11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference
6 - 8 September 2006, Portsmouth, Virginia

AIAA 2006-7088

Parameter Screening Using Impact Factors and Surrogate-based ANOVA Techniques

Wenbin Song* and Andy Keane†
University of Southampton, Southampton, Hampshire, SO17 1BJ, UK

This paper introduces the concept of parameter impact factors in order to screen important parameters in high dimensional design optimization problems which make use of computationally expensive high fidelity simulation models. Based on a snapshot dataset obtained by evaluating design points produced by Design of Experiments techniques, a simple concept of parameter impact factors is introduced and calculated to obtain preliminary estimates on the importance of parameters in the simulation results. Combined with parallel tuning of hyperparameters used in Gaussian process surrogate models and ANOVA techniques using the progressively built surrogate models, a more accurate estimation on the impact of different parameters can be achieved. Less important parameters can then be fixed in order to reduce the dimensionality of the problem to make the problem more tractable within given computational budget and time constraints.

I. Introduction

High fidelity simulations are increasingly relied upon in multidisciplinary optimization studies to solve design problems of complex systems.[1,2] Very often such problems feature a high number of variables. This adds further difficulties to solving such problems. One of the popular approaches in dealing with high computational cost under these circumstances is the use of surrogate models.[3] There exist a variety of techniques for building approximate surrogate models for use with optimization.[4] These include first or second order polynomial regression models; radial basis function based interpolation models; or Gaussian process based statistical models. Among these models, radial basis function interpolations and Gaussian process models, also known as Kriging,[5] are believed to be able to produce a better fit for the data in the presence of nonlinear relations between responses and variables.

Surrogate models can be built either globally in the whole design space[6] or locally within a trusted region.[7,8] The number of data points required to produce an accurate global surrogate models is typically high, particularly, when the number of design variables is large and the objective function is highly multimodal. One the other hand, local surrogate models, valid only within a small region, often need to be coupled with an appropriate surrogate management strategy. The Kriging model is a global model in nature, and the more data that is available, the better the model will be, provided that the hyperparameters are properly tuned. It is therefore critically important to obtain a set of good estimates for the hyperparameters.

Parameter screening and sensitivity analysis are often used along with design of experiment studies to investigate the influence of variables on the objective functions. These techniques, coupled with surrogate models, can provide useful methods for solving high dimensional multitidsciplinary optimization problems.

This paper first introduces the concept of parameter impact factors, a simple preprocessing technique, which can be used as an indication on whether the amount of data available is sufficient to characterize the design space, or along with surrogate-based ANOVA-like (ANalysis Of VAriance) techniques, to remove parameters that are less important. Combined with the parallel tuning of hyperparameters, problem dimensionality can be reduced to a level affordable within a limited computational budget and time scale. In addition, these techniques can be applied offline with the data available to provide further guidance in the design process while allowing simulations to carry on at the same time.

*Research Fellow, School of Engineering Sciences, University of Southampton, SO17 1BJ, UK
†Professor, School of Engineering Sciences, University of Southampton, SO17 1BJ, UK

II. Parameter Impact Factors

As a simple, preliminary technique, parameter impact factors can be used to illustrate the importance of different variables using snapshot datasets obtained from Design of Experiment methods. The parameter impact factor is computed as follows. First, the dataset is standardized using its mean and variance, and the range for each variable is divided into a number of grids. The summed objective functions within each grid is calculated and then summed against each variable. This quantity is termed the parameter impact factor as it indicates how much this variable contributes to the objective function.

This technique may not be effective when the number of design points is small; it is also affected by the locations of the design points. However, it offers a simple way to preprocess the data set to gain some basic insight into the problem and can be at least used to eliminate variables whose impact factors are consistently small when the data set is augmented with more points.

For example, Fig. 1 shows the impact factors of 33 design variables in an aeroengine nacelle shape optimization problem using different numbers of points. The 33 variables are all geometry parameters used to define the CAD model of the nacelle and total pressure recovery at fanface is used as the objective function of the problem. It is clear that after 200 points are evaluated the four most dominant variables remain unchanged in the example.

