UNIVERSITY OF SOUTHAMPTON

FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

Aeronautics and Astronautics Computational Engineering

Volume 1 of 1

MODEL VALIDATION AND UNCERTAINTY QUANTIFICATION FOR THE PRELIMINARY AERO-ENGINE DESIGN PROCESS

by

BOGDAN PROFIR

Thesis for the degree of ENGINEERING DOCTORATE

05/2019

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND THE ENVIRONMENT

Computational Engineering Design Research Group

Thesis for the degree of Engineering Doctorate

MODEL VALIDATION AND UNCERTAINTY QUANTIFICATION FOR THE PRELIMINARY AERO-ENGINE DESIGN PROCESS

Bogdan Profir

This thesis investigates the design decisions taken during the preliminary aero-engine design process where the amount of knowledge is limited, although deciding on fuel efficiency, noise, emissions, weight and overall performance occurs within this stage. In order to not commit all resources during this phase, those decisions are made using low fidelity models. Unfortunately, the results from low fidelity models lack accuracy, so there is a natural need to take this into account. Improving those low fidelity methods via uncertainty quantification methods is the main theme of this thesis.

In order to create accurate models for the preliminary design stage of the aero engine, a probabilistic framework was created and implemented. This framework is based upon suggestions from the literature and was constructed from two main components: expert systems as well as Bayesian inference. The way in which this was developed will be shown in detail later. This framework was applied to three aero-engine related case studies which reflect the need to have more knowledge available during the preliminary stage.

The results obtained from the framework have the form of predictions which offer information which was not available otherwise. For the fan-blade off case study, the posterior predictive distributions show what the characteristics of the most likely events are, and this information can be

used to make the detailed design stage less expensive by doing fewer finite element analysis simulations. For the grain growth case study, the results show what probability distributions the manufacturing inputs should have in order to keep the size of the grain within certain limits in order to maximize the life cycles. Finally, the results obtained with regards to fatigue failure due to non-metallic particle inclusions show how component life as well as failure cause can be obtained. This type of knowledge was not previously available in the literature, and making use of it can avoid removing components from their service too early.

Table of Contents

Table of Tables

Table of Figures

Academic Thesis: Declaration Of Authorship

I, Bogdan Profir

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Model Validation and Uncertainty Quantification for the Preliminary Aero-Engine Design Process

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published in three conference papers for the AIAA (American Institute of Aeronautics and Astronautics) Forum and Exposition conferences in 2016, 2017 and 2018 [1], [2], [3], one conference paper for the ASME TurboExpo conference in 2017 [4] and one Rolls-Royce plc technical design report [5].

Signed:

Date:

Acknowledgements

Completing this EngD would not have been possible without the support of both academic and industrial supervisory teams who offered me their support regarding the carrying out of this project. First and foremost, I would like to thank my main supervisor, Dr. Hakki Eres, whose continuous support and positive attitude had a tremendous impact on my research direction as well as on my overall experience. I would also like to express my thanks towards my secondary supervisor, Professor James Scanlan, who although I have not met very often, did not hesitate to offer me timely advice and feedback whenever I was intending to submit a publication.

This research work has also seen a great amount of support from Rolls-Royce plc. More specifically, my industrial supervisor Dr. Ron Bates was always keen to offer me important technical advice especially in the field of statistics. In addition to that, I would like to thank Dr. Christos Argyrakis with whom I have worked and who, besides a plethora of valuable input, always tried to put me in contact with various Rolls-Royce plc employees who could potentially benefit from my work. I am certain that a large part of this work would not have been possible without Dr. Michael Moss, who suggested a variety of case studies in order to develop my framework, and also introduced me to the right people when I arrived at the company.

Besides the academic and industrial teams mentioned above, I would also like to express my thanks towards my parents, Silviu and Gabriela, who offered me their unconditional support during this EngD and well beyond that. Moreover, none of this would have been possible without the person who ignited my passion towards physics and without whom my life at the moment would have been completely different: my high school physics teacher Vasile Ciuchina. Last but not least, I would like to thank all my friends without whom this whole journey would have been very lonely.

Nomenclature

Abbreviations

BIC	Bayesian information criterion
CDF	Cumulative distribution function
CFD	Computational Fluid Dynamics
DReX	Dynamically recrystallized grain
EFSA	European Food Safety Agency
FBO	Fan blade-off
FEA	Finite element analysis
GEM	General Expertise Measure
MBA	Master of Business Administration
MCMC	Markov Chain Monte Carlo
OOB	Out of balance
PDF	Probability distribution function
RLNL	Red line speed
SHELF	Sheffield Elicitation Framework
TSFC	Thrust specific fuel consumption

Mathematical Symbols

c	Cause of failure
dmg	Damage
f	Function/Likelihood function/PDF
F	CDF
i, j	Variable subscripts

k	Variable subscript/Number of outcomes
n	Total number of outcomes
C_n^k	Binomial coefficient
q	Distribution quantile/Proposal distribution
T	Temperature
T^*	Non-dimensional temperature
v	Rotational speed
x	Argument of the function f
$\boldsymbol{x}$	Argument vector for the function f
X	Random variable
$E(X)$	Expected value of X
$P(X = k)$	Probability of X being equal to k
$\mathrm{Var}(X)$	Variance of X
α, β	Beta distribution parameters
ε	Error/Strain
ε_h	Error due to numerical approximation
ε_m	Error due to modelling
ε_d	Error due to data uncertainty
Φ	CDF of the standard normal distribution
Γ	Gamma function
θ	Calibration parameter
$\boldsymbol{\theta}$	Vector of calibration parameters
μ	Mean (same as $E(X)$)
π	PDF (this form is used with the Bayes' rule)
σ	Standard deviation

Chapter 1 Introduction

It is a sensible assumption that the aero-engine is one of the most impressive and complex human achievements, as it encapsulates the simple idea which started it all: the desire to fly. As it can be expected, designing one is a multi-faceted task as there is a high number of factors which need to be considered. One of the influencing factors is the current market that needs to be understood, which is why a brief overview regarding the aero-engine market needs to be provided. Afterwards, the spotlight will be put on the preliminary stage of the design process and its importance will be discussed.

1.1 The Aero-Engine Market

The growth of the airline industry has been uninterrupted over the last decade and a half. This can be easily deduced by the fact that the revenue has increased from $379 billion in 2004 to $812 billion in 2018, and by the end of 2019 it is predicted to increase by another 6.5%, which would let it reach the considerable amount of $865 billion [6]. In addition to that, The International Air Transport Association (IATA) forecasts that by 2035 the number of passengers will virtually double compared to 2016 [7]. Although this sounds promising, it still does not show the entire picture. For example the profit margin has only recently started to increase from below 1% to around 7%-8% [8]. This is due to the fact that in order to be cost efficient, airlines have to employ various cost reduction strategies in a variety of operations, spanning from crew management all the way to ticket sales. Even so, in order to achieve a sizeable cost reduction, it is necessary to have a well-designed management strategy. Ultimately, this would allow cost reductions in the most important areas such as maintenance and fuel consumption.

Even so, the factors listed above are not nearly close to forming an exhaustive list that airlines need to take into consideration. Besides what has already been mentioned, the most important factor – safety needs to be given proper thought. In addition to that, there are environmental regulation constraints such as fuel efficiency and noise reduction. In order to improve on those, it is required to invest in research and development in order to look into novel technologies. Those can range from investigating novel materials to developing new computational fluid dynamics methods (and ultimately solving the Navier-Stokes equations!). In the end, it is crucial to remember that all of those areas are building blocks for what

constitutes the full picture that is aircraft design. For it is aircraft design that is the main research driver for the aforementioned fields as well as for this project.

As the theme of the project refers to aero-engines (more specifically commercial engines), it is worth narrowing down our scope to this topic. Consequently, it is worth looking at the market which is specific to the aero-engines themselves. This is done in order to put the work in the context of the market as well as to be able to motivate the chosen direction of research.

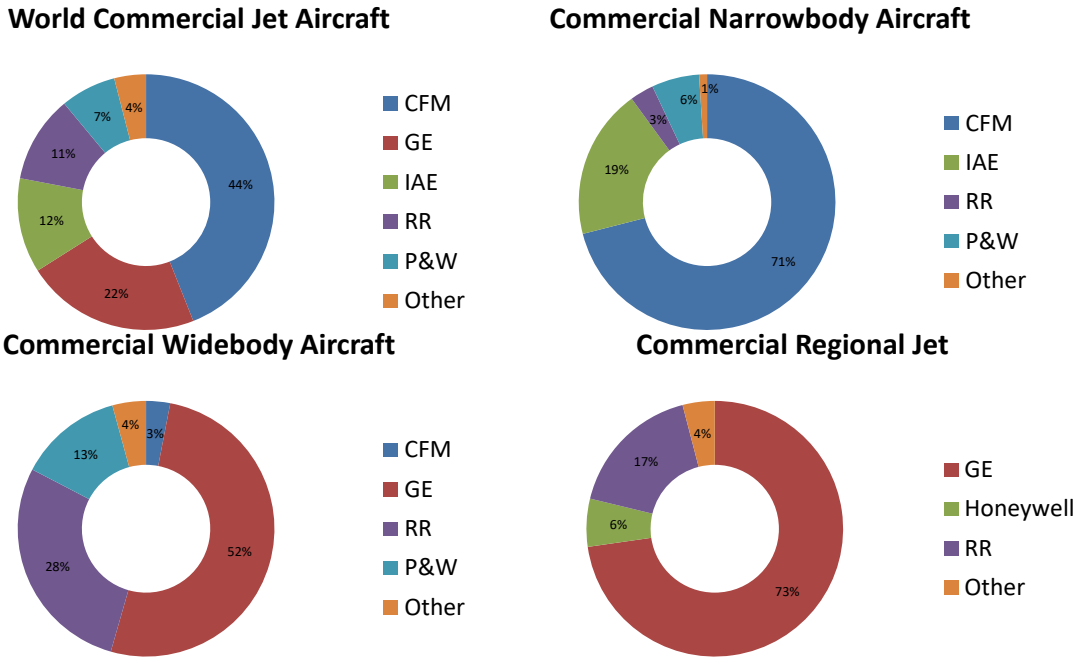


Figure 1: Market share for the various engine manufacturers (2018) [9]

Figure 1 shows three main engine manufacturers which dominate the market: General Electric Aviation (GE), Pratt & Whitney (P&W), as well as Rolls-Royce plc (RR). In addition, those companies also form consortia in order to both share development effort as well as increase the probability of securing orders. The main three consortia are therefore: CFM International (CFM), International Aero Engines (IAE) and Engine Alliance (EA). By inspecting Figure 1, it can be inferred straightaway that CFM seems to be the dominating force in terms of World Commercial Jet Aircraft. However, it is also worth having a closer look at each individual sector in order to better understand the specific contributions of each manufacturer. For example, while CFM still operates 71% of the narrowbody aircraft, GE dominates the market both in the widebody and regional jet market groups, with the specific shares of 52% and 73% respectively. Rolls-Royce plc on the other hand does not

hold a majority of the market share in any of the above groups, although it still holds shares of 28% and 17% both in widebody as well as regional jet sectors, which is enough for it to be considered a major aero-engine manufacturer.

In order to have a clearer picture regarding the aero-engine market, it is worthwhile to look not just at the current market, but also at how it is predicted to change over the next timeframe of 10 years. By 2028, the commercial fleet is expected to grow at a rate of 3.8% per year [9], eventually increasing from 30,000 to 44,000 aircrafts during this time span. The actual number of aircraft forecasted to be delivered between now and 2028 is expected to be around 22,000 while the engine delivery forecast shows 44,000 engines that are expected to enter operation during this time. The principal aircraft types which are driving this growth are Airbus 319, Airbus 320, Boeing 737 MAX and Boeing 787, while the relevant engines which will go into service are the Trent XWB, Trent 7000, Trent 1000 as well as the CFM56-LEAP-1B [9]. The substantial contribution of Rolls-Royce plc in manufacturing engines over the next 10 years is backed up by their confidence that their Trent XWB is going to "set a new record for reliability" [9], and also under the UltraFan programme, the new engine they are developing is set to reduce fuel consumption by 25% relative to the current Trent XWB, and is scheduled to enter service in 2025. Additionally, Rolls-Royce plc. also attains engineering excellence by investing more than £1.3 billion in research and development spread across 31 different University technology centres and 7 advanced manufacturing centres [10].

It is undeniable that the aero-engine market is constantly growing, especially when looking at predictions made for the following 10 years. One of the crucial aspects for any manufacturer to secure their market share is to be able to constantly be able to provide new technologies whose ultimate purpose is to reduce cost while maintaining high safety standards. For instance, in the Rolls-Royce plc vision, shifting the entire market towards electric propulsion is not a matter of "if", but rather a matter of "when" [9]. "Electrifying the market" will have the effect of disrupting current business models, especially those which do not yet give electric power the proper credit. This vision is not unique to Rolls-Royce plc; they actually share it with its competitors: General Electric Aviation, Pratt & Whitney and Honeywell have all disclosed their intention for their power plants to be electrically driven [9]. In the end, integrating any technology requires a well thought out design process.

The type of aero-engine that is considered in this thesis is the turbofan, as that is the one used for civil aircraft across the industry. The reason why the turbofan

engine is preferred traces back to the idea that it is more efficient to slowly accelerate a large amount of air (this is achieved via the bypass area) rather than to sharply accelerate a smaller amount (which is done by turbojet engines generally used on military aircraft). A simplified schematic of a turbofan engine is shown below:

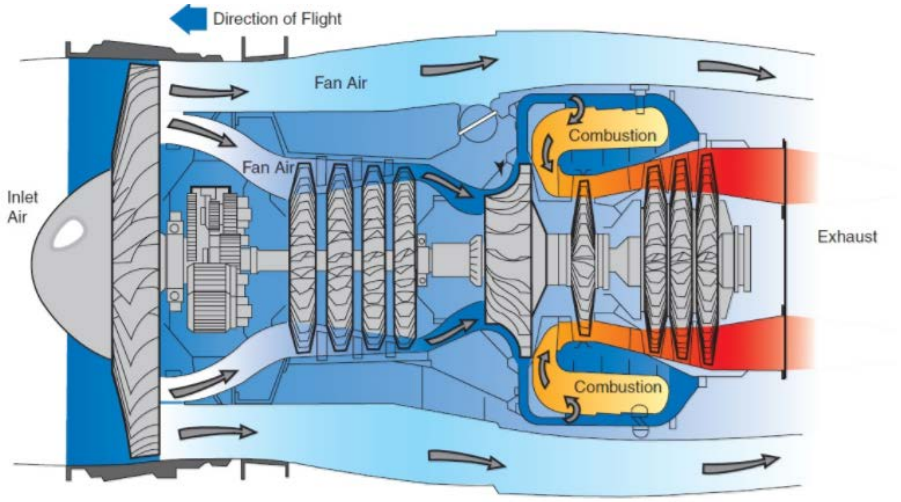


Figure 2: Diagram of a turbofan engine [11]

The principle behind thrust generation is the increase in the velocity of the gas which is achieved by increasing the temperature of the mixture passing through the core of the engine, which in turn does work on the turbine that is connected to the compressor and the fan. The fan has main role of compressing the air passing through the bypass area, which contributes the most to the total amount of thrust.

To understand why an optimum design process is needed, it would be necessary to examine each of its stages. Therefore, the scope of the following section is not only to do that, but to also highlight that the importance of the preliminary stage of the design process should not be underestimated, as that is a paramount component of the workflow which eventually leads to a successful state-of-the-art aero-engine.

1.2 The Engine Preliminary Design Process

Before specifically referring to the aero-engine design process, it would be useful to have a definition for the general design process itself as well as the steps contained by it. A rigorous definition which has been given by the Engineering Accreditation Commission is as follows: “Engineering design is the process of devising a system, component or process to meet desired needs. It is a decision-making process (often iterative), in which the basic sciences, mathematics and the engineering sciences are applied to convert resources optimally to meet these stated needs” [12]. One recurring feature which appears in most definitions of a design process refers to “decision-making”. This is emphasized by Keane and Nair [13] who state that the decision-making process needs to be on the forefront whenever a design is being made. At its very core, the decision-making process can simply be split into two main parts: exploring the design space and choosing the best design. The main problem with this description is that it is very generic and does not give enough information regarding the actual steps taken until the final design is reached.

One useful method of quantifying the design process is proposed by Verhagen et al. [14]. They separated the design process into three main phases: conceptual design, preliminary design and detailed design, each of those serving a different role. The conceptual phase involves a top-level view which includes general design requirements related to the final purpose of the product. In the case of the aero-engine, it refers to considering the life, the emissions and the overall cost. It can also refer to more specific parameters such as the turbine entry temperature as well as the compression ratio. The output from the conceptual stage will feed into the preliminary one. At this point, the designer’s focus will be on the individual components such as the fan, compressor, combustor and turbine. Due to the iterative nature of the whole design process, at this point it needs to be investigated whether the findings from the preliminary stage agree with the requirements from the conceptual stage. If that is not the case, then the conceptual stage will need to be reassessed. The same applies to the detailed design stage, where deeper analysis is being done. In the case of an aero-engine, high fidelity Computational Fluid Dynamics (CFD) and Finite Element Analysis (FEA) can be done on the various components, and this process is usually time consuming. Therefore, the designer needs to make sure that the conceptual and preliminary stages have been analysed in a robust manner. Even so, it may be the case that the converged design at the end of the detailed phase will still not satisfy all the design requirements due to

conflicts with either of the two previous design stages (i.e.: the preliminary and the detailed design stages could yield different values for the turbine entry temperature). In this case, there is a need to return to the problematic stage in order to re-analyse and eventually create a final design which successfully passes all design stages and ultimately checks all requirements.

The three main design stages can be split even further. Once an aero engine project has started, six different stages can be considered to take place along the said project's life. These are requirements specification, conceptual design, preliminary design, detailed design, development, and production. From the identification of the customer requirements to the eventual production of the engine, the three variables (cost, knowledge and freedom) are closely linked to each of the stages shown in Figure 3. What can be inferred straightaway is the fact that those variables might differ greatly across all those stages. Because of the shapes of the graphs below, it comes as no surprise that most if not all the decisions taken during the life of the project have to be based on trade-offs. In the most ideal case, the design space should be fully explored before committing to one particular engine design. As this is hardly practical, it means that the knowledge which the designer has up to and during the preliminary design process is below 25% [14]. Although the incurred cost is arguably low (around 10%), in reality the committed cost almost peaks at this point (80% - 90%) and consequently, there might not be enough freedom to adjust the design subsequent to the preliminary design stage. This fact stands to emphasize that taking the correct decisions during this stage is crucial. However, according to Verhagen et al. [14], because knowledge is limited in this region, taking the appropriate decisions can represent a challenge in itself.

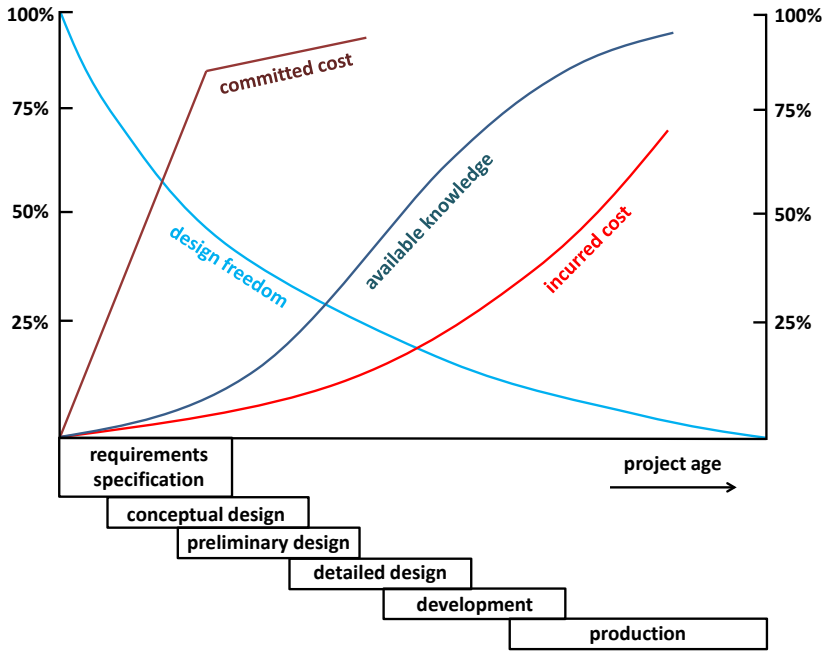


Figure 3: Cost, design knowledge and freedom along the different stages for any preliminary design process [14]

The decisions which need to be taken during the preliminary design process can generally be twofold: doing low fidelity or high fidelity analysis. The first method is able to explore the design space extensively; however it does not necessarily give an accurate picture of it, as it sacrifices accuracy for speed. Consequently, the results it gives should at most only be viewed as trends and not as exact. Conversely, high fidelity methods work in the opposite way: they focus on exploitation rather than exploration. Due to the relatively high cost of those methods, they can only focus on one part of the design space and can identify local optima in that region. However, there is no reason to assume that there are no other designs in other parts of the design space which are more suitable than the one identified in this way. Unfortunately, as high fidelity analysis is expensive, exploring more than a small area of the space might be unfeasible.

What can immediately be concluded is that by solely relying on either of the above techniques, it is unlikely that an optimum engine design will be reached [13]. Developing this idea further, it also means that it will be difficult for the respective company to secure market share. As a result, a different strategy needs to be used. Although deterministic low fidelity techniques are usually preferred for the preliminary design stage as they can indicate relevant features of the design space, the claim here is that there is a way to improve this. Namely, by using a probabilistic framework which combines the benefits of low fidelity and high fidelity methods as a “best of both worlds” approach, it is possible to save time during the

preliminary stage. Details about uncertainty quantification methods will be discussed in Chapter 2, while the next section summarizes the research hypothesis.

1.3 The Research Hypothesis

The research hypothesis of this project is motivated by the above statements. To sum up, the preliminary stage of the design process of the aero-engine is delicate, as a balance between cost, knowledge and design freedom needs to be reached. In order to do this, techniques for investigating the design space need to be employed which have to strike equilibrium between exploration and exploitation. Knowing this, the research hypothesis can be formulated:

"Using probability-based methods during the preliminary stage of the aero-engine design process can allow fast and accurate investigation of the design space in order to aid the generation of an optimum design which can afterwards be passed to the detailed design phase. Ultimately, this can have the effect of making the entire design process faster and less expensive."

The following section has the goal to make the original contributions clear to the reader. Afterwards, the next chapter describes the literature survey which covers the theory used within this work.

1.4 Summary of Original Contributions

The principal contribution was the development of a framework which is deemed suitable to use during the preliminary aero-engine design process. This framework combines a few ideas: knowledge gathering techniques that can be used especially when data is scarce, Bayesian calibration, as well as putting everything together within an expert system which can be applied to a large variety of aero-engine design studies. Details regarding the framework are presented in Chapter 3.

The Bayesian-elicitation framework was written by using custom code. This is also claimed to be a novelty, as other Bayesian software packages found in the literature do not have the capability of being meshed with the elicitation component, which is also a major part of this EngD project. The code is presented throughout Chapter 3 in the form of pseudocode snippets.

The application area of this framework is also considered to be new, as there are no journal articles in the literature which describe Bayesian networks being applied to problems related to the aero-engine such as the ones shown in Chapter 4 of this thesis. It should also be noted here that the proposed framework can be applied to other aspects of gas turbine design. If a new case study needs to be considered (whose application area is a different part of the engine), as long as uncertainty quantification remains a top priority, the framework maintains its usefulness. The reason is because the code has the capability of accommodating the new physical model in order to subsequently do predictions.

Additionally, an alternative way of gaining feedback during expert elicitations will be presented. Most literature sources hint towards obtaining feedback regarding the probability density functions (PDFs) themselves which serve as Bayesian priors. However, another way of doing this is to obtain feedback regarding the prior predictive distribution. A clearer description regarding how this works can be found both in section 3.3 and section 4.3.3 during the outline of one of the case studies.

Chapter 2 Literature Review

While in the previous chapter the goal was to set the stage and put the project in context of the current aero-engine market, the purpose of this chapter is to explore the literature regarding probability techniques which have been considered while developing the framework which is going to be described in detail in Chapter 3. The literature review covers the following topics: uncertainty quantification, Bayesian updating, expert systems as well as expert elicitation techniques.

2.1 Uncertainty of the Preliminary Aero-Engine Design Process

The motivation for exploring this particular challenge will be given first in order to comprehend why this matter is relevant to the aero-engine design process. Typically, this motivation arises from the fact that in any real-life engineering system, it is virtually guaranteed that some form of uncertainty appears, so using deterministic values for the different parameters is far from ideal. In the case of a structural system, there will be uncertainties appearing from different sources: uncertainties due to material properties (e.g. Young's Modulus and Poisson's Ratio) caused by small scale defects, uncertainties in dimensions due to tolerances in the manufacturing process or uncertainties due to the operating environment itself. Rolls-Royce plc. does recognise the need for uncertainty quantification, therefore the engineers involved in design use Monte Carlo simulations in order to propagate variation in parameters whenever they analyse processes such as blade design, manufacture and blade testing [10].

It is common practice for the nominal values to be used in engineering in order to describe parameters which may or may not be deterministic in the first place. The problem with doing so is that the performance can degrade fast because of perturbations in the design variables. This behaviour is further exacerbated if the design had undergone some form of optimisation beforehand; any perturbation can move the design away from the extrema, which can lead to a catastrophic drop in performance or a violation of constraints. This behaviour in particular can be visualized in Figure 4 below.

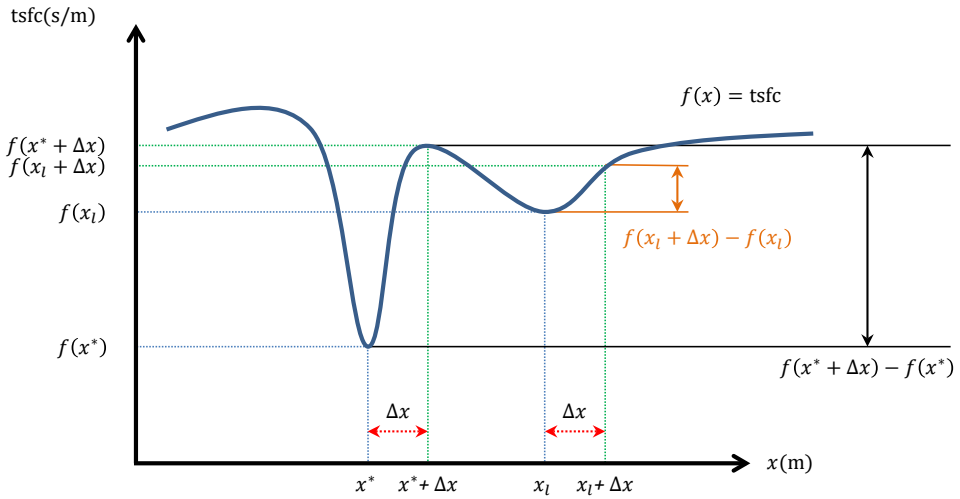


Figure 4: Increased sensitivity arising from design perturbations

Assuming that $f(x)$ is a function of some dimension within the engine (x could represent the fan diameter for instance) which gives the thrust specific fuel consumption (tsfc), it can be seen that there are two different minima on the graph in Figure 4 (This example is completely fictitious and it was chosen for illustrative purposes only). Although x^* is a better choice if the goal is to minimize thrust specific fuel consumption, because of the steep slopes of the graph around it, a small perturbation can lead to a high performance drop compared to the case in which x_l had been chosen as the design point. This simple yet effective example shows that finding the deterministic value of the optimal solution of any real world problem is not enough; the full picture needs to be taken into account and that includes the effects of perturbations. In structural design, perturbations might lead to stress levels which are close to the yield strength of the material, which is something that needs to be avoided at all costs. One solution to circumvent this would be to impose safety factors to ensure that the stress level gets nowhere near the critical region. However, this can lead to over engineering which can translate to an increase in overall weight and consequently to an increase in manufacturing and operating costs.

Creating a design which is insensitive to perturbations from different sources equates to creating a robust design. The concept of robustness can be traced back to the Japanese engineer Genichi Taguchi who believed that “one should design a product in such a way so as to make its performance insensitive to variation in variables beyond the control of the engineer” [13]. Figure 5 shows a graphical way of interpreting robustness in a probabilistic fashion.

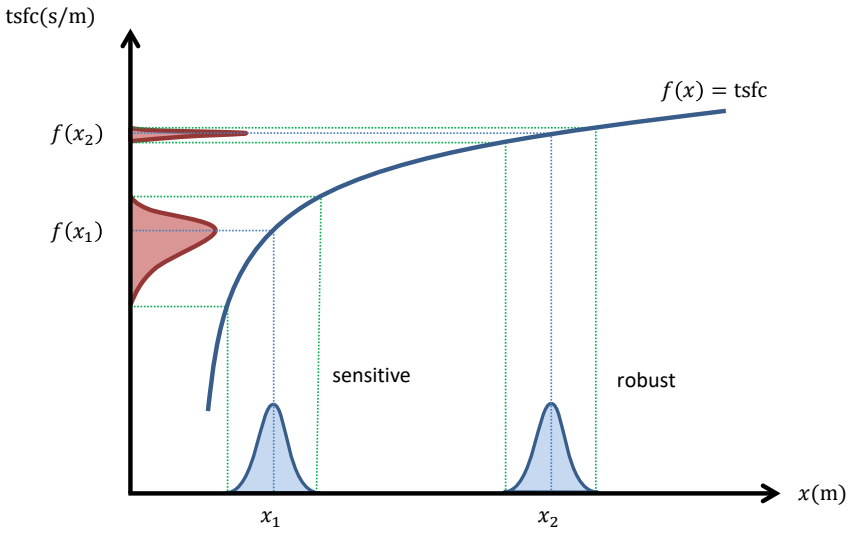


Figure 5: Comparison between a sensitive and a robust design

A fictitious function representing thrust specific fuel consumption is once again used as an example to illustrate the idea of robustness. Although the design point x_1 is better than x_2 from a deterministic point of view, it is more sensitive to perturbations, so choosing between the two requires a more detailed analysis.

Taguchi classified the design parameters of any general design in two main groups:

- Control parameters: those are the parameters which can be easily controlled by the engineer
- Noise parameters: they are much harder and expensive to control

For example, in an aircraft wing design problem, the control parameters could be the wingspan/chord, while the Mach number or the geometric uncertainties due to manufacturing constitute the noise parameters.

Keane and Nair [13] identified three main steps which need to be taken in order to be able to efficiently optimize a design in the presence of uncertainty:

- Identify all sources of uncertainty and convert them into mathematical models.
- Use uncertainty propagation techniques in order to quantify their effect on the output.
- Formulate an optimization problem in such a way so that the robustness of the optimum solution is ensured.

Those steps first require a more detailed accounting of the types of uncertainties present in the design process. Oberkampf *et al.* classify uncertainties as being aleatory or epistemic [13], [15]. Aleatory uncertainties, also known as *type A* or *stochastic uncertainties* [16] generally occur due to variations of a parameter, and if enough data is available, they can be characterized by a probability density function (PDF). Epistemic uncertainties (also known as *type B* or *cognitive uncertainties* [16]) on the other hand represent a lack of knowledge of a certain aspect of the problem. Variables about which not enough information is known can be represented by intervals instead of PDFs. A significant difference between the two types is that if the amount of data becomes sufficient, then the epistemic uncertainties can be removed completely; the same behaviour does not apply for the aleatory type [15].

A much more detailed and practical classification of uncertainties is made by Thunnissen [17] who classifies them into ambiguity uncertainties (caused by imprecision or vagueness), epistemic and aleatory uncertainties (which have the same definition as above) as well as interaction uncertainties due to unforeseeable sequences of events. It might be more convenient to categorize uncertainties in a way which depends on the application undergoing research. For instance, Walton [18] focused on the lifecycle of a space system and the uncertainties were split into development, operational and model uncertainties. In his doctoral thesis, Padmanabhan [19] considered that generally, uncertainties arise due to variations of the design parameters themselves, the formulation of the problem and also due to numerical errors and approximations. Keane and Nair [13] also focused on those three main sources of error, and he also illustrated an expression which highlights this idea:

$$\varepsilon = \varepsilon_h + \varepsilon_m + \varepsilon_d \quad [13]$$

The total error is given by ε, while the constituent terms are:

- ε_h is the error due to numerical approximations.
- ε_m is the modelling error which could be reduced by using a more detailed model (in fluid mechanics for instance, it can be reduced by using a Large-Eddy Simulation model instead of the Reynolds-averaged Navier-Stokes equations).
- Finally ε_d is the error due to data uncertainty and it can be quantified by using several methods which will be explored further in this section.

Uncertainty modelling is mostly done by using probabilistic techniques as plenty of research has been invested in this area and as a result it is better known than other non-probabilistic techniques which include evidence theory, possibility theory and interval analysis [13]. The framework which constitutes the basis of this project is focused on probabilistic techniques only, which is why they will be the main focus of the thesis.

