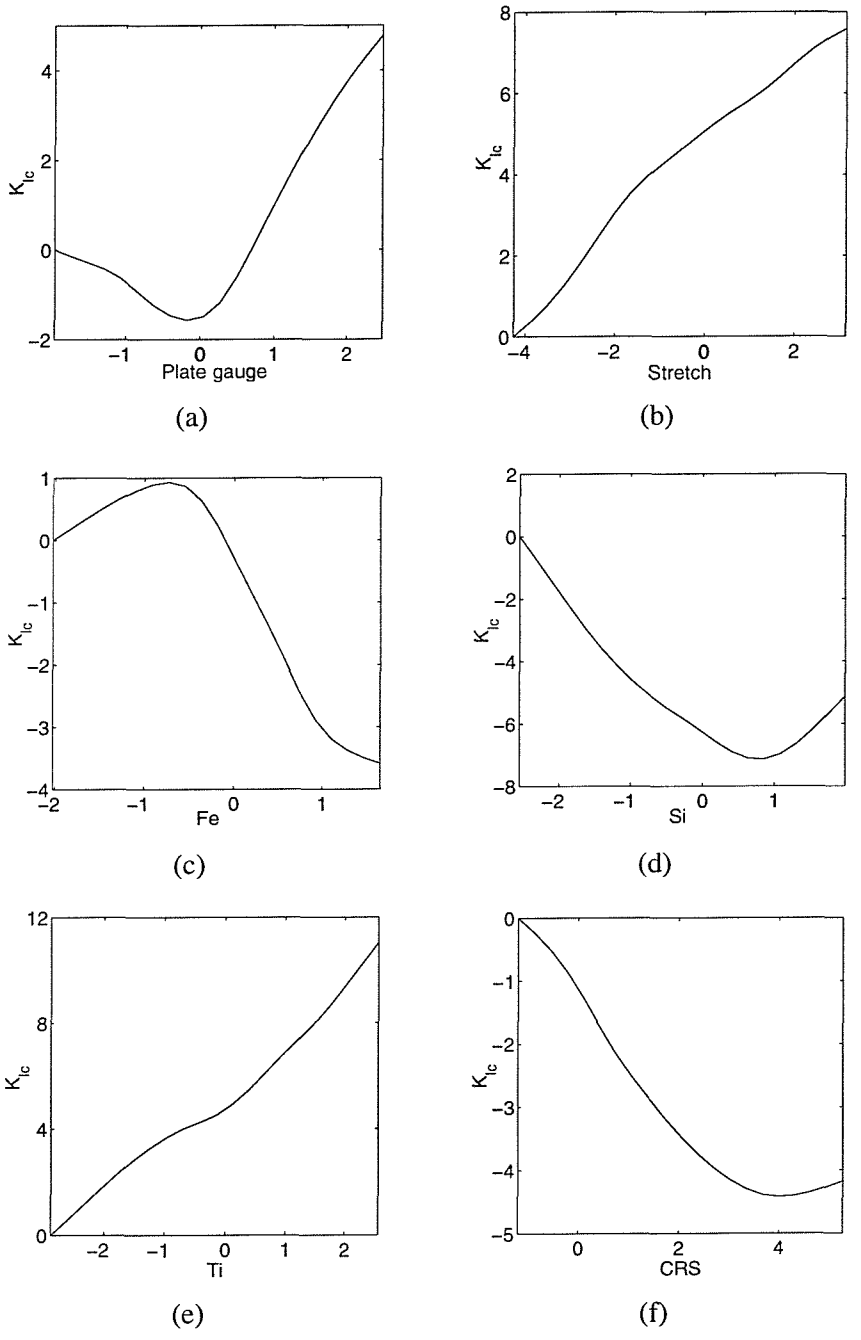


Figure C.27: General form of the univariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP4LTC5W data set.

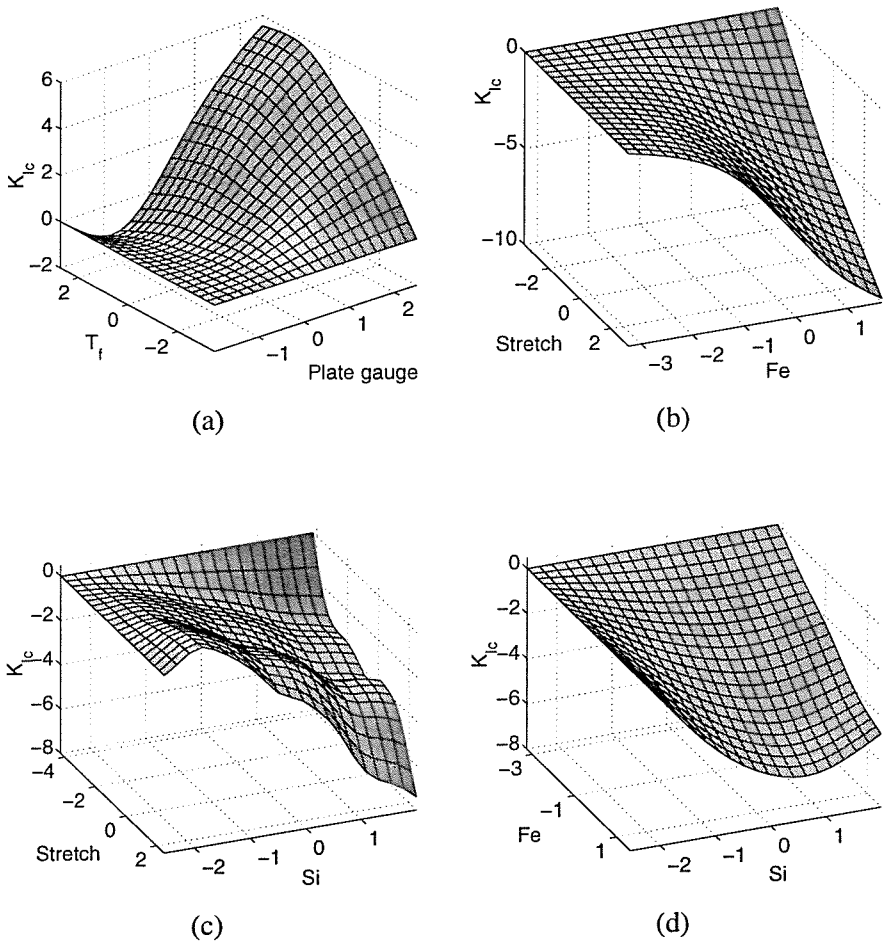


Figure C.28: General form of the bivariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP4LTC5W data set.

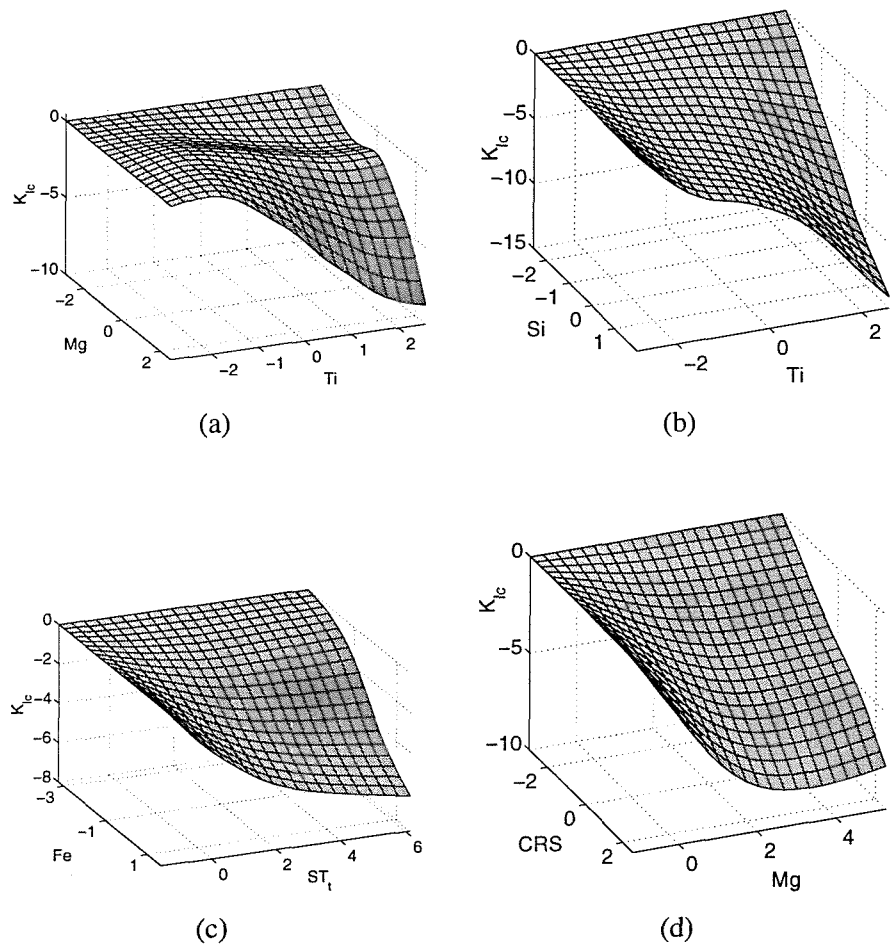


Figure C.29: General form of the bivariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP4LTC5W data set.