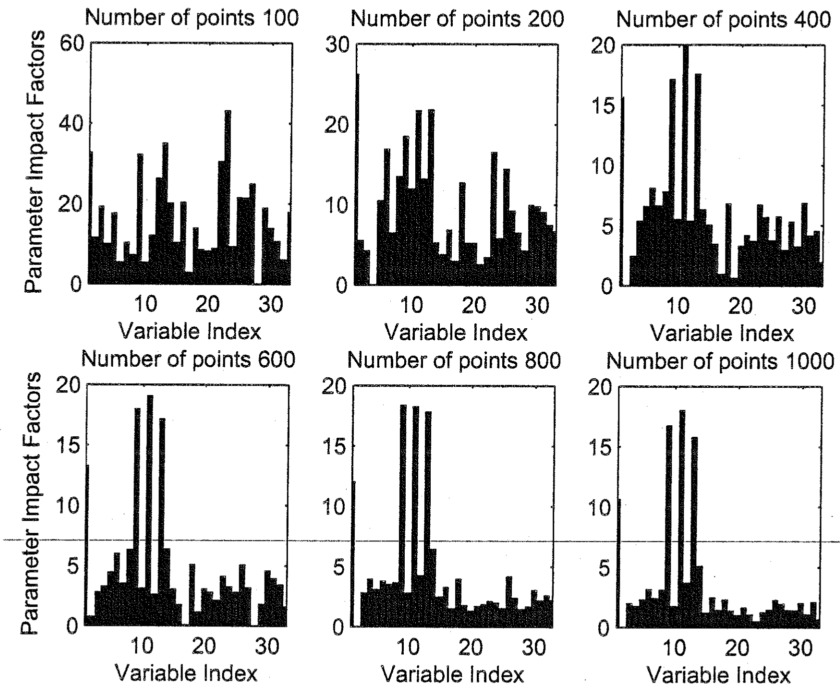

Figure 1. Parameter Impact Factors Estimated Using Different Number of Design Points

III. Gaussian Process Based Surrogate Modelling

The interpolative Kriging model is here expressed as

$$y(\mathbf{x}) = \beta + Z(\mathbf{x}) \tag{1}$$

where β represents a constant term in the model, and $Z(\mathbf{x})$ is a Gaussian random process with zero mean and variance of σ^2. The covariance matrix of $Z(\mathbf{x})$ is given by

$$Cov(Z(\mathbf{x}^i), Z(\mathbf{x}^j)) = \sigma^2 R(\mathbf{x}^i, \mathbf{x}^j) \tag{2}$$

are illustrated in Fig. 2. It can be seen that although hyperparameters change from one case to another, the predominant variable is always reasonably identified. It may be suggested that the magnitude of the hyperparameters can be used as a means to identify important variables. Fig. 3 compares mean square prediction errors achieved using different sets of hyperparameters obtained using different numbers of data points. It can be seen from the comparison that using a small number of data points can sometimes produce a good approximation in terms of the mean square error at unused data points. However, it is not clear how the data points should be chosen in order to produce the best approximations; it is also not clear how many local searches should be carried out or how to choose the start values for different hyperparameters. In addition, multiple approximation models can be built simultaneously for multiple objective functions or constraints. This will further reduce the cost of using surrogates for optimisation. Here in both generating the data set and initial values for hyperparameters, a random Latin hypercube method is used.