When trying to represent parameter uncertainty which occurs during the design process, there are generally three main types of probabilistic models which can be used: random variable, random field as well as time-dependent stochastic processes [13]. It can be argued that it is most appropriate to use probabilistic models when sufficient computational or empirical data regarding the parameters of interest exists. This approach belongs to one of the two main schools of thought in statistics called the frequentist approach. This assumes that the variables of interest have fixed unknown values which can be quantified using confidence intervals. However, this method is not accurate unless data is easy to be obtained. Indeed, for the case studies presented in Chapter 4, it will become clear that obtaining data is time consuming and expensive, therefore another method of quantifying uncertainties is preferred. The alternative lies in the Bayesian approach, whose basic philosophy is that probability is the only measure of uncertainty [20]. As a consequence, even if not enough data is available, modeling the uncertainties can include past experiences as well as expert opinions. Lindley [21] even makes a very bold statement, saying that “Every statistician would be a Bayesian if he took the trouble to read the literature thoroughly and was honest enough to admit that he might have been wrong.”

In a real situation, there is limited data in the early stages of the design process, meaning that finding distribution moments based on real data can be extremely difficult, which means that employing the frequentist approach is not appropriate. As a result, the next section will look into the theory regarding Bayesian inference.

2.2 Bayesian Analysis

Bayesian theory provides an approach to statistical inference which differs from the well-known classical frequentist approach [21]. This difference lies in the form of the unknown parameters. In the frequentist context, those parameters are considered to be fixed meaning that they can be captured within confidence intervals given that a significant number of experiments are performed. The philosophy behind Bayesian analysis is that probability is the only measure of uncertainty and consequently, the unknown parameters from the problem should only be modelled as probability distribution functions instead of assigning them fixed unknown values. In the Bayesian context, those PDFs are also known as priors [22] (denoted by $\pi(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the vector containing the parameters of interest) and they capture knowledge about the parameters before the data has been observed. They can have the form of expert knowledge which is subsequently updated via the likelihood function which contains data that has been observed. It is common in the literature to refer to the likelihood function as "probability of the data", and this formulation can cause confusion because it is not immediately apparent what it means. Technically speaking, this expression is incomplete, as the Bayesian statistician is not interested in the probability of the data itself; the greatest interest lies in how probable the data is *given* the initial belief of the values taken by parameter $\boldsymbol{\theta}$. As a result, it is common to mathematically express the likelihood function as $f(\boldsymbol{x}|\boldsymbol{\theta})$. This formulation of the likelihood is analogous to the way in which conditional probabilities are expressed, which stands to emphasize the ideas presented before. After the data has been observed and captured within the likelihood function, the prior is updated into a posterior denoted by $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ which can be translated as "the belief regarding parameter $\boldsymbol{\theta}$ given the observed data $\boldsymbol{x}$. If the values which parameter $\boldsymbol{\theta}$ can be captured within an interval, then mathematically, the posterior can be written as it is shown in Equation (1):

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{-\infty}^{+\infty} f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}} \tag{1}$$

If on the other hand, $\boldsymbol{\theta}$ takes discrete values, then the expression for the posterior will have the equivalent form shown below:

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\sum_{-\infty}^{+\infty} f(\boldsymbol{x}|\boldsymbol{\theta_i})\pi(\boldsymbol{\theta_i})} \quad (2)$$

The formulations of the posterior from Equations (1) and (2) constitute what is known as the Bayes' rule. One of its advantages stands in the fact that the technique can be used in an iterative fashion. Assuming that the first posterior had been obtained based on the data available at the moment, whenever a newer set of data becomes accessible, a re-updated posterior can be obtained. This concept is demonstrated below.

Assuming two data points x_1 and x_2 forming vector $\boldsymbol{x}$ have been obtained independently and sequentially, then the resulting posterior can be written in the following way [20]:

$$\pi(\boldsymbol{\theta}|x_1, x_2) = \frac{f(x_1, x_2|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{f(x_1, x_2)} = \frac{f(x_1|\boldsymbol{\theta})f(x_2|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{f(x_1)f(x_2)} \quad (3)$$

$$\pi(\boldsymbol{\theta}|x_1, x_2) = \frac{f(x_2|\boldsymbol{\theta})}{f(x_2)}\left(\frac{f(x_1|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{f(x_1)}\right) = \frac{f(x_2|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|x_1)}{f(x_2)} \quad (4)$$

Equation (4) shows that the posterior density after the two data points have been obtained can also be found by first calculating the posterior corresponding to the first data point x_1 and then using this as the prior in order to find the final posterior $\pi(\boldsymbol{\theta}|x_1, x_2)$.

In order to illustrate how Bayesian updating works, an example will be given in section 3.1 of the Methodology chapter which will also illustrate the concept of conjugacy.

This "specific manner" of choosing priors and likelihoods in order to obtain conjugacy can be seen in Figure 6 and Figure 7 below [23]:

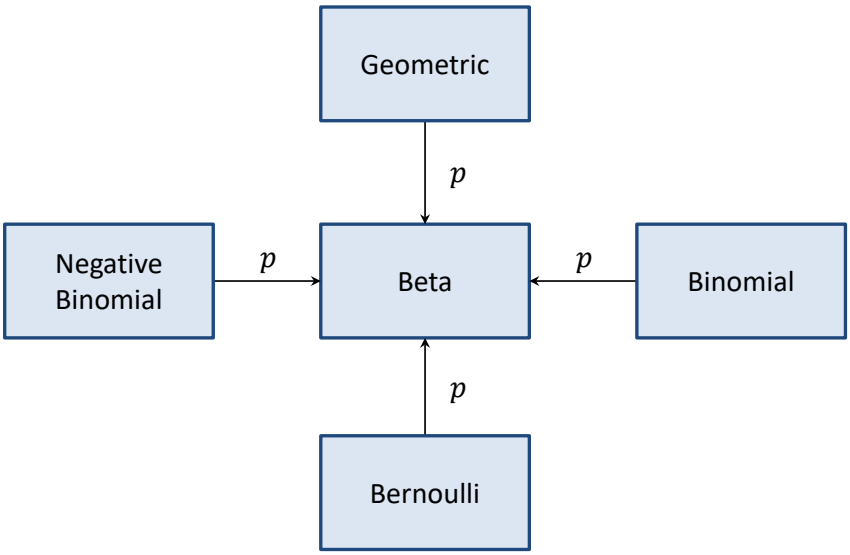


Figure 6: Natural conjugacies regarding the beta prior (arrows start from boxes representing the likelihood function and point to the box showing the conjugate prior). The parameter on each arrow shows the parameter for which the prior is used to quantify uncertainty [23].

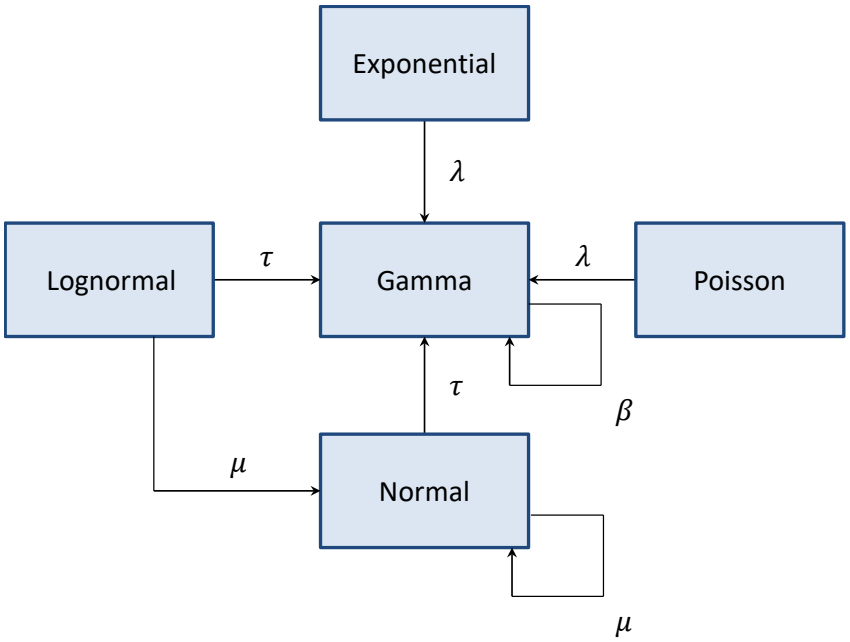


Figure 7: Natural conjugacies regarding the gamma and normal priors (arrows start from boxes representing the likelihood function and point to the box showing the conjugate prior). The parameter on each arrow shows the parameter for which the prior is used to quantify uncertainty [23].

The way in which the flowcharts above are read is in the following way: the arrows leave from the box which represents the PDF of the likelihood function and they point towards a box which represents the PDF of both the prior and the posterior. In some cases, an arrow can loop which means that both the prior and the likelihood for the particular distribution help forming a conjugate relationship (as it is the case with the normal and gamma distributions). The different Greek letters on each arrow represent the parameters from each likelihood distribution which have their own PDFs given by the boxes that the arrows point to.

In most real situations however, there is no natural conjugate relationship between the Bayesian parameters as it was described above. This means that in order to calculate the posterior distribution, it is needed to calculate the denominator from Equation (1) which integrates over all possible values for the parameter $\boldsymbol{\theta}$. Even when dealing with conjugate priors, solving the integral can prove to be time consuming, although it can be done analytically. For more complex examples however this stops being the case, meaning that other ways of computing the posterior distribution need to be found. For instance, Smith et al [24] presented a method of computing the integral by approximating it with a sum of Gauss-Hermite quadrature terms. This numerical method is suitable for problems which have up to six dimensions, and as Naylor and Smith [25] pointed out, it only works for a class of problems which require the product between prior and likelihood to have the form of "polynomial function multiplied by a normal distribution". A more general method is preferred in this case; preferably one which does not impose restrictions on the shape of either the prior or the likelihood and the method itself should be applicable to an arbitrary number of dimensions. Several such methods were identified in the literature and they are as follows:

- The CDF inverse method involves knowing the analytical expression of the CDF and then calculating its inverse. A number u is chosen randomly between 0 and 1, meaning that the drawn sample from the distribution of interest will be equal to $\text{CDF}^{-1}(u)$. As straightforward as this method seems, it becomes virtually impossible to use as the complexity of the problem starts increasing because of the difficulty associated to calculating the inverse.

- The rejection sampling method involves finding a PDF q and a constant c so that the PDF given by cq should be greater than the target PDF π at all points. Subsequently, a sample is drawn from the distribution q which is then accepted with probability π/cq. Several problems which are associated to this method have to do with the fact that in higher dimensions, the potential samples which are generated will have low values of likelihood associated to them, therefore the acceptance probabilities are expected to be low as well [26]. One way to go around this issue would be to try to maximize the likelihood which more often than not requires using sophisticated tools. This usually defeats the purpose of using rejection sampling as a computationally cheap method, which is why other sampling methods based on iterative Monte Carlo might be preferred.
- The weighted bootstrap method involves drawing a number of samples n $(\theta_1, \theta_2, \dots, \theta_n)$ from the proposal distribution q, and for each of them a weight is calculated. The value of weight i is given by the following expression [26]:

$$w_i = \frac{\frac{\pi(\theta_i)}{q(\theta_i)}}{\sum_{i=1}^{n} \frac{\pi(\theta_i)}{q(\theta_i)}} \tag{5}$$

  Subsequently, a resampling procedure is set up, this time however the samples are drawn from the discrete distribution $\{\theta_1, \theta_2, \dots, \theta_n\}$ with corresponding probabilities $\{w_1, w_2, \dots, w_n\}$. A random sample which is drawn from this last distribution is assumed to be an approximate sample from the target distribution π. As it was the case with the rejection sampling algorithm, problems which involve a high number of dimensions become difficult to solve using the weighted bootstrap method as posterior regions become harder to cover.
- The Gibbs sampler is different than all the other methods presented above with regards to the proposal distribution; the Gibbs sampling algorithm does not use one. Instead, the samples are being drawn from the conditional target distribution with respect to each variable (provided that the conditional distributions are easy to sample from). Because a kernel function q does not need to be found, there is therefore no error introduced when drawing the respective samples. In other words, the acceptance probability corresponding to those samples is always going to

be equal to 1. Given that a sample θ^i at step i exists, the sample at step $i+1$ can be obtained in the following way:

$$\theta_1^{i+1} \sim \pi\left(\theta_1^i | \theta_2^i, \theta_3^i, \dots, \theta_n^i\right)$$
$$\theta_2^{i+1} \sim \pi\left(\theta_2^i | \theta_1^{i+1}, \theta_3^i, \dots, \theta_n^i\right)$$
$$\theta_3^{i+1} \sim \pi\left(\theta_3^i | \theta_1^{i+1}, \theta_2^{i+1}, \dots, \theta_n^i\right)$$
$$\vdots \qquad \vdots \qquad \vdots$$
$$\theta_n^{i+1} \sim \pi\left(\theta_n^i | \theta_1^{i+1}, \theta_2^{i+1}, \dots, \theta_{n-1}^{i+1}\right)$$

One important feature to notice is that each conditional distribution makes use of the previously updated sample element at step $i+1$. The main shortcoming of this method is the fact that the conditional distributions have to be accessible and easy to sample from. For complex problems it might happen that only a few components will have their conditional distributions available [26]. One way to solve this problem, as Muller suggested [27], would be by making use of a Metropolis-Hastings sub-chain within the Gibbs algorithm that will only draw samples corresponding to the components for which the conditionals are not known.

- The Metropolis-Hastings algorithm is predominantly used in the literature in the cases when the distribution the distributions which need to be sampled from are complex and also not completely known (generally, the denominator from Equation (1) is unknown) [26]. As opposed to the Gibbs sampler, it can be used in instances when conditional distributions are not known [26]. One drawback of this method is that its success depends on the acceptance probability. For example, if the step size is too low, then although the acceptance probability will be close to 1, the search process will be very slow, meaning that a large number of iterations are needed. Conversely, if the step size is large, the samples will be most likely taken from the tails of the posterior distribution, resulting in a very low acceptance ratio.

The following section is dedicated to a brief introduction to Bayesian networks as they had been identified as a formal tool with a great amount of potential regarding the visualization of the Bayesian processes.

2.3 Expert Systems and Bayesian Networks

This section briefly introduces the idea of expert systems and Bayesian networks (BNs) by first reviewing the relevant literature. The reason why Bayesian networks are investigated is twofold: they represent a formal way of implementing Bayesian inference to a problem where making predictions is a priority, and also their graphical format can help visualise the way they propagate information throughout the model. Before going into further detail, the term "expert" which lies at the foundation of this chapter will be analysed.

Knowledge is usually linked to the idea of intelligence, which is generally not as easy to define as it is to identify [28]. Usually the intelligence of a person is quantified based not only on the knowledge they possess, but also what they can apply that knowledge to. This knowledge is then associated to the term *"expert"*; according to the *Concise Oxford English Dictionary*, it is defined as "a person having special skill or knowledge". In simple terms, people might use the word *"expert"* in order to associate it to someone who they turn to whenever they encounter a problem which is either too difficult for them or outside their own area of specialization. In the same context, an expert system can be considered to be a tool whose aim is to condense the knowledge of one or multiple experts so as to make it accessible for people who don't have that specialized knowledge. A more formal definition of an expert system has been given by Welbank [29] in the following way:

"An expert system is a program which has a wide base of knowledge in a restricted domain, and uses complex inferential reasoning to perform tasks which a human expert could do."

Cowell [30] summarizes the above phrase by using the following qualitative expression:

Expert System = Knowledge Base + Inference Engine

Cooper [31] investigated the current research directions regarding the development of Bayesian network based expert systems. The domains in which

expert systems are applied are often complex and characterised by uncertain knowledge and data. In the beginning, probability approaches have been avoided whenever uncertainty had to be represented because of concerns that those kinds of approaches are usually not mathematically rigorous. Recently however, probability has started being revisited in the context of artificial intelligence [32]. Belief networks have been identified in recent research as a pragmatic method of acquiring knowledge. More details about them will be given in this section.

Sakellaropoulos and Nikiforidis [33] used a Bayesian belief network in order to create a decision support system for the prognosis of head-injured patients at an intensive care unit (ICU). They compared its performance to other systems which include: a simpler belief network which assumes conditional independence among the results, as well as a human expert. The results tend to show that performance is not only better than the performance of the independence model, but is also comparable to that of the expert neurosurgeon [33], which by itself stands to emphasize the power of using this approach. The authors also stated that although there are other techniques such as 3-nearest neighbours which are suitable for their problem, Bayesian approaches have been implemented for other reasons. Namely, Bayesian networks have rigorous mathematical descriptions, while at the same time they mirror the knowledge structures in the human mind [34]. This has the effect of facilitating the interpretation of knowledge and optimising decision making.

Guo [35] implemented the Bayesian belief network framework to a safety assessment expert system. The first step taken in this work was to identify the problems happening during the application of safety standards. It is the case that those problems are related to the uncertainty present in the safety assessment, which makes the Bayesian networks a suitable tool due to their probabilistic nature. Consequently, due to the solid mathematical foundation behind them, they can allow for a quantitative evaluation of safety. In addition, they can also be viewed as expert systems which allow people to make decisions.

Pitchforth and Mengersen [36] referred to Bayesian networks in terms of robust methods of validating them in the case when there are no objective sets of data available. Their aim was to demonstrate that even when there is no data at all, validation is still possible. The traditional validation methods for expert elicited networks involve either comparing the model predictions with available data, or by directly asking the experts who provided the data in the first place. However, unless

objective data is available, Bayesian networks cannot be properly validated using test data. As a result, it cannot be said whether a test has been passed in a conclusive fashion. Consequently, they claim that for Bayesian networks based either solely on expert elicitation or on a combination of expert elicitation and data, a comprehensive framework is needed for the validation process. The framework proposed, although not comprehensive, is simply formed from a number of steps which can be followed in order to allow the designer to gain confidence in the validity of the model. Their main proposal is that when actual data cannot be used for the validation process, the posterior predictive distribution (the prediction) should be shown to an expert who will be able to tell whether he thinks the predictions match their expectations.

Jongsawat and Premchaiswadi [37] put forward a methodology which focuses on obtaining the structure of a Bayesian network from experts. More specifically, their method is based on transforming qualitative expert knowledge into rigorous prior distributions which can be used to build a Bayesian probabilistic model. It is proposed by the authors that this method is useful for group decision making. To be more specific, the qualitative knowledge extracted from a group can be transformed into probabilities used further in the Bayesian network.

Karandikar et al. [38] used a Bayesian approach in order to be able to predict the life for a cutting tool. The empirical model which is used to obtain life contains several calibration parameters for which prior distributions were given. As a first iteration, only knowledge regarding the range of the parameters was assumed (which resulted in uniform prior distributions being used). The likelihood function only used three experimental values and was assumed to have a normal distribution. The Bayesian updating was done using the component-wise Metropolis-Hastings method, the authors mentioning that the alternative updating procedure (block-wise updating) would be more computationally expensive. Comparison with a deterministic model lead to the conclusion that using an informed prior would result in a more accurate prediction of the final result, without the need to use a large amount of experimental data. As a result, normal distributions were used for the priors, and also different values for the standard deviation were used for the likelihood. The case study stands to illustrate the power of the Metropolis-Hastings algorithm as a tool used to sample from multivariate distributions, and also that the use of optimal priors and Bayesian inference is a robust method which can be used to quantify uncertainties.

Karandikar et al. [39] also used Bayesian inference in order to predict the remaining useful life for structures that had been damaged due to fatigue. They used the Paris' law, for which the calibration parameters were assigned prior distributions that were subsequently updated using experimental data. Although the fundamental theory is the same as for the previous case study, the actual algorithm employed this time is the random walk approach. A future direction for Karandikar's research would be to combine it with decision analysis in order to assign a monetary value to the information gained from experiments prior to performing them. In this fashion, it is possible to determine the optimum number of performance cycles before the structure has to be placed under maintenance.

Zhu et al. [40] applied MCMC methods in order to predict the life of turbine disks due to low cycle fatigue. Besides predicting life, the Bayesian framework developed by Zhu et al. also has the aim of quantifying uncertainty of material properties. The results of the Bayesian framework were afterwards compared against experimental data as well as four different fatigue models, and they showed that the probabilistic predictions are in agreement with the experimental data. According to the author, this makes the framework a preferred alternative to deterministic methods for low cycle fatigue life prediction.

Yeratapally [41] used an MCMC approach in order to study the fatigue life as well as the crack propagation based on microstructure, and the MATLAB code used in this respect (which is currently owned by Rolls-Royce plc) was one of the starting points of this research. This is especially true when looking into the second case study shown in Chapter 4.

The knowledge base is considered to represent the core of any expert system, in the sense that regardless of how advanced the inference engine itself is, as long as the level of knowledge is poor, the consequent inferences will not be very useful.

Expert systems are characterized by their ability to deal with uncertainty which is also the main scope of this thesis. The way in which it does so is by making use of the Bayes' theorem explained in the previous section and by representing the probabilistic expert system onto a graph; the resulting structure has the potential to become a Bayesian network. Before discussing Bayesian networks however, a

brief overview will be given regarding similar graphical methods of taking decisions.

- *Diagnostic decision trees*. They form a structured sequence of questions where each response determines the next set of questions which should be asked [30]. This particular system is similar to a flowchart in that after answering a question, the appropriate path will have to be taken according to the answers given to the questions which are contained by the nodes of the tree. Franklin et al. [42] have taken into consideration a diagnostic decision tree whose purpose was to diagnose heart problems in newborns. The particular algorithm has been applied to around 400 cases. Although this type of decision tree is easy to implement and explain, they have several shortcomings. Because of their form, they are completely reliant upon the knowledge base in the sense that a single erroneous choice can lead to a completely wrong conclusion. In addition to that, their usage becomes extremely limited if there is any information missing.
- *Production systems.* Those are generally more flexible than the diagnostic decision trees and their goal is to perform reasoning by using logical rules. An example to illustrate this idea is shown with an example from Winston [43]:

    - IF the animal has hair THEN it is a mammal.
    - IF the animal gives milk THEN it is a mammal.
    - IF the animal has feathers THEN it is a bird.
    - IF the animal flies and it lays eggs THEN it is a bird.

    Production systems do rely on rules, however they allow for more flexibility unlike diagnostic decision trees. The main reason is that they can make deductions from a given set of assertions (known as forward chaining) as well as determine whether assertions exist given a specific conjecture (backward chaining). On the other hand, they don't allow for multiple choice questions as much as they focus on the assertions themselves. The consequence of that is that production systems can't handle negations very well [43]. In particular, they cannot differentiate between formulations such

as “found to be false” and “not found to be true”, meaning that extra conditioning would need to be applied.

Although the above examples are tightly related to logical deductions, most of the time the available data can be scarce, meaning that a method of coping with uncertainty needs to be devised. It has been the case, especially in the Artificial Intelligence (AI) community, that probability theory is not a suitable tool in order to quantify uncertainty having to do with expert systems. As a result, alternative measures of measuring uncertainty have been developed such as fuzzy logic [44] as well as belief functions [45]. In the next paragraph the reasoning will be given regarding why there is a good reason for using probability theory for studying expert systems and also why the initial beliefs about its unsuitability might not be necessarily true. The reasons why probability was considered useless, according to Cowell [30] were as follows:

- The kind of uncertainty present in an expert system is different from the very idea of probability which is strongly related to observable events.
- A joint probability distribution of many quantities is impractical.
- Those two ideas stem from the fact that the interpretation of probability has always been a hot debate topic. There have been two main approaches regarding the understanding of probability: the objective and the epistemological approach. One of the most relevant objective interpretations has been the frequentist one regarding to which probabilities have been viewed as quantities characteristic of any event, the main belief being that their true values become observable as the number of experiments able to cause the said event becomes very large (if n is the number of experiments, then frequency of a particular event to occur becomes close to its innate probability p as n tends to infinity). However, this approach implies that the meaning of probability is strictly related to events which are repeatable; their corresponding probability can be found if enough observations are being made. Unfortunately, this view severely limits the usage of probability theory in various contexts such as expert systems, and it was this particular view which caused people to adopt non-probabilistic approaches and discard probability as a potential method which could be used in this respect [30].

The second view on probability refrains from describing it as a deterministic unknown value characteristic of the events being studied, and instead, it uses the concept of conditional probability in order to potentially refer to a state of "mental uncertainty" [30] which can change as more information is being provided. This view is consistent with the Bayesian formulation regarding uncertainty quantification which has been studied in the previous section. The Bayesian approach is also used for the intents and purposes of the research conducted for this project. Seeing that Bayesian theory has been proposed as a consistent framework for different probabilistic applications, it might not be unreasonable to wonder why it is not used more often. According to Lindley [21], the reason for this lies first of all in the fact that only recently has there been a trend of this subject being taught thoroughly at universities throughout the world, which can also be indicative of people starting to realize its strength. Coupled with this fact, Bayesian theory might not be an accessible tool especially for someone who had been trained to make use of the frequentist approach. There is even an arguably blunt remark that Lindley makes by saying that "Every statistician would be a Bayesian if he took the trouble to read the literature thoroughly and was honest enough to admit that he might have been wrong." [21].

The Bayesian belief network is the third way of quantifying uncertainties related to expert systems and in broad terms, they represent a probability distribution in the shape of a Directed Acyclic Graph (DAG) which has the following mathematical formulation [46]:

$$f(x_1, x_2, \dots, x_n) = \prod_{i=1}^{N} f\big(x_i|\mathrm{pa}(x_i)\big) \tag{6}$$

The notation $\mathrm{pa}(x_i)$ stands for the "parent" node of random variable x_i. Also, $f(x_i)$ is the probability distribution of random variable x_i which can be either discrete or continuous depending on the context. An example which can illustrate those ideas and also expand on them is shown below in Figure 8:

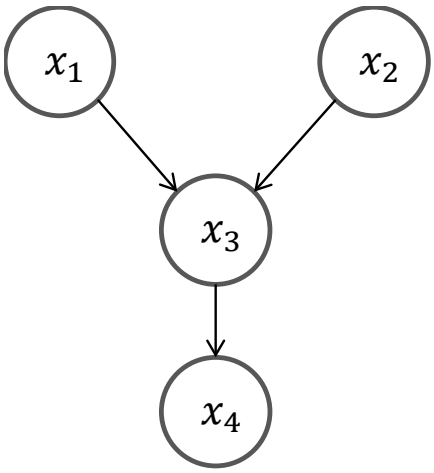


Figure 8: Example of a Bayesian Network consisting of four variables.

The network shown in Figure 8 is considered to contain two independent nodes given by x_1 and x_2. As a result, they do not have any parent nodes associated with them (there are no arrows pointed to either of them). They are however the parent nodes of node x_3 and although x_4 also depends on x_1 and x_2 (through x_3), the only parent node of x_4 is going to be x_3. As the dependence of x_4 on x_1 and x_2 is manifested through x_3 it is therefore considered that x_4 is independent of either of them. Equation (6) (which is a particularization of the law of total probability [22]) can be rewritten as:

$$f(x_1, x_2, x_3, x_4) = f(x_4|x_3, x_2, x_1) f(x_3, x_2, x_1) \tag{7}$$

By applying the same rule to the second factor on the right hand side, Equation (7) becomes:

$$f(x_1, x_2, x_3, x_4) = f(x_4|x_3, x_2, x_1) f(x_3|x_2, x_1) f(x_2|x_1) f(x_1) \tag{8}$$

The independence relations between the variables x_1 and x_2 will result in Equation (8) to become:

$$f(x_1, x_2, x_3, x_4) = f(x_4|x_3)f(x_3|x_2, x_1)f(x_2)f(x_1) \quad (9)$$

There are two particular types of variables which can represent the nodes of a Bayesian network; discrete or continuous. There are a great number of simple discrete Bayesian networks which can be found in the literature (such as the "wet grass" example [46] or the "burglar model" [47]), however in most cases, continuous variables are used in order to quantify uncertainty surrounding various processes. A very simple example of a continuous Bayesian network is given below in Figure 9.

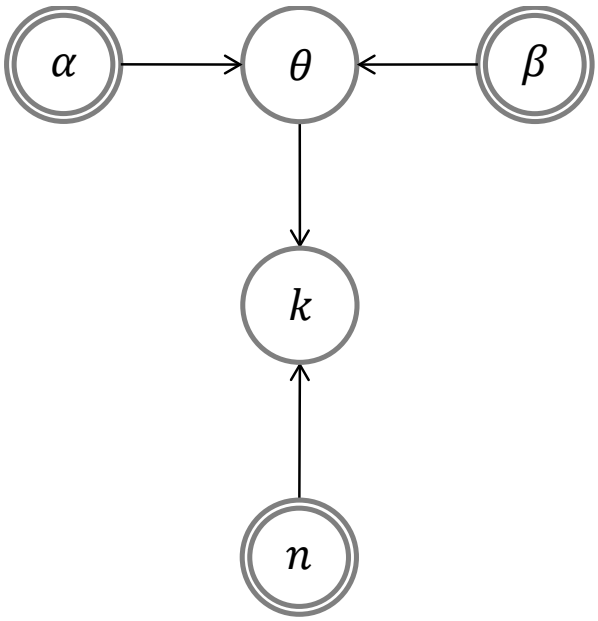


Figure 9: Bayesian network showing two continuous variables and a constant.

The distributions which are used for the different parameters above are as follows:

$$\pi(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} \quad (10)$$

$$f(k|\theta) = C_n^k \theta^k (1-\theta)^{n-k} \tag{11}$$

The parameter θ has a prior beta distribution with parameters α and β. The notation $\Gamma(x)$ stands for the gamma function calculated in point x. The first term from the formula of the binomial distribution of k given θ stands for "n choose k" and it is expressed in the usual way $C_n^k = \frac{n!}{(n-k)!k!}$. Also, in Figure 9, the nodes for n, α and β are represented by double circles, which means that those parameters are deterministic. Even more, in this particular case, they are assumed to be constant for explanatory purposes. As a result, there are several operations which can be done on a Bayesian Network, and consequently, the one in Figure 9 will be used as an example to illustrate them.

First of all, a method needs to be used in order to calculate the prior predictive distribution. Mathematically, this refers to the denominator of Equation (1). An important reason why doing this is a good idea is in order to see how the model "thinks" the observed data should look like before actually obtaining any real data. In addition, by knowing the PDF of the prior predictive, synthetic data can be generated (by taking samples from it) in order to perform the posterior inference (which will be described shortly) to test the inference procedures used in the network. If subsequently it is found that the inference performs properly, then it can be expected that the inference should also work for any actual data. As it was explained above however, analytically computing this integral is no easy task and unless the problem being studied has a simple form (for instance using conjugate priors), then it can become virtually impossible. Using Bayesian networks however, allows for the integral to be computed numerically using what is known as *ancestral sampling*. The method is easy to understand and implement and it is only based on Monte Carlo sampling (instead of MCMC as it is the case when sampling from the posterior). The first step consists of drawing samples from all the independent nodes (nodes that have no parents). After that, samples are drawn from the distributions conditional on the already sampled parent nodes, and this top-to-bottom process is continued until samples have been drawn from all relevant nodes. In the case shown in Figure 9, the first sample is drawn from the prior distribution of θ, and given the value obtained, a second sample is drawn from the distribution $k|\theta$. This second sample is considered to come from the prior predictive distribution of k (also known as $f(k)$). n is considered to be a constant

so it will be kept at its original value all of the time. The whole process is repeated for a pre-set number of iterations (the exact number depends on the problem) until the designer is confident that enough samples have been drawn in order to obtain an accurate representation of the prior predictive.

For the example given in Figure 9, the prior predictive distribution has been computed in two different ways: first, a Monte Carlo analysis was used to find 10000 samples and plot those on a histogram, and for the second one, the integral giving $f(k)$ has been computed analytically. The values for α and β have been randomly chosen as 5 and 2, while the total number of trials n has been set to 10. The evaluation of the integral gave an expression for $f(k)$ as follows:

$$f(k) = C_n^k \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+k)\Gamma(\beta+n-k)}{\Gamma(\alpha+\beta+n)} \tag{12}$$

The resulting histogram found from the algorithm described above as well as the plot of $f(k)$ for integer values of k between 0 and 10 can be seen below in Figure 10:

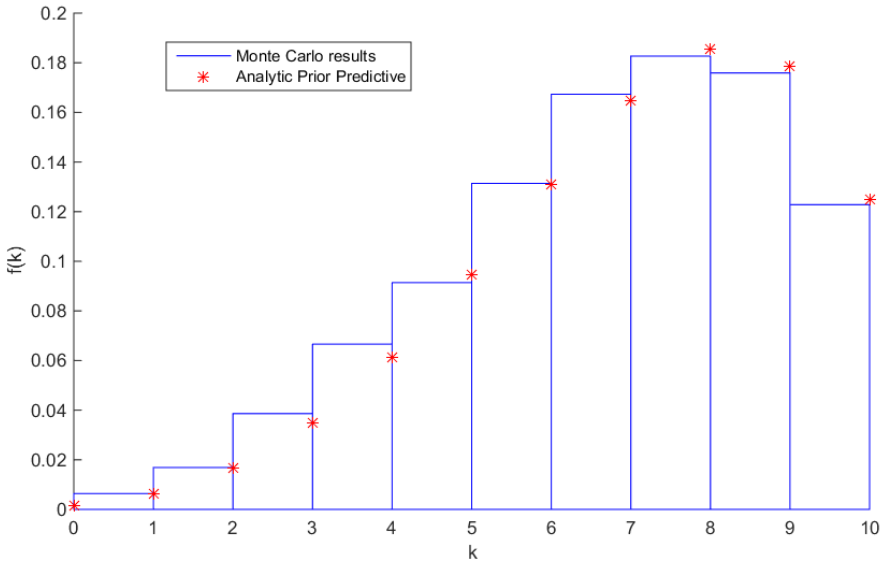


Figure 10: Comparison between numerical and analytic calculations of the prior predictive distribution

It can be seen that when 10,000 samples are taken, the result from the numerical method follows the analytical solution. This concept has been illustrated to a very simple example, however it could be applied to more complex cases as well.