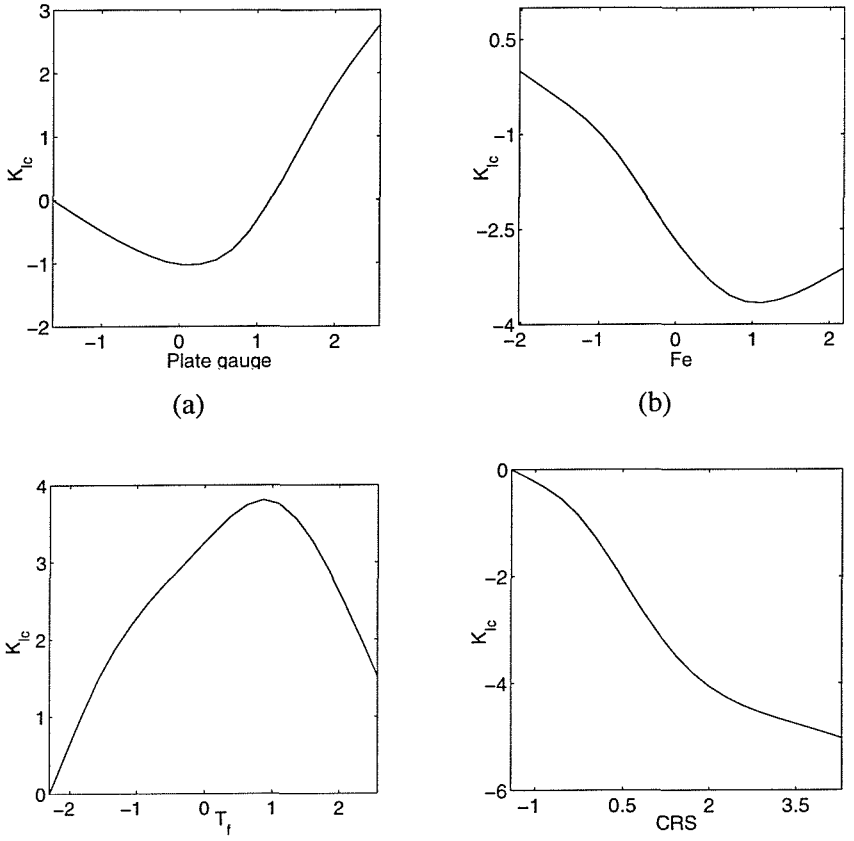


Figure C.30: General form of the univariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP4SLC5W data set.

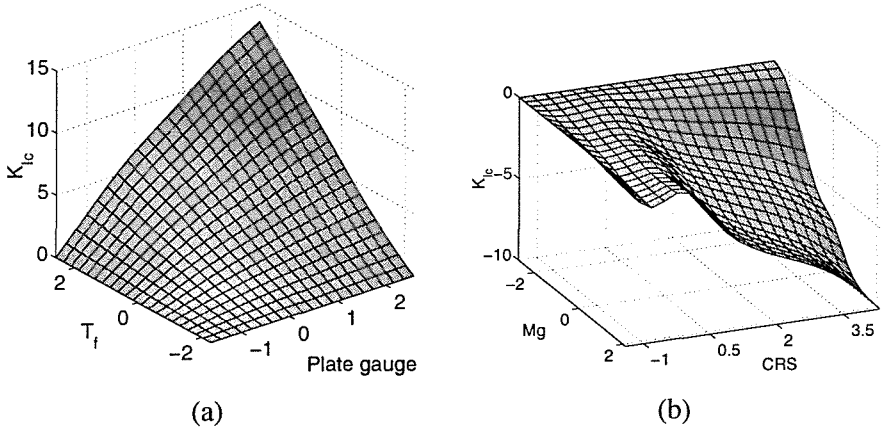


Figure C.31: General form of the bivariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP4SLC5W data set.

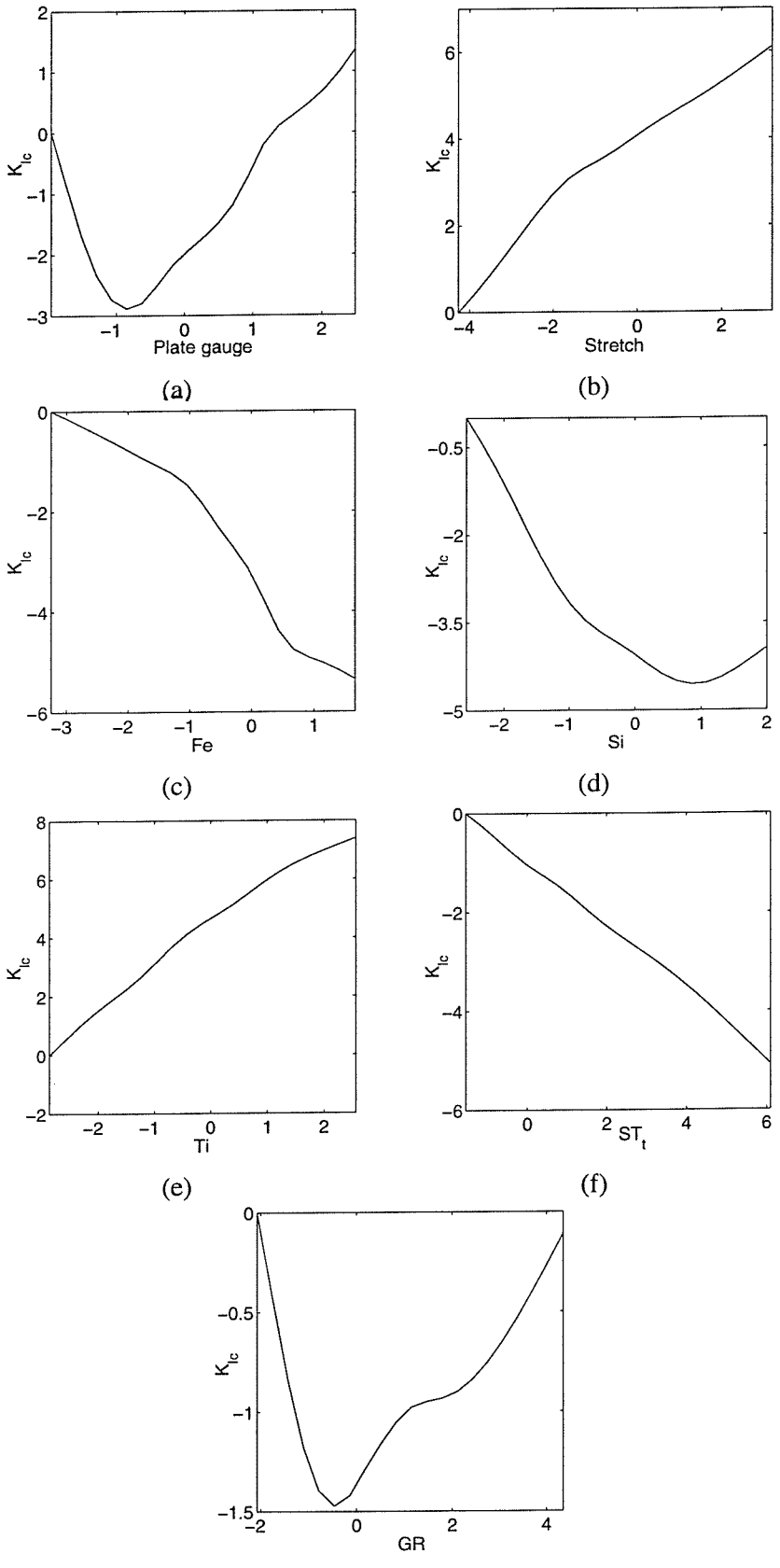


Figure C.32: General form of the univariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP4TLC5W data set.

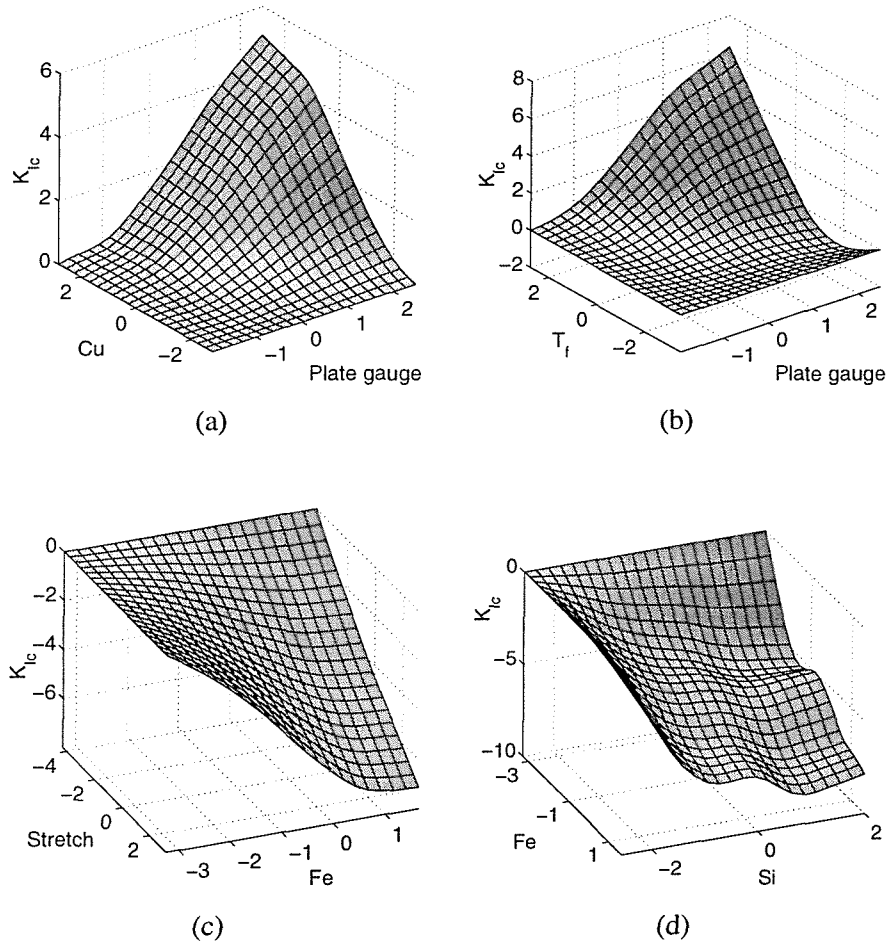


Figure C.33: General form of the bivariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP4TLC5W data set.