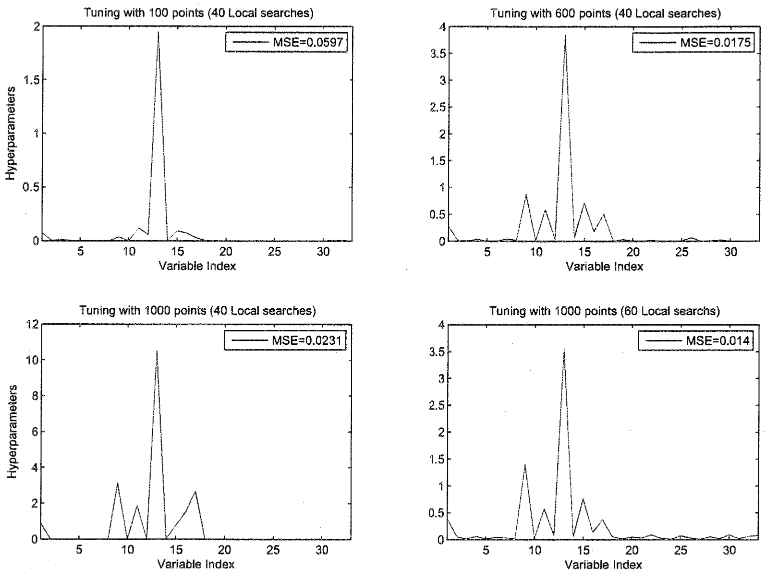

Figure 2. Hyperparameters returned from various searches

V. Parameter Screening using ANOVA techniques

Parameter impact factors can give preliminary information on the influence of each variable. A more sophisticated, ANOVA-like technique can be used to obtain more reliable information. However, this relies on a reasonably accurate estimation of the objective function using the surrogate models. This is the reason why hyperparamter tuning is critically important in such approaches.

If the surrogate model is represented by $\hat{y}$, the main effect of variable x_i is:

$$\hat{\mu}_i(x_i) = \int \cdots \int \hat{y}(x_1, \cdots, x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1} \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n - \hat{\mu} \tag{9}$$

where $\hat{\mu}$ is the total mean of the model, computed as follows:

$$\hat{\mu} = \int \hat{y}(x_1, \cdots, x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n \tag{10}$$

The variance due to the design variable x_i is

$$\int [\hat{\mu}_i(x_i)]^2 \mathrm{d}x_i \tag{11}$$

The main effect of each variable is computed from the surrogate model built using the best set of hyperparameters and the results shown in Fig. 4.

where σ^2 is the variance of the stochastic process and $R(.,.)$ is a correlation function between $\mathbf{x}^i$ and $\mathbf{x}^j$. Different types of correlation function can be employed as noted in Jones et al.[5] A commonly used type of correlation function can be expressed as

$$R(\mathbf{x}^i, \mathbf{x}^j) = \prod_{k=1}^{n} exp(-\theta_k |x_k^i - x_k^j|_k^p) \tag{3}$$

where $\theta_k > 0$ and $1 \leq p_k \leq 2$ are the hyperparameters. Note that the above equation asserts that there is a complete correlation of a point with itself and this correlation decreases rapidly as two points move away from each other in the parameter space. The choice of $p_k = 2$ would provide enough flexibility for modelling smooth but highly non-linear functions for most cases. The hyperparameters θ_k are estimated by maximizing the log-likelihood function given by

$$-\frac{1}{2}[n \ln \sigma^2 + \ln |\mathbf{R}| + \frac{1}{\sigma^2}(\mathbf{y} - \mathbf{1}\beta)^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\beta)] \tag{4}$$

where σ^2 and β can be derived using the following equations once the θ_k are given

$$\hat{\beta} = (\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}, \tag{5}$$

$$\hat{\delta}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{1}\beta)^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\beta). \tag{6}$$

A numerical optimisation procedure is required to obtain the Maximum Likelihood Estimates (MLE) of the hyperparameters. Once the hyperparameters are obtained from the training data, the function value at a new point can be predicted by

$$\hat{y}(\mathbf{x}^*) = \hat{\beta} + \mathbf{r}^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\beta}) \tag{7}$$

along with the posterior variance $s^2(\mathbf{x}^*)$ given by

$$s^2(\mathbf{x}^*) = \sigma^2[1 - \mathbf{r}\mathbf{R}^{-1}\mathbf{r} + \frac{(1 - \mathbf{1}^T \mathbf{R}^{-1}\mathbf{r})^2}{(\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1})}] \tag{8}$$

where $\mathbf{r}(\mathbf{x}) = R(\mathbf{x}, \mathbf{x}^1), ..., R(\mathbf{x}, \mathbf{x}^n)$ is the correlation vector between the new point $\mathbf{x}$ and the training dataset. This quantity provides a good indication on the accuracy of the prediction at new points and can be used to compare the quality of the surrogate models and to decide whether and where further exact analyses are required.