The other operation which might have to be done on a Bayesian network is related to finding the posterior distribution given that some data has been observed. The way to solve this problem for continuous variables is to make use of an MCMC algorithm. The only difference however, is that this time the samples drawn will always be conditioned upon the observed values. The same example shown in Figure 8 is going to be used to illustrate this concept. After the numerical algorithm is applied, the result will be compared to the analytical one. In this instance, it is easy to compute the posterior because a quick look at Figure 6 shows that the beta prior is conjugate if a binomial likelihood is used. As a result, the posterior will have its distribution given by:

$$\theta|k \sim \text{Beta}(\alpha + k, \beta + n - k) \tag{13}$$

It will be assumed that the observed value is $k = 3$. The Metropolis-Hastings algorithm has been used in order to draw samples from the posterior which has been conditioned on the observed value of k. The results have been plotted on the same graph in Figure 11 below and it is clear that in this particular case the MCMC method has been accurate in finding the shape of the posterior. The main reason for that of course is that the case under consideration has been relatively simple.

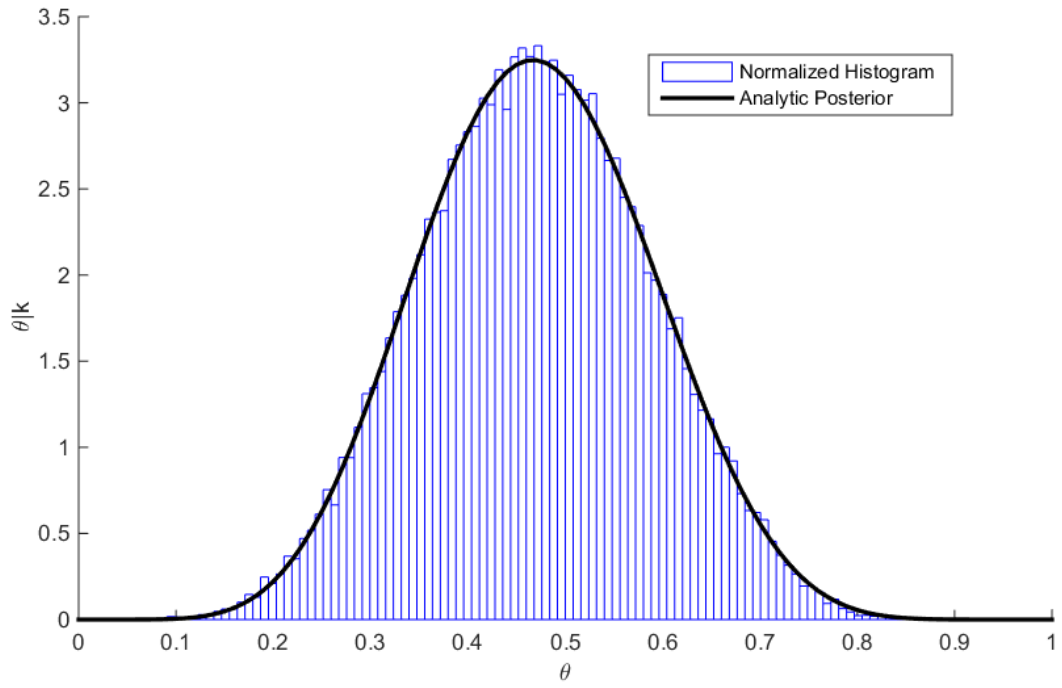


Figure 11: Analytic posterior and histogram for an observed value of k = 3

The examples shown above have the purpose of demonstrating one way in which Bayesian networks could be implemented within an expert system. For instance, some of the nodes in Figure 9 could be based on expert judgements regarding the distribution of various parameters such as θ or even α and β if they were not considered fixed or deterministic. The same kind of philosophy could be applied in the same manner to more complex problems as it will be demonstrated on the case studies presented later on. Before that however, it would be reasonable to ask how the expert knowledge, which lies at the core of any expert system, is going to be elicited. It therefore comes as no surprise that this topic itself needs to be analysed. As a result, the following section will look into various considerations which have to be taken when trying to elicit expert knowledge.

2.4 Elicitation of Expert Judgements

There is a great body of literature which is concerned with uncertainty quantification via elicitation of expert judgements. Several authors such as O'Hagan, Kadane, and Winkler published a number of papers which explore this subject in great detail and will be discussed in this section. The main points which have to be addressed have to do with the structure of the elicitation process as well as with the methods needed to be used in order to draw the relevant judgements from the experts. In addition to that, an elicitation has already been performed within Rolls-Royce plc for the purposes of this project. For more details about that, the reader is referred to section 4.3, which outlines the details of the third case study.

Before going further, it is important to consider the meaning of the term "elicitation". Generally, it refers to information needed to be drawn in order to fulfil a certain purpose. More often than not, the information being elicited has one out of two forms:

- Exact knowledge possessed by the experts can be simply drawn out by asking a simple question. In this case, if multiple experts have the same particular piece of knowledge, they should all give the same value [48].
- Judgements regarding an uncertain quantity involve a more complex process, since this time, the elicitor is not asking for specific pieces of knowledge; the aim is to represent the expert's uncertainty about some parameters of interest in the form of PDFs. The elicitor is able to draw this out by asking for some estimates. More details about this will be given in this section.

At the end of section 2.3 it was stated that in a Bayesian context, expert judgements are the ones giving the prior for the expert systems taken into consideration. Although there are several reviews and books on prior elicitation such as the ones by Kadane and Wolfson [49] as well as Meyer and Booker [50], according to O'Hagan [51] the topic deserves more attention. O'Hagan states that *"elicitation is a key component of any serious Bayesian analysis in which the data are not so numerous as to swamp whatever prior information there might be"*. Because of its importance and complexity, a rigorous framework needs to be set up within which the needed

judgements have to be drawn out with as little bias as possible. Before talking about the potential biases which can occur, it would be useful to give a short outline of how a generic elicitation process looks like. Some steps towards outlining this have been identified by O'Hagan and are shown in this section. In order to gain a deeper understanding into the whole elicitation process, the reader is referred to O'Hagan [48]. A brief summary of the process can be seen below:

1) *The objectives*. Obviously, the prime objective of any elicitation process is to gather enough information in order to solve a problem. This information mainly refers to quantities which need to be elicited to solve the problem, and whose values are uncertain. It is also important to clearly define the quantities of interest to the expert so the chances of any ambiguity are reduced as much as possible.

2) *The format*. There is flexibility in the manner in which the elicitation is conducted; it can take the form of a questionnaire which can be sent by mail/email, uploaded on a website, or it can take a more personal form either by telephone or face-to-face meeting.

3) *The experts*. The number of experts taking part in the elicitation process can vary; it usually depends on availability, resources as well as the quantities of interest which are elicited. For instance, in the case of the European Food Safety Agency (EFSA), the experts whose knowledge is elicited are either external experts or members of their relevant working group [48].

4) *The elicitor*. In the case when the format is either a telephone elicitation or a face-to-face meeting, an obvious participant is the elicitor who has the purpose of controlling the flow of the elicitation. If the elicitation has the form of a questionnaire, the elicitor is still able to control the direction of the elicitation process, however in this case, this is only limited to designing the questionnaire.

5) *Other participants*. If the elicitation has the form of a meeting, there can be other participants as well. Usually, there can be field experts that have general knowledge, and their role can be to offer background data as well as clarify the objectives to the experts.

6) *Preparation and training*. The experts need to be aware of the probabilistic judgements which they need to provide, meaning that unless already familiar with the statistics aspects, enough preparatory material has to be provided prior to the elicitation.

7) *The sequence of questions*. As mentioned in the section above, the ultimate goal of any elicitation process is to represent the experts' uncertainty regarding different parameters in the form of PDFs. Although the types of questions asked are definitely of key importance, the exact sequence and wording of questions must also be taken into account.

8) *Fitting*. The judgements are used in order to fit a PDF (which can refer to one variable or several, in which case the PDF will be a joint distribution). Although relatively straightforward, an issue arises when there are multiple experts giving their judgements.

9) *Documentation and reporting*. After the fitting step, the obtained PDFs are reported back to the experts who can then provide feedback which depends on whether the uncertainty regarding the parameters of interest. The elicitor then uses this new information to update the distributions which are once again shown to the experts. Several such iterations can take place until both parties are satisfied with the result.

A simplified diagram illustrating the main outline of an elicitation process is shown below in Figure 12.

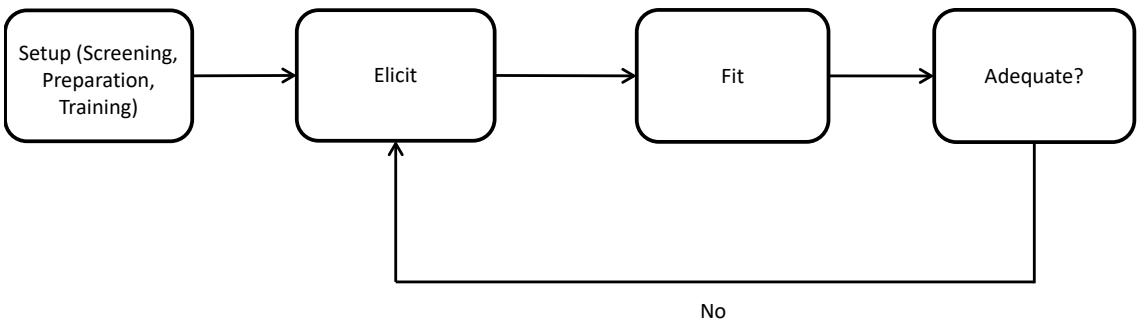


Figure 12: Simplified chart of the elicitation procedure [52]

It becomes clear that elicitation itself is an iterative process; the elicitor asks the expert about estimates which are used to build PDFs, and then the PDFs are presented to the experts who can then choose to update their estimates if necessary. This process continues until both parties are satisfied with the results (illustrated by the feedback loop at the bottom of Figure 12).

2.4.1 Elicitation frameworks

Landquist et al. [53] used the SHELF method [54] in order to perform an expert elicitation workshop in order to quantify uncertainties regarding risk assessment of shipwrecks. The input coming from experts is mainly in the form of probabilities of an opening in a shipwreck occurring due to a variety of causes. The motivation for using experts for this case study was that information necessary for a shipwreck risk assessment may be incomplete or simply non-existent. The SHELF quartile method has been implemented in the elicitation workshop in order to obtain Bayesian priors. Available data is used to update them, giving posterior distributions, which allow the Bayesian network to estimate the annual discharge probability from shipwrecks and consequently provide decision support.

An elicitation framework different from SHELF has been used by Johnson et al. [55] in order to look at methods of controlling invasive species (in particular the tegu lizards from southern Florida) by using expert judgements. The type of elicitation workshop used was based on the 3-point process [56] which involved eliciting a median and a 95% confidence interval from each expert in a group. The judgements were drawn separately and were subsequently shown to everyone who were afterwards allowed to revise their opinions subjected to group discussions. However, consensus was not enforced because, similarly to what O'Hagan [48] suggests, the final result would not be based on genuine agreement, but on group pressure instead. This goes against the main interest of performing the elicitation which was to assess the full range of the experts' opinions.

One important consideration to take is that due to the personal nature of the procedure, there is a high probability of biases appearing for instance if the expert has some kind of personal interest in the outcome of the elicitation. Those can impact the outcome in a negative way which makes it paramount for the elicitor to be aware of them and know how to account for them. Before moving on to the next section and talking about biases, it would be worth defining what an expert is. The

most popular idea that people have about this concept is that an expert is a person who has special knowledge regarding the problem for which the elicitation is being performed. However, this is not always the case. For instance, some surveys were concerned with studying adolescent decision making around risky behaviours [52]. In this case, the elicitor might want to ask the teenagers themselves how they perceive risk. It becomes clear that in this situation the lack of expertise is the whole point of the elicitation process. Consequently, the title of "expert" is proposed to refer just to a person whose knowledge is being elicited and nothing beyond that. In the section below, several results of psychological research which have to be taken into account when designing a method of quantifying expert knowledge are presented.

2.4.2 Psychological Considerations

It is agreed that any elicitation method in general has the purpose of connecting the expert's opinions to their representation in statistical form. As a natural consequence, understanding both the psychological and the statistical parts of the elicitation process is needed in order to develop an efficient elicitation method. This idea is reinforced by Hogarth [57] in the following statement:

"Assessment techniques should be designed both to be compatible with man's abilities and to counteract his deficiencies."

The amount of psychological research that should go into elicitation is an enormous topic by itself, so only some key findings will be presented here. The way and which a person is able to assess the probability of an event or judge various statistical parameters. Those judgements are found to be based upon heuristics which generally can prove to be effective [52], however there is the risk that they can cause significant levels of bias. Some of the most common types of biases that can occur are listed below:

- *Judgement by representativeness* is a common heuristic which is relevant to questions of the form: "What is the probability that event A will generate event B?" [52]. In statistical terms, this asks about the conditional probability $P(B|A)$, and in general people tend to compare the main features of A with

those of B when giving an answer. What many people do on the other hand (which is where the bias stems from) is that they don't take into account the unconditional probability $P(B)$ when making the judgement [58] (i.e. little to no attention is given to the marginal probability of event B).

- *Judgement by availability* refers to the case when a person is asked about the frequency of an event occurring and they make judgements by using examples which come to mind. It is inevitable that some examples are recalled better than others, and although others might seem unlikely, they can prove to occur more often than the more obvious ones. For instance, someone might be asked whether a randomly chosen English word is more likely to have the letter "r" as its starting or as its third letter. It is much easier to recall words which begin with "r", meaning that most probably the expert will skew the probability in favour of words that start with "r", although the exact opposite is true [58].

- *Judgement by anchoring and adjustment* is one of the most predominant types of biases which can occur in elicitation processes especially when the expert is asked to estimate a value based on an initial starting point which is called an *anchor*. The bias usually comes from the fact that the person fails to adjust sufficiently in order to obtain a better result; this phenomenon is known as *anchoring* [59]. There was an experiment performed by Tversky and Kahneman [60] in which people were asked to estimate the percentage of African countries in the United Nations using different initial starting values. As it turned out, most people failed to adjust enough; subjects with higher starting values arrived at much greater final values than those who were given small starting values.

- Tversky and Kahneman refer to the *Law of small numbers* [60] as the bias which governs some people's belief that a small sample from a population should contain all important characteristics of the population as a whole. The simplest example possible is flipping a coin and writing down the outcomes. Most people expect the sequence HTHTHTH (H-heads, T-tails) to be more likely than HTTTHHH and even more likely than HHHHTHH, as the last two sequences don't seem to be an accurate representation of the fairness of the coin. Most of the time, chance is fallaciously viewed as a self-correcting process, the belief being that a deviation in one direction (more

instances of heads) should result in a deviation in the opposite direction. This idea is also known as the "gambler's fallacy". What actually happens in reality is that those deviations will simply be diluted in the long run, causing the probability of a head to get closer to 0.5.

- *Hindsight bias* is a certain type of bias which occurs whenever people are asked to estimate the prior probability of an event *after* its occurrence. Because the subjects already know what has occurred, they have a tendency to overestimate the prior probability, even if the occurrence of the event might have seemed unlikely before it actually took place. In the same manner, the prior probability can be underestimated for events that did not occur [52].

- *Conservatism* is a type of bias which occurs in the Bayesian context when the experts are made aware of the prior probabilities. This is also similar to the anchoring bias discussed above in the sense that the judged posterior values are close to the original prior. An experiment which is usually done in order to illustrate this idea is by drawing red and blue poker chips from a bag and asking about the posterior probability that the bag from which the chips were drawn contains mostly red chips. For example, there could be two bags containing 70% red and 30% blue chips, while another contains 30% red and 70% blue (those percentages are unknown to the expert). Samples are drawn from a single bag which is chosen via a coin flip (each bag has a 50% probability of being selected – this is the prior probability). Assuming that the sample contains 8 red and 4 blue chips, the expert is asked to assess the probability that the chips were drawn from the "red" bag. It has been observed that the posterior probability given by the experts is around 75% which is a lot less than the true probability found by using the Bayes' theorem which is approximately 96.7%.

- *Conjunction fallacy* is discussed in Tversky and Kahneman [61] and it is based on the idea that the joint probability $P(A,B)$ cannot be higher than either $P(A)$ and $P(B)$. The bias itself occurs when a high probability is assigned to an event which in turn is a subset of an event with a lower probability [49]. For example, the class of 7-letter English words ending in "-ing" is contained within the class of English words that have "n" as the 6th

letter, the former is usually judged to be more frequent simply because it is easier to think of words ending in "-ing". This is deemed to be a serious mistake which, by definition, violates the fundamental laws of probability [62].

- *Overconfidence* is a problem which can occur especially when assessing the tails of a distribution. More details about this however will be given in Chapter 4.

According to Kadane [49] , one way to take into account the above-mentioned types of biases is to ask the expert about observable quantities rather than about moments of the distribution. There have been some common views in the literature (for instance in Wilson [62], Kadane [63], Murphy and Winkler [64]) regarding the way in which an elicitation should be carried out. Some common points which are made are:

- Expert opinions are the most worthwhile to elicit.
- Experts should be asked to only assess observable quantities (such as quantiles) and not be asked about the moments of the distribution.
- Feedback should be given to the expert in a continuous fashion (see Figure 12 illustrating the iterative nature of the process).
- Experts should be asked to give their estimates both before and after seeing the data.
- The elicitor should avoid using leading questions (such as questions which expect a specific answer from the expert).
- Neutral language should be used instead of emotive expression whenever asking questions.
- Compound questions should not be asked, as it is better to ask questions regarding one feature at a time (avoid asking about both health and safety for instance).
- All the technical terms should be properly defined so that misinterpretations do not become an issue.
- When proposing an example, general usage of numbers should be avoided as it can cause anchoring (described above).

In addition to those points, it is considered good practice to ask the experts to express their answers in numerical terms whenever possible. For instance, EFSA uses a semi-quantitative method as the most popular elicitation procedure. This involves a list of verbal descriptions given alongside a list of numerical values; however conversion from verbal to numerical can be a source of hidden traps. On the other hand, using purely quantitative scales can also pose various problems, as most experts can be reluctant to give numbers in order to represent uncertainty due to a variety of reasons:

- Some experts can simply find it difficult to express uncertainties using numbers; probably they are unused to do so in their everyday lives. This is one occurrence when appropriate training can potentially solve the problem.
- Some experts are reluctant to give numbers so as not to imply an unjustified level of precision.
- Some experts are concerned that their estimates could be misinterpreted by decision-makers and the public [48].

As a compromise is needed, a semi-quantitative method is used in many cases. One of the best-known elicitation scales is that used by the Intergovernmental Panel on Climate Change (IPCC) who recommends using a scale consisting of seven verbal terms which get translated into probabilities as it is shown below in Table 1:

Table 1: Verbal probabilities description, as used by IPCC [48]

Verbal Description	**Probability Interpretations**
Virtually certain	>0.99
Very likely	>0.9
Likely	>0.66
About as likely as not	0.33-0.66
Unlikely	<0.33
Very unlikely	<0.1
Exceptionally unlikely	<0.01

This subsection had the purpose of introducing some of the most common ways in which people might judge numerical quantities during an elicitation process. It is considered that the elicitor should be aware of the psychological biases which the expert could have so that he can devise a method to quantify the error in the responses he receives. Also, Hogarth [57] provided a much more detailed list of such heuristics and biases in his papers.

2.4.3 Identifying Experts

Depending on the application for which an elicitation has to be conducted, one or more experts are needed. As a result, several considerations related to finding experts have to be employed, and for that, several questions have to be answered by the elicitors themselves such as: "How many experts is optimal?", "What is the kind of expertise that the experts should have?", "How to actually find the experts we need?", "How to persuade them to join the project?", "How to evaluate their expertise?" [48]. At this point it would be worth remembering that the kinds of judgement which are elicited involve numerical quantities which are used in order to construct the experts' uncertainty about a specific parameter in the form of a PDF.

O'Hagan [48] suggests several indicators which can point to whether a potential candidate has the relevant kind of expertise for the problem having to be solved, such as:

- Formal qualifications and training
- Job experience in terms of:
    - Amount (years of experience)
    - Quality of experience (industrial vs. academic); it is important to note that the environment where the expert gathered experience is also a factor in deciding the suitability for the elicitation process.
- Useful outputs (papers, reports, speeches)
- Awards received
- References from other experts: this feature can prove whether an expert has credibility within his or her field.

Alternately, there are several unreliable indicators which should not be followed when assessing an individual's expertise such as:

- Name of his/her job as well as the social role
- Projected confidence
- Being verbose
- Presence in the media

The best way of quantifying the expertise of a candidate up to date is asking for a resume as most of the good indicators of expertise are found there. However, the resume has to be complemented by at least two other documents. In order to gather further information about the actual job experience of the candidate as well as whether he/she is suited for the particular elicitation exercise, an expertise questionnaire should be filled in by the expert and then sent back to the elicitor. O'Hagan suggests a general template for such a questionnaire; a slightly more specific version was created for Rolls-Royce plc purposes which can be used to address the Fan Blade Off (FBO) related issues and it can be found in Appendix A at the end of the thesis. The second document consists of what is known as a Generalized Expertise Measure (GEM) [48] whose purpose is to identify if the experts possess the following characteristics:

- Specific training and education
- Qualifications required
- Ability to assess work-related issues
- Capability of self-improvement
- Intuition
- Confidence in the knowledge they possess

The last several characteristics described were previously also listed as "unreliable indicators of expertise", which is also the reason why the purpose of the GEM is to be completed not by the experts themselves, but instead either by the experts' peers or by the person who referenced them to the elicitor (who is also assumed to be an expert). There is also the issue that people may find it difficult to give impartial answers to slightly more personal questions. An example of the GEM form which could be used for screening experts can be found in Appendix B.

The purpose of the next section will develop on one of the ideas briefly mentioned which is related to the quantities that experts should be asked about. Before that however, this section will conclude with a summarized process of selecting experts which also covers the points above.

1. The elicitor should identify the desired characteristics of the experts who may be able to answer the questions related to the problem of interest.
2. Identify the experts who possess the characteristics from step 1.
3. Contact the experts either by email/letter/telephone and mention that:
    - "*Our company has an important problem and we have identified you as an expert*"
    - Also, offer a brief overview of the said problem
4. If there is a positive reply, the expert should be sent the expertise questionnaire to fill it in. If the expert would like to invite a peer, the GEM should also be completed by the expert.
5. If an expert is shortlisted, he should be sent an invitation letter which should also contain:
    - More detailed information regarding the problem as well as the motivation for the elicitation taking place
    - The type of activities performed during the elicitation as well as any kind of compensation
    - The date of the elicitation
    - The reasons why the particular expert was shortlisted
6. If the expert accepts the invitation, a second detailed information letter is sent, which as its name suggests, contains further detailed (and most likely confidential information) about the elicitation such as:
    - The exact location
    - A more detailed timetable for the activities occurring during the elicitation day
    - Details on the parameters being elicited
    - Clause of confidentiality
    - Ask for any additional information which the expert might need to share

2.4.4 Quantities to Elicit

Whenever an elicitation procedure is designed, more often than not the elicitor can choose which quantities to ask the expert about. Preferably, these quantities should be easy to elicit without intensive prior training. In the past, several experiments have been performed whose purpose was to examine people's ability to estimate basic statistical quantities [52]. Some of those experiments focused on the experts' capability of estimating probabilities/proportions. For instance, the experiments by Erlick [65], Nash [66], Pitz [67], Shuford [68] generally involved showing binary data to the subjects who were then asked to estimate a proportion. More specifically, Shuford [68] displayed on a screen several 20x20 matrices which contained blue and red squares, after which the subjects were asked to estimate the proportion of blue within each matrix. The conclusions were that the proportions given by the experts were assessed very accurately, the error being less than 0.05 in most cases.

Other experiments were conducted in order to assess people's capability of accurately estimating central tendencies within a sequence. Beach and Swenson [69] as well as Spencer [70] found that when a symmetric distribution is being used, the estimates given by the subjects contain a high degree of accuracy. This was not the case for asymmetric distribution however; Peterson and Miller [71] concluded that for a skewed distribution the subjects became confused about the significance of the mean, giving values which were more appropriate to those the mode and median instead. Several other experiments performed by Hofstatter [72] arrived at the conclusion that people are not familiar with the idea of variance either.

Several basic principles regarding the elicitation process itself which should dictate its flow have been suggested by O'Hagan [48]:

- It should not be assumed that the expert has experience in making probabilistic judgements even though he might have years of experience in their domain. As a result, appropriate training will have to be given.
- On the other hand, even if enough training is given, the expert can still find it difficult to understand the concepts of expected value and variance. Consequently, it is advised against eliciting the moments of the distribution.

- It needs to be emphasized that the goal of the elicitation process is not to obtain a point estimate, but a probability distribution which represents the expert's uncertainty about a certain variable.

Taking into consideration the aforementioned points, several intuitive ways of eliciting estimates had to be devised. Those methods are based upon the fixed and variable interval methodologies [48] which are described below.

The fixed interval approach is based upon first choosing a value of x (Figure 13) and then asking about the probability that the true value of x is smaller than the value given by the expert. This can be seen below in a visual format.

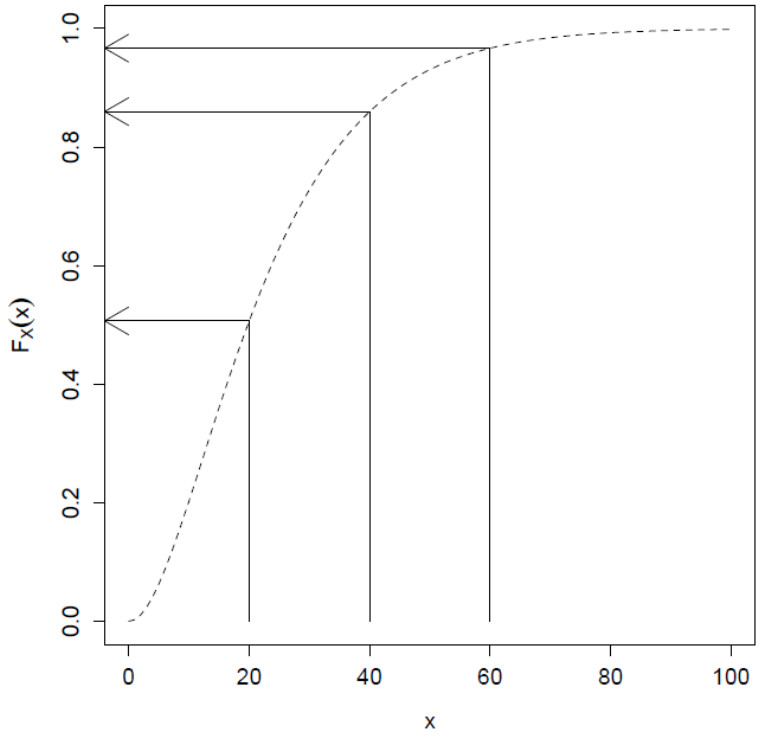


Figure 13: Graphical representation of the fixed interval method [48]

The variable interval method on the other hand, is considered to be an inverse operation in that the elicitor asks the expert for a value of x such that the true unknown x is less than the required value with a specified probability. In statistical terms this is known as "eliciting quantiles"; this concept which is illustrated in Figure 14 will be developed below.

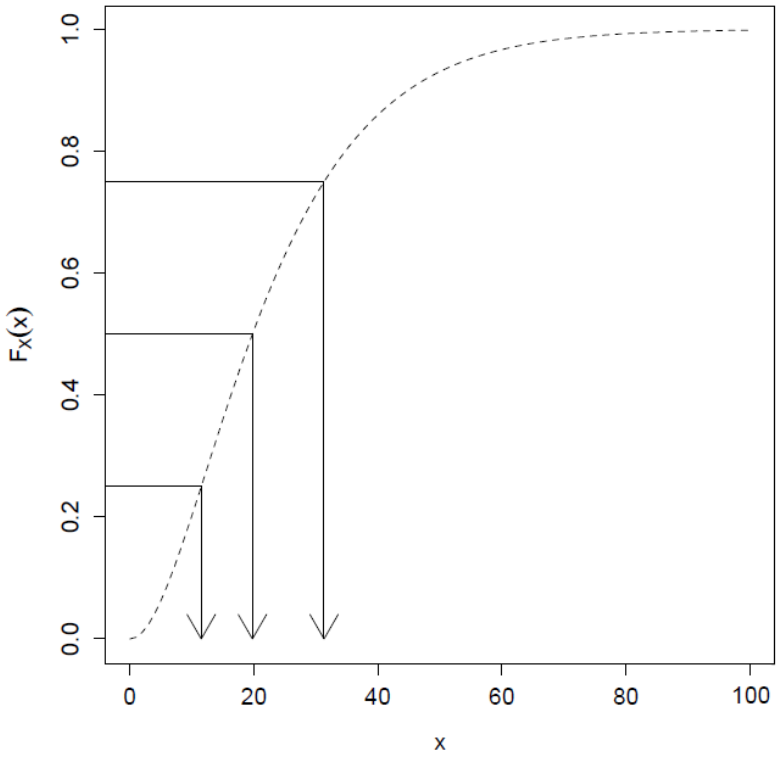


Figure 14: Graphical representation of the variable interval method [48].

The main suggestion from O'Hagan regarding the variable interval method is to first elicit the expert's range [48]. This has the impact of preventing the expert from anchoring to any value; if the expert gives the range at first, the tendency to anchor to both sides of the range will have the effect of negating anchoring in the first place. O'Hagan also suggests several methods of eliciting quantiles [48] as follows:

- Elicit the quartiles: the median (0.5 quantile), upper and lower quartiles (0.25 and 0.75 quantiles)
- Elicit the tertiles: the 0.33 and 0.66 quantiles
- Elicit the median, the 0.05 quantile and the 0.95 quantile
- Elicit the median, the 0.17 quantile and the 0.83 quantile

Two issues regarding the choice of quantiles refer to interpretability and overconfidence. Interpretability can be described in the following way. Assuming that the quartiles have been elicited, the range can be divided into four intervals of equal probability. For the sake of argument, it can be assumed that those intervals are: [0, 20], [20, 40], [40, 80], [80, 100]. The expert might be asked to place an imaginary bet on which of those intervals is the most likely to contain the real

unknown random variable X. Theoretically, the expert should have no preference as to which interval is most likely to contain X. However, if the expert believes that [0, 20] has a higher likelihood than [20, 40], it most likely means the expert believes that 20 and 40 are not the actual lower quartile and median.

Overconfidence is the other issue which has been investigated especially by O'Hagan *et al.* [48] who suggested that experts shouldn't be asked about extreme quantiles and instead resort to moderate quantiles such as tertiles or quartiles. An experiment was performed by Soll and Klayman [56] which reinforced this finding: they elicited an 80% interval in two different ways: directly and then by eliciting the 0.1 and 0.9 quantiles. The second approach was found to work better as it forced the experts to consider low and high values simultaneously which, as mentioned before, had the consequence of reducing the anchoring effect.

In the variable interval case, a bisection method can be applied in order to elicit the summaries regardless of the quantile method chosen. For instance, in the case of quartile, the following set of questions can be used to elicit for the median as well as the lower and upper quartiles [52]:

- "Can you determine a value such that X is equally likely to be less than or greater than this point?" - This is an intuitive way of asking about the median which the expert might not be fully familiar with.
- "Suppose you were told that X is below your assessed median. Can you now determine a new value such that it is equally likely that X is less than or greater than this value?" - Lower quartile.
- "Suppose you were told that X is above your assessed median. Can you now determine a new value such that it is equally likely that X is less than or greater than this value?" - Upper quartile.

The obvious advantage of this way of questioning is that the expert is asked about 50-50 intervals, and intuitively, it is much easier than being asked about intervals which have 25% chance of containing the real value of X.