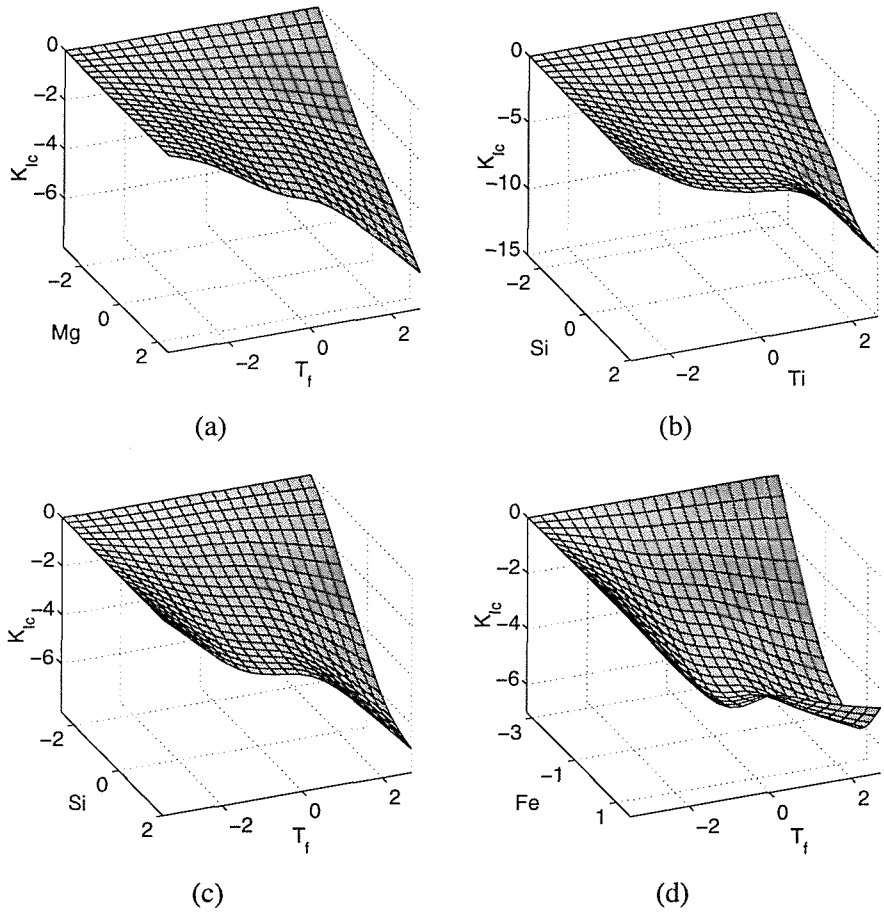


Figure C.34: General form of the bivariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP4TLC5W data set.

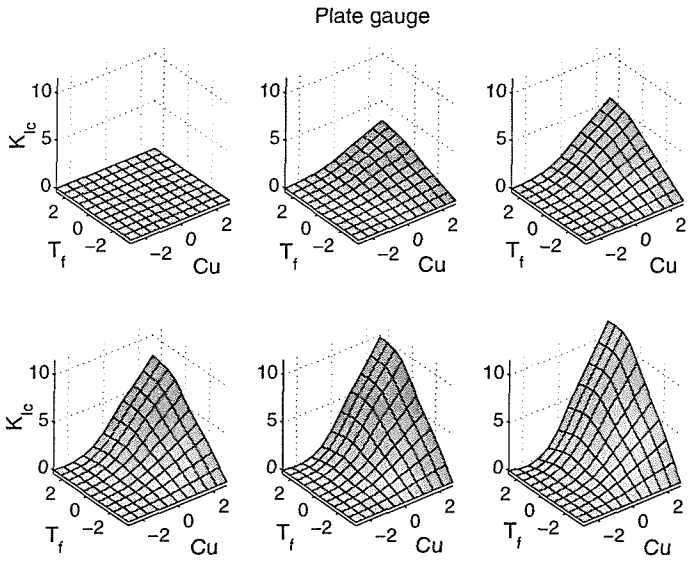


Figure C.35: General form of the trivariate kernel approximations exhibited by the ANOVA terms most consistently selected in the sparse basis selection for the BAP4TLC5W data set.

References

Adler, P., R. DeIasi, and G. Geschwind (1972). Influence of microstructure on the mechanical properties and stress corrosion susceptibility of 7075 aluminum alloy. *Metallurgical and Material Transactions 3*, 3191–3200.

Aikin, B. and T. Courtney (1993). Modeling of particle size evolution during mechanical milling. *Metallurgical and Material Transactions 24A*, 2465–2471.

Allen, D. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics 16*, 125–127.

Anderson, K. (1994). US Patent no.5312498.

Arafeh, L., H. Singh, and S. Putatunda (1999). A neuro fuzzy logic approach to material processing. *IEEE Transactions on Systems, Man, and Cybernetics 29*(3), 362–370.

Archambault, P., J. Chevrier, G. Beck, and J. Bouvaist (1980). A contribution to the optimization of the 7075 heat treatment. *Materials Science and Engineering 43*, 1–6.

Ardell, A. (1985). Precipitation hardening. *Metallurgical and Material Transactions A 16*, 2131–2165.

Atkinson, A. (1986). Masking unmasked. *Biometrika 73*, 533–541.

Atkinson, A. and A. Donev (1992). *Optimum Experimental Designs*. Oxford: Clarindon Press.

Badmos, A. and H. Bhadeshia (1998). Tensile properties of mechanically alloyed oxide dispersion strengthened iron alloys part 2 - physical interpretation of yield strength. *Materials Science and Technology 14*, 1221–1226.

Badmos, A., H. Bhadeshia, and D. MacKay (1998). Tensile properties of mechanically alloyed oxide dispersion strengthened iron alloys part 1 - neural network models. *Materials Science and Technology 14*, 793–809.

Bailer-Jones, C., H. Bhadeshia, and D. MacKay (1999). Gaussian process modelling of austenite formation in steel. *Materials Science and Technology 15*, 287–294.

Bailer-Jones, C., T. Sabin, D. MacKay, and P. Withers (1997). Prediction of deformed and annealed microstructures using Bayesian neural networks and gaussian processes. In *Intelligent Processing and Manufacturing of Materials.*

BAP (2000). Private communication with proprietors of the data. British Aluminium Plate.

Barrett, B. and R. Ling (1992). General classes of influence measures for multivariate regression. *Journal of the Americal Statistical Association 87*(417), 184–191.

Bartlett, P. (1997). For valid generalization, the size of the weights is more important than the size of the network. In M. Mozer, M. Jordan, and T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9*, pp. 134–140.

Bates, C. and G. Totten (1988). Procedure for quenching media selection to maximise tensile properties and minimise distortion in aluminium-alloy parts. *Heat Treatment of Metals 15*(4), 89–97.

Bellman, R. (1961). *Adaptive Control Processes: A Guided Tour.* New York: Princeton University Press.

Belsley, D. (1984). Demeaning conditioning diagnostics through centering. *The American Statistician 38*, 73–93.

Belsley, D. (1991). *Conditioning Diagnostics: Collinearity and Weak Data in Regression.* John Wiley and Sons.

Belsley, D., E. Kuh, and R. Welsch (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* John Wiley and Sons.

Benjamin, J. and M. Bomford (1977). Dispersion strengthened aluminum made by mechanical alloying. *Metallurgical and Material Transactions A 8*, 1301–1305.

Benjamin, J. and R. Schelleng (1981). Dispersion strengthened aluminium-4 pct magnesium alloy made by mechanical alloying. *Metallurgical and Material Transactions A 12*, 1827–1832.

Benjamin, J. and T. Violin (1974). The mechanism of mechanical alloying. *Metallurgical and Material Transactions A 5*, 1929–1934.

Berk, K. (1977). Tolerance and condition in regression computations. *Journal of the Americal Statistical Association 72*, 863–866.

Berk, K. (1978). Comparing subset regression procedures. *Technometrics 20*, 1–6.

Bhadeshia, H. (1999). Neural networks in materials science. *ISIJ International 39*(10), 966–979.

Bishop, C. (1995). *Neural networks for pattern recognition.* Clarendon Press.

Bossley, K. (1997). *Neurofuzzy Modelling Approaches in System Identification.* Ph.D. thesis, University of Southampton.

Box, G. (1966). Use and abuse of regression. *Technometrics 8*(4), 625–629.

Box, G. (1976). Science and statistics. *Journal of the Americal Statistical Association 71*, 791–799.

Box, G. and N. Draper (1987). *Empirical Model-Building and Response Surfaces.* New York: John Wiley and Sons.

Box, G., W. Hunter, and J. Hunter (1978). *Statistics for Experimenters (An introduction to Design, Data Analysis, and Model Building).* John Wiley and Sons.

Breiman, L. (1992). The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error. *Journal of the Americal Statistical Association 87*, 738–754.

Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics 37*, 373–384.