IV. Parallel Tuning of Krig Hyperparamters

With the increasing availability of high-performance clusters, such as the one deployed at University of Southampton which includes 600 2.2Ghz AMD Opterons and 214 1.8Ghz Intel Xeons, a large number of standalone jobs can be executed in parallel. This provides an ideal platform for implementing both massively parallel standalone jobs and MPI jobs. A strategy of parallel hyperparameter tuning for the Gaussian process approximation models is implemented on this platform. The strategy adopted is effectively a multi-start gradient-descent method which use points generated by a Design of Experiment approach.

In the process of building a global Krig model, tuning of the hyperparameters is the most computationally expensive part. This will become even difficult, if not impossible, when the response function contains a large number of peaks and valleys for which a large number of data points need to be sampled. However, it may be argued that this tends not to happen very often for many real-world engineering design problems. It is also known that the maximum likelihood estimate of the Gaussian process model is highly nonlinear and multimodel. Therefore appropriate methods need to be used to find optimised hyperparameters which give good agreement in the prediction of function values in the whole design space. Here, a parallel approach making use of a multi-start gradient descent search method is used to tune the hyperparameters. The starting values for the gradient descent search have been generated using a random Latin hypercube. The best point returned by all the gradient descent searches is used in subsequent predictions. This process produces reasonably good estimates of hyperparameters within a time scale which can otherwise only produce a single set of hyperparameters coming from a local optimum. The Hyperparameters found using this method

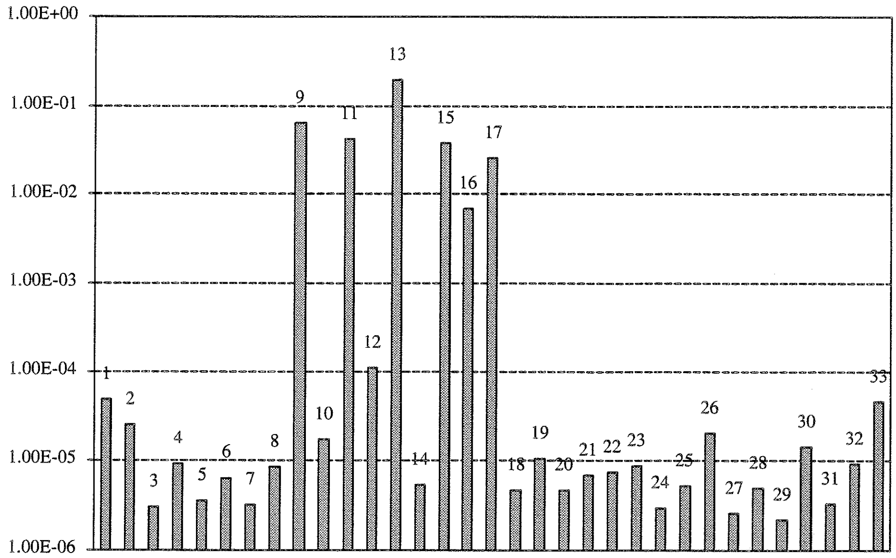

Figure 4. Main effect analysis of design variables using ANOVA techniques and Kriging model