The next logical step after the elicitor obtains the quantiles is to decide what type of distribution to use depending on the possible values of the unknown variable X. The most common ones used in plenty of applications are as follows:

- If X represents a proportion (such as a probability), it must lie between 0 and 1. In this case, a Beta distribution might be preferable in order to quantify uncertainty.
- If X has a lower limit, a gamma, Weibull or a log-normal distribution can be expected to fit the data well.
- If the PDF of X is expected to be symmetric about the median, then a Gaussian or a t-distribution can be used. If X has both upper and lower bounds, a Beta distribution could also be used.

Once the distribution is chosen and the expert's data is fitted to it, the next step consists of checking with the expert whether the distribution matches his uncertainty about the data. This is usually done in two ways [48]:

1. *Feedback method*. Summaries from the fitted distribution are given to the expert who is subsequently invited to give his opinion.
2. *Overfitting method*. The expert is asked to provide additional judgements in order to be compared to those from the distribution fitted by the elicitor.

This is the stage where the expert has the opportunity to modify his or her initial judgements. If this happens, the elicitor has to re-fit the distribution and to re-check with the expert once more. This iterative process continues until both parties are satisfied with the outcome (in the same way suggested by Figure 12).

2.4.5 Multivariate Elicitation

In reality, it is rare for a problem to only have one input or one variable; in general, tasks are much more complex as it is actually needed to elicit judgements regarding a joint probability distribution involving several variables. This ultimately translates in having to ask questions about other aspects of the problem.

In the simplest case, the variables of the problem can be independent, in which case PDFs about each individual variable can be elicited using the techniques described above, and the joint PDF is therefore obtained by multiplying all marginal PDFs. However, when variables are dependent upon each other, the complexity surrounding multivariate elicitation can no longer be by-passed. Although eliciting summaries for individual variables is still possible, it may no longer give the elicitor meaningful results. Therefore, a means of eliciting correlation has to be put forward. Research in this direction has been mainly focused on eliciting correlation between variables as they are drawn from a population [52]. For instance, Clemen et al. proposed six methods which can be used to elicit the correlation between height and weight of a population of MBA students [73]:

- Verbal description of the correlation strength using a 7-point scale ranging from “very strong negative relationship” to “very strong relationship”. In the original papers, the author made several assumptions which are used to convert those into numerical values.
- Direct estimation of the correlation by asking the expert to specify a value between -1 and 1. -1 implies a strong negative correlation, 0 stands for a lack of correlation (this suggests that independence of variables is a particular case of correlation) and 1 implies a strong positive correlation.
- Ask the expert to imagine that a person from the population has been selected at random; the expert is then given the percentile for the first variable and is then asked about the percentile for the second variable.
- Ask the expert to imagine that two people (A and B) have been selected. If A is greater than B for the first variable, ask about the probability that A is greater than B for the second variable as well.
- Ask the expert to imagine that a person has been picked at random. The expert is then asked to estimate the probability that the person is below a specific percentile for both variables.

- Ask the expert to imagine that a person has been selected randomly from the population. Considering that the person is below a certain percentile for one of the variables, ask the expert the probability that the certain person is below the same percentile for the second variable.

In his report, Clemen *et al.* found that the second method performed best [73]. This however seems to disagree with the other experiments previously mentioned which discourage direct elicitation of statistical moments. It is also important to note that several cases of multivariate elicitation are not comparable to draws from a population. If the problem of interest revolves around eliciting judgements about the correlation between fuel consumption and maximum acceleration of a car, methods 4 and 6 suggested by Clemen *et al.* clearly do not apply; the two variables are not single draws from a population anymore. For one of the case studies that are presented in this work however, the elicitation was structured in such a way so that it would be performed on each set of variables by conditioning on the values of the previous variables which had been elicited. A more detailed analysis will be shown in Chapter 4. Before moving on however, there is a need to talk about the case in which the elicitation is performed with multiple experts, because most of the time, it might be desirable to have more than one opinion taken into account when trying to represent a probability distribution.

2.4.6 Number of Experts

There are several psychological factors which should be accounted for when more than one expert is involved. Generally, there are two different kinds of approaches which can be used when eliciting the experts' knowledge. First, there is the situation when the experts do not interact and separate probability distributions are elicited separately and are then synthesized into a single distribution. The question to be asked as a result of that is how to combine those distributions into a single one [52].

The most popular methods are based upon "opinion pools" which in general terms represent either a weighted average of all the probability distributions which form it (linear pool) or the weighted average of the logarithms of each probability distribution (logarithmic pool). The advantage of using those techniques is the fact that the values of the weights attributed to the experts can be modified based on

how accurate the probability distribution from each expert is believed to be [74]. However, pooling methods could lead to a final PDF which does not explicitly show whose opinion is captured most; it may be the case that no one's opinion is captured in the final PDF. This technique is also known as *mathematical aggregation*.

Behavioural aggregation is another way of eliciting judgements from a group of experts and it involves allowing the experts to discuss which in the end should result in a single PDF. In this case, the elicitor should act as a moderator and encourage all experts to contribute to the group discussion, bearing in mind that some of them may be less likely to express their opinions than others who have stronger personalities and will show a tendency to dominate the group. As a result, the elicitor should be aware of this situation and encourage everyone to give their opinions, while remaining aware that some people may be naturally quiet as they may not have anything to say on a topic.

An essential factor is heterogeneity of opinion which should exist in any group where there is more than one expert. The prime reason why the elicitor wishes to increase the number of experts is to improve the quality of the judgements. As a result, it can be disadvantageous to add experts with the same opinion as in this way, even though the overall confidence increases, the accuracy might not [48]. This means that as long as the experts are able to discuss freely and exchange opinions, a net benefit should exist which should be greater than in the case when experts give individual opinions. Also, caution needs to be taken when an expert proposes another expert from his or her area of expertise to take part in the elicitation; it is often the case that experts have the tendency to propose others who have the same opinions as themselves.

One of the methods to manage interaction within a group is also known as the Delphi method. It involves eliciting each PDF individually, after which each expert's opinion is fed back to all the others while explaining the reasoning behind it. After that, each expert is allowed to revise their reasoning and hence the individual PDFs. Each revision is passed to all other experts and this process continues in an iterative fashion. In a way this is a hybrid between the mathematical and behavioural aggregation methods and is not likely to promote a very efficient sharing of knowledge as the latter. A review of the Delphi method (which has been used frequently in political sciences) has been made by Pill [75].

There is no general consensus regarding which aggregation method is better. Even more, it is downright difficult to compare the two and the reason for that is intuitive. It is impossible to test the two methods with the same group of experts simply because once a judgement, the said judgement already exists in their head meaning they cannot give it a second time. It can be possible however to test the two methods on different groups of experts, however this will only give statistically significant results on which method is better and to obtain an estimate, a large number of those tests need to be made, meaning that the whole process can become extremely expensive.

A popular method of drawing judgements is called the Sheffield Elicitation Framework (SHELF) and it has been created by Prof. O'Hagan from University of Sheffield. The method uses the behavioural aggregation technique and moreover, it uses the idea of a "rational impartial observer" (RIO). This technique makes the consensus which is obtained following the interactions between all experts to have the form of a PDF which would come from a hypothetical rational and impartial observer who has attended the whole meeting and has formed an opinion based upon everything that the experts said. It should be noted though, that the idea of "consensus" does not imply that the experts agree that the final distribution represents the opinion of either of them. In this manner, the judgements given by an expert who had been dominating the discussion and whose opinions were not agreed upon by the majority of the others will not reflect too much in the RIO's final PDF. This has an opposite effect on people who are quiet and whose judgements are closer to the others' beliefs, thus making their opinion stand out more in the resulting PDF.

An important point which has to be addressed has to do with the actual number of experts whose judgements should be elicited. Theoretically, adding experts to the elicitation exercise should increase the odds to have the actual values of interest elicited. In addition to that, more judgements would result in a reduction in the potential error, meaning that apparently, the more experts involved, the better the results of the elicitation would seem to be [48]. However, this is usually the case when the elicitation has the form of a questionnaire; in a workshop type of meeting it may be impractical to have a large number of attending experts as multiple sessions would be required. This means that a trade-off between quality and quantity will have to be made. Obviously, this doesn't take into account availability which is a prime factor that decides the difficulty of finding the experts and as a

result, the maximum number of experts able to attend the elicitation. Regarding the actual question *"How many experts?"*, there are several suggestions in the literature regarding the number of experts which would be appropriate. For instance, Aspinall [76] shares his experience with more than 20 panels of experts, saying that generally 8 up to 15 experts is an appropriate range to consider whenever setting up an elicitation. Having more than that, not only offers little improvement to the accuracy, but also increases cost and time. Meyer and Booker [50] answer the same question in a different manner; they are of the opinion that having less than 5 experts does not address the point mentioned above referring to heterogeneity of opinion, and does not offer as much diversity to the whole process.

2.5 Literature Review Summary

This chapter had the purpose of critically reviewing the concepts which sit at the foundation of this project. Those concepts are first and foremost placed within an uncertainty quantification context and refer to Bayesian updating, expert systems, as well as judgements elicitation. As mentioned in section 2.3, there is a large variety of expert systems which have been developed, and the idea of using probabilities to characterise them has recently started to be implemented [31]. Bayesian belief networks have been used in order to create a decision support system for the prognosis of head-injured patients of the intensive care unit (ICU). The power of this kind of network became apparent as its performance was deemed similar to that of a human expert in the field.

One of the main hypotheses within this thesis is the fact that such a network could be improved if the experts' opinions themselves could be added. Jongsawat and Premchaiswadi [37] actually put forward a methodology which can obtain the structure of a Bayesian network from experts, however it seems to be basic and is applicable to discrete distributions. Karandikar et al. [38] employed the strength of Bayesian inference via Markov Chain Monte Carlo in order to predict the tool life of a cutting tool. However, they did not use any expert input, as their research was solely based upon the updating procedure. There have been other sources where MCMC methods have been applied in order to do predictions, such as Zhu et al. [40], but they do not make use of a proper expert elicitation framework. In addition

to that, while exploring the literature on expert elicitation, it became apparent that the feedback process is very simplistic, because the distributions which are shown to the experts after the fitting procedure do not contain information regarding the physics of the problem. This is counterintuitive, as generally the experts have more experience in their field rather than in statistics, meaning that more can be gained by converting the purely statistical outputs, which literature suggests, into tangible ones.

As there is a common theme in the literature regarding research which mainly investigates one of the two components (and no application regarding the preliminary engine design process), unifying them becomes natural. Moreover, as it was mentioned in section 2.1, the type of uncertainty quantification that Rolls-Royce does is just related to using Monte Carlo simulations on the variables present in various problems in order to propagate their variability, and check the probability density functions of the outputs [10]. The claim is that this process can be improved by going further and using Markov Chain Monte Carlo in order to also be able to predict outputs when real data is scarce.

Therefore, by integrating a robust expert elicitation procedure together with the capabilities allowed by Bayesian inference it is possible to create a framework that uses a "best of both worlds" approach in order to analyse the preliminary design process of the aero-engine. This relates back to Chapter 1 where it was mentioned that by doing so it is possible to strike a balance between cost and flexibility, as this methodology would combine both high and low fidelity techniques in order to improve the design process. A more detailed description of this is going to be outlined in the following chapter. Before delving into Chapter 3 however, a project aim should be stated as well as a set of objectives which are going to be followed in order to satisfy the corresponding aim.

2.6 Aim and Objectives

The aim of this project can be considered to be quantified by the research hypothesis stated in section 1.3, which is:

"Using probability-based methods during the preliminary stage of the aero-engine design process can allow fast and accurate investigation of the design space in order to aid the generation of an optimum design which can afterwards be passed to the detailed design phase. Ultimately, this can have the effect of making the entire design process faster and less expensive."

The objectives which can satisfy this aim are based upon gaps that were found in the literature, and they are expressed as follows:

- **Objective 1:** Build the expert elicitation component, including the new feedback procedure which is deemed to be more informative for the experts who took part in the elicitation.
- **Objective 2:** Develop the Bayesian inference component which is used to make predictions on how a given model should behave in conditions that were not previously tested.
- **Objective 3:** Combine the two components into a framework which is used to update prior judgements regarding a physical problem by using real data.
- **Objective 4:** Validate the framework.
- **Objective 5:** Identify aero-engine case studies of industrial relevance which can benefit from this framework, and solve them by directly applying the framework to them.

As the novel aspects presented in section 1.4 are closely related to the objectives above, they are listed once more as follows:

- Building the Bayesian-elicitation framework
- Writing the custom code for the framework
- Applying the framework to the specific case studies from Chapter 4
- Developing the new elicitation feedback procedure

Chapter 3 is going to lay the foundations of the framework by mainly looking into the components of the framework, while Chapter 4 will apply the framework to three different case studies.

Chapter 3 Methodology

This is the part of the thesis which describes how all the literature investigated in Chapter 2 comes together in order to build the framework which was used for various case studies related to the preliminary stage of the design process. The framework has been constructed in MATLAB, and in this section, its various components (including snippets of pseudocode) are described separately in order to give the reader an in-depth understanding regarding the way in which it works. The reason why the MATLAB code was written from the ground up instead of using existing tools is because although there are software packages which can do Bayesian updating, the priority was to be able to easily link the Bayesian component to the elicitation one. Writing custom code allows that by offering more flexibility and also removes the issue of treating any other software package as a black box. Before showing the entire framework, an example will be given regarding how Bayesian inference can be applied to a simple problem. The reason for showing this is because the philosophy behind this toy problem is the same as the same as the one behind the real case studies presented in Chapter 4.

3.1 Bayesian Updating Example

The following problem illustrates both the main idea behind Bayesian inference as well as the idea of conjugacy which was briefly touched upon in section 2.2.

The turbine blades of a certain type of aircraft engine are examined during a maintenance procedure for potential crack initiation points after the engine spent several years in service. The goal of the Bayesian statistician in this circumstance is to update the already existing model which predicts the number of crack initiation points in a turbine blade by using the actual results which come out of the maintenance procedure.

It should first of all be made clear that this example is fictitious and its only purpose is to illustrate some of the potential benefits of performing Bayesian analysis. It is known that discrete events that have a low probability of occurring can be represented by a Poisson distribution [22], which is also the distribution of choice used here in order to describe the likelihood function. As a result, the number of crack initiation points within the blade could be represented in the following way:

$$f(x|\theta) = \frac{e^{-\theta}\theta^{x}}{x!} \tag{14}$$

Here, x is the number of crack initiation points which appear on a turbine blade after a certain amount of time, while θ is defined as the mean of the Poisson distribution. In addition to that, judgements regarding the parameter θ have to be made and at first they are captured in a prior which for instance can take the following form:

$$\pi(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta} \tag{15}$$

It has been assumed that the Gamma distribution is flexible enough in order to capture prior beliefs regarding the mean of the Poisson distribution. The parameters α and β could for instance be found from previous cases or they could be elicited from experts by setting up an elicitation workshop in this respect. For illustrative purposes, it will be assumed that both α and β are known and it will be examined how the posterior distribution changes as data is obtained.

The following nomenclature will be used in order to define several parameters:

- n represents the number of blades which are examined
- x_i is the number of crack initiation points on blade i, where $i = \overline{1,n}$

The likelihood function corresponding to all blades (assuming that the crack initiation on each blade is independent of all the rest – in a real situation this is debatable) will therefore have the following form:

$$f(x_1, x_2, \dots, x_n|\theta) = \prod_{i=1}^{n}\frac{e^{-\theta}\theta^{x_i}}{x_i!} = \frac{e^{-n\theta}\theta^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!} \tag{16}$$

According to the Bayes' rule in Equation (1), the posterior will have the following form:

$$\pi(\theta|x_1, x_2, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_n|\theta)\pi(\theta)}{\int_0^{+\infty} f(x_1, x_2, \dots, x_n|\theta)\pi(\theta)\mathrm{d}\theta} \tag{17}$$

The integral in the denominator is constant with respect to the posterior's argument θ as the integral takes into consideration all possible values of θ (and for a Gamma distribution, $\theta \in (0, \infty)$). By ignoring the terms which do not contain θ, the posterior can be rewritten as:

$$\pi(\theta|x_1, x_2, \dots, x_n) \sim e^{-n\theta}\theta^{\sum_{i=1}^{n} x_i}\theta^{\alpha-1}e^{-\beta\theta} = \theta^{(\sum_{i=1}^{n} x_i+\alpha)-1}e^{-\theta(\beta+n)} \tag{18}$$

It can be observed that the posterior has the form of another Gamma distribution, which can be fully verified by computing the integral in the denominator. Whereas the prior had a Gamma distribution with parameters α and β, the new posterior will have a distribution given by:

$$\theta|x_1, x_2, \dots, x_n \sim \mathrm{Gamma}\left(\alpha + \sum_{i=1}^{n} x_i, \beta + n\right) \tag{19}$$

In order to understand this better, a numerical example will be given for the above problem. First, values for the α and β parameters for the prior will have to be considered. For instance, if α and β are equal to 5 and 2.5 respectively, then the shape of the resulting prior distribution can be observed below in

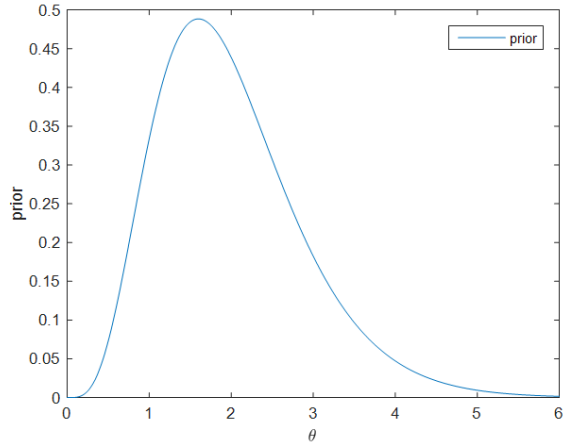


Figure 15: Shape of the prior distribution for the example problem

Also, the prior distribution can be mathematically written as:

$$\theta \sim \text{Gamma}(5, 2.5) \tag{20}$$

For illustration purposes it can be assumed that the number of blades n is equal to 20, while the total number of crack initiation points found on these blades will be equal to 10. After applying the Bayesian updating procedure, and by using the numbers and the posterior shape given in Equation (19) above, the resulting distribution will have the following form:

$$\theta | x_1, x_2, \dots, x_n \sim \text{Gamma}(5 + 10, 2.5 + 20) \tag{21}$$

The PDFs corresponding to both the prior and the posterior distribution are shown below in Figure 16:

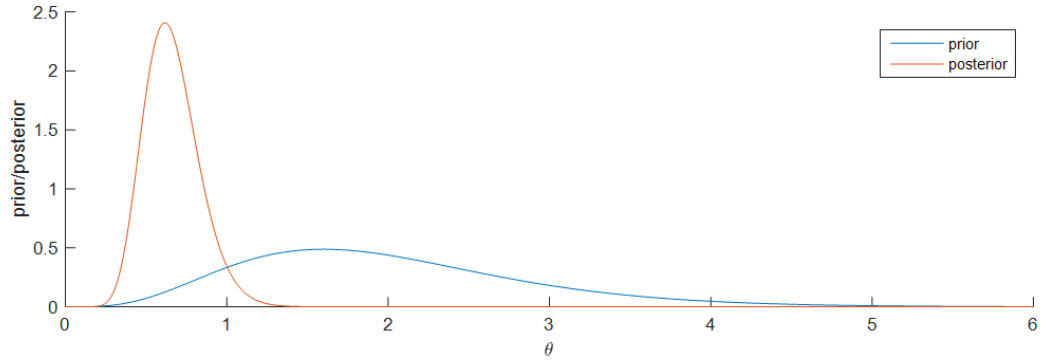


Figure 16: Prior and posterior shapes for the given numerical example

In the first instance however, a top-level view will be adopted, the purpose of which is to put everything into perspective. This will also show how the previous sections have been integrated into the framework.

As a brief observation, it could be inferred about the graphs in Figure 16 that the likelihood function had the effect of causing the posterior to cover a smaller range of values (decrease its variance). Also, the fact that the posterior has the same distribution as the prior (or in other words, $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ belongs to the same parametric family as $\pi(\boldsymbol{\theta})$ under the likelihood $f(\boldsymbol{x}|\boldsymbol{\theta})$) means that $\pi(\boldsymbol{\theta})$ is a conjugate prior for the parameter $\boldsymbol{\theta}$. This means that in general, if the prior and the likelihood are chosen in a specific manner, the full analytic calculation of the posterior becomes redundant as its distribution will be the same as that of the prior (albeit with different parameters).

For complex problems with multiple variables, making use of the concept of conjugate priors is impossible, which means that although the basic philosophy behind Bayesian updating shown in this simple example remains untouched, a numerical method needs to be used. The following section has the role of discussion the particular type of numerical Bayesian inference method used to solve the case studies presented later in this thesis.

3.2 The Markov Chain Monte Carlo Algorithm

After analysing the literature on numerical Bayesian updating methods in section 2.2, the particular method used here is the Metropolis-Hastings algorithm. The reason is because it is the most widely used [26] and in general it is a suitable algorithm for problems where no further information is known beforehand (such

as having to know the conditional PDFs of certain variables, in which case the Gibbs sampler would be more efficient). This subsection will therefore have the goal of describing this algorithm. The algorithm is part of a class of methods which combines the concepts of Monte Carlo analysis and Markov Chains, hence being named Markov Chain Monte Carlo (MCMC). The fundamental philosophy behind the Monte Carlo aspect is that the important features (mean and variance for instance) of a probability distribution which is unknown (in the Bayesian context that is the posterior) can be discovered given that a large number of samples have been obtained for the said distribution. The obvious question that rises has to do with how samples are drawn from a distribution which is unknown in the first place. In most sources from the literature, the algorithm is given without too much proof so the following steps have the purpose of offering some insight regarding the foundations of MCMC by starting from the very basic Bayes' rule. Before that however, there is a need to emphasize some basic concepts regarding the Monte Carlo method.

In mathematical form, the expected value of a random variable (which could be one of the features that Monte Carlo can be used to estimate) has one of the following two forms depending on whether the random variable X is discrete or continuous:

$$E(X) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x_i \tag{22}$$

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\, \mathrm{d}x \tag{23}$$

In this context, N is the total number of samples, x_i is a particular sample taken from the distribution of interest, and in the continuous case $f(x)$ is the PDF. Although it would be attractive to be able to compute the mean (and other statistical moments – such as the variance) in an analytical fashion using the integral from Equation (23), it becomes at least impractical to do so as the problem being studied increases in complexity. As a result, even for continuous random variables it is often the case that the mean is approximated by using Equation (22)

as it is easier to implement. Even if computers are used to perform the calculations, there is still a glaring problem with Equation (22). Although it is mathematically correct, computers cannot deal with the concept of infinity; simply put an infinite time and infinite memory is needed to obtain an infinite number of samples. Seeing that none of those specifications can ever be viable in practice, a compromise will have to be made. Consequently, the term "sample mean" is used whenever such computations have to be performed. Its expression is given below in Equation (24):

$$E(X) \cong \bar{X} = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{24}$$

Simply put, the limit is removed and instead of finding an exact solution, only N samples are taken in order to obtain an approximate value for the mean. However, the inevitable question which arises in this situation is "How to take samples?". In other words, there has to be a method which takes into account the shape of the probability distribution and takes samples from it. If the distribution has a simple form, samples can be obtained using the method of inverting the cumulative distribution function (CDF) and sampling from a uniform distribution between 0 and 1. Usually this cannot be done as most distributions which are of interest in real life have a more convoluted form, meaning that sampling needs to be done using a different technique.

The Markov Chain Monte Carlo technique mentioned in a previous paragraph makes use of the concept of a Markov Chain that is characteristic of stochastic processes. In general terms, a Markov Chain is a sequence of events or states for which the probability of moving between the current and the next state only depends on what the current state is and does not depend on any previous states of the chain. In a sense, the chain "forgets" everything that happened until the current state. By considering an arbitrary state "i" as the current state, and "j" as a potential next state, Bayes' rule can be applied in the following way:

$$P(X_{n+1} = j | X_n = i) = \frac{P(X_n = i | X_{n+1} = j) P(X_{n+1} = j)}{P(X_n = i)} \tag{25}$$

The conditional probabilities shown above represent transition probabilities between states "i" and "j", while the marginal probabilities represent probabilities of being in a certain state. For simplicity, the transition and state probabilities are going to be denoted by p and π respectively. Equation (25) therefore becomes:

$$p_{ij} = \frac{p_{ji}\pi_j}{\pi_i} \tag{26}$$

After multiplying Equation (17) by π_i, the reversal theorem can be obtained:

$$p_{ij}\pi_i = p_{ji}\pi_j \tag{27}$$

The way in which this is linked to a potential sampling routine is by considering π as the target distribution (unknown posterior), and by further splitting the transition probabilities in two more terms:

$$(q_{ij}\alpha_{ij})\pi_i = (q_{ji}\alpha_{ji})\pi_j \tag{28}$$

q_{ij} is the probability that if the current state is i, a tentative next state j is going to be selected. On the other hand α_{ij} represents the acceptance probability of state j as the next state of the Markov Chain. The acceptance probability can be rewritten as:

$$\alpha_{ij} = \left(\frac{q_{ji}\pi_j}{q_{ij}\pi_i}\right)\alpha_{ji} \tag{29}$$

If the ratio from Equation (29) is greater than 1, then the value of α_{ij} becomes 1 (because α_{ij} is a probability and if the said ratio is greater than 1 it means that state j will be accepted for sure). If on the other hand the ratio is smaller than 1,

then by inverting Equation (29) it is obtained that α_{ji} is equal to 1 because of the same reasoning. As a result, α_{ij} becomes equal to the ratio $\left(\frac{q_{ji}\pi_j}{q_{ij}\pi_i}\right)$ itself. The solution for the acceptance probability α_{ij} can be written as:

$$\alpha_{ij} = \min\left(1, \frac{q_{ji}\pi_j}{q_{ij}\pi_i}\right) \tag{30}$$

This is the core of what is known as the Metropolis-Hastings algorithm which is used in order to draw samples from a target distribution π which, as Equation (30) suggests, is only sufficient to be known up to a constant of proportionality. This ultimately means that the problem concerning the integral from Equation (1) has been completely by-passed. The Metropolis-Hastings algorithm itself can be outlined as follows [47]:

- Set iteration step $t = 1$ and generate an initial value for the chain (the process can be done many times with different starting points and they should give the same distribution in the end).
- Set $t = t + 1$
- Generate a proposal value θ^* from the proposal distribution $q(\theta|\theta^{(t)})$
- Compute the probability of acceptance $\alpha = \min\left(1, \frac{q(\theta^{(t)}|\theta^*)\pi(\theta^*)}{q(\theta^*|\theta^{(t)})\pi(\theta^{(t)})}\right)$
- Independently sample u from the uniform distribution $U(0,1)$
- If $u \leq \alpha$, then the proposal value is accepted and hence $\theta^{(t+1)} = \theta^*$. Otherwise, $\theta^{(t+1)} = \theta^{(t)}$
- The iteration step is once again increased by 1, and the process is repeated.

It might not be obvious at an intuitive level why this works. As a result, a graphical interpretation of this is suitable in order to grasp those concepts better.

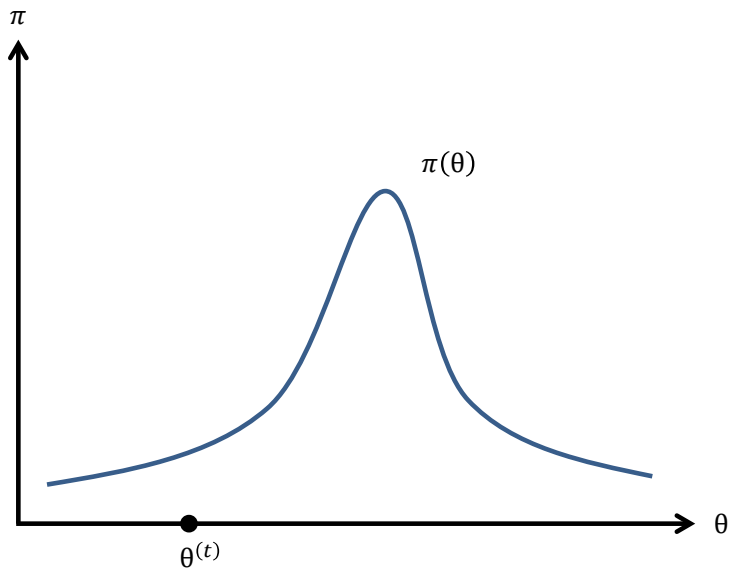


Figure 17: State of the Markov Chain at step t as well as the target distribution (which might be known, but not easy to sample from) [47].

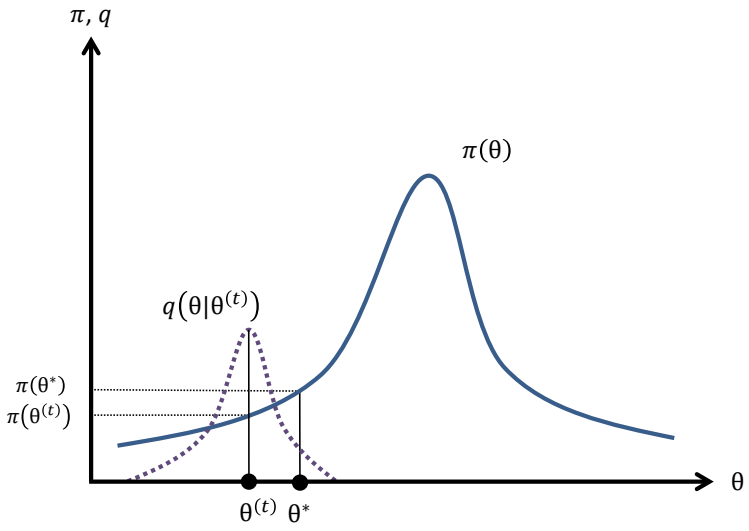


Figure 18: Choosing a proposed sample from the proposal distribution q and comparing its corresponding probability to the probability of the sample at step t [47].

Figure 18 stands to illustrate the process of choosing a proposal value from the distribution q. The value of probability corresponding to the target distribution is then calculated for the new proposed value ($\pi(\theta^*)$). The way in which it is decided

if the next state of the chain becomes equal to the proposed value is simply by comparing the "heights" of the points $\pi(\theta^{(t)})$ and $\pi(\theta^*)$. Given that the proposal distribution is symmetric, if $\pi(\theta^*)$ is higher on the graph than $\pi(\theta^{(t)})$, then it will be accepted 100% of the time. Otherwise, the new proposed value will only be accepted with probability $\pi(\theta^*)/\pi(\theta^{(t)})$. The reason why points with lower probability values are not automatically discarded is because in this manner, the Markov chain has the potential to move away from local maxima and find the global one (in the case when the distribution is multi-modal, and there is no reason to assume otherwise in general). In the case shown in Figure 18 the proposal value is accepted, meaning that the new proposal distribution will have its mean at θ^* and the whole process will be repeated. It should also be mentioned that the term $q(\theta^{(t)}|\theta^*)/q(\theta^*|\theta^{(t)})$ accounts for the potential asymmetry of the proposal distribution; if the proposal is symmetric, then this ratio becomes equal to 1 (and the general Metropolis-Hastings algorithm gets reduced to the more particular Metropolis algorithm).

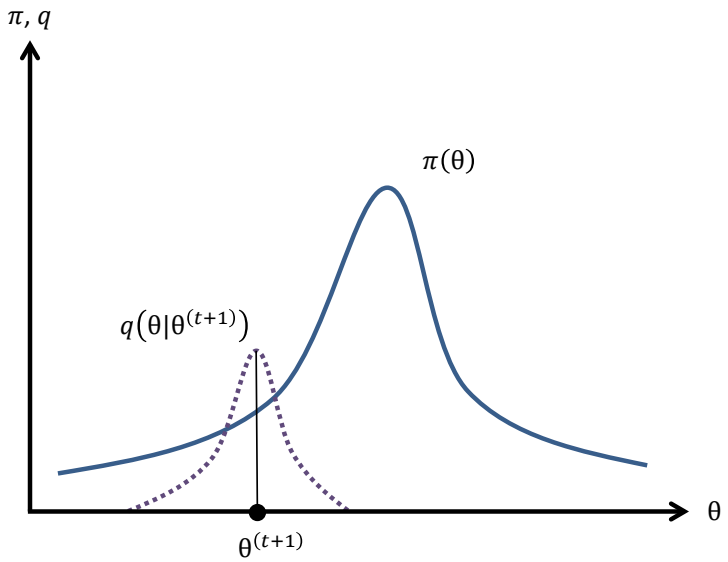


Figure 19: Generating the proposal distribution based on the next state of the Markov Chain (in this case, the proposed value has been accepted for the next step) [47].