Breiman, L. (1996a). Bagging predictors. *Machine Learning 24*, 123–140.

Breiman, L. (1996b). Heuristics of instability and stabilization in model selection. *The Annals of Statistics 24*, 2350–2383.

Breiman, L. and J. Friedman (1997). Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society B 59*, 3–54.

Breiman, L., J. Friedman, R. Olshen, and C. Stone (1984). *Classification and Regression Trees.* Chapman and Hall.

Breiman, L. and P. Spector (1992). Submodel selection and evaluation in regression. The X-random case. *International Statistical Review 60*, 291–319.

Bridges, P., J. Brooks, and P. Gilman (1985). An aluminium-magnesium-lithium alloy made by mechanical alloying. In *Advanced Materials Research and Developments for Transport: Light Metals 1985, 7^{th} Symposium*, pp. 85–92. Les Editions de Physique, Les Ulis, Fr.

Brown, M., K. Bossley, and C. Harris (1996). Neurofuzzy algorithms for model identification: Structure and parameter determination. In *Computational Engineering in Systems Applications '96: Symposium on Control, Optimization and Supervision*, Lille, France, pp. 1061–1066. volume 2.

Brown, M., K. Bossley, D. Mills, and C. Harris (1995). High dimensional neurofuzzy systems: Overcoming the curse of dimensionality. In *Int. Joint Conf. of the 4th Int. Conf. on Fuzzy Systems and the 2nd Int. Fuzzy Engineering Symp.*, Yokohama, Japan, pp. 2139–2146.

Brown, M. and C. Harris (1994). *Neurofuzzy adaptive modelling and control.* Prentice-Hall.

Brown, R. and L. Willey (1967). *Aluminium: Properties, Physical Metallurgy and Phase Diagrams*, Chapter 2, pp. 31–54. Ohio: ASM.

Buja, A., T. Hastie, and R. Tibshirani (1989). Linear smoothers and additive models. *The Annals of Statistics 17*(2), 453–555.

Buntine, W. and A. Weigend (1991). Bayesian back-propagation. *Complex Systems 5*, 603–643.

Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery 2*(2), 1–43.

Carroll, R. and D. Ruppert (1988). *Transformation and Weighting in Regression.* Chapman and Hall.

Caruana, R. and D. Freitag (1994). Greedy attribute selection. In *Proceedings 11^{th} International Conference on Machine Learning*, pp. 28–36.

Chakrabarti, D., H. Weiland, B. Cheney, and J. Staley (1996). Through thickness property variations in 7050 plate. In *Materials Science Forum*, Volume 217-222, pp. 1085–1090.

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society A 158*(3), 419–466.

Chatterjee, S. and A. Hadi (1986). Influential observations, high leverage points, and outliers in linear regression. *Statistical Science 1*, 379–416.

Chatterjee, S. and A. Hadi (1988). Impact of simultaneous omission of a variable and an observation on a linear regression equation. *Computational Statistics and Data Analysis 6*, 129–144.

Chen, C. and J. Knott (1981). Effects of dispersoid particles on toughness of high-strength aluminium alloys. *Metal Science 15*, 357–364.

Chen, J., T. Sun, R. Viswanadham, and J. Green (1977). Grain boundary segregation of an Al-Zn-Mg ternary alloy. *Metallurgical and Material Transactions 8*, 1935–1940.

Chen, S. (1995). *Basis Pursuit.* Ph.D. thesis, Department of Statistics, Stanford University.

Chen, S., E. Chng, and K. Alkadhimi (1996). Regularized orthogonal least squares algorithm for constructing radial basis function networks. *Int. J. Control 64*, 829–837.

Cherkassky, V. and F. Mulier (1998). *Learning from Data: Concepts, Theory, and Methods*. John Wiley and Sons.

Cherkassky, V., X. Shao, F. Mulier, and V. Vapnik (1999). Model complexity control for regression using VC generalization bounds. *IEEE Transactions on Neural Networks 10*(5), 1077–1089.

Cho, S., Y. Cho, and S. Toon (1997). Reliable roll force prediction in cold mill using multiple neural networks. *IEEE Transactions on Neural Networks 8*(4), 874–882.

Chun, M., J. Biglou, J. Lenard, and J. Kim (1999). Using neural networks to predict parameters in the hot working of aluminum alloys. *Journal of Materials Processing Technology 86*, 245–251.

Cios, K., G. Baaklini, and A. Vary (1995). Soft computing in design and manufacturing of advanced materials. *Journal of engineering for gas turbines and power 117*, 161–165.

Clinch, M., S. Christensen, P. Reed, I. Sinclair, W. Hepples, and N. Holroyd (2000). The use of modelling techniques to describe recrystallized grain size of commercially processed 6000 series aluminium alloy components. In N. Hansen, X. Huang, J. D.J., E. Luridsen, T. Leffers, W. Pantleon, T. Sabin, and J. Wert (Eds.), *Proceedings of the 21st Riso International Symposium on Materials Science: Recrystallization - Fundamental Aspects and Relations to Deformation Microstructure*, pp. 297–302.

Cook, R. (1979). Influential observations in linear regression. *Journal of the Americal Statistical Association 74*, 169–174.

De Jong, H. (1980). Thickness direction inhomogeneity of mechanical properties and fracture toughness as observed in aluminium 7075-T651 plate material. *Engineering fracture mechanics 13*, 175–192.

Derksen, S. and H. Keselman (1992). Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authenic and noise variables. *British Journal of Mathematical and Statistical Psychology 45*, 265–282.

Deshais, G. and S. Newcomb (2000). The influence of microstructure on the formation of stress corrosion cracks in 7xxx series aluminium alloys. In E. J. Starke, T. Sanders, and W. Cassada (Eds.), *Materials Science Forum*, Volume 331-337 of *Proceedings of the 7th International Conference on Aluminum Alloys (ICAA-7)*, Charlottesville, VA, pp. 1635–1640.

Deshamps, A. and Y. Brechet (1999). Influence of predeformation and ageing of Al-Zn-Mg alloy - II. modeling of precipitation kinetics and yield stress. *Acta Metallurgica 47*, 293–305.

Domingos, P. (1999). The role of Occam's razor in knowledge discovery. *Data Mining and Knowledge Discovery 3*(4), 1–19.

Donachie, S. and P. Gilman (1983). The microstructure and properties of Al-Mg-Li alloys prepared by mechanical alloying. In T. J. Sanders and E. J. Starke (Eds.), *Aluminium-Lithium II*, pp. 507–516. The Metallurgical Society of AIME.

Donachie, S. and P. Gilman (1986). US patent no. 4,600,556.

Dorward, R. (1999). Precipitate coarsening during overaging of Al-Zn-Mg-Cu alloy. *Materials Science and Technology 15*, 1133–1138.

Dorward, R. and D. Beerntsen (1995). Grain structure and quench-rate effects on strength and toughness of AA7050 Al-Zn-Mg-Cu-Zr alloy plate. *Metallurgical and Material Transactions 26A*, 2481–2484.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society B 57*(1), 45–97.

Draper, N. and H. Smith (1981). *Applied Regression Analysis* (Second ed.). John Wiley and Sons.

Draper, N. and R. Van Nostrand (1979). Ridge regression and James-Stein estimation: Review and comments. *Technometrics 21*, 451–466.

Dubost, B. (1993). Industrial applications and determination of equilibrium phase diagrams for light alloys. Progress and prospects. *Revue de Metallurgie - CIT 90*(2), 195–209.

Edwards, A. (1984). *An introduction to Linear Regression and Correlation* (Second ed.). W.H. Freeman and Company.

Efron, B. and R. Tibshirani (1993). *An Introduction to the Bootstrap*. Chapman and Hall.

Ehrstrom, J., P. Achon, J. Hebert, and P. A. (1996). Microstructural modelling of fracture toughness of Al alloys. In *Materials Science Forum*, Volume 217-222, pp. 1539–1544.

Elder, J. I. and D. Pregibon (1996). *Advances in Knowledge Discovery and Data Mining*, Chapter 4, pp. 83–116. AAAI/MIT Press.

Embury, J., D. Loyyd, and T. Ramachandran (1989). *Aluminum alloys - contemporary research and applications*, Volume 31 of *Treatise on Materials Science and Technology*, Chapter 22, pp. 579–601. Academic Press, Inc.

England, R., J. Pickens, K. Kumar, and T. Langan (1988). Improved weldable Al-Mg alloys made by rapid solidification powder metallurgy technology. In Y.-W. Kim and W. Griffith (Eds.), *Dispersion Strengthened Aluminum Alloys*, pp. 371–393. TMS Publications.

Evancho, J. and J. Staley (1974). Kinetics of precipitation in aluminum alloys during continuous cooling. *Metallurgical and Material Transactions 5*(1), 43–47.

Ewalds, H. and R. Wanhill (1984). *Fracture mechanics*. Arnold.

Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth (1996). *Advances in Knowledge Discovery and Data Mining*, Chapter 1, pp. 1–36. AAAI/MIT Press.

Fox, J. (1984). *Linear statistical Models and Related Methods*. John Wiley and Sons.

Frank, I. and J. Friedman (1993). A statistical view of some chemometrics regression tools (with discussion). *Technometrics 35*, 109–148.

Friedman, J. (1991). Multivariate adaptive regression splines. *The Annals of Statistics 19*, 1–141.

Friedman, J. (1993). Estimating functions of mixed ordinal and categorical variables using adaptive splines. In S. Morgenthaler, E. Ronchetti, and W. Stahel (Eds.), *New Directions in Statistical Data Anaysis and Robustness*, Basel, pp. 73–113. Birkhauser Verlag.