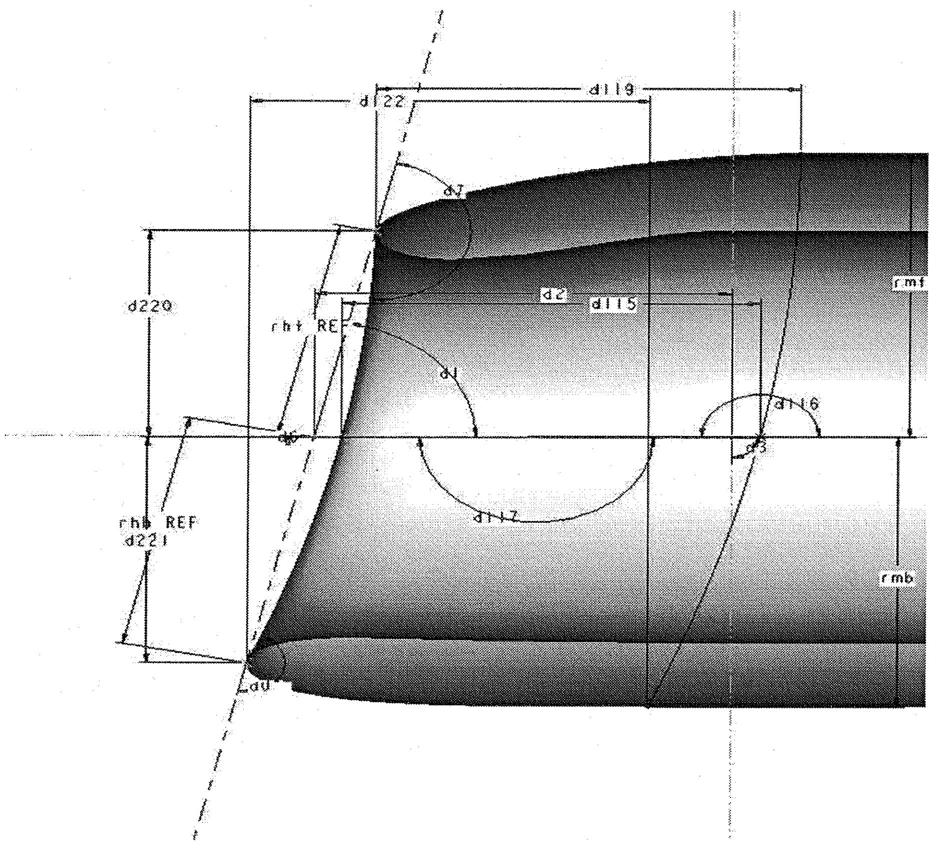

Figure 5. Negative scarf nacelle geometry and some of its parameters

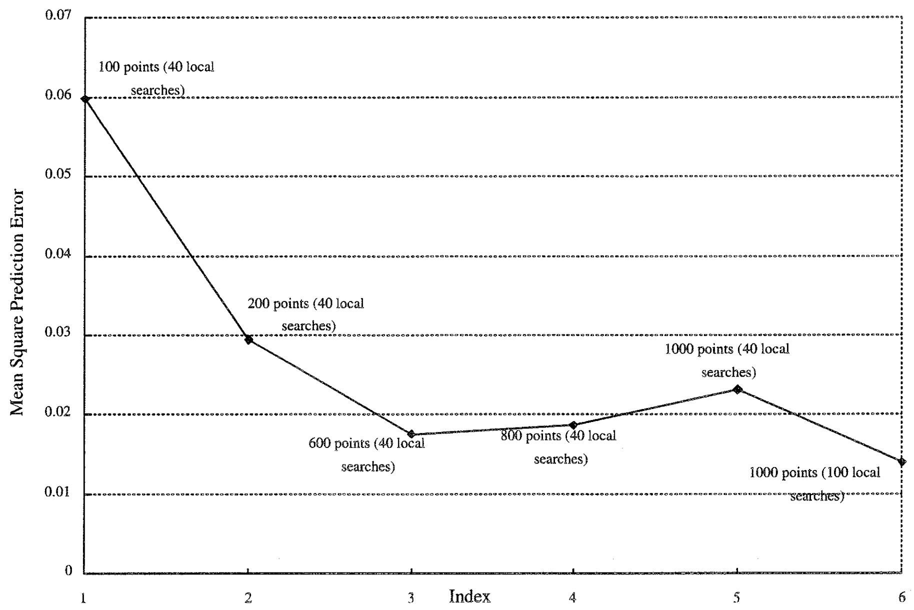

Figure 3. Mean square prediction errors obtained using different set of hyperparamter

Comparing this result with results obtained in Fig. 1 and Fig. 2, although all three techniques offer consistent results, the ANOVA technique offers a much clearer picture and can also be used to other types of surrogate models. However, all these techniques should be used to complement each other. It should also be noted that for problems with parameters of similar effect, it is often down to the designer to filter out certain groups of parameters, and these types of techniques are not very helpful. Also, it is sometimes necessary to look at interactions between variables before any variables can eliminated from the optimization study.