It is generally the case however that parameter $\boldsymbol{\theta}$ is actually a vector, meaning that in order to obtain samples from the posterior distribution, the algorithm shown above needs to be slightly modified in order to accommodate this. This method

illustrates a sampling procedure which is implemented assuming that $\boldsymbol{\theta}$ has just two components; however the method could be extended for as many components as needed [47].

- Set iteration step $t = 1$ and generate an initial value for the chain
- Set $t = t + 1$
- Generate a proposal value θ_1^* from $q(\theta_1|{\theta_1}^{(t)})$
- Compute the probability of acceptance $\alpha = \min\left(1, \frac{q\left({\theta_1}^{(t)}|\theta_1^*\right)\pi\left(\theta_1^*, {\theta_2}^{(t)}\right)}{q\left(\theta_1^*|{\theta_1}^{(t)}\right)\pi\left({\theta_1}^{(t)}, {\theta_2}^{(t)}\right)}\right)$
- Independently sample u from the uniform distribution $U(0,1)$
- If $u \leq \alpha$, then the proposal value is accepted and hence ${\theta_1}^{(t+1)} = \theta_1^*$. Otherwise, ${\theta_1}^{(t+1)} = {\theta_1}^{(t)}$
- Generate a proposal value θ_2^* from $q(\theta_2|{\theta_2}^{(t)})$
- Compute the probability of acceptance $\alpha = \min\left(1, \frac{q\left({\theta_2}^{(t)}|\theta_2^*\right)\pi\left({\theta_1}^{(t+1)}, \theta_2^*\right)}{q\left(\theta_2^*|{\theta_2}^{(t)}\right)\pi\left({\theta_1}^{(t+1)}, {\theta_2}^{(t)}\right)}\right)$ (note that the argument of the target distribution uses the new updated value for θ_1
- Independently sample u from the uniform distribution $U(0,1)$
- If $u \leq \alpha$, then the proposal value is accepted and hence ${\theta_2}^{(t+1)} = \theta_2^*$. Otherwise, ${\theta_2}^{(t+1)} = {\theta_2}^{(t)}$
- The iteration step is once again increased by 1, and the process is repeated. For k different variables, the updating step for each iteration will simply be performed k times, and each update procedure uses the new values of the variables updated prior to that.

This particular method is a branch of the Metropolis-Hastings algorithm also known as *component-wise sampling*. Its name obviously comes from the fact that the components of the vector $\boldsymbol{\theta}$ are updated one at a time. There is another similar method of drawing samples from a target distribution called *block-wise sampling* whose algorithm is similar to what has been shown above; the difference lies in the fact that instead of updating each component at a time, a multi-dimensional proposal distribution is used (having the same number of dimensions as the vector $\boldsymbol{\theta}$) so that all vector components are updated simultaneously. The main problem with that however is the fact that in high dimensional problems it might be difficult to find a suitable proposal distribution, and also the method is generally associated with high rejection rates.

As the updating algorithm behind this framework was outlined above, the following section has the role of presenting the new method for obtaining feedback from experts, which is also the last step before presenting the overarching framework.

3.3 A new alternative way of obtaining expert feedback

The main suggestion found in the literature described in section 2.4 is to fit the experts' judgements to a PDF $\pi(\theta)$, after which to show it back to the experts in order to check whether it captures their uncertainty regarding the parameters of the problem. Whereas this method is quite simple to understand and apply, in most cases, the experts who attend the elicitation workshop are arguably more knowledgeable in their fields of expertise rather than when it comes to interpreting the statistics that are a direct result of the elicitation procedure. This means that it would be more useful if their judgements could be converted to a metric that is directly related to the problem undergoing elicitation, and thus more suitable for them to interpret. Fortunately, there is a way of doing this, and it requires one extra step: the PDF containing the first iteration of elicited quantiles is then converted to a prior predictive distribution, which is subsequently shown to the experts who can therefore change their original beliefs based on that.

By using the usual meanings of the variables presented in Section 2.2, the expression for the prior predictive distribution is as follows:

$$f(x_1, x_2, \dots, x_n) = \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta} \tag{31}$$

The above expression contains information regarding the physical problem for which elicitation is done in the first place. Therefore, by analysing $f(x_1, x_2, \dots, x_n)$, the experts can have a better understanding regarding whether and how they need to rethink their original judgements. In order to better understand the difference between the two feedback methods, it would be useful to highlight this idea using graphs:

Figure 20: Original method for obtaining expert feedback

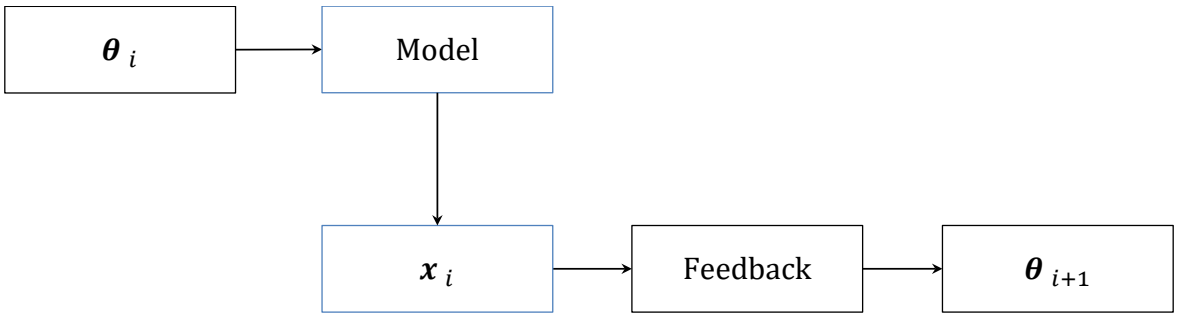


Figure 21: Proposed method for obtaining expert feedback

The difference between Figure 20 and Figure 21 has already been emphasized above: in order to make it easier for the experts, the statistical moments are converted to physical parameters via the physical model. The way in which this idea is implemented into the current framework is going to be illustrated in more detail in Chapter 4, section 4.3.

Finally, the entire framework is going to be described over the following section, which should make the reader understand how its individual components fit together.

3.4 A Top Level-View of the Framework

As previously mentioned, the framework constitutes of two main components: an expert elicitation component as well as a Bayesian inference one, both of which create the overarching expert system which is at the root of the whole project. Although they have been explained in detail in the sections above, it would be useful to briefly summarize them in order to set the scene for the rest of this chapter.

The expert elicitation framework from section 2.4 can be represented as a flowchart as follows:

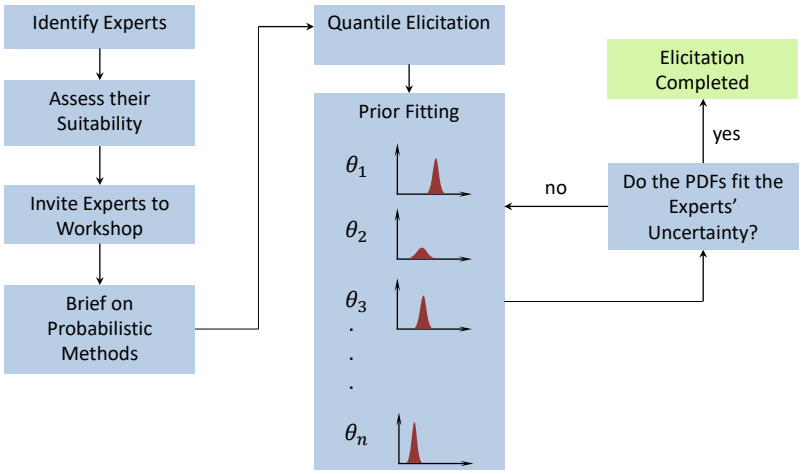


Figure 22: Flowchart of the elicitation process

The process shown in Figure 22 uses as a starting point the natural step of identifying the suitable experts for the problem requiring investigation. Afterwards, the experts are invited to the workshop where they should be briefed on probabilistic methods. Their judgements are subsequently elicited in the form of quantiles which are then fitted to PDFs. The feedback procedure commences after that. This can either be direct by asking the experts about the PDFs themselves, or indirect by using the method proposed in section 3.3 which involves converting the statistical moments to physical quantities. If the experts consider their beliefs need re-updating, the process goes back to the previous step. Otherwise, the elicitation is successful. If this is the case, the whole flowchart from Figure 22 becomes a component of the overall framework that is illustrated further.

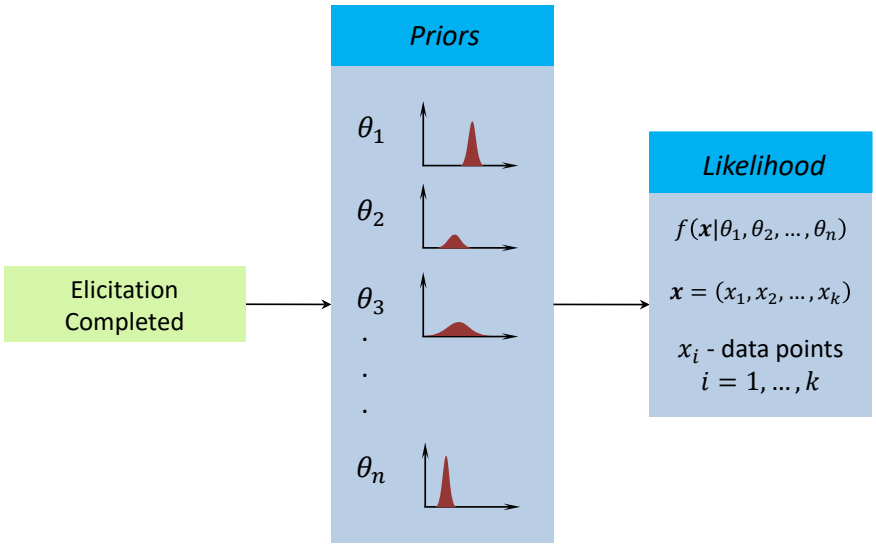


Figure 23: First part of the framework (Constructing the likelihood function based on the experts' priors)

The likelihood function simply shows how the physical model "thinks" the data should look like given the priors. The next step is using Bayesian inference in order to update the priors by using available high fidelity data (such as data from experiments or simulations):

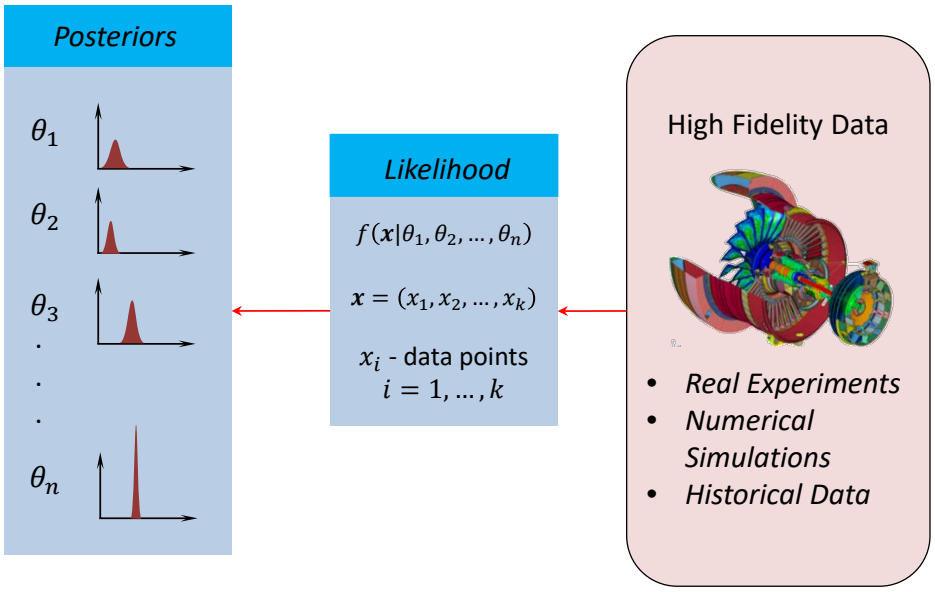


Figure 24: Second part of the framework (Updating the priors based on available high fidelity data)

Finally, the posteriors from Figure 24 can be used in order to generate the posterior predictive distribution, which ultimately allows the designer to predict how the subsystem in question will behave at operating conditions different than the ones for which data is available. The actual steps taken in order to be able to do accurate predictions are going to be shown in detail in the next section. Their purpose is to offer the reader insight into the inner workings of the framework by highlighting the code used in this respect.

3.5 A Detailed View of the Framework

This section is going to be split into several parts, each of them explaining in detail the various components of the framework. All the theory from Chapter 2 is going to be brought together at this point, so various references will be made to that section of the thesis. The framework description is split into further subsections, each of them dealing with one part of the MATLAB code used to put all the ideas shown in Chapter 2 in practice. Therefore the first subsection describes the expert elicitation component as well as how it fits into the overall picture.

3.5.1 The Expert Elicitation Component

The overarching framework was built to be flexible. In other words, its individual components can potentially be swapped with alternative ones if this procedure is deemed suitable for any specific case study undergoing analysis. For now however, considering the literature survey described in section 2.4, the elicitation method of choice that was chosen to be implemented was SHELF [54]. Thus, after the elicitation process (as shown in Figure 22) is successful, the data can be collected into a table. Then this table can be used as an input for a MATLAB script.

Due to the nature of the elicitation itself, the elicited quantities are the quantiles of n different distributions (as shown in Figure 22). The framework is able to accept four different distributions types which are: uniform, normal, lognormal and beta (although more distributions can be added). In addition to that, due to using the SHELF framework, the quantiles that can be elicited can be of various forms:

- Quartiles (the 25% and 75% quantiles are elicited)
- Tertiles (the 33% and 66% quantiles are elicited)
- Extreme (either the 1% and 99% or the 5% and 95% quantiles are elicited)

The type of the elicited quantiles depends on the nature of the calibration parameters, which are in turn dependent on the case study itself. It is important to note that in each case, besides the two quantiles shown above, the lower and upper bounds as well as the 50% quantile should also be elicited. As it has already been suggested in section 2.4.4, the reason for introducing this apparent redundancy is due to the idea of *overfitting*, which is linked to psychology. Although a PDF is uniquely defined by only two quantiles, the reason of asking for more is that it is very unlikely that just two quantiles are able to accurately depict the expert's "true" distribution [54].

The results from the elicitation can be placed within a table which assigns a lower bound (LB), an upper bound (UB) as well as three quantiles q_1, q_2, q_3 to each parameter θ_i. An example of how the collated data looks like can be seen in Table 2:

Table 2: Template of the table containing elicitation data used for the framework

Parameter	Distribution type	LB	q_1	q_2	q_3	UB
θ_1						
θ_2						
θ_3						
θ_4						

Assuming independence between the priors, all n sets of quantiles are read into a MATLAB script, and by considering their values as well as the "Distribution type", a set of n different means and standard deviations are created, and this will cause the prior distributions to be fully determinate. A few lines of pseudocode will be shown in order to illustrate the logic of the script.

```
read elicitation_table
for rows = 1:n
    if "Distribution type"=="normal";
        calculate mean and variance for normal distribution
    if "Distribution type"=="lognormal";
        calculate mean and variance for lognormal distribution
    if "Distribution type"=="beta";
        calculate mean and variance for beta distribution
    if "Distribution type"=="uniform";
        calculate mean and variance for uniform distribution
return a vector with n values for mean
return a vector with n values for variance
```

A point worth making here is the way in which the final values for mean and standard deviation (or variance) were selected. In essence, due to the fact that each distribution is uniquely defined by only two quantiles, it is possible to find three different pairs of means and variances. Quantiles q_1, q_2 and q_3 correspond to probabilities p_1, p_2 and p_3. Any two of those taken at a time are able to give a value

of mean and standard deviation. For example, (q_1, p_1) and (q_2, p_2) give (μ_1, σ_1); (q_2, p_2) and (q_3, p_3) give (μ_2, σ_2), while (q_1, p_1) and (q_3, p_3) give (μ_3, σ_3). This means that at the end, there will be three sets of expected values and standard deviations: (μ_1, σ_1), (μ_2, σ_2) and (μ_3, σ_3). The final values of the mean and standard deviation are simply found by averaging:

$$\mu_{\text{final}} = \frac{\mu_1 + \mu_2 + \mu_3}{3} \tag{32}$$

$$\sigma_{\text{final}} = \frac{\sigma_1 + \sigma_2 + \sigma_3}{3} \tag{33}$$

The pseudocode snippet shown above also describes that once the type of distribution is established, the mean and variance can be computed based on the quantile values. The algorithm used to do that is therefore illustrated below as doing so is considered to benefit the reader. This was done using the method proposed by Cook [77] which converts the quantiles and corresponding probabilities into pairs of (μ, σ). For a normal and lognormal distribution, this is done by introducing the function Φ, which is the CDF for the standard normal distribution (that is a normal distribution with an expected value of 0 and standard deviation of 1). The equations for finding μ and σ can therefore be written as:

$$\mu = \frac{q_i \Phi^{-1}(p_j) - q_j \Phi^{-1}(p_i)}{\Phi^{-1}(p_j) - \Phi^{-1}(p_i)} \tag{34}$$

$$\sigma = \frac{q_j - q_i}{\Phi^{-1}(p_j) - \Phi^{-1}(p_i)} \tag{35}$$

The subscripts i and j are simply used to represent the two different quantiles/probabilities used, while Φ^{-1} stands for the inverse of the standard normal CDF. If a lognormal distribution is deemed suitable, a similar approach is used. However, in this case it will not be the mean and standard deviation of the lognormal distribution that will be computed, but instead the mean and standard deviation of the associated normal distribution. In other words, if $X \sim \text{lognormal}(\mu, \sigma)$,

then $Y = \ln(X) \sim \text{normal}(\mu, \sigma)$ [77], and Equations (34) and (35) are simply modified by replacing q_i with $\ln(q_i)$:

$$\mu = \frac{\ln(q_i)\Phi^{-1}(p_j) - \ln(q_j)\Phi^{-1}(p_i)}{\Phi^{-1}(p_j) - \Phi^{-1}(p_i)} \tag{36}$$

$$\sigma = \frac{\ln(q_j) - \ln(q_i)}{\Phi^{-1}(p_j) - \Phi^{-1}(p_i)} \tag{37}$$

In the case of a beta distribution, the situation gets slightly more complex because there is no MATLAB function (or explicit equation) which can compute the moments of the distribution as directly as for the normal and lognormal distributions. Three different methods have been identified which are able to convert quantiles into moments of a beta distribution. The first one is based on SHELF [54] itself, which has an intuitive Graphics User Interface that is able to fit various PDFs based on quantiles elicited from experts. An example of that can be seen below:

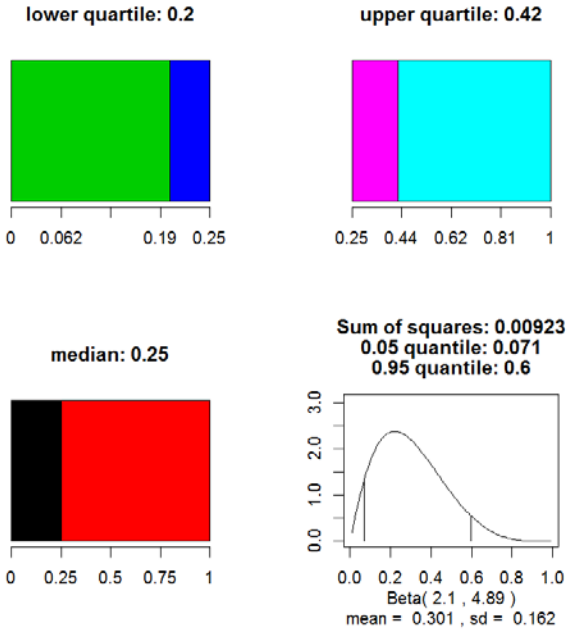


Figure 25: Example of a beta distribution created in SHELF

Cook [77] proposes an optimisation algorithm that can find the distribution parameters based on just two quantiles. The optimization problem can be formulated in the following way:

$$\min\big(f(\alpha,\beta)\big) = \min((\mathrm{CDF}(q_1) - p_1)^2 + (\mathrm{CDF}(q_2) - p_2)^2) \tag{38}$$

The terms $\mathrm{CDF}(x_1)$ and $\mathrm{CDF}(x_2)$ are functions of q_1, q_2 (which are the elicited quantiles), as well as α and β which are the unknowns. The algorithm requires starting values for both α and β as well as a direction of search. As before, p_1 and p_2 are the known probabilities for which the quantiles are elicited. It can be easily verified that the theoretical minimum of the function from Equation (38) is clearly zero, which coincides with the minimum value of a sum of squares of real numbers. This also means that the combination of α and β which brings the function $f(\alpha,\beta)$ as close to 0 as possible is the solution to the optimization problem. Unfortunately, this particular algorithm is very slow, and because it also needs to run multiple times it leads to an overall running time that is too large to be practical. One way to overcome this would be to use an approximate closed form expression for the median of the beta distribution, as suggested by Kerman [78]:

$$\text{median} = q_2 \cong \frac{\alpha - \frac{1}{3}}{\alpha + \beta - \frac{2}{3}} \tag{39}$$

Equation (39) involves the value of q_2 which is the elicited value of the median (i.e.: the 50% quantile – that is always elicited regardless of the elicitation type). This means that a relationship between α and β can be explicitly formulated, which in turn means that the optimization algorithm from Equation (38) can simply use just one independent variable (either α and β), while the other one can be deduced by using Equation (39). Although this reduces the computational time exponentially, there is yet one more method which is more accurate as well as elegant and quick.

The final method for solving the backwards problem is based on a paper by van Dorp and Mazzuchi [79] which takes into account the monotony of the incomplete beta function in terms of α and β in order to produce a fast algorithm which solves for the parameters. The first part of their analysis was based on analytically proving that only two quantile constraints are needed in order to obtain a solution for α and β (although this solution might not necessarily be unique). The existence of this proof also seems to answer the uncertainty identified in the paper by Cook [77] regarding whether the proof exists in the first place. The essential feature of the algorithm is its high degree of accuracy as well as fast computing time, which was

the reason why this particular procedure was implemented. It is based on a bisection optimization algorithm, although the full details are beyond the scope of this work, so the reader is simply referred to the relevant paper [79].

As the beta distributions are determined using the method described above, the framework then converts α and β into a mean and variance:

$$\mu = \frac{\alpha}{\alpha + \beta} \tag{40}$$

$$\sigma = \sqrt{\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}} \tag{41}$$

Finally, in the simplest case when a uniform distribution with lower and upper bounds a and b is used, the mean and standard deviation can be derived very easily, and they have the following form:

$$\mu = \frac{a + b}{2} \tag{42}$$

$$\sigma = \sqrt{\frac{(b - a)^2}{12}} \tag{43}$$

The pseudocode described above contains the logic within a MATLAB script called **PriorFit.m**. In order to put everything into perspective, it would be useful to represent the pseudocode in the form of a flowchart:

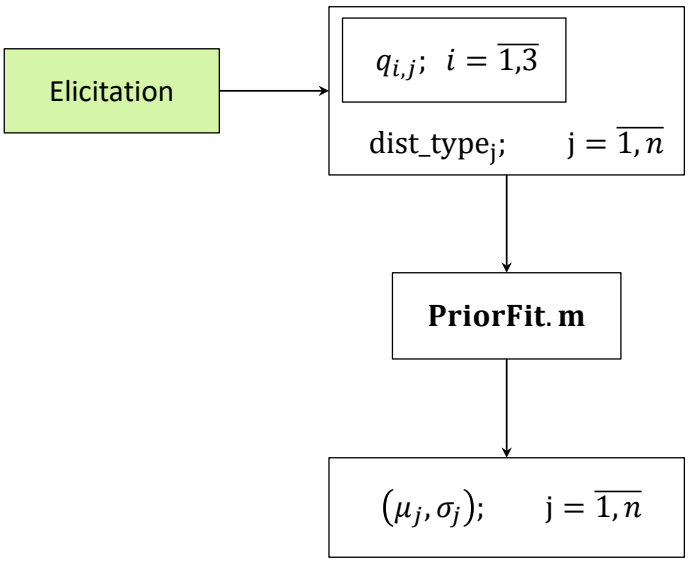


Figure 26: Determining the prior distributions from the elicitation process

The next subsection will therefore delve into the next step of the framework, which has to do with analysing the likelihood function.

3.5.2 The Likelihood Function

The likelihood function has been first defined in section 2.2. In general, it is written as $f(\boldsymbol{x}|\boldsymbol{\theta})$, and in qualitative terms, it tells "how the model thinks the data $\boldsymbol{x}$ should look like, given a prior distribution for the parameter $\boldsymbol{\theta}$". The available data is contained within a table which associates a set of physical parameters $(\text{Physical\_param}_{i,j})$ to a data point Data_i. A template of how such a table can look like is shown below in Table 3.

Table 3: Template of the table containing data for the updating procedure

Physical_param1	Physical_param2	Physical_param3	*Data*

The subscript i refers to the number of data points, while j stands for the number of physical parameters (in the simple case from Table 3, j takes values from 1 to 3). Of course in general, the number of the physical parameters can be extended indefinitely. They can refer to the known physical experimental inputs (e.g.: temperature, stress, speed), and together with the unknown calibration parameters (which are approximated via the prior distributions) are used to obtain the data. In

addition to that, the distribution type for the likelihood has to be selected. This has been done by using a particular script named **allfitdist.m**. It works by simply fitting the data to multiple distributions and then comparing their Bayesian Information Criterion (BIC) to each other, and then uses the one with the lowest BIC as the most suitable distribution [80]. This simple process looks as follows:

Figure 27: Setting up the distribution type for the likelihood function

One important point to discuss is the physical model itself. In other words, there also needs to be a relationship between the physical parameters representing the inputs (shown by the first three columns in Table 3) as well as the physical parameter representing the output (which can be more than one). A general expression of this relationship is usually contained within a model, and it can have the following general form:

$$h(\mathbf{phys\_param}_1, \mathbf{phys\_param}_2, \dots, \mathbf{phys\_param}_m) = \boldsymbol{x} \tag{44}$$

In Equation (44), h is the model itself, while m is the total number of calibration parameters. The bold font emphasizes the fact that the parameters are vector quantities.

The script used to compute the likelihood function is called **LikelihoodValue.m**. It simply finds a PDF value in the point data_k by considering the type of distribution found via **allfitdist.m**, as well as a mean (which is effectively the value of the function h when the values of the physical parameters from Table 3 correspond to the k-th data point), and a standard deviation. The manner in which the standard deviation is defined uses the concept of measurement uncertainty. In other words, the standard deviation is a fraction of the measured data:

$$\sigma_k = f * \mathrm{data}_k \tag{45}$$

f is a number between 0 and 1, whose value is related to the amount of confidence in the available data. The following section will take into consideration one of the

crucial elements used in the numerical updating procedure, that is the nature of the proposal distribution q.

3.5.3 The Numerical Algorithm for the Bayesian Updating

The numerical method used in order to update the experts' belief with real data has been described in section 3.2 as the "single component Metropolis-Hastings algorithm". The block-wise sampling method was not chosen because generally it is associated with high rejection rates [38]. The joint prior distribution (which for now has been considered to be made of independent variables) is updated one variable at a time, using a number of proposal distributions equal to the total number of variables.

The framework allows the proposal distribution to be of various types (normal, lognormal, beta, uniform – which are the same distribution types as the prior). At the moment, the proposal distribution is assigned the same type as the prior (although this can be easily changed). The mean of the proposal distribution is always equal to the current state of the chain, while the standard deviation was chosen to be the same as that of the prior. Choosing a suitable standard deviation is generally not an easy task because if it is too small, the algorithm tends to only focus on a small area of the design space, whereas if it is too large, it will explore the space without exploiting potentially suitable areas [38]. It has been observed that choosing a standard deviation for the proposal distribution of a similar magnitude (or at least the same order of magnitude) as the prior distribution creates a good compromise between exploration and exploitation. This can be expressed as follows:

$$q\left(\theta_j^* | \theta_j^{(t)}, \sigma_j\right) \tag{46}$$

Equation (46) gives the value of the proposal distribution q in the point θ_j^* (which is the proposed value for variable θ_j), $\theta_j^{(t)}$ is the current value, and σ_j is the standard deviation. As presented in section 3.2, the mean of the proposal distribution always changes as long as newly proposed values are accepted.

Another important point to make is regarding the likelihood function; more specifically regarding the way in the likelihood function uses the data. There are

two different ways in which updating can be performed: simultaneously or sequentially. Sequential updating allows the likelihood function to use each data point at a time and update the priors in this manner. Therefore, the first posteriors obtained in this way become the priors for the second data point, and this process continues until all data points have been used. The second method (i.e.: simultaneous updating) implies a likelihood function which contains information about the entire data set. Although the results from the two methods are mathematically equivalent, when doing a numerical simulation simultaneous updating can be cheaper and more accurate. The reason is that for simultaneous updating, the MCMC procedure is only done once, while for sequential updating the total number of iterations is done for each data point. In the early stages of developing the framework, while analysing the available options, sequential updating has been looked into (and also used for the first case study). Later, the simultaneous updating has been investigated, and has then been implemented for case studies two and three. In the end, simultaneous updating has been used as the default method of performing Bayesian inference within the framework.

It is worth showing how the simultaneous updating algorithm works. To do so, Equation (1) can be rewritten by expanding vectors $\boldsymbol{\theta}$ and $\boldsymbol{x}$:

$$\pi(\theta_1, \theta_2, \dots, \theta_n | x_1, x_2, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_n | \theta_1, \theta_2, \dots, \theta_n)\pi(\theta_1, \theta_2, \dots, \theta_n)}{\iiint_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n | \theta_1, \theta_2, \dots, \theta_n)\pi(\theta_1, \theta_2, \dots, \theta_n)\mathrm{d}\theta_1 \mathrm{d}\theta_2 \dots \mathrm{d}\theta_n} \tag{47}$$

Although Equation (18) looks complex, there are several assumptions which simplify both the equation and the expression for the likelihood function given by:

$$L = f(x_1, x_2, \dots, x_n | \theta_1, \theta_2, \dots, \theta_n) \tag{48}$$

The first assumption is that each experiment is independent of all the others; in other words x_j does not depend on x_i. Knowing this, the likelihood function can be rewritten in the following way:

$$L = f(x_1 | \theta_1, \theta_2, \dots, \theta_n) * \dots * f(x_n | \theta_1, \theta_2, \dots, \theta_n) = \prod_{i=1}^{n} f(x_i | \theta_1, \theta_2, \dots, \theta_n) \tag{49}$$

The other assumption is that each data point is solely dependent on its corresponding model prediction (i.e.: x_i only depends on θ_i). As a result, the total likelihood can finally be simplified as follows:

$$L = f(x_1|\theta_1) * f(x_2|\theta_2) * \dots * f(x_n|\theta_n) = \prod_{i=1}^{n} f(x_i|\theta_i) \tag{50}$$

Computationally, Equation (50) will be easier to use, especially when finding the acceptance ratio for the Metropolis-Hastings algorithm. As a reminder, the acceptance ratio for the i^{th} parameter is given by:

$$\alpha = \min\left(1, \frac{q\left(\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_n^{(t)} | \theta_i^*\right) \pi\left(\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_i^*, \dots, \theta_n^{(t)}\right)}{q\left(\theta_i^* | \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_n^{(t)}\right) \pi\left(\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_i^{(t)}, \dots, \theta_n^{(t)}\right)}\right) \tag{51}$$

As it may be deduced, t stands for the current iteration step. In some cases, the proposal distribution q is symmetric, which simplifies the above expression to the simpler Metropolis algorithm:

$$\alpha = \min\left(1, \frac{\pi\left(\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_i^*, \dots, \theta_n^{(t)}\right)}{\pi\left(\theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_i^{(t)}, \dots, \theta_n^{(t)}\right)}\right) \tag{52}$$

The equation simplifies such that the only remaining term from the acceptance probability expression is the posterior distribution. Depending on the prior distribution, Equation (52) could be simplified even further. For instance, if the prior is non-informative (such as a uniform distribution), then the acceptance ratio will simply become equal to the ratio of the likelihoods:

$$\alpha = \min\left(1, \frac{f\left(x_1, x_2, \dots, x_i, \dots x_n \middle| \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_i^*, \dots, \theta_n^{(t)}\right)}{f\left(x_1, x_2, \dots, x_i, \dots x_n \middle| \theta_1^{(t+1)}, \theta_2^{(t+1)}, \dots, \theta_i^{(t)}, \dots, \theta_n^{(t)}\right)}\right) \tag{53}$$

Both Equations (52) and (53) are computationally easy to solve. At the end of the entire MCMC simulation, the posterior predictive distribution can be computed; which in general has the following expression:

$$f(x_{n+1}|x_1, x_2, \dots, x_n) = \iiint_{-\infty}^{+\infty} f(x_{n+1}|\theta_1, \theta_2, \dots, \theta_n)\pi(\theta_1, \theta_2, \dots, \theta_n|x_1, x_2, \dots, x_n)\mathrm{d}\theta_1\mathrm{d}\theta_2 \dots \mathrm{d}\theta_n \quad (54)$$

The numerical equivalent of Equation (54) was described as ancestral sampling in section 2.3. The manner in which all the ideas illustrated above are put together is depicted to the reader in two ways: by using pseudocode as well as in a more graphical fashion by condensing all the information into several flowcharts. The pseudocode looks as follows:

```
read prior mean, standard deviation and distribution type
read data points
read physical parameters
set number of MCMC iterations
set uncertainty of the likelihood function (parameter f)
set P = vector with starting points for the prior
set standard deviation of proposal distribution q
set a flag variable to a value of 0

for i=1:number of MCMC iterations
    while j<=number of calibration parameters
        create a new vector Q = P
        create a sample from the proposal distribution
        assign the proposed value to Q(j)
        compute the prior based on the proposed value
        compute the prior based on the current value
        compute the proposal q based on the proposed value
        compute the proposal q based on the current value
```

```
compute the likelihood for each data point
obtain the total likelihood by multiplying all the above

perform a logic check on the model values
if the test fails, assign the flag a value of 1
    else j=j+1
end

find the acceptance probability a
sample a variable u from the uniform distribution (0,1)
if u<a
    assign the proposed value to the next step
else
    assign the current value to the next step
end
```

The results obtained using the code illustrated by the above lines give a joint posterior distribution. In order to compute the posterior predictive, an ancestral sampling routine is performed on the posterior. The pseudocode above can be put a more visual format as follows:

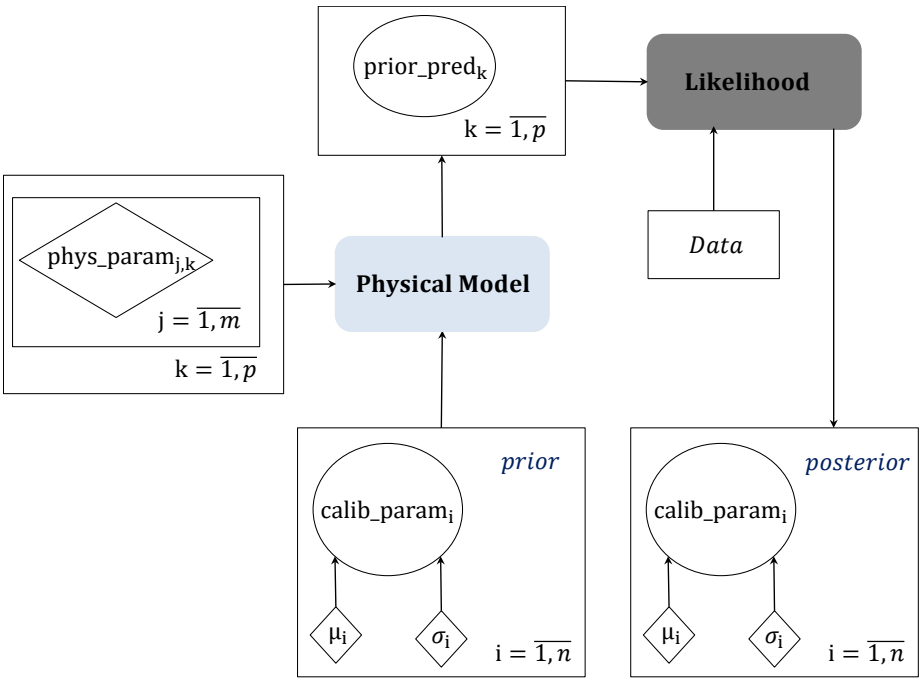


Figure 28: Framework of the updating process (simultaneous updating)

In Figure 28, the circles represent probabilistic variables (such as the calibration parameters), while the diamonds refer to deterministic ones. Here, n stands for the total number of calibration parameters, m is the number of physical parameters used by the physical model, while p is the number of data points. The likelihood function uses the actual data as well as the "predicted data" in order to update the original calibration parameters and hence find their posterior distribution.