Friedman, J. (1994). An overview of predictive learning and function approximation. In V. Cherkassky, J. Friedman, and H. Wechsler (Eds.), *From Statistics to Neural Networks: Theory and Pattern Recognition Applications*, Volume 136 of *NATO ASI Series F: Computer and System Sciences*. Springer-Verlag.

Fujii, H. and H. Bhadeshia (1999). Preface to the special issue on "Application of Neural Network Anaysis in Materials Science". *ISIJ International 39*(10), 965.

Fujii, H., D. MacKay, and H. Bhadeshia (1996). Bayesian neural network analysis of fatigue crack growth rate in nickel base superalloys. *ISIJ International 36*, 1373–1382.

Furu, T., H. Shercliff, C. Sellars, and M. Ashby (1996). Physically-based modelling of strength, microstructure and recrystallisation during thermomechanical processing of Al-Mg alloys. In *Materials Science Forum*, Volume 217-222, pp. 453–458.

Garrett, G. and J. Knott (1978). The influence of compositional and microstructural variations on the mechanism of static fracture in aluminum alloys. *Metallurgical and Material Transactions 9A*, 1187–1201.

Garson, G. (1991). Interpreting neural-network connection weights. *AI Expert 6*(4), 47–51.

Gavard, L., H. Bhadeshia, D. MacKay, and S. Suzuki (1996). Bayesian neural network model for austenite formation in steels. *Mat. Sci. Tech. 12*, 453–463.

Gedeon, T. (1997). Data mining of inputs: analysing magnitude and functional measures. *International Journal of Neural Systems 8*(2), 209–218.

Geman, S., E. Bienenstock, and R. Doursat (1992). Neural networks and the bias/variance dilemma. *Neural Computation 4*, 1–58.

Gibbons, D. and G. McDonald (1984). A rational interpretation of the ridge trace. *Technometrics 26*, 339–346.

Gill, P., W. Murray, and M. Wright (1993). *Practical Optimization.* San Diego: Academic Press.

Gilman, P. (1983). The physical metallurgy of mechanically-alloyed, dispersion-strengthened Al-Li-Mg and Al-Li-Cu alloys. In T. J. Sanders and E. J. Starke (Eds.), *Aluminium-Lithium II*, pp. 485–506. The Metallurgical Society of AIME.

Gilman, P. and J. Benjamin (1983). Mechanical Alloying. *Ann. Rev. Mater. Sci. 13*, 279–300.

Gilman, P., J. Brooks, and P. Bridges (1985). High temperature tensile properties of mechanically alloyed Al-Mg-Li alloys. In C. Baker, P. Gregson, S. Harris, and C. Peel (Eds.), *Aluminium-Lithium III*, pp. 112–120. The Institute of Metals.

Gilman, P. and S. Donachie (1987). US Patent no. 4,643,780.

Gilman, P. and W. Nix (1981). The structure and properties of aluminium alloys produced by mechanical alloying: Powder processing and resultant powder structures. *Metallurgical and Material Transactions 12*, 813–824.

Gokhale, A., N. Deshpande, D. Denzer, and J. Liu (1998). Relationship between fracture toughness, fracture path, and microstructure of 7050 aluminum alloy: Part II. multiple micromechanisms-based fracture toughness model. *Metallurgical and Material Transactions 29A*, 1203–1210.

Gomiero, P., Y. Brechet, F. Louchet, A. Tourabi, and B. Wack (1992). Microstructure and mechanical properties of a 2091 AlLi alloy - II. mechanical properties: Yield stress and work hardening. *Acta Metallurgica 40*, 857–861.

Granger, D. (1989). *Aluminum alloys - contemporary research and applications*, Volume 31 of *Treatise on Materials Science and Technology*, Chapter 4, pp. 109–135. Academic Press, Inc.

Gregson, P. (1995). *High Performance Materials in Aerospace*, Chapter 2, pp. 49–84. Chapman and Hall.

Gull, S. (1989). Developments in maximum entropy and data analysis. In J. Skilling (Ed.), *Maximum Entropy and Bayesian Methods*, pp. 53–71. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Gunn, S. and M. Brown (1999). SUPANOVA - a sparse, transparent modelling approach. In *Proc. IEEE International Workshop on Neural Networks for Signal Processing*, Madison, Wisconsin.

Gunn, S., M. Brown, and K. Bossley (1997). Network performance assessment for neurofuzzy data modelling. In X. Liu, P. Cohen, and M. Berthold (Eds.), *Lecture Notes in Computer Science n.1208*, Berlin Heidelberg, pp. 313–323. Springer-Verlag.

Gunst, R. (1983). Regression analysis with multicollinear predictor variables: Definition, detection, and effects. *Communications in Statistics Theory and Methods 12*, 2217–2260.

Guyot, P. and L. Cottignies (1996). Precipitation kinetics, mechanical strength and electrical conductivity of AlZnMgCu alloys. *Acta Metallurgica 44*(10), 4161–4167.

Hadi, A. and J. Simonoff (1993). Procedures for the identification of multiple outliers in linear models. *Journal of the Americal Statistical Association 88*, 1264–1272.

Hadi, A. S. (1992). A new influence measure of overall potential influence in linear regression. *Computational Statistics and Data Analysis 14*, 1–27.

Hahn, G. and A. Rosenfield (1968). Sources of fracture toughness: The relation between K_{Ic} and the ordinary tensile properties of metals. In *Applications Related Phenomena in Titanium Alloys*, Volume ASTM STP, pp. 5–32. American Society for Testing Materials.

Hahn, G. and A. Rosenfield (1975). Metallurgical factors affecting fracture toughness of aluminum alloys. *Metallurgical and Material Transactions 6A*, 653–668.

Hamerton, R., H. Cama, and M. Meredith (2000). Development of the coarse intermetallic particle population in wrought aluminium alloys during ingot casting and thermomechanical processing. In E. J. Starke, T. Sanders, and W. Cassada (Eds.), *Materials Science Forum*, Volume 331-337 of *Proceedings of the 7^{th} International Conference on Aluminum Alloys (ICAA-7)*, Charlottesville, VA, pp. 143–154.

Hastie, T. and R. Tibshirani (1990). *Generalized Additive Models*. Chapman and Hall.

Hatch, J. (1984). *Aluminum: Properties and Physical Metallurgy*. ASM.

Haykin, S. (1998). *Neural Networks: A comprehensive Foundation*. Prentice Hall.

Hepples, W. (1987). *Environment-Sensitive Cracking of 7000 Series Aluminium Alloys*. Ph.D. thesis, University of Newcastle upon Tyne.

Hildeman, G. and M. Koczak (1989). *Aluminum alloys - contemporary research and applications*, Volume 31 of *Treatise on Materials Science and Technology*, Chapter 11, pp. 323–364. Academic Press, Inc.

Hirsch, J., K. Karhausen, and P. Wagner (2000). Practical application of modeling in the industrial sheet production. In E. J. Starke, T. Sanders, and W. Cassada (Eds.), *Materials Science Forum*, Volume 331-337 of *Proceedings of the 7^{th} International Conference on Aluminum Alloys (ICAA-7)*, Charlottesville, VA, pp. 421–430.

Hoaglin, D. and R. Welsch (1978). The hat matrix in regression and ANOVA. *The American Statistician 32*, 17–22.

Hocking, R. (1976). The analysis and selection of variables in linear regression. *Biometrics 32*, 1–49.

Hocking, R. (1983a). Developments in linear regression methodology: 1959-1982. *Technometrics 25*, 219–230.

Hocking, R. (1983b). Developments in linear regression methodology: 1959-1982 - response. *Technometrics 25*, 248–249.

Hoerl, A. and R. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics 12*, 55–67.

Hoerl, R., J. Shuenemeyer, and A. Hoerl (1986). A simulation of biased estimation and subset selection regression techniques. *Technometrics 28*, 369–380.

Hoeting, J., A. Raftery, and D. Madigan (1996). A method for simultaneous variable selection and outlier identification in linear regression. *Computational Statistics and Data Analysis 22*, 251–270.

Hornbogen, E. and E. A. J. Starke (1993). OVERVIEW NO. 102 theory assisted design of high strength low alloy aluminum. *Acta metall. mater. 41*(1), 1–16.

Huang, J., J. Conley, and P. Callau (1998). A study of neural networks for porosity prediction in aluminum alloy 356 castings. In *Modeling of Casting, Welding and Solidification Processes*, Volume VIII of *TMS*. San Diego: TMS.

Huber, P. (1981). *Robust Statistics*. Wiley.

Hunt, E. (1992). *Structure Property Relationships in Fibre Reinforced Aluminium Laminates and SiC Particulate Aluminium Composites*. Ph.D. thesis, University of Southampton.

Jang, J.-S., C. Sun, and E. Mizutani (1997). *Neuro-fuzzy modelling and soft computing*. Prentice-Hall.

Jensen, D. and P. Cohen (2000). Multiple comparisons in induction algorithms. *Machine Learning 38*(3), 309–338.

Jonas, J., C. Sellars, and W. Tegart (1969). Strength and structure under hot-working conditions. *Metallurgical Reviews, Review 130 15*, 1–23.

Jones, J. (1997). *Neural Network Modelling of the Tensile Properties of Ni-Base superalloys.* Ph.D. thesis, Selwyn College, University of Cambridge.

Kandola, J., S. Gunn, I. Sinclair, and P. Reed (1999). On the use of advanced inductive methods for knowledge extraction from complex datasets. submitted to *Machine Learning.*

Kavli, T. (1993). ASMOD - an algorithm for adaptive spline modelling of observation data. *Int. J. Control 58*, 947–967.