VI. Application on Engine Nacelle Shape Optimisation

An aeroengine nacelle shape optimization problem has been formulated as a 33 variable unconstrained design problem. The design variables are geometry parameters used to define the parametric geometry in ProEngineer. Two objective functions are investigated in the study, the first is the total pressure recovery at the fan face and the second objective is the negative scarf angle which is used to indicate the level of noise radiated to the ground. A Fluent model is built and solved to give total pressure recovery at the fan face. The geometry model is shown in Fig. 5

In order to obtain a more complete picture regarding the trade off between scarf angle and pressure recovery, a multi-objective Genetic Algorithm (GA), NSGA II[9] is used in this study to build a Pareto front between these two objectives. To overcome the challenge of using 33 variables in the problem, an initial study was made to identify the most important design variables to reduce the computing time required to achieve a result. Based on the techniques described in previous sections, seven of the 33 variables including the scarf angle have been chosen to formulate a reduced problem. Two approximated Pareto fronts have been obtained for the two objective functions, as shown in Fig. 6. The second Pareto front is produced by using a surrogate built from augmented data points suggested by the first Pareto front. Exact evaluations using CFD for points on the second Pareto front show improvements over points on the first Pareto front. It should be noted that the use of scarf angle as an indication of ground noise level is very preliminary and that designs with very high scarf angle can result, as shown in this case.

To illustrate the effectiveness of the approach used, a full search using 33 variables is also carried out and three Pareto fronts are shown in Fig. 7, it can be seen that the final set of points from studies using reduced models and full model are very similar.

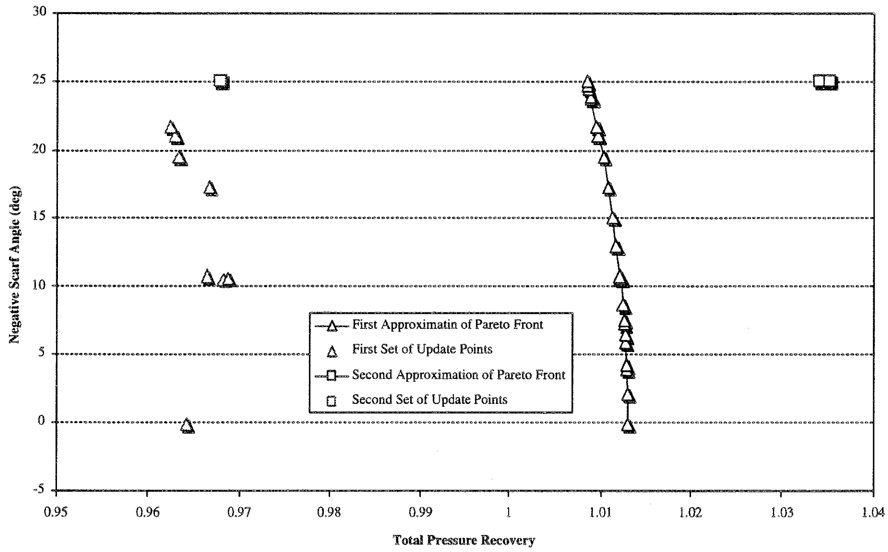

Figure 6. Updated points and approximated Pareto fronts for the six variable problem