Earlier, it was mentioned that the updating procedure of choice was the simultaneous one from Figure 28, however for completion purposes Figure 29 is used to show sequential updating that was used in a previous iteration of the framework.

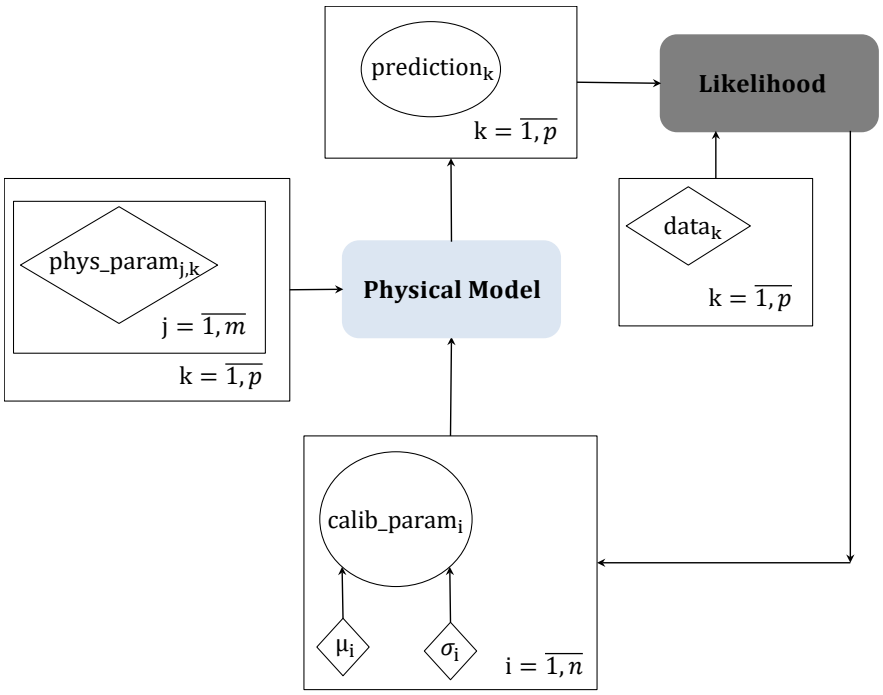


Figure 29: Framework of the updating process (sequential updating)

Finally, when the posterior distribution has been calculated in either way, the ancestral sampling routine is then ran using a simple Monte Carlo simulation which draws samples from the posterior distribution, which together with the deterministic values of the physical parameters results in a distribution at the end which shows how the model estimates the uncertainty of a new data point given all previously known data. This algorithm is once again shown graphically in Figure 30:

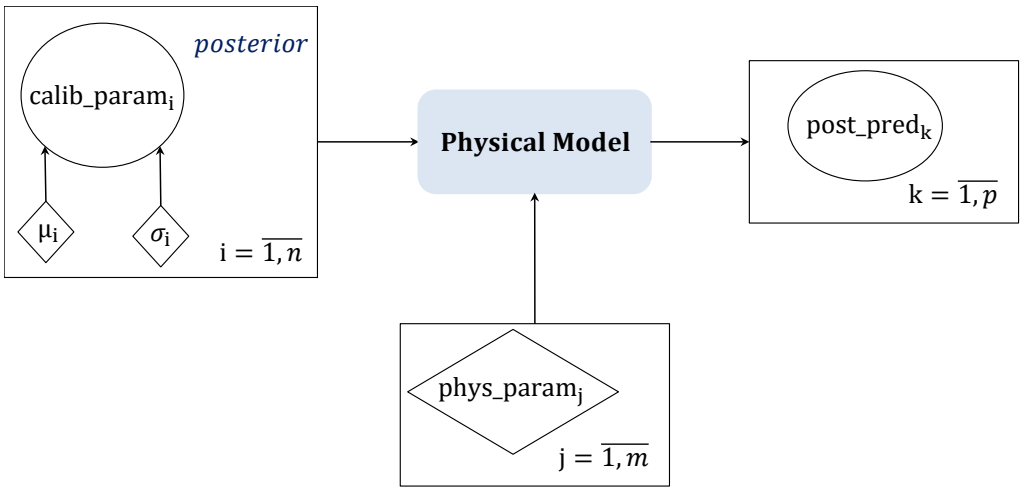


Figure 30: Ancestral sampling used in order to perform predictions

Although all the theory above holds from a purely mathematical perspective, its power should become apparent as it is put into practice. Its application to real-life case studies should make this hypothesis clear for the reader, and this is the main subject of the following Chapter 4. However, there is a major limitation that the framework is subject to. Because the prior distributions are based on expert judgements, it is automatically implied that the experts need to have experience regarding the problem that is being discussed. The issue here is that if one aspect of the problem is completely new (for instance if elicitation is done in order to gather judgements on the properties of a turbine blade made from a completely new material), then by definition the elicitation and hence the framework become invalid. However, in the aerospace sector (and in Rolls-Royce in particular) a great amount of research needs to be done before something completely new is implemented, which gives experts time to gather knowledge on the topic, and in those circumstances, the framework keeps its validity.

Before moving on to Chapter 4, there is one more step which should be taken in order to confirm the suitability of the framework for any applications: the validation process.

3.6 The Validation Process

It is imperative for any new approach to solving problems to undergo a validation process in order to gain confidence that the procedure used is sound. Consequently, this framework has been used in order to solve a problem identified in a research paper written by Karandikar et al [38] who have investigated the life of a cutting tool using a Bayesian approach. In other words, by using the authors' inputs and the framework presented here, that problem has been looked into, and the results of the framework were compared to theirs. The physical model which they used was the following:

$$T = \frac{C}{V^p f^q} \tag{55}$$

In Equation (55), T is the tool life itself, V is the cutting speed measured in m/min, while f is the feed rate in mm/revolution. C, p and q are the three calibration parameters for which Bayesian inference is made. By considering the three priors as uniform and independent of each other, the posterior predictive distribution (which is equivalent to the predicted life) for a combination of speed and feed rate for which data was not available can be determined. Next to that is the same prediction done with the Bayesian-elicitation framework presented in this thesis. The same prior and proposal distributions as well as the same likelihood function have been considered, and as Figure 31 proves, the similarity between the two graphs is undeniable. This is supported by doing a two tailed F-test on the two PDFs which yields that the test statistic is within the 95% confidence interval, therefore that the hypothesis that the two distributions are similar is accepted.

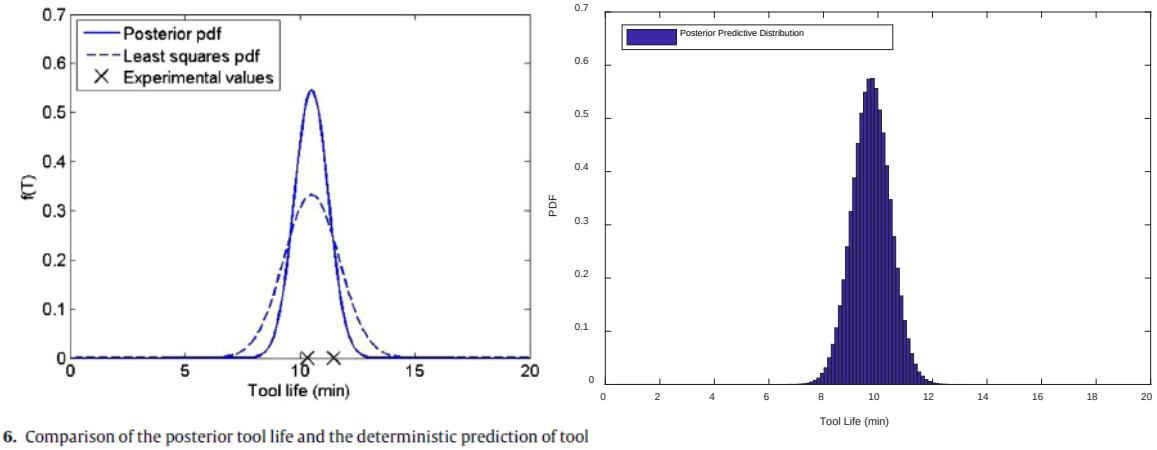


Fig. 6. Comparison of the posterior tool life and the deterministic prediction of tool life at $V = 192.01$ m/min and $f_r = 0.51$ mm/rev. The 'x' symbols denote experimental values and the 'o' symbol represents the curve fit prediction.

Figure 31: Comparison of predicted tool life between the journal paper results (left) [38] and the framework (right)

The additional step which needs to be taken in order to convert the histogram into a PDF was done via the previously mentioned MATLAB "allfitdist" function which compares the Bayesian Information Criterion (BIC) [80] for various models and considers the one with the lowest BIC as the most suitable one. Another prediction done for different inputs can be seen in Figure 32, where it is once again obvious that the two results are nearly identical:

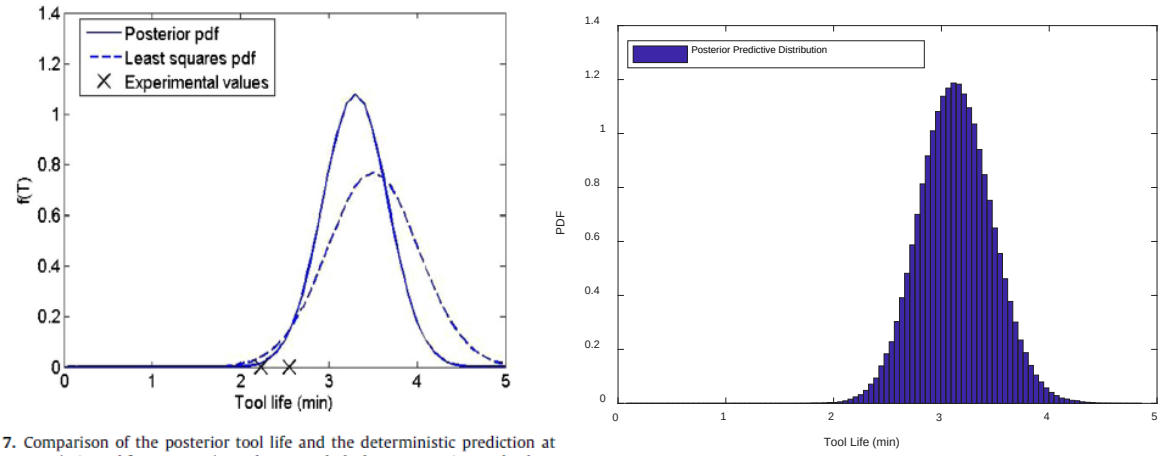
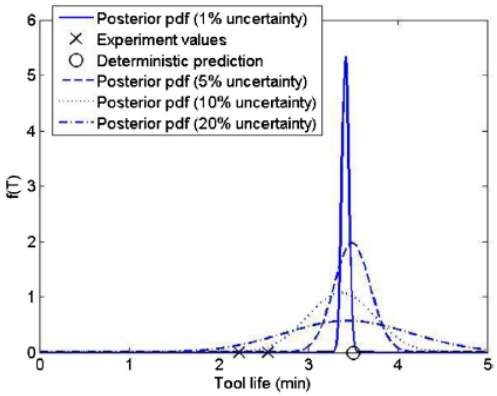


Fig. 7. Comparison of the posterior tool life and the deterministic prediction at $V = 230.42$ m/min and $f_r = 0.61$ mm/rev. The 'x' symbols denote experimental values and the 'o' symbol represents the curve fit prediction.

Figure 32: Comparison of predicted tool life between the journal paper results (left) [38] and the framework (right)

In addition to just finding the posterior distributions, the authors also tried using different uncertainty values for the likelihood function (similarly to what is shown in Equation (45)). As a general rule, as the standard deviation tends to 0, the PDF converges to a single point. This behaviour can be observed in Figure 33:



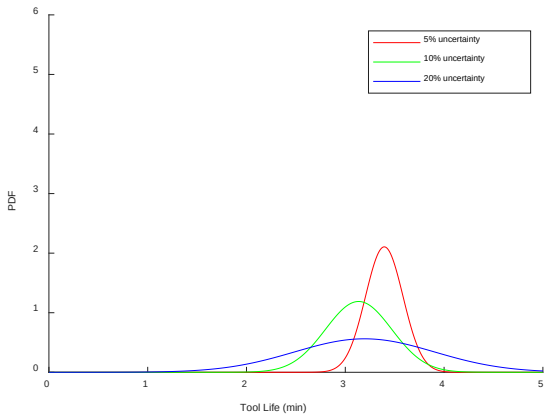


Fig. 12. Tool life predictions at $V = 230.42$/min and $f_t = 0.61$ mm/rev for different likelihood uncertainty levels. Note that the posterior pdf at 1% likelihood uncertainty is not normalized for demonstration purposes.

Figure 33: Comparison between predicted tool life for various likelihood uncertainty levels [38]

It is becoming more and more clear that the results obtained from the current framework once again display a large degree of similarity to the ones from Karandikar et al. [38]. Once again, the F-test yielded that the PDFs are the same at the 95% confidence level. The 1% uncertainty PDF has not been tested because the comparison would have been meaningless (the 1% likelihood uncertainty from the journal paper is not to scale).

This shows that the Bayesian component of the framework has been validated. However, as it was highlighted in this chapter, the framework also relies on an elicitation component, meaning that validation of the elicitation component also needs to be done. According to Pitchforth and Mengersen [36], when it comes to frameworks which rely on expert judgements, the validation process is not done in the traditional way by comparing the results with others' results, simply because sometimes the outputs from such frameworks are predictions regarding the future behaviour of a model, and that data may be either non-existent or very expensive to obtain. Instead, the posterior predictive distributions are shown to an expert who can affirm whether the predictions match their beliefs and experience. In addition to that, the feedback loop of the elicitation process ensures that the expert has a chance to change their judgements before they are passed on to the Bayesian component. In this respect, the feedback procedure is considered to be a validation process.

The results from this section indicate that the Bayesian-elicitation framework presented is valid. Consequently, as it was mentioned at the end of section 3.5, Chapter 4 is fully dedicated to demonstrate how this framework can be applied in

the context of the aero-engine preliminary design process, by illustrating three different case studies.

Chapter 4 : Application of the Framework to Real Life Aero-Engine Related Case Studies

The purpose of this chapter is to illustrate how the methodology described in the previous chapter (which in turn is based upon all the theory investigated in Chapter 2) can be applied to real-life case studies that are relevant to the preliminary stage of the aero-engine design process. One important thing to take note of is that those case studies were being looked into as the framework was being developed. This means that each individual case study will offer the reader the opportunity to see how the framework has evolved during the timespan in which this doctoral project has been undertaken. This is due to the fact that the framework and the case studies have not been analysed at separate times, but rather simultaneously. The main reason was that a great part of the work has been done in collaboration with Rolls-Royce plc and by the time the first case studies were looked into, the framework was in its incipient stages. Additionally, those three particular case studies were selected (as opposed to something else) because at the time, quantifying uncertainties regarding those problems presented a high degree of interest to the engineers from Rolls-Royce plc, as those were areas where not enough knowledge was available during the preliminary design stage.

The first case study investigated the disc on which the turbine blades are mounted. In this particular case, the interest was to model the grain size as that in turn is related to the fatigue life of the component. This particular case study did not deal with the first component of the framework (the expert elicitation part), however it played a paramount role regarding the development of the main Bayesian inference algorithm.

The second case study shows how knowing the likelihood of Fan Blade Off (FBO) events can allow the designer to reduce costs by eventually having to do fewer CFD and FEA simulations. As this project deals with rare data, an expert elicitation workshop had been held before the case study started to be investigated as part of the EngD project. Actually, the idea of using expert judgements as Bayesian priors came about while investigating FBO events simply due to their rare occurrence rate and the fact that obtaining experimental data is extremely expensive as it involves building a fan for the sole purpose of destroying it.

Finally, the third case study, also related to the turbine disc, looked into reasons why fatigue failure can occur (either due to crystallographic or inclusion failure). It is at this point that the framework was fully developed, and enough experience regarding using it had been gathered as a result of working on the previous two case studies. It would therefore be useful to present each individual case study for which the Bayesian-elicitation framework has been applied, starting with the one that was the main inspiration for using expert judgements in the first place.

4.1 Grain Size Growth within a Turbine Disk

This case study was primarily investigated in order to develop the Bayesian updating procedure. As a result, instead of drawing expert judgements, Bayesian priors have been obtained using a MATLAB code which simulates the grain size during an isothermal forging process as well as grain growth during heat treatment by using physics-based models [81]. Subsequently, the Bayesian updating component from the framework has been applied in order to make predictions regarding the life of a turbine disc based on the grain size.

4.1.1 Overview of the Physics of the Case Study

The typical disc application material for gas turbine engines consists of Nickel-based high temperature superalloys, with Ni acting as the base material and a variety of alloying elements coming from the IVB, VB and VIB groups of the periodic table [82]. The microstructure of the material during the manufacturing stages (which include alloy gas atomisation, hot isostatic pressing, extrusion, isothermal forging, heat treatment and machining) changes according to the applied temperature, strain and strain rate that the material experiences. It has been observed that the forging and heat treatment stages are the ones which dictate the final grain size distribution of the material [83]. Consequently, modelling correctly the two processes would provide a direct benefit to any aero-engine company, while minimising the experimental costs as well as concessions due to non-conforming microstructure.

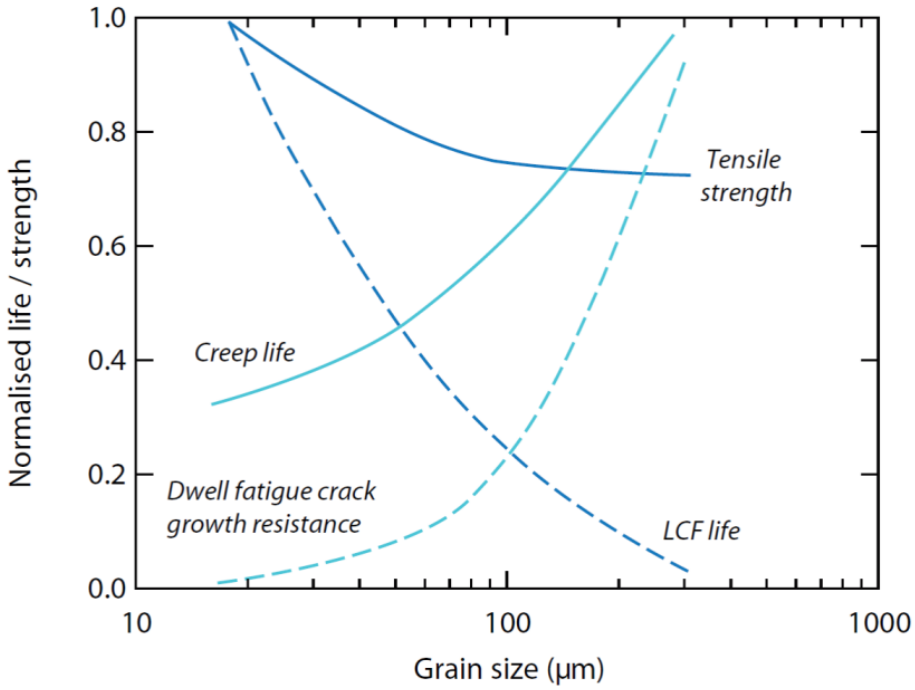


Figure 34: Modelled relationship between life/strength and grain size [84]

Randomness plays a role especially during the isothermal forging stage of the manufacturing process when the strain and strain rate are not uniform within the forging stock the disk is manufactured from. As a consequence, it was considered that Bayesian inference should be used in order to be able to define the limits of strain and strain rate allowable within the forging that should lead to an acceptable grain size distribution and optimal die design. The whole modelling process covers the following three steps:

1. Modelling isothermal forging and dynamic recrystallization
2. Using the initial grain size found at step 1, it is studied how the grain grows during the heat treatment phase
3. The MCMC component is used to generate samples from the posterior distributions of different parameters of interest

The first step is realized using a physics-based thermo-statistical model developed by Galindo-Nava and Rivera-Diaz [85], which takes as inputs the initial billet grain size, forging temperature, strain and strain rate in order to return the grain size at the end of the forging process (in this context it is also called "DReX", which stands for "dynamically recrystallized grain"). The second step involves using a model which predicts the evolution of the DReX during heat treatment by taking into account certain inputs such as grain boundary energy, grain boundary mobility, carbide dispersion as well as the heat treatment

temperature [81]. The reason behind the importance of grain size modelling and manufacturing conformity is shown in Figure 34. While the bore of the turbine disc experiences lower operating temperatures, the circumferential load due to rotation creates a greater stress range than those of the rim. Thus, higher material strength is needed in this region in order to accommodate those stresses, whereas the increased temperatures observed in the rim area require a larger grain size distribution which would provide much better creep and fatigue crack propagation performance.

4.1.2 Modelling the Uncertainty behind the Grain Growth Process

The way in which the framework was applied to the grain growth model (which includes both the DReX model as well as the grain growth during solution heat treatment) can be seen below in Figure 35 [5]:

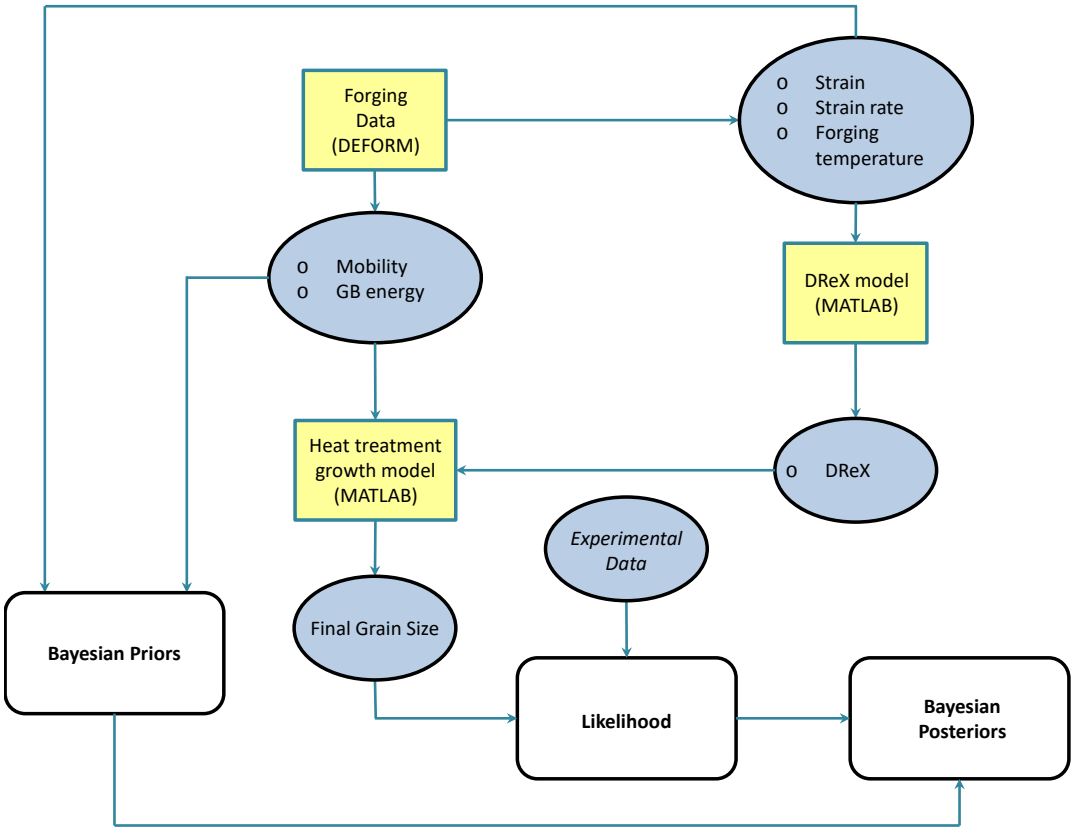


Figure 35: Flowchart of the Bayesian updating for the grain growth case study [85], [86]

As a reminder, the type of the MCMC algorithm used for finding the posterior distribution for this problem is based upon the multivariate component-wise updating algorithm similar to the one presented in section 3.2. The steps followed by the methodology illustrated in Figure 35 are described below:

- There are five different prior distributions corresponding to five inputs that are obtained from forging data.
- Three of those (namely strain, strain rate and forging temperature are fed into the DReX model in order to find the DReX grain size after the isothermal forging process).
- The other two inputs (mobility and grain boundary energy) as well as the DReX are used in order to compute the final grain size after the heat treatment process.
- The likelihood function is based upon experimental data (which is based on approximately 1000 experiments) and together with the five priors mentioned above, it is used to find the posterior distributions for strain, strain rate, forging temperature, mobility, grain boundary energy and final grain size.

The plots showing the posterior distributions of the above-mentioned parameters have been obtained by fitting PDFs on the histograms representing posterior distributions. Here, the priors and posteriors for the strain levels as well as grain size are shown in order to compare them.

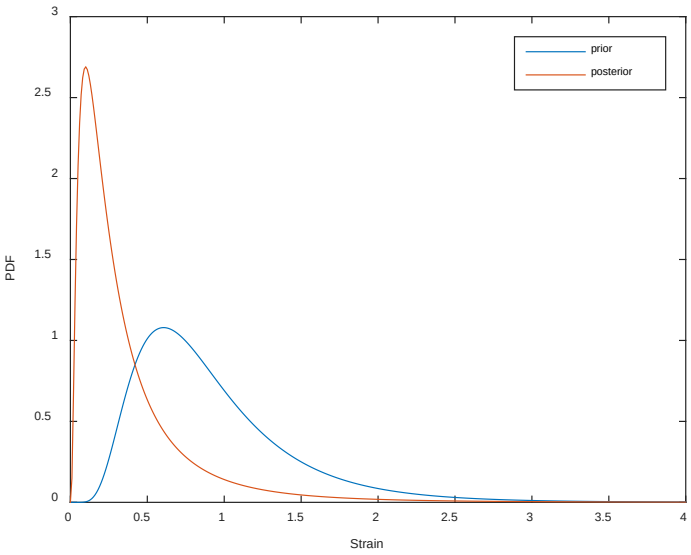


Figure 36: Prior and posterior of strain distribution during the forging process

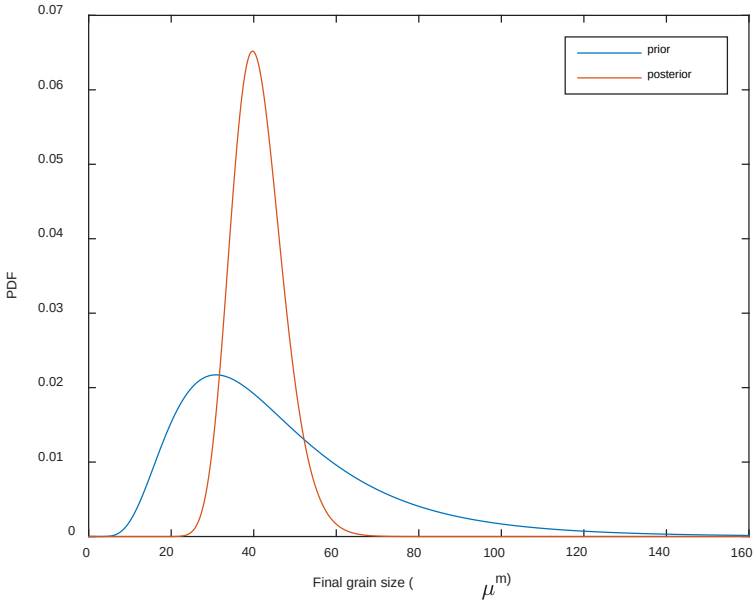


Figure 37: Prior and posterior of the grain size distribution after the heat treatment process

The plots from Figure 36 and Figure 37 show the differences between the prior and posterior distributions obtained after real experimental data had been input to the model, and since the relationship between all the parameters was not at all evident, the posteriors were computed using the MCMC approach described in section 2.3. After the posteriors have been obtained, further operations could be done on them. For instance, in order to ensure a consistent and large number of cycles to failure (which is the low cycle fatigue), it is necessary to specify a range for the final grain size (whose posterior is shown in Figure 37). Therefore, by fixing the range for the grain size, it is possible to work backwards in order to check which values of strain, strain rate as well as temperature should be used during the manufacturing process so that the corresponding grain size, and hence service life should be satisfied.