Kirman, I. (1971). The relation between microstructure and toughness in 7075 aluminum alloys. *Metallurgical and Material Transactions 2*, 1761–1770.

Kolby, P. (1996). Applications of thermodynamically modelled phase diagram data in mathematical modelling of industrial processes. In *Materials Science Forum*, Volume 217-222, pp. 661–666.

Kovacs, I., J. Lendvai, T. Ungar, T. Turmezey, and G. Groma (1977). The properties of AlZnMg alloy deformation strengthened by Guinier-Preston zones. *Acta Metallurgica 25*, 673–680.

Krogh, A. and J. Hertz (1992). A simple weight decay can improve generalization. In J. Moody, S. Hanson, and R. Lippmann (Eds.), *Advances in Neural Information Processing Systems 4*, San Mateo, CA, pp. 950–957. Morgan Kauffmann.

Kwok, J.-Y. (1999). Integrating the evidence framework and the support vector machine. In *Proceedings of the European Symposium on Artificial Neural Networks (ESANN)*, Bruges, Belgium, pp. 177–182.

Larkiola, J., J. Nylander, P. Myllykoski, and A. Korhonene (1996). Determining tandem cold rolling efficiency using physical models and neural computing. In *Proc. of the* 2^{nd} *Int. Conf. on Modelling of Metal Rolling Processes*, pp. 389–399.

Last, H. and J. R. Garrett (1996). Mechanical behaviour and properties of mechanically alloyed aluminum alloys. *Metallurgical and Materials Transactions, A 27*, 737–745.

Leger, C. and N. Altman (1993). Assessing influence in variable selection problems. *Journal of the Americal Statistical Association 88*, 547–556.

Li, X. and M. Starink (2000). Analysis of precipitation and dissolution in overaged 7xxx aluminium alloys using DSC. In E. J. Starke, T. Sanders, and W. Cassada (Eds.), *Materials Science Forum*, Volume 331-337 of *Proceedings of the 7th International Conference on Aluminum Alloys (ICAA-7)*, Charlottesville, VA, pp. 1071–1076.

Liang, H., S.-L. Chen, and Y. Chang (1997). A thermodynamic description of the Al-Mg-Zn system. *Metallurgical and Material Transactions A 28*, 1725–1734.

Ludtka, G. and D. Laughlin (1982). The influence of microstructure and strength on the fracture mode and toughness of 7xxx series aluminum alloys. *Metallurgical Transactions A 13A*, 411–425.

MacKay, D. (1992a). Bayesian interpolation. *Neural Computation 4(3)*, 415–447.

MacKay, D. (1992b). A practical bayesian framework for backprop networks. *Neural Computation 4(3)*, 589–603.

MacKay, D. (1995). Probable networks and plausible predictions - a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems 6*, 469–505.

MacKay, D. (1999). Comparison of approximate methods for handling hyperparameters. *Neural Computation 11*(5), 1035–1068.

Madigan, D. and A. Raftery (1994). Model selection and accounting for model uncertainty in graphical models using Occam's window. *Journal of the Americal Statistical Association 89*, 1535–1546.

Maloney, S., I. Polmear, and S. Ringer (2000). Effects of Cu on precipitation in Al-Zn-Mg alloys. In E. J. Starke, T. Sanders, and W. Cassada (Eds.), *Materials Science Forum*, Volume 331-337 of *Proceedings of the 7th International Conference on Aluminum Alloys (ICAA-7)*, Charlottesville, VA, pp. 1055–1060.

Mantini, L. and T. Prucher (1993). Predicting powder metal mechanical properties - a statistical approach. *Advances in Powder Metallurgy and Particulate Materials 3*, 139–153.

Marenbach, P. and M. Brown (1997). Evolutionary versus inductive construction of neuro-fuzzy syetems for bioprocess modelling. In *Galesia'97*, Glasgow, UK, pp. 320–325.

Marquardt, D. and R. Snee (1975). Ridge regression in practice. *The American Statistician 29*, 3–20.

Martin, J. (1998). *Precipitation Hardening* (Second ed.). Butterworth-Heinemann.

Mason, R. and R. Gunst (1985). Outlier-induced collinearities. *Technometrics 27*, 401–407.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models*. Chapman and Hall.

Melander, A. and P. Persson (1978). The strength of a precipitation hardened AlZnMg alloy. *Acta Metallurgica 26*, 267–278.

Michot, G. and G. Champier (1991). Physical metallurgy of P/M aluminium alloys. *Memoires et Etudes Scientifiques Revue de Metallurgie 88*, 425–439.

Miller, A. (1984). Selection of subsets of regression variables. *Journal of the Royal Statistical Society A 147*, 389–425.

Miller, M., E. Harley, T. Turner, A. Beaudoin, and W. Cassada (2000). Mechanical behaviour of thin sheets machined from AA7050-T7451. In E. J. Starke, T. Sanders, and W. Cassada (Eds.), *Materials Science Forum*, Volume 331-337 of *Proceedings of the 7^{th} International Conference on Aluminum Alloys (ICAA-7)*, Charlottesville, VA, pp. 1243–1248.

Mills, D., K. Bossley, M. Brown, and C. Harris (1995). Towards parsimonious high-dimensional neurofuzzy systems. In *World Congress on Neural Networks*, Volume 2, Washington D.C., pp. 717–720.

Milne, L. (1995). Feature selection using neural networks with contribution measures. In *Proceedings of the Australian Conference on Artificial Intelligence (AI'95)*, Canberra.

Mondolfo, L. (1971). Structure of the aluminium:magnesium:zinc alloys. *Metallurgical Reviews Review 153*, 95–124.

Mondolfo, L. (1976). *Aluminium Alloys*. London: Butterworths and Co. Ltd.

Montgomery, D. and E. Peck (1992). *Introduction to linear regression analysis* (Second ed.). John Wiley and Sons.

Moody, J. (1994). Prediction risk and architecture selection for neural networks. In V. Cherkassky, J. Friedman, and H. Wechsler (Eds.), *From Statistics to Neural Networks: Theory and Pattern Recognition Applications*. Springer-Verlag.

Morere, B., J.-C. Ehrstrom, P. Gregson, and I. Sinclair (2000). Microstructural effects on fracture toughness in AA7010 plate. *Accepted for publication in Metallurgical Transactions A*.

Mukai, T., K. Ishikawa, and K. Higashi (1995). Influence of strain rate on the mechanical properties in fine-grained aluminum alloys. *Materials Science and Engineering A204*, 12–18.

Narayanan, G., B. Wilson, and W. Quist (1983). P/M aluminum-lithium alloys by the mechanical alloying process. In T. J. Sanders and E. J. Starke (Eds.), *Aluminium-Lithium II*, pp. 517–541. The Metallurgical Society of AIME.

Neal, R. (1994). Priors for infinite networks. Technical Report CRG-TR-94-1, Dept. of Computer Science, University of Toronto.

Neal, R. (1996). *Bayesian Learning for Neural Networks*. Springer-Verlag.

Neter, J., M. Kutner, C. Nachtsheim, and W. Wasserman (1996). *Applied Linear Statistical Models* (Fourth ed.). Irwin.

Noble, B., S. Harris, and K. Dinsdale (1982). Yield characteristics of aluminium-lithium alloys. *Metal Science 16*, 425–430.

Olafsson, P., R. Sandstrom, and A. Karlsson (1996). Electrical conductivity of aluminium alloys. In *Materials Science Forum*, Volume 217-222, pp. 981–986.

Orr, M. (1995a). Local smoothing of radial basis function networks. In *Proc. International Symposium on Artificial Neural Networks*, Hsinchu, Taiwan.

Orr, M. (1995b). Regularisation in the selection of radial basis function centres. *Neural Computation 7*, 606–623.

Orr, M. (1996). Introduction to radial basis function networks. Technical report, University of Edinburgh.

Papazian, J. and P. Gilman (1990). Age hardening in mechanically alloyed Al-Mg-Li-C-O. *Materials Science and Engineering A125*, 121–127.

Park, J. and A. Ardell (1983). Microstructures of the commercial 7075 Al alloy in the T651 and T7 tempers. *Metallurgical and Material Transactions 14A*, 1957–1965.

Park, J. and A. Ardell (1988). Precipitate microstructure of peak-aged 7075 Al. *Scripta Metallurgica 22*, 1115–1119.

Peel, C. and P. Gregson (1995). *High Performance Materials in Aerospace*, Chapter 1, pp. 1–48. Chapman and Hall.

Penny, W. and S. Roberts (1998a). Bayesian neural networks for classification: how useful is the evidence framework. *Neural Networks 12*, 877–892.

Penny, W. and S. Roberts (1998b). Error bars for linear and nonlinear neural network regression models. Technical report, Department of Electrical Engineering, Imperical College, London.

Pickens, J. (1981). Review aluminium powder metallurgy technology for high-strength applications. *Journal of Materials Science 16*, 1437–1457.

Pickens, J. (1985). US Patent no. 4,532,106.

Pickens, J., R. Schelleng, S. Donachie, and T. Nichol (1981a). US Patent no. 4,292,079.

Pickens, J., R. Schelleng, S. Donachie, and T. Nichol (1981b). US Patent no. 4,297,136.

Pitcher, P. (1998). Optimisation of strength and electrical conductivity in 7010 plate. DER-A/MSS2/CR980546/1.0.

Pitcher, P. (1999). Ageing trials in 7xxx alloys (DERA Report).