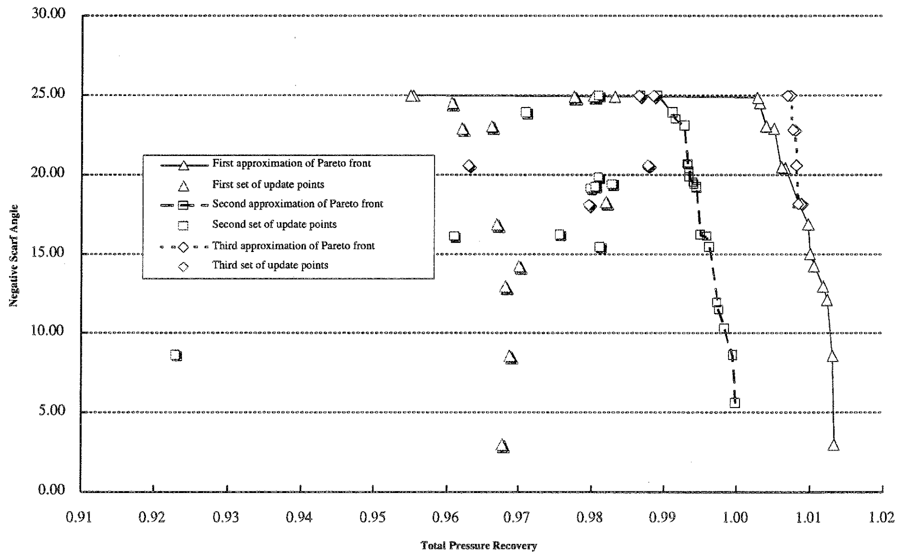

Figure 7. Update points and approximated Pareto fronts using 33 variables

VII. Conclusions

Based on a snapshot data sets obtained by evaluating design points decided by Design of Experiments techniques, a simple concept of parameter impact factors is introduced and calculated to obtain preliminary estimates on the importance of parameters in simulation results. Combined with parallel tuning of hyperparameters used in Gaussian process surrogate models and ANOVA techniques using the progressively refined surrogate models, a more accurate estimation on the impact of different parameters can be achieved. Less important parameters can then be fixed in order to reduce the dimensionality of a problem to make the problem more tractable within given computational budget and time constraints.

References

[1]Livne, E., "Special issue: Multidisciplinary design optimization of aerospace systems," Journal of Aircraft, Vol. 36, No. 1, 1999, pp. 9-10.

[2]Keane, A.J. and Nair, P.B., "Computational Approaches for Aerospace Design", John-Wiley and Sons, ISBN 0-470-85540-1, 2005.

[3]Agarwal, H. and Renaud, J. E., "Reliability based design optimization using response surfaces in application to multidisciplinary systems," Engineering Optimization, Vol. 36, No. 3, 2004, pp. 291-311

[4]Simpson, T. W., Booker, A. J., Ghosh, D., Giunta, A. A., Koch, P. N., and Yang, R. J., "Approximation methods in multidisciplinary analysis and optimization: a panel discussion," Structural and Multidisciplinary Optimization, Vol. 27, No. 5, 2004, pp. 302-313.

[5]Jones, D. R., Schinlau, M., and Welch, W. J., "Efficient Global Optimisation of Expensive Black-Box Functions," Journal of Global Optimization, Vol. 13, 1998, pp. 455-492.

[6]Simpson, T. W. and Mistree, F., "Kriging models for global approximation in simulation-based multidisciplinary design optimization," AIAA Journal, Vol. 39, No. 12, 2001, pp. 2233-2241.

[7]Rodriguez, J. F., Renaud, J. E., Wujek, B. A., and Tappeta, R. V., "Trust region model management in multidisciplinary design optimization," Journal of Computational and Applied Mathematics, Vol. 124, No. 1-2, 2000, pp. 139-154.

[8]Ong, Y. S. and Keane, A. J., "Meta-Lamarckian Learning in Memetic Algorithm", IEEE Transactions on Evolutionary Computation, Vol. 8, No. 2, 2004, pp. 99-110.

[9]Deb, K., Pratap. A, Agargal, S. and Meyarivan, T. "A fast and elistist multi-objective genetic algorithm: NSGA-II", IEEE Transactions on Evolutionary Computation, 6(2), 181-197. 2002.