This procedure is illustrated in the following example. It can be assumed that the manufacturer does not have an idea on which temperature from a typical range is the best in order to obtain an optimum grain size, so in the preliminary stage they might have no preference over any temperatures, for instance between the Nickel superalloy forging temperatures (1055℃ up to 1155℃). However, by using the Bayesian techniques shown above, the designer can compute the posterior distribution of the grain size which is shown in Figure 38:

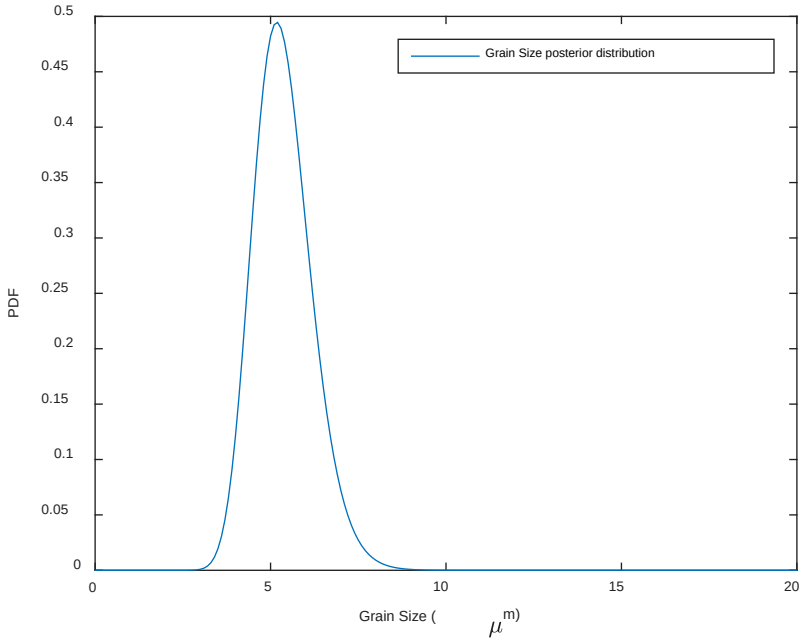


Figure 38: Posterior distribution of the grain size

Finally, the PDF for the forging temperature can be computed by constraining the grain size between 3 and 5 microns. The resulting posterior distribution for the temperature together with the designer's initial belief (which was a uniform distribution of temperature) are shown below in Figure 39.

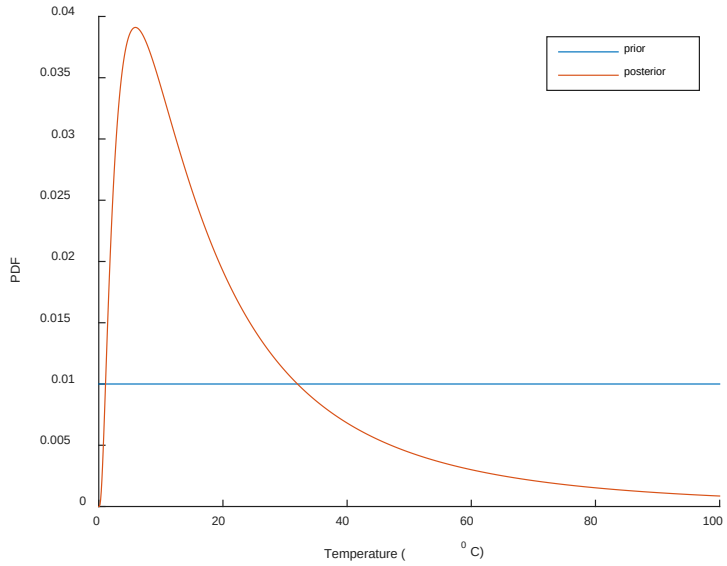


Figure 39: Prior and posterior of the temperature difference between the actual forging temperature and the lower bound of the uniform distribution (which is 1055 Celsius).

Note that what is shown in Figure 39 is not actually the forging temperature itself, but the difference between the forging temperature and the lower bound of the

uniform temperature distribution, which is 1055℃, as it is mentioned above. Therefore, in order to obtain the actual temperature used in forging, all there needs to be done is to add 1055℃ to the posterior from Figure 39 above. In this way, the engineer can control the final grain size to be within the specified range of 3 to 5 microns. The ultimate goal of this is that the stress analyst should be able to get a much better estimate of how many flight cycles the engine can withstand safely in operation and avoid having to remove it from service too early based on location-specific material properties which have grain size as their main independent variable.

The purpose of this case study was mainly to develop the core Bayesian algorithm which is integrated within the Bayesian-elicitation framework. The way in which this was done was by trying to solve a real-life problem that deals with uncertainties regarding the service life of a turbine disc. By investigating the grain size distribution at the end of the manufacturing process, and by using the framework proposed, the designer can potentially be able to make predictions regarding the life of the component. The expert elicitation component can potentially be implemented if a new material is planned to be used for the turbine disc for which no prior data is available.

The second case study which applies the probabilistic framework to a real-life scenario is illustrated in the following section and it involves modelling fan blade-off events. As it was mentioned at the beginning of Chapter 4, this case study will make use of the updating technique developed in this section, while also focusing on expert judgements.

4.2 Investigation of Rare In Service Fan Blade Off (FBO) Events

At the point when the Fan Blade Off (FBO) events became part of the Engine Certification Requirements for turbofan engines, there was also a need to provide evidence showing that the "worst case" FBO event would be safe. This makes use of the assumption that the heaviest fan blade is released at Low Pressure (LP) redline speed (which is also known as the maximum operational limit revolutions per minute (RPM); generally greater than 100%) from its outermost retention feature into a Fan Containment Casing which is manufactured to its minimum dimensions. This is considered to be the worst Fan Blade containment event. If the engine can withstand the worst case FBO event, it is reasonable to assume that it can also withstand lesser such events.

However, a further Certification Requirement needed aero-engine companies to prove that the flight phase following an FBO event is safe. The reason for that is because once a fan blade detaches from the fan, a phenomenon known as "windmilling" starts occurring. In broad terms, when the engine stops working, the aircraft continues to fly and the incoming flow of air will cause the fan to rotate and because the axisymmetry is lost, there will be large out of balance forces and moments acting on the fan. This also has the effect of reducing nominal clearances between the blade and the casing, and consequently, different types of interactions between the rotor and the stator start appearing which are related to friction and impacting [87].

A model has been created in MATLAB which is able to simulate any number of random FBO events and then assess their effect on the released and trailing blades and can also help quantify the effect of the windmilling process. The principal outputs of interest which are given by the model refer to the characteristics of the most likely FBO events that can be expected to occur in service. It is at this point that the Bayesian-elicitation framework shines: it can be integrated into the model in order to obtain more realistic outputs that are then used to do more efficient physical FEA simulations. This process is illustrated below in a visual fashion:

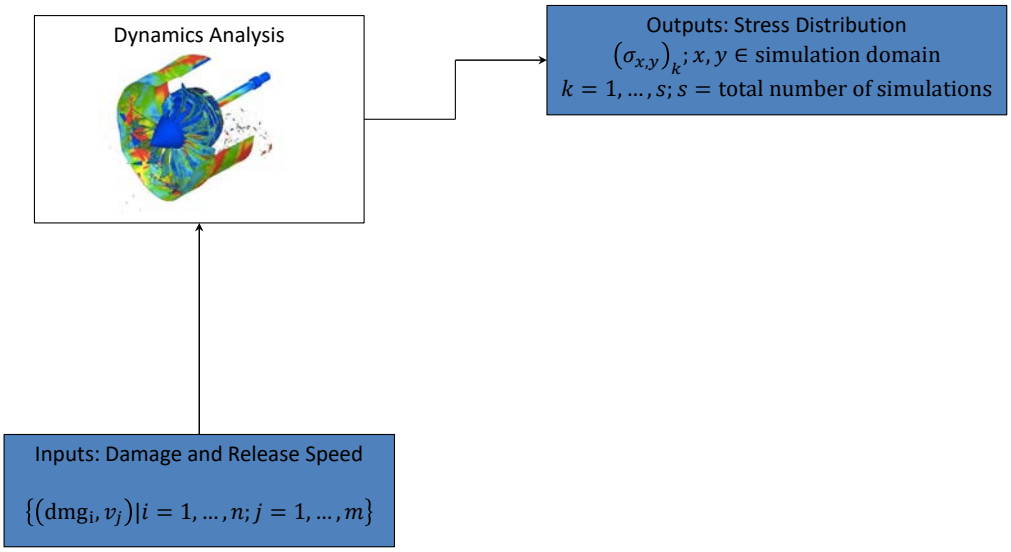


Figure 40: Obtaining physical outputs from damage and release speed.

One physical parameter of interest to the engineer is the stress distribution over the fan, as this can provide crucial information regarding whether or not the subsystem can withstand windmilling subsequent to the FBO event. In order to analyse an FBO event, information regarding damage and release speed (those will be explained in detail further in this chapter) need to be provided. However, there is a chance that the designer has little information on which FBO events are most likely to occur, hence the number of FBO events that will be simulated can potentially be massive (around 2000-3000). Figure 40 illustrates this by considering $n * m$ pairs of damage and release speed as the inputs. Therefore, if this product is complex, there needs to be an equally large amount of FEA simulations performed, which can end up being costly. One way to circumvent this issue would be to use the Bayesian-elicitation framework. By using this framework, Figure 40 will change in the following way:

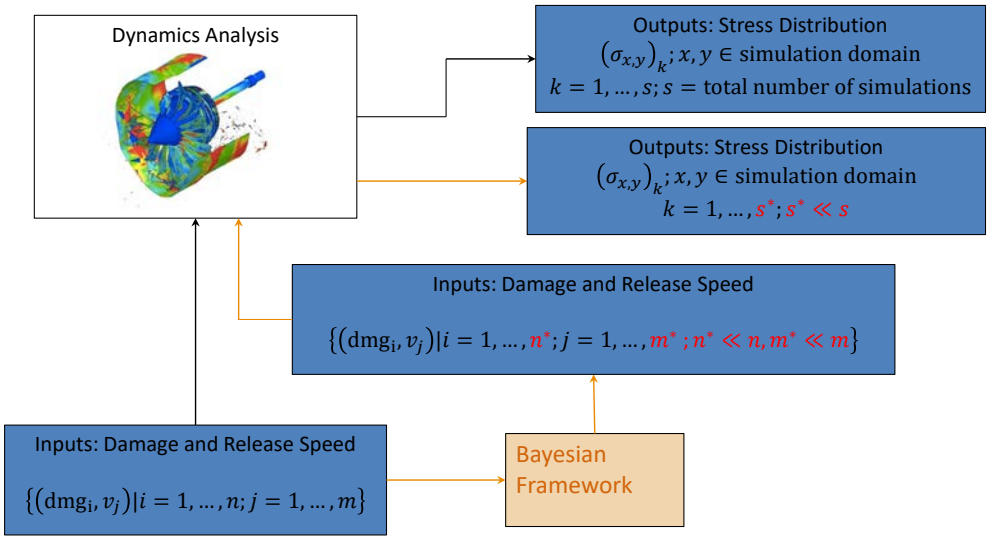


Figure 41: Obtaining physical outputs from damage and release speed (after the update).

The main consequence of using the framework is that a higher level of understanding can be gained regarding the characteristics of the most likely FBO events. This will not only result in realistic scenarios being analysed further with FEA, but also as a much lower number of simulations need to be done, the total cost of the detailed analysis stage will decrease. In order to understand this better, it would be useful to first illustrate the outline of the FBO model, and this is the subject of the following section.

4.2.1 Overview of the FBO Model

Several inputs have been identified as needing to be represented in the form of PDFs [1], meaning that uncertainty surrounding them has to be quantified. Those inputs refer to features having to do with three different aspects of the FBO event: the initial released blade, the trailing blade and the windmill. The inputs associated to those aspects are listed below. Before going into detail, it is worth explaining the notation used subsequently in this section. The expression $P(X = x|Y = y)$ represents the conditional probability for the random variable X to take the value x given that the random variable Y has the value y. Conditional probabilities of this type make up for a large proportion of the number of inputs used for the FBO event modelling. As a result, the list with all the different inputs appearing is presented below.

Initial Released Blade

- INPUT: Causes of failure: There are 8 different causes of FBO events identified, and their names (similarly to all of the other inputs to the model) are based on expert judgements. The acronyms used in some of them are as follows:
    - Foreign object damage (FOD)
    - High Cycle Fatigue (HCF)
    - Low Cycle Fatigue (LCF)

- INPUT: Probability of being at a certain speed given that a blade is released:

$$P\big(v = v_j|F = 1\big)$$

The notation $F = 1$ emphasizes the binary nature of a potential FBO event; it either happens (in which case $F = 1$), or it does not (therefore $F = 0$). However, as it will shortly become evident, the model is only concerned with the phenomena which occur during an FBO event, meaning that in all cases from now on, the binary variable F will always be assigned the value of 1.

- INPUT: Probability of having a specific failure cause given a certain speed and the fact that failure happened:

$$P\big(c = c_k | v = v_j, F = 1\big)$$

- INPUT: Percentage of the released blade (known as the phase 1 damage) given a specific failure cause, a certain speed and the fact that failure happened:

$$P\big(\mathrm{dmg} = \mathrm{dmg}_i | c = c_k,\ v = v_j, F = 1\big)$$

Trailing Blade

- INPUT: Percentage of damage to the trailing blade given the initial amount of released blade. The cumulative damage of the released and the trailing blade will be referred to from now on as the phase 2 damage.

$$P\big(\mathrm{dmg}_t = \mathrm{dmg}_{t,l} | \mathrm{dmg} = \mathrm{dmg}_i\big)$$

The Windmill

- INPUT: Inclusion of "Other" effects. Those effects are actually taking place during the windmilling phase which happens after the release of the trailing blade. The time scale is also worth taking into account. Although the initial and trailing blades can get released in a matter of seconds, a longer period of time (of the order of minutes) might pass until windmilling starts. There are several phenomena associated to the windmill which have an effect on the overall amount of "out-of-balance" (OOB) associated with the fan. The OOB term which the model refers to, is directly linked to the amount of damage which in reality, is the same as the percentage of the blade which gets released. The difference between the OOB and the amount of damage having to do with it, stands in the fact that the OOB is considered to be a vector defined by two features: a magnitude which is exactly equal to the amount of damage and an orientation which is defined by an angle θ with respect to the initial released blade. This idea is illustrated in Figure 42 below:

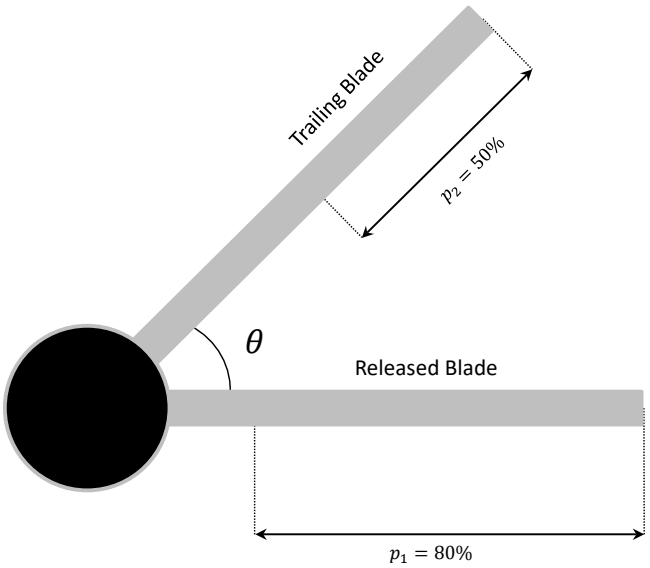


Figure 42: Visual representation of the released and trailing blades from the fan

Figure 42 shows the configuration of two blades from a fan, which are at an angle θ with respect to each other. The values p_1 and p_2 represent the percentages of the blades which are released after the FBO event. Their values are also equal to the magnitudes of the OOB vectors which can be added in order to find the overall level of OOB in a manner similarly to the one shown below in Figure 43:

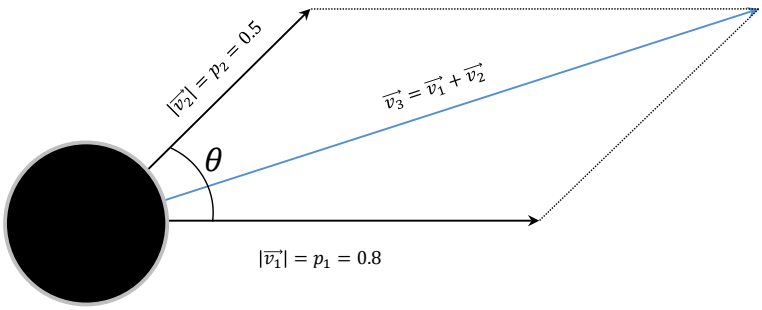


Figure 43: Out-of-balance levels represented as vectors

The way in which the final OOB vector v_3 is obtained is by doing a simple vector addition, where v_1 is the OOB vector associated to the initial released blade, while v_2 corresponds to the trailing blade. A physical explanation regarding the reason why the OOB levels are best described as vectors can be associated with the fact that the damage done to a particular blade (which is the same as that percentage of the blade which is missing) is equivalent to "missing mass", and it is this particular feature that is represented by a vector which has the same direction as the particular blade that had suffered the damage. The resulting OOB due to different damaged blades therefore needs to be obtained not only by adding up the amounts of damage for each blade, but also by taking into account the angles with respect to the initial released blade at which those other fan blades are situated. It is also important to mention that the fact that the blades break in the manner described in Figure 42 and Figure 43 is just an assumption; in reality there is no constraint which forces the blade to break off in this fashion.

- The effects associated to the windmill are input by the experts as well as their effect on the overall OOB and the phase. The phase is considered to be the angle with respect to the initial released blade where the OOB corresponding to any of the "other" effects occurs. A simple example which illustrates this idea is as follows: one of the damages that happen during windmilling is secondary blade damage. Physically, it can be described as what takes place when the rest of the fan starts impacting the inner casing: a certain percentage of the blades that come in contact with it may get released, and those blades are obviously at an angle with respect to the blade which gets released first when the FBO events starts. This angle is the actual phase and its possible values (or rather its distribution) are elicited from experts. The amount of secondary blade damage (or the damage due to any effect in general) is also in the form of a PDF which is elicited from experts. In mathematical form, the two inputs described above are as follows:

$$P\left(\mathrm{dmg}_o = \mathrm{dmg}_{o,m} | e = e_n, F = 1\right)$$

This is the probability for the amount of damage due to the other effects to be equal to $\mathrm{dmg_{o,m}}$ given windmilling effect $\mathrm{e_n}$ and obviously given that failure occurred. The probability for the phase to be equal to $\alpha_{\mathrm{o,m}}$ given windmilling effect $\mathrm{e_n}$ and failure has the following expression:

$$P(\theta_o = \theta_{o,m} | e = e_n, F = 1)$$

- The actual level of OOB does not only depend on the particular windmilling effect but also on the speed of blade release as well as on the cumulative damage done to the released and the trailing blades prior to windmilling (also known as the phase 2 damage). The dependence on those two inputs is expressed in terms of factors (which are essentially numbers between 0 and 1) which are also elicited from the experts. The way in which this works is in the following way. Assuming that secondary blade damage has a certain PDF associated to it and a sample was taken from that particular PDF which for illustration purposes is considered to be equal to 55%. In other words, the amount of damage (which is linked to OOB) due to secondary blade damage for a particular FBO event might be equal to 55%, however this does not yet take into account the release speed nor how much phase 2 damage exists in the first place. Instead, the particular value sampled from the PDF (55% in this case) is considered to be the worst case scenario which is associated to the maximum release speed and the maximum phase 2 damage, and consequently the corresponding release speed and OOB factors for this particular FBO event are both equal to 1. In order to find the final value of OOB, the sample from the PDF needs to be multiplied by both of the factors described above. It therefore becomes clear that for release speeds and phase 2 damage levels lower than the maximum, the factors take non-negative values below 1, which in the end will cause the sample (55%) to have a lower value.

- The reason why the phase is elicited as well is because the amount of OOB is considered to be equal to the magnitude of the associated OOB vector (equivalently, it can be considered to be a complex number). Therefore, the real part of the OOB level (found after the

factorization process outlined above) is multiplied by the cosine of the angle found during step 7 after sampling from the phase distribution. Similarly, the imaginary part is found by multiplying the same OOB level by the sine of the same angle. The reason this is done is because the cumulative damage of the released and trailing blade damage (the phase 2 damage) is also considered to be a complex number with a real and an imaginary part. As a result, in order to add the OOB of the windmilling effects to the phase 2 damage, the addition must be done separately for the real parts and the imaginary parts respectively. This particular step has been explained in more detail in step 6 above and in Figure 42 and Figure 43.

The way in which the elicited judgements are input to the model also needs to be taken into consideration. A Microsoft Excel spreadsheet has been set up which uses the elicited quantities associated to the three types of inputs described above (released blade, trailing blade and windmilling) and propagates them through the MATLAB model in order to obtain outputs which will be presented later in this section. Although Table 4 contains inputs regarding the released blade, it should be pointed out that the numbers are randomly generated and do not represent any Rolls-Royce plc actual data or trends in any way. Instead, they were simply chosen for illustrative purposes in order to emphasize the methodology behind the actual FBO model.

Table 4: Variables related to the released blade.

	100%RLNL				95%RLNL -> 75%RLNL			
CAUSE of Failure	Prob	NLT	ML	NGT	Prob	NLT	ML	NGT
Cause 1	10	40	60	80	p1	q11	q12	q13
Cause 2	9	50	70	90	p2	q21	q22	q23
Cause 3	11	60	65	70	p3	q31	q23	q33
Cause 4	12	50	70	90	p4	q41	q24	q34
Cause 5	3	45	55	65	p5	q51	q25	q35
Cause 6	9	60	65	70	p6	q61	q26	q36
Cause 7	14	70	80	90	p7	q71	q27	q37
Cause 8	6	63	65	67	p8	q81	q28	q38
Speed Probability	17				83			

At this point, it should be mentioned that during its initial stage, the model used triangular distributions in order to quantify uncertainty regarding the parameters in Table 4. The way in which this was done was by considering the Not Less Than (NLT) and Not Greater Than (NGT) to be the lower and upper bounds of the possible values taken by each variable, while Most Likely (ML) represented the peak of the

triangle (statistically known as the mode). However, recent work on elicitation has shown that using triangular distributions is not preferable due to two main reasons. First of all, the elicitor fails to elicit anything meaningful other than a range and a mode. More importantly, there is virtually no engineering quantity that can be accurately represented by a triangular distribution [88] (as the function itself is not smooth as it is made of straight line segments).

By taking into account suggestions from elicitation experts after following a detailed literature survey, it was decided that beta distributions should be used in order to quantify the potential amounts of damage done to the released blade. Some of the benefits of using beta distribution have been described previously in section 2.4.4. In the original input spreadsheet, there are 7 large columns, each of them corresponding to a value of rotational speed as a percentage of the redline speed in multiples of 5%, starting from 100% down to 70%. The redline speed (abbreviated as RLNL) is considered to be that particular speed level above which, the operation of the engine could become critical for the different components within it. For simplicity, in this case study, the redline speed is simply considered to be the maximum rotational speed that the fan can attain. Also, in this example, only the 100% redline speed column has been filled in with numbers, while all the other ones have been combined into a single one. The reason for this is due to periodicity; each speed columns is filled in exactly the same way as the 100% RLNL one, the only difference being the actual numbers used as inputs. The column denoted with "95% RLNL -> 70%RLNL" shows the generic structure according to which all speed columns are filled in. The smaller column titled "Probability" contains conditional probabilities which had been described in more detail above, while the other three columns (NLT, ML, NGT) represent the quantiles of the beta distribution which corresponds to each of the 8 causes listed above. To be more precise, NLT refers to the 1% quantile, ML refers to the 50% quantile, while NGT refers to the 99% quantile. The way in which those quantiles have been used in order to create the corresponding beta distributions was described in detail in Chapter 3.

The general methodology which forms the foundation of the FBO model is going to be explained by describing the outputs of interest. Those outputs are closely linked to the inputs shown at the beginning of section 4.2.1:

- A joint PDF as a function of both the phase 1 damage (ranging from 0% to 110%) and speed has to be plotted. The reason for that is because in order to obtain all the other outputs, a number of samples will have to be drawn from this particular distribution. Its actual shape will be shown later.
- A joint PDF as a function of both the phase 2 damage (ranging from 0% to 200%) and speed has to be plotted next. As a reminder, if the phase 1 damage is the damage done to the initial released blade, the phase 2 damage is the cumulative damage to both the released and the trailing blades. Samples from this distribution are then drawn in order to construct two more outputs.
- A joint PDF as a function of both the phase 3 damage (ranging from 0% to 400%) and speed is plotted next. Once again, the phase 3 damage is the cumulative damage which includes the phase 2 as well as the damage due to the other effects occurring during windmilling. This is one of the two final output plots.

The steps which have to be taken in order to produce those outputs are outlined as follows:

1) The first step of the algorithm involves plotting the joint PDF of the phase 1 damage and speed. In order to do so, the values in Table 4 are used along with the Bayesian rule in order to find probabilities of having a specific *damage* and a specific *speed* given failure. This uses directly the inputs listed at the beginning of this section and it is expressed as follows:

$$P\big(\mathrm{dmg} = \mathrm{dmg}_i, v = v_j | F = 1\big) = \frac{P\big(\mathrm{dmg} = \mathrm{dmg}_i, v = v_j, F = 1\big)}{P(F = 1)} \tag{56}$$

$$P\big(\mathrm{dmg} = \mathrm{dmg}_i, v = v_j | F = 1\big) = \frac{\sum_{k=1}^{n} P\big(\mathrm{dmg} = \mathrm{dmg}_i, c = c_k, v = v_j, F = 1\big)}{P(F = 1)} \tag{57}$$

$$P(\mathrm{dmg} = \mathrm{dmg}_i, v = v_j | F = 1)$$
$$= \frac{\sum_{k=1}^{n} P(\mathrm{dmg} = \mathrm{dmg}_i | c = c_k, v = v_j, F = 1)\, P(c = c_k | v = v_j, F = 1) P(v = v_j | F = 1) P(F = 1)}{P(F = 1)} \quad (58)$$

$$P(\mathrm{dmg} = \mathrm{dmg}_i, v = v_j | F = 1) = \sum_{k=1}^{n} P(\mathrm{dmg} = \mathrm{dmg}_i | c = c_k, v = v_j, F = 1)\, P(c = c_k | v = v_j, F = 1) P(v = v_j | F = 1) \quad (59)$$

All quantities in Equation (59) above are directly found using the values from Table 4. Also, although the values for speed in Table 4 are multiples of 5 ranging from 70% to 100% redline speed, Equation (59) was applied for all integer speed percentages in this range (such as 72%, 83%, 98% for instance). The way in which this was done was as follows: for each speed value v_j from Table 4 (which is not 70% or 100%), the exact same set of quantiles which applies to that particular speed, is also used for all integer speed values in the interval $[v_j - 2\%, v_j + 2\%]$. If v_j takes the value of either of the extreme speed values, then its characteristic set of quantiles is used for integer speed values in the interval $[v_j, v_j + 2\%]$ (if v_j is equal to 70%) or $[v_j - 2\%, v_j]$ (if v_j is equal to 100%). The reason this was done was because it was considered that there is not enough information to justify using other values of quantiles for intermediate speed values other than those which are present in Table 4.

In addition to that, Equation (59) was applied only to integer damage levels dmg_i ranging from 0% up to 110% (which once again is the range of phase 1 damage). The idea of only considering integer values was implemented as it had been decided that going beyond this and considering real values of damage as well would not give a significant increase in accuracy, while the computational time would see an increase. Therefore, by applying Equation (59) to all relevant speeds and damage levels, a joint PDF can be obtained which is used further in the analysis (Figure 44).

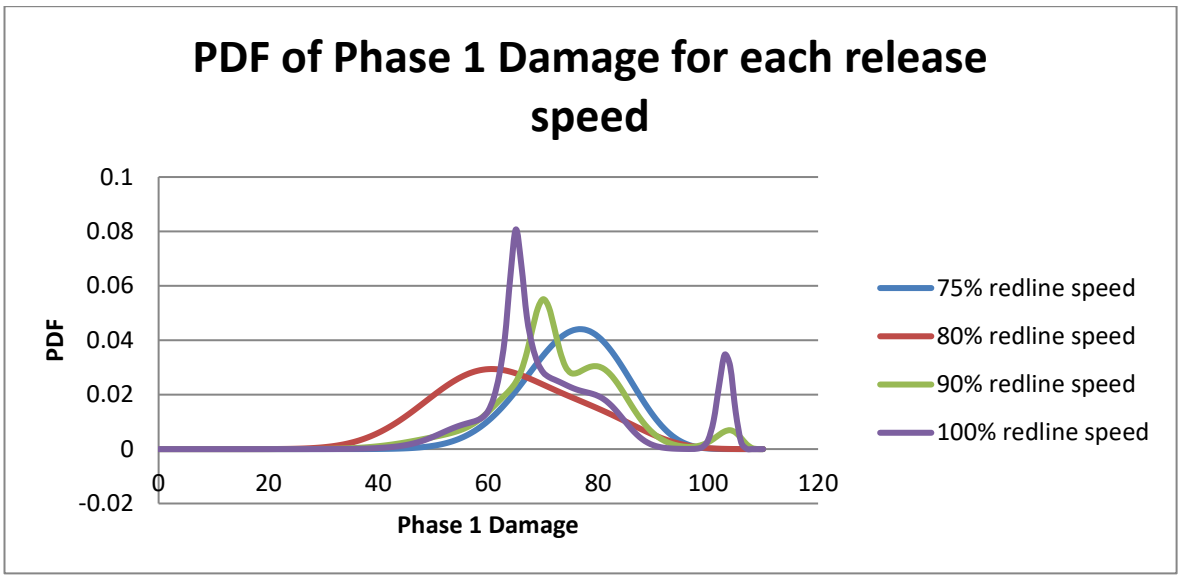


Figure 44: Joint PDF of Phase 1 Damage and Release Speed.

2) The joint PDF found in step 1 is used to perform a Monte Carlo (MC) analysis. The number of iterations of the MC algorithm coincides with the number of simulated FBO events. An FBO event can simply be considered to be a mathematical entity uniquely defined by two parameters: a value of the phase 1 damage as well as a speed. Obviously, in order to properly obtain any such event, there is the need to sample from the distribution obtained in step 1, which is a relatively simple task, seeing that the distribution itself is discrete. Therefore, the MC algorithm always starts with two variables: a value of the phase 1 damage as well as a value of speed.
3) Once the values of dmg_i and v_j have been sampled, the amount of damage done to the trailing blade $\mathrm{dmg}_{t,l}$ is calculated based on expert judgements. As before, the judgements from the experts are directly related to the quantiles of the beta distributions corresponding for each combination of phase 1 damage (as a multiple of 5) and speed (also as a multiple of 5). Therefore, the phase 2 damage is the cumulative effect of the phase 1 damage as well as the amount of damaged trailing blade. All values of the phase 2 damage obtained in this manner, as well as the original release speeds from all iterations of the MC algorithm are used in order to create a joint PDF (see Figure 45). Additionally, the value of the phase 2 damage after each iteration is propagated further in order to assess the effects of windmilling.

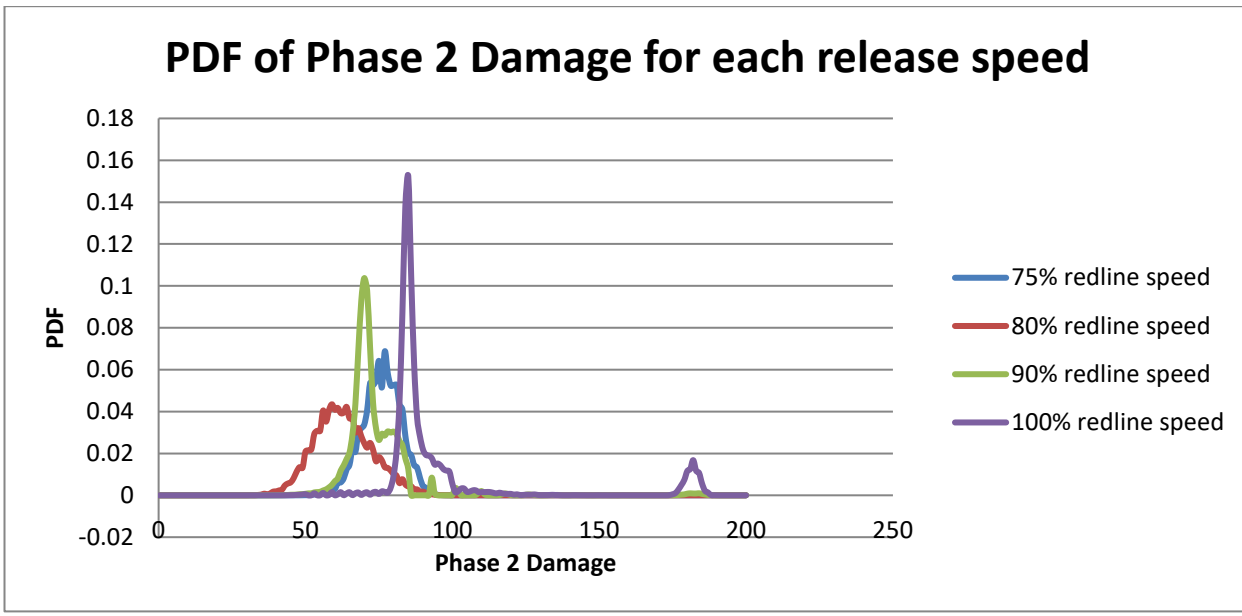


Figure 45: Joint PDF of Phase 2 Damage and Release Speed.