Plate, T., J. Bert, J. Grace, and P. Band (1998). Visualizing the function computed by a feedforward neural network. Technical Report CS-TR-98-5, School of Mathematical and Computing Science, Victoria University of Wellington.

Polmear, I. (1981). *Light alloys (Metallurgy of the light metals), Third edition.* Arnold.

Press, W., S. Teukolsky, W. Vetterling, and B. Flannery (1992). *Numerical Recipes in C* (Second ed.). Cambridge University Press.

Raftery, A., D. Madigan, and J. Hoeting (1993). Model selection and accounting for model uncertainty in models. Technical Report 262, University of Washington.

Reed, R., R. Marks II, and S. Oh (1995). Similarities of error regularization, sigmoid gain scaling, target smoothing, and training with jitter. *IEEE Transactions on Neural Networks 6*, 529–538.

Rendigs, K.-H. (1997). Aluminium structures used in aerospace - status and prospects. In *Materials Science Forum*, Volume 242, pp. 11–24.

Sabin, T., C. Bailer-Jones, S. Roberts, D. MacKay, and P. Withers (1997). Modelling the evolution of microstructures in cold-worked and annealed aluminium alloy. In *International Conference on Thermomechanical Processing of Steels and Other Materials*, pp. 1043–1049.

Sabin, T., P. Withers, C. Bailer-Jones, and D. MacKay (1998). Static and dynamic modelling of materials forging. *Australian Journal of Intelligent Information Processing Systems 5*(1), 10–17.

Sainfort, P., C. Sigli, G. Raynaud, and P. Gomiero (1997). Structure and property control of aerospace alloys. In *Materials Science Forum*, Volume 242, pp. 25–32.

Sanders, T. J. and E. J. Starke (1982). The effect of slip distribution on the monotonic and cyclic ductility of Al-Li binary alloys. *Acta Metallurgica 30*, 927–939.

Sarkar, B., M. Marek, and E. J. Starke (1981). The effect of copper content and heat treatment on the stress corrosion characteristics of Al-6Zn-2Mg-XCu alloys. *Metallurgical Transactions A 12A*, 1939–1943.

Sarle, W. (1997). Neural networks FAQ. ftp://ftp.sas.com/pub/neural/FAQ.html.

Saunders, N. (1996). Phase diagram calculations for commercial Al-alloys. In *Materials Science Forum*, Volume 217-222, pp. 667–672.

Scamans, G., N. Holroyd, and C. Tuck (1987). The role of magnesium segregation in the intergranular stress corrosion cracking of aluminium alloys. *Corrosion Science 27*(4), 329–347.

Schmidt, B. (1998). Investigation of amorphous Al-alloys with neural networks. In T. Sato, S. Kumai, T. Kobayashi, and Y. Murakami (Eds.), *Proceedings of the 6^{th} International Conference on Aluminium Alloys (ICAA-6)*, Toyohashi, Japan, pp. 1761–1764.

Scholkopf, B., C. Burges, and A. Smola (Eds.) (1999). *Advances in Kernel Methods: Support Vector Learning*. Cambridge, MA: MIT Press.

Schooling, J. (1997). *The modelling of fatigue in Ni-base superalloys*. Ph.D. thesis, Darwin College, University of Cambridge.

Schooling, J., M. Brown, and P. Reed (1999). An example of the use of neural computing techniques in materials science - the modelling of fatigue thresholds in Ni-base superalloys. *Materials Science and Engineering 260*, 222–239.

Seaver, B., K. Triantis, and C. Reeves (1999). The identification of influential subsets in regression using a fuzzy clustering strategy. *Technometrics 41*(4), 340–351.

Seibi, A. and S. Al-Alawi (1997). Prediction of fracture toughness using artificial neural networks (ANNs). *Engineering fracture mechanics 56*, 311–319.

Sellars, C. (1987). Influence of changing strain rate on microstructure during hot deformation. *Acta Metallurgica 35*, 2649–2657.

Sellars, C. (1992). Microstructural modelling of thermomechanical processing. In *Proc. of the 3^{rd} International Conference on Aluminium Alloys*, Volume 3, pp. 89–105.

Shahani, R., T. Warner, C. Sigli, P. Lassince, and P. Lequeu (1998). High strength 7xxx alloys for ultra-thick aerospace plate: Optimisation of alloy composition. In T. Sato, S. Kumai, T. Kobayashi, and Y. Murakami (Eds.), *Proceedings of the 6^{th} International Conference on ALuminum Alloys (ICAA-6)*, Volume 2, pp. 1105–1110.

Shavlik, J. and G. Towell (1994). Knowledge-based artificial neural networks. *Artificial Intelligence 70*, 119–165.

Shercliff, H. and M. Ashby (1990). A process model for age hardening of aluminium alloys - I. the model. *Acta Metallurgica 38*, 1789–1802.

Shercliff, H. and A. Lovatt (1999). Modelling of microstructure evolution in hot defomation. *Phil. Trans. R. Soc. Lond. A 357*, 1621–1643.

Sigli, C. (2000). Nucleation, growth and coarsening of spherical precipitates in aluminum alloys. In E. J. Starke, T. Sanders, and W. Cassada (Eds.), *Materials Science Forum*, Volume 331-337 of *Proceedings of the 7^{th} International Conference on Aluminum Alloys (ICAA-7)*, Charlottesville, VA, pp. 513–518.

Sigli, C., L. Maenner, C. Sztur, and R. Shahani (1998). Phase diagram, solidification and heat treatment of aluminum alloys. In T. Sato, S. Kumai, T. Kobayashi, and Y. Murakami (Eds.), *Proceedings of the 6^{th} International Conference on Aluminum Alloys (ICAA-6)*, Volume 3, pp. 87–98.

Sigli, C., H. Vichery, and B. Grange (1996). Computer assisted metallurgy for non heat treatable aluminum alloys. In *5^{th} Int. Conf. on Aluminium Alloys*, Volume 217-222, pp. 391–396.

Singer, R., W. Oliver, and W. Nix (1980). Identification of dispersoid phases created in aluminum during mechanical alloying. *Metallurgical and Material Transactions A 11*, 1895–1901.

Singh, S., H. Bhadeshia, D. MacKay, H. Carey, and I. Martin (1998). Neural network analysis of steel plate processing. *Ironmaking and Steelmaking 25*, 355–365.

Smith, A. (1992). Current status of aluminium-lithium alloys in the construction of the Westland-Agusta EH101 helicopter. In M. Peters and P.-J. Winkler (Eds.), *Aluminium-Lithium*, pp. 1305–1310. DGM Informationsgesellschaft mbH. Papers presented at the Sixth International Aluminium-Lithium Conference, October, 1991, Germany.

Smith, G. and F. Campbell (1980). A critique of some ridge regression methods. *Journal of the Americal Statistical Association 75*, 74–103.

Smola, A., B. Scholkopf, and K.-R. Muller (1998). The connection between regularization operators and support vector kernels. *Neural Networks 11*(4), 637–649.

Snee, R. (1983). Developments in linear regression methodology: 1959-1982 - discussion. *Technometrics 25*, 230–237.

Sollich, P. (2000). Bayesian methods for support vector machines: Evidence and error bars. Submitted to *Machine Learning*.

Song, R.-G., Q.-Z. Zhang, M. Tseng, and B.-J. Zhang (1995). The application of artificial neural networks to the investigation of aging dynamics in 7175 aluminium alloys. *Materials Science and Engineering: C 3*, 39–41.

Speidel, M. (1975). Stress corrosion cracking of aluminum alloys. *Metallurgical Transactions A 6A*, 631–651.

Speidel, M. and M. Hyatt (1972). *Advances in Corrosion Science and Technology*, Volume 2, Chapter 3, pp. 115–305. Plenium Press.

Staley, J. (1976). Properties related to fracture toughness. ASTM STP 605.

Staley, J. (1987). Quench factor analysis of aluminium alloys. *Materials Science and Technology 3*(11), 923–935.

Staley, J. (1992). Metallurgical aspects affecting strength of heat-treatable alloy products used in the aerospace industry. In *3rd Int. Conf. on Aluminium Alloys*, Trondheim, pp. 107–143.

Starink, M. and P. Gregson (1995). A quantitative interpretation of DSC experiments on quenched and aged SiCp reinforced 8090 alloys. *Scripta Metallurgica 33*, 893–900.

Starink, M. and P. Gregson (1996). Thermodynamics and precipitation in 8090 (Al-Li-Cu-Mg-Zr) alloys studied by DSC and TEM. In *Mater Sci. Forum*, Volume 217-222, pp. 673–678.

Starink, M., A. Hobson, and P. Gregson (2000). Modelling of strengthening of Al-Li-Cu-Mg alloys. In E. J. Starke, T. Sanders, and W. Cassada (Eds.), *Materials Science Forum*, Volume 331-337 of *Proceedings of the 7th International Conference on Aluminum Alloys (ICAA-7)*, Charlottesville, VA, pp. 1321–1326.

Starink, M., I. Sinclair, P. Reed, and P. Gregson (2000). Predicting the structural performance of heat-treatable Al-alloys. In *Materials Science Forum*, Volume 331-337, pp. 97–110.

Starink, M., P. Wang, I. Sinclair, and P. Gregson (1999). Microstructure and strengthening of Al-Li-Cu-Mg alloys and MMCs: II. modelling of yield strength. *Acta Metallurgica 47*(14), 3855–3868.

Starink, M. and A.-M. Zahra (1998). β' and β precipitation in an Al-Mg alloy studied by DSC and TEM. *Acta Metallurgica 46*, 3381–3397.

Starke, E. J. (1989). *Aluminum alloys - contemporary research and applications*, Volume 31 of *Treatise on Materials Science and Technology*, Chapter 2, pp. 35–63. Academic Press, Inc.

Starke, E. J. and J. Staley (1996). Application of modern aluminum alloys to aircraft. *Prog. Aerospace Sci. 32*, 131–172.

Stone, M. (1974). Cross-validatory choice of statistical predictors (with discussion). *Journal of the Royal Statistical Society B 36*, 111–147.

Strawbridge, D., W. Hume-Rothery, and A. Little (1948). The constitution of aluminium-copper-magnesium-zinc alloys at 460°C. *J. Inst. Metals 74*, 191–225.

Styles, C. and P. Pitcher (1998). Forging behaviour and properties of metal matrix composites based on mechanically alloyed Al-Mg-Li alloy. *Materials Science and Technology 14*, 913–919.

Sugamata, M., C. Blankenship Jr., and E. Starke Jr. (1993). Predicting plane strain fracture toughness of Al-Li-Cu-Mg alloys. *Materials Science and Engineering A163*, 1–10.

Sugamata, M., J. Kaneko, and H. Higuchi (1998). Solid state reactions in mechanically alloyed Al-Li and Al-Li-Mg alloys with various metal oxides. In *Materials Science Forum*, Volume 269-272, pp. 157–162.

Sumpter, B. and D. Noid (1996). On the design, analysis, and characterization of materials using computational neural networks. *Annu. Rev. Mater. Sci. 26*, 223–277.

Taniguchi, M. and V. Tresp (1997). Averaging regularized estimators. *Neural Computation 9*, 1163–1178.

Tarrant, A. (1998). Private communication with proprietor of the data. Aerospace Metal Composites, (AMC) U.K.

Thodberg, H. (1993). Ace of bayes: Application of neural networks with pruning. Technical Report 1132E, The Danish Meat Research Institute, Magleggaardsvej 2, DK-4000 Roskilde.

Thompson, D. (1975). Metallurgical factors affecting high strength aluminum alloy production. *Metallurgical and Material Transactions A 6*, 671–683.

Thomson, R., F. Perez-Perez, A. Warters, and G. Thewlis (1999). A neural network approach to the prediction of submerged arc weld metal chemistry. *ISIJ International 39*(10), 1096–1105.

Tibshirani, R. and G. Hinton (1998). Coaching variables for regression and classification. *Statistics and Computing 8*, 25–33.

Tresp, V., J. Hollatz, and S. Ahmad (1993). Network structuring and training using rule-based knowledge. In *Advances in Neural Information Processing Systems 5*. Morgan Kaufman.

Tresp, V. and M. Taniguchi (1995). Combining estimators using non-constant weighting functions. In *Advances in Neural Information Processing Systems 7*, pp. 419–426.

Tsai, T. and T. Chuang (1996). Relationship between electrical conductivity and stress corrosion cracking susceptibility of al 7075 and al 7475 alloys. *Corrosion Science 52*, 414–416.

Vafaie, H. and K. DeJong (1993). Robust feature selection algorithms. In *Proceedings of the International Conference on Tools with AI*, Boston, MA, pp. 356–364.

Vafaie, H. and I. Imam (1994). Feature selection methods: Genetic algorithms vs. greedy-like search. In *In Proceedings of the International Conference on Fuzzy and Intelligent Control Systems.*

Van de Laar, P. (1999). *Selection in Neural Information Processing.* Ph.D. thesis, University of Nijmegen, The Netherlands.

Van de Laar, P., S. Gielen, and T. Heskes (1997). Input selection with partial retraining. In W. Gerstner, A. Germond, M. Hasler, and J. Nicoud (Eds.), *Artificial Neural Networks - ICANN'97*, pp. 469–474.

Van de Laar, P., T. Heskes, and S. Gielen (1999). Partial retraining: A new approach to input relevance determination. *International Journal of Neural Systems 9*, 75–85.

van Rijckevorsel, J. (1988). *Component and Correspondence Analysis*, Chapter 2: Fuzzy Coding and B-Splines, pp. 33–54. John Wiley and Sons.

Van Stone, R., R. Merchant, and J. J. Low (1974). Investigation of the plastic fracture of high-strength aluminum alloys. *Fatigue amd Fracture Toughness - Cryogenic Behaviour ASTM STP 556*, 93–124.

Van Stone, R. and J. Psioda (1975). Discussion of "metallurgical factors affecting fracture toughness of aluminum alloys". *Metallurgical Transactions A 6A*, 668–670.

Vapnik, V. (1979). *Estimation of Dependencies Based on Empirical Data [in Russian].* Nauka, Moskow: English translation: Springer Verlag, New York, 1982.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory.* New York: Springer-Verlag.

Vapnik, V. (1998). *Statistical Learning Theory.* New York: John Wiley and Sons.

Vapnik, V., S. Golowich, and A. Smola (1997). Support vector method for function approximation, regression estimation, and signal processing. In M. Mozer, M. Jordan, and T. Petsche (Eds.), *Neural Information Processing Systems*, Cambridge, MA. MIT Press.

Vatne, H.-E., F. Perocheau, H.-E. Ekstrom, L. Poizat, K. Nord-Varhaug, K. Marthinsen, E. Lindh, J. Hagstrom, and T. Furu (2000). Industrial verification of microstructural models for thermomechanical processing by application to hot rolling of AA3104. In E. J. Starke, T. Sanders, and W. Cassada (Eds.), *Materials Science Forum*, Volume 331-337 of *Proceedings of the 7^{th} International Conference on Aluminum Alloys (ICAA-7)*, Charlottesville, VA, pp. 551–556.

Vermeulen, W., P. Morris, A. de Weijer, and S. van der Zwaag (1998). Prediction of martensite start temperature using artificial neural networks. *Ironmaking and Steelmaking 25*, 433–437.

Viswanadham, R., T. Sun, and J. Green (1980). Grain boundary segregation in Al-Zn-Mg alloys - implications to stress corrosion cracking. *Metallurgical and Material Transactions 11*, 85–89.

Vitiaz, P., F. Lovshenko, Z. Lovshenko, and G. Lovshenko (1997). Fundamental formation laws of phase composition, structure and properties of aluminium materials manufactured by mechanical alloying. *Advanced Performance Materials 4*, 325–336.

Vooijs, S., B. Davenport, and S. van der Zwaag (2000). Monitoring the precipitation reactions in a cold rolled Al-Mn-Mg-Cu alloy using thermoelectric power and resistivity measurements. In E. J. Starke, T. Sanders, and W. Cassada (Eds.), *Materials Science Forum*, Volume 331-337 of *Proceedings of the 7^{th} International Conference on Aluminum Alloys (ICAA-7)*, Charlottesville, VA, pp. 933–938.

Wagner, J. and R. Shenoy (1991). The effect of copper, chromium, and zirconium on the microstructure and mechanical properties of Al-Zn-Mg-Cu alloys. *Metallurgical and Material Transactions A 22*, 2809–2818.

Wang, C., S. Venkatesh, and J. Judd (1994). Optimal stopping and effective machine complexity in learning. In J. Cowan, G. Tesauro, and J. Alspector (Eds.), *Advances in Neural Information Processing Systems 6*, pp. 303–310. Morgan Kaufmann Publishers, Inc.

Wang, L.-X. (1992). Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man, and Cybernetics 22*, 1414–1427.

Warner, T., R. Shahani, P. Lassince, and G. Raynaud (1997). Aluminium alloy developments for affordable airframe structures. In *3^{rd} ASM Conference on Synthethis, Processing and Modelling of Advanced Materials*, Paris.

Weber, J. and M. Phillips (1993). Evolution of properties in a powder metallurgy aluminum-lithium alloy. In *Powder Metallurgy in Aerospace, Defense and Demanding Applications*, pp. 259–266.

Weisberg, S. (1995). *Applied Linear Regression* (Second ed.). John Wiley and Sons.

Wichmann, N. and E. Bartlett (1997). Ranking input variables using general regression neural networks. In Dagli, Akay, Ersoy, Fernandez, and Smith (Eds.), *Intelligent Engineering Systems Through Artificial Neural Networks*, Volume 7, pp. 959–964.

Williams, C. (1997). Computing with infinite networks. In M. Mozer, M. Jordan, and T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9*. MIT Press.

Williams, P. (1995). Bayesian regularisation and pruning using a laplace prior. *Neural Computation 7*(1), 117–143.

Wolpert, D. (1993). On the use of evidence in neural networks. In S. Hanson, J. Cowan, and C. Giles (Eds.), *Advances in Neural Information Processing Systems (NIPS) 5*, pp. 539–546. Morgan Kaufmann.

Wu, J., G. Cheng, and X. Liu (1997). Reasoning about outliers by modelling noisy data. In X. Liu, P. Cohen, and M. Berthold (Eds.), *Lecture Notes in Computer Science n.1208*, Berlin Heidelberg, pp. 549–558. Springer-Verlag.

Zadeh, L. (1965). Fuzzy sets. *Information and Control 8*, 338–353.

Zadeh, L. (1973). Outline for a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics 3*, 28–44.

Zaidi, M. and J. Wert (1989). *Aluminum alloys - contemporary research and applications*, Volume 31 of *Treatise on Materials Science and Technology*, Chapter 5, pp. 137–170. Academic Press, Inc.

Zhang, Y., L. Lu, and S. Yap (1999). Prediction of the amount of PCA for mechanical milling. *Journal of Materials Processing Technology 89-90*, 260–265.