4) The cumulative phase 2 damage $\mathrm{dmg}_{t,l}$ and the release speed v_j are used (as it was described at the beginning of this section) in order to obtain a value of the damage (equivalent to the amount of unbalance) due to the windmilling effects. The name "phase 3" damage has been therefore given to the cumulative effect of phase 2 damage and the amount of damage due to windmilling. At this point, a third joint PDF of phase 3 damage and speed can be created (Figure 46).

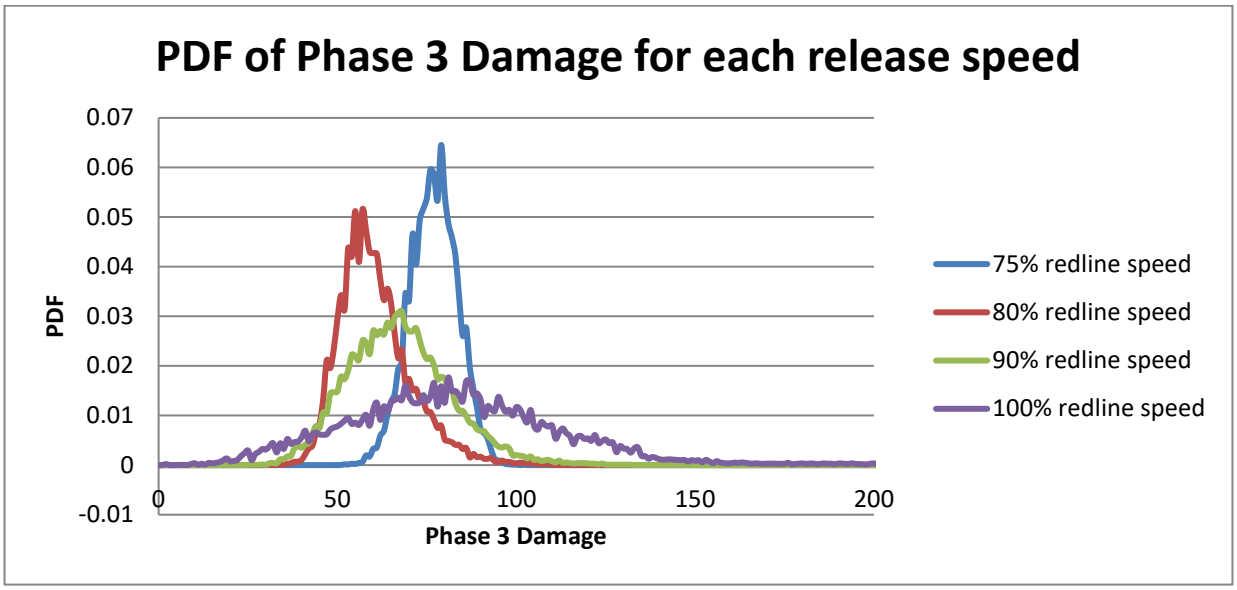


Figure 46: Joint PDF of Phase 3 Damage and Release Speed.

As it has already been mentioned, those outputs are based upon the aforementioned inputs, which in turn have been obtained from expert judgements. Those expert judgements have been assigned prior PDFs, which in a Bayesian context, are able to be updated via a likelihood function in order to obtain posterior PDFs which should depict reality in a more accurate fashion. In order to achieve this, the likelihood function should contain information regarding real events. The reader should note that at the moment, fictitious data is being used as the interest lies more in illustrating the concept of updating expert judgements with data, rather than showing what the real data is. Therefore, if the methodology is correct, the updating procedure could be done using any values for the observed data.

4.2.2 Updating Expert Judgements

The updating procedure has been done to four different inputs: the release speed, the cause of release, the amount of damage done to the blade as well as the phase (applicable only to the "other" effects). The priors as well as the likelihood functions for each of the parameters are shown below:

1) Each release speed is assigned a marginal prior distribution on which subsequent updating is performed [2]. Those priors are beta distributions; the reason being that each probability value is within the interval $[0,1]$ which coincides with the range of the beta distributions. However, it is worth noting that the marginal priors for each speed probability are NOT independent of one another. The easiest way to understand this is by recalling that the sum of all the speed probabilities has to be equal to 1. As a result, the joint prior distribution cannot be found by simply multiplying all the marginal PDFs. Fortunately, the correlation between the priors is taken care of if a Dirichlet distribution is used. One method in which multiple beta distributions are converted into a single Dirichlet distribution can be found in Vazquez [89]. In addition to that, all the speed probabilities are part of a categorical distribution (which can be thought of as a multi-dimensional Bernoulli distribution) which plays the role of a likelihood function. A table containing the quantiles for the speed probabilities is shown below in Table 5.

Table 5: Quantiles for the speed probabilities.

Release Speed (%RLNL)	70		75		80		85		90		95		100	
Quantiles	25%	50%	25%	50%	25%	50%	25%	50%	25%	50%	25%	50%	25%	50%
	0	0	3.2	4	6.4	8	6.4	8	21.6	27	28.8	36	13.6	17

The marginal beta distributions are first found by employing the algorithm shown by van Dorp [79], while the Dirichlet distribution is found by using the method in Appendix C. The updating procedure is easy to understand considering that the Dirichlet – categorical distributions form a conjugate pair [86], meaning that the posterior has a Dirichlet distribution as well. The prior and the likelihood function used to find the posterior are shown below in Equation (60) and Equation (61):

$$\pi(p_1, p_2, \dots, p_k) \sim \mathrm{Di}(\alpha_1, \alpha_2, \dots, \alpha_k) \tag{60}$$

$$f(x_1, x_2, \dots, x_k | p_1, p_2, \dots, p_k) \sim \mathrm{Cat}(p_1, p_2, \dots, p_k) \tag{61}$$

The parameters from Equations (60) and (61) have the following meanings: p_i stands for the i^{th} speed probability, α_i is the i^{th} parameter for the joint Dirichlet distribution, and x_i represents the number of FBO events which occurred at release speed v_i. After using Equation (1), the posterior, which also has a Dirichlet distribution due to conjugacy, can be easily found to have the following parameters:

$$\pi(p_1, p_2, \dots, p_k | x_1, x_2, \dots, x_k) \sim \mathrm{Di}(\alpha_1', \alpha_2', \dots, \alpha_k') \tag{62}$$

$$\alpha_i' = \alpha_i + x_i \tag{63}$$

It becomes obvious that each parameter from the posterior distribution is equal to the sum between the corresponding Dirichlet parameter of the prior distribution as well as the number of FBO events which occurred at the corresponding release speed v_i. Besides the posterior, another relevant quantity to consider is the posterior predictive distribution. In short, given the updated posterior for the speed probabilities, it would be necessary to know what the likelihood function for a new data point looks like. In a more intuitive sense, asking for the posterior predictive distribution is the same as asking "How does the model think the distribution of the data looks like, given the shape of the posterior?". The posterior predictive can generally be found using the following equation:

$$f(x_{k+1}|x_1, x_2, \dots, x_k) = \int_{\theta_0}^{\theta_n} f(x_{k+1}|\theta)\, \pi(\theta|x_1, x_2, \dots, x_k) \mathrm{d}\theta \tag{64}$$

In Equation (64), the general variable θ can be replaced with the parameter vector α used in the current case study. According to Johnson [90], the posterior predictive distribution of a Dirichlet-categorical conjugate pair is equal to the expected value of the posterior distribution. In other words, for the current problem, Equation (64) can be rewritten as:

$$f(x_{k+1}|x_1, x_2, \dots, x_k) = \frac{\alpha_{k+1} + x_{k+1}}{\sum_{i=1}^{k}(\alpha_i + x_i)} \tag{65}$$

The consequence of Equation (65) is that by obtaining the posterior predictive distribution using actual data, the quantiles in Table 5 are most likely going to change. In order to illustrate this idea, an example will be given shortly.

2) Each cause of release is also assigned a marginal beta prior distribution, while all causes for a particular speed are given a joint Dirichlet distribution. The likelihood function is still a categorical distribution, meaning that the same kind of conjugacy is achieved in this case as well. An example of a table showing some of the cause probabilities for the 95% redline speed is shown as follows:

Table 6: Quantiles for the cause probabilities at 95% redline speed

Cause of release index	**1**		**2**		**3**		**4**	
Quantiles	**25%**	**50%**	**25%**	**50%**	**25%**	**50%**	**25%**	**50%**
	10.4	**13**	**3.2**	**4**	**6.4**	**8**	**8.8**	**11**

The actual names for the causes can be seen in Table 4. Because of the same Dirichlet-categorical conjugate pair, both the posterior and the posterior predictive distributions have the same form as the ones shown in equations (62) and (65). The example shown shortly will illustrate how the values in Table 6 are going to be updated in light of new data.

3) The damage done to the released blade due to any specific cause is considered to be a percentage of the actual blade length. Knowing this, a scaled beta distribution is used as the likelihood function for the damage level on the released blade. The prior distribution on the other hand has been elicited from experts in an indirect fashion. To illustrate what is meant by this, it is first worth looking at the quantiles.

Table 7: Quantiles for the blade damage given the second failure cause and 95% redline speed.

	95% RLNL		
Quantiles	1%	50%	99%
	60	70	80

The reason why three quantiles (1%, 50% and 99%) had been elicited for the damage level was in order to achieve "overfitting". In short, it refers to eliciting more quantiles than it is needed in order to fit a distribution. In the case displayed in Table 7, where the fitted distribution is a scaled beta, there is exactly one more quantile than necessary in order to fit a beta distribution. Consequently, a reason for employing overfitting is to capture the experts' range regarding the distribution parameters involved. For the case in Table 7, using the 1% and 50% quantiles gives a beta distribution with parameters $\alpha = 101.795$ and $\beta = 58.3113$. The 50% and 99% quantiles give $\alpha = 87.89$ and $\beta = 50.37$, while using the 1% and 99% quantiles gives $\alpha = 95.089$ and $\beta = 53.696$. It now becomes clear that the three quantiles *do not* correspond to a single beta distribution. Instead, using the results above, it can be said that the intervals $\alpha \in [87.89, 101.795]$ and $\beta \in [50.37, 58.3113]$ quantify the expert's uncertainty regarding the distribution parameters. In this respect, two prior distributions have been constructed for each parameter shown above. Two normal distributions have been chosen for each of the beta parameters, the means and variances of which have been identified by using the technique found in Cook [77]. Also, the parameters of the beta distributions have been computed from quantiles using the algorithm from van Dorp [79].

The prior distribution and the likelihood function for the phases related to the "other" effects are also a normalized beta and a normal distribution respectively,

as it is the case for the damage levels. The reason for this similarity is because the phase has a similar structure: it is bound between 0 and 360 degrees, which is why it has been modelled using a scaled beta distribution (the likelihood), while the beta parameters have been assigned one normal distribution each (the prior).

Table 8: Quantiles for the phase and damage corresponding to secondary blade damage

	Damage	Angle
NLT	1	30
ML	60	120
NGT	105	345

The priors and likelihood functions for the parameters above are used in order to obtain updated posterior distributions as data becomes available. Below, an example is shown whose purpose is to illustrate the effect of real data on the FBO model inputs [2]. Figure 47 shows the initial speed probabilities from Table 5 by using a bar chart (once again, speed probability p_i refers to the probability that failure occurs at release speed v_i).

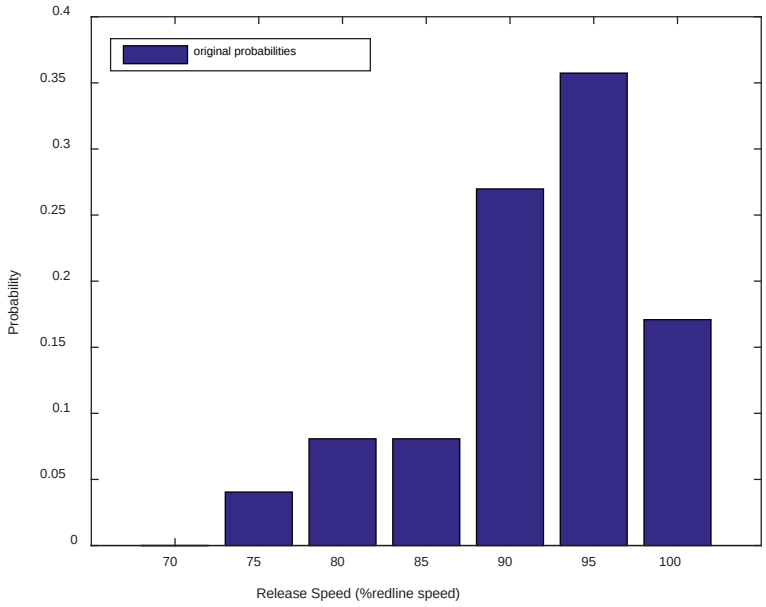


Figure 47: Original speed probabilities.

Assuming there have been seven different FBO events, each of them occurring at the following speeds: 90,90,95,95,90,85,90 (all of them are percentages of the redline speed), the goal is to investigate how the probabilities in Figure 47 change in light of new data. After applying the techniques described previously in this section, the posterior predictive distribution for the speed probabilities can be seen below in Figure 48:

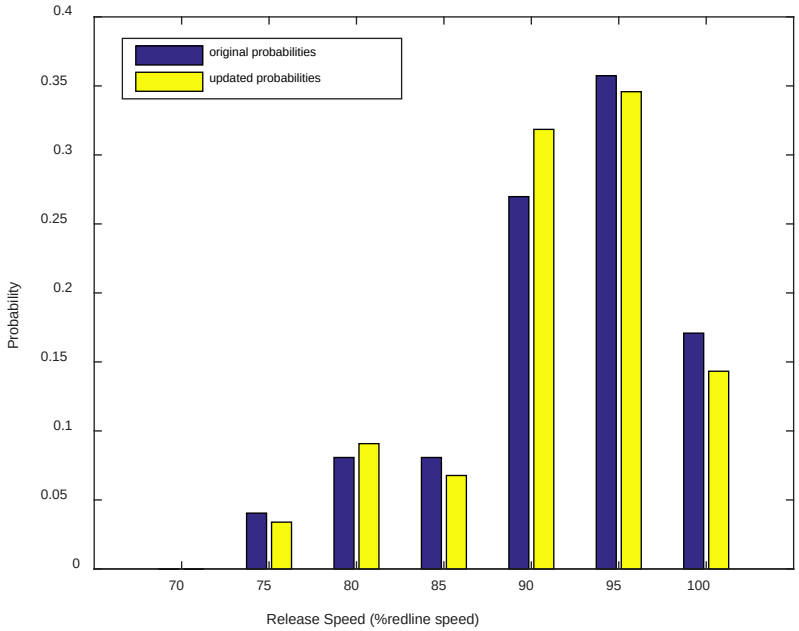


Figure 48: Updated speed probabilities

Using a Dirichlet distribution has the implicit effect of accounting for the correlations between all the speed probabilities. This can also be observed above in Figure 48: as the probability for certain speeds increases due to the addition of evidence, the other probabilities for which there is little or no data will decrease such that the overall sum remains equal to 1.

The cause probabilities at 95% redline speed before and after data has been added can be seen below in Figure 49:

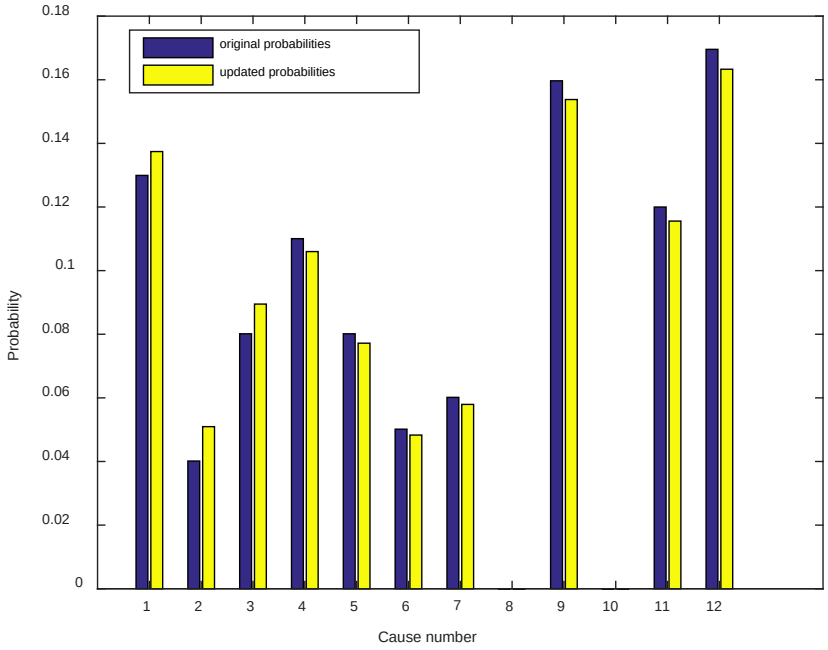


Figure 49: Updated cause probabilities for 95% redline speed.

The causes which have been input as evidence into the model are: bird ingestion, foreign object damage (FOD) as well as low cycle fatigue. The effect of updating can be observed in Figure 49; just as it is the case with Figure 48, some probability values increased while others decreased in order to keep the total sum equal to 1.

Finally, the effect of updating the damage level due to an FBO event at 95% redline speed which occurred due to FOD (cause number 2, as it is shown in Table 6) is shown below in Table 9:

Table 9: Updated quantiles for the damage level due to FOD at 95% redline speed

	95% RLNL		
Original	1%	50%	99%
	60	70	80
Updated	62.6	73.46	82.68

Also, the beta PDFs showing the original as well the updated damage quantiles can be seen below in Figure 50:

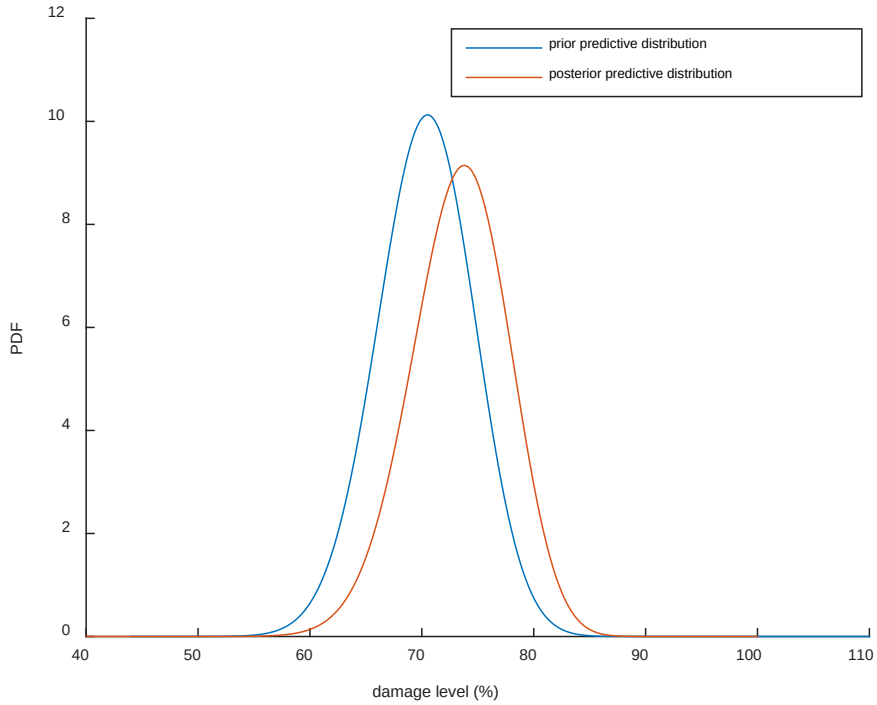


Figure 50: Comparison between the prior and the posterior predictive distributions.

The damage levels which were used as observations were 82%, 85%, 81%, 84%, and they are outside the original range expressed in Table 9. The resulting posterior predictive will therefore be represented by a beta distribution shifted to the right (or in general, in the direction where the mean of the observations lies). As one can imagine, introducing even more evidence will cause the posterior predictive to shift even more. In order to show this, seven more damage levels with values close to the four above were added in order to produce another similar graph.

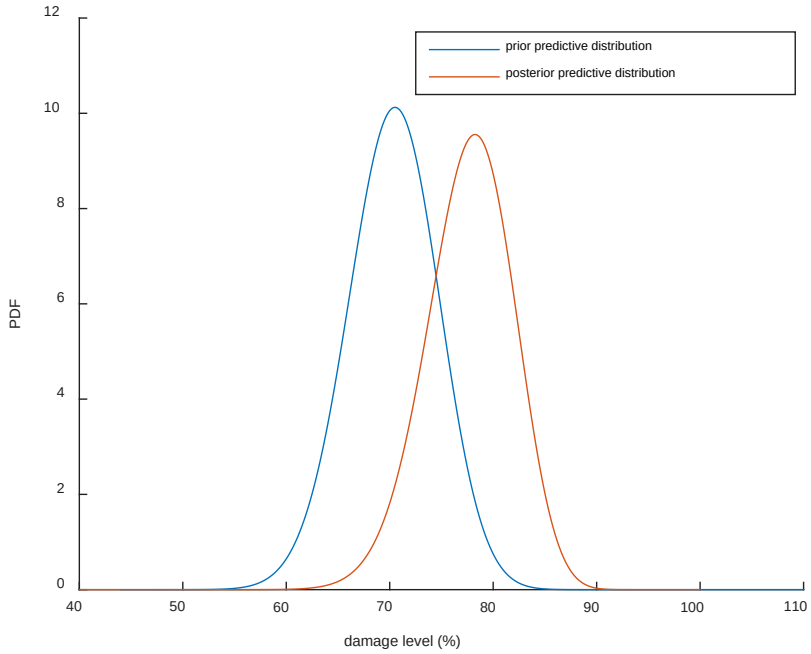


Figure 51: Comparison between the prior and the posterior predictive distributions (with more data added)

In order to analyse the effect that updating has on the outputs of interest, it is first worth noting what the data used looks like. The observed values can be seen below in Table 10 [4].

Table 10: Data used in order to update the expert judgements

Release Speed (%RLNL)	Release Cause	Released Blade Damage (%Total Blade)	Trailing Blade Damage (%Total Blade)
90	2	86	0
95	1	72	0
100	3	71	25
95	2	88	77
90	1	95	82
95	12	75	12

The joint prior distribution of the phase 1 damage and release speed from Figure 44 is shown once again below as a contour plot in order to compare it with its posterior predictive distribution in Figure 53.

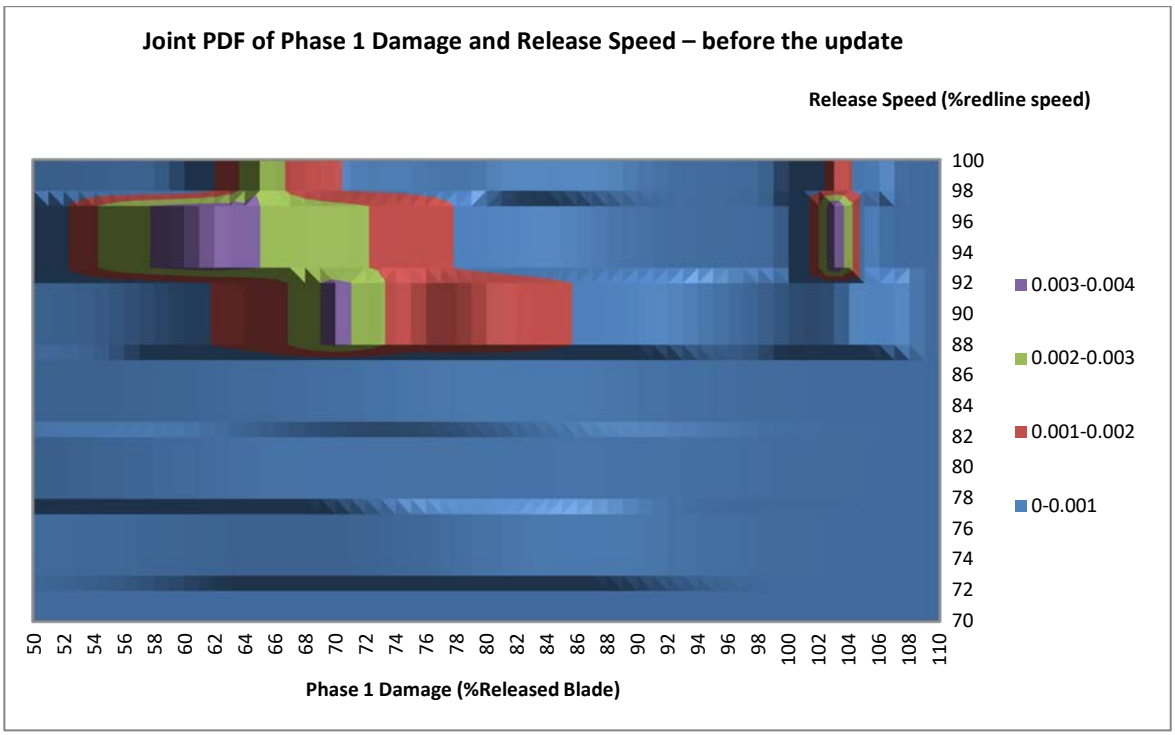


Figure 52: Prior predictive joint distribution for phase 1 damage and release speed

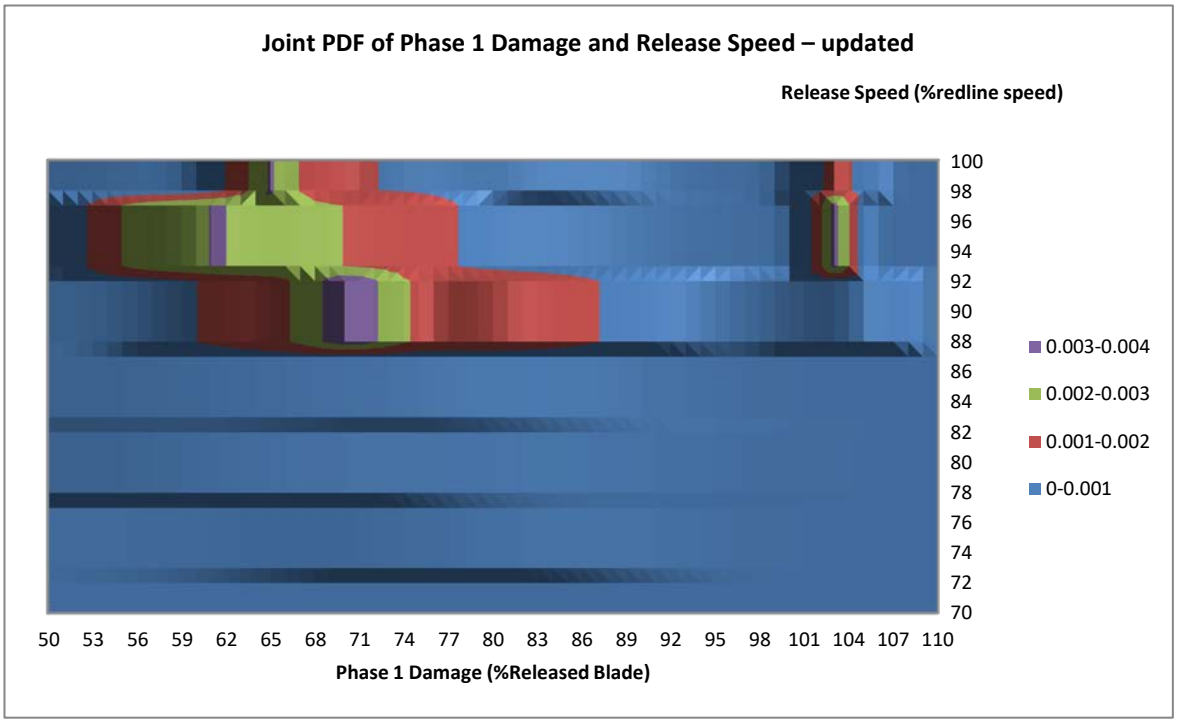


Figure 53: Posterior predictive joint distribution for phase 1 damage and release speed

Before comparing the priors and the posteriors, there is a particular aspect which should be brought to light, namely the fact that it might be suggested that despite

the evidence being added, there is still not much difference between Figure 52 and Figure 53. However, displaying a large difference is not actually the purpose of the exercise presented here. The actual goal is to create a flexible model which changes every time data becomes available in order to reflect both the original expert judgements as well as the data itself [4].

It can be seen that after the update, the peaks of the PDFs have changed in magnitude. Before the update, it can be inferred from Figure 52 that the most likely FBO events correspond to a damage range of 56% up to 65% of released blade damage and 95% release speed. Those are only the prior distributions as they are solely based upon expert judgements. After using the data in order to perform the update, the most likely FBO events become centred in the region corresponding to 68% up to 75% blade damage and 90% release speed. As it has been mentioned in the introduction, this information is going to be further used in order to perform a dynamics analysis of the fan in order to study the behaviour during the windmilling stage and design it such that the integrity of the engine during this phenomenon should be maintained. The joint prior and posterior distributions of phase 2 damage (which is simply the cumulative trailing blade and released blade damage) and speed of release have been created by using a Monte Carlo simulation with around 1,000,000 iterations. This simulation has been run three times and it was found that the inherent randomness had virtually no effect on the final shapes of the graphs which can be seen below in Figure 54 and Figure 55.

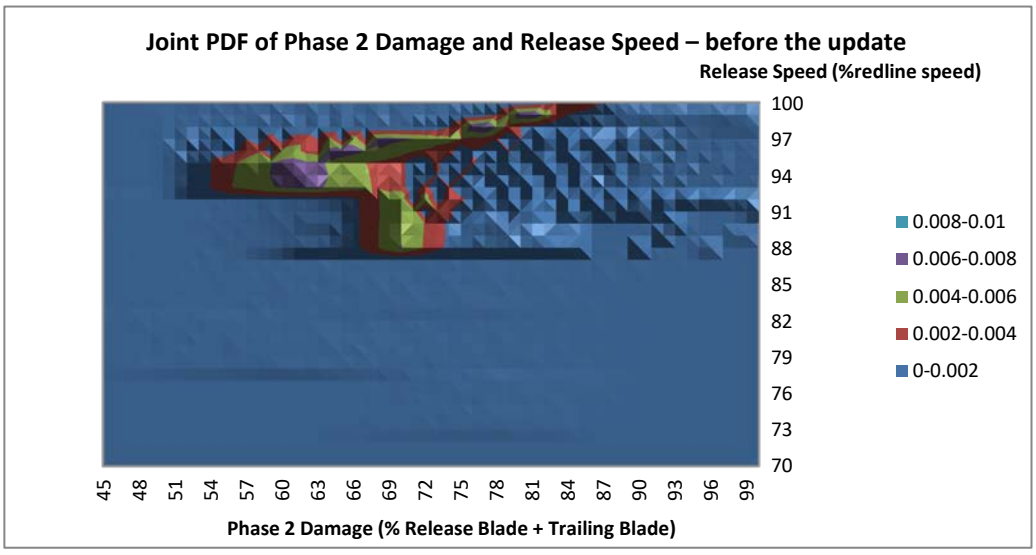


Figure 54: Prior predictive distribution of phase 2 damage and release speed

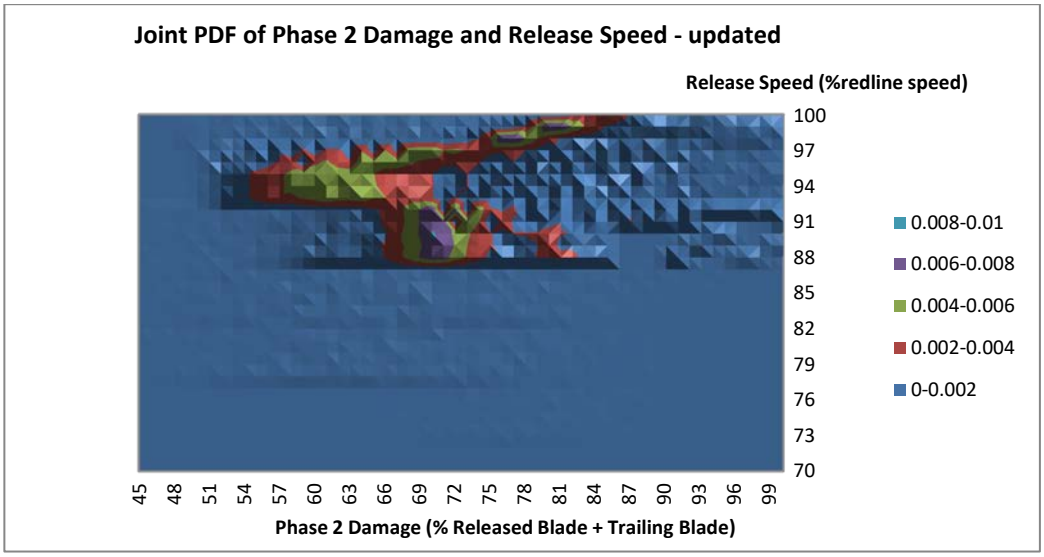


Figure 55: Posterior predictive distribution of phase 2 damage and release speed

Although Figure 54 and Figure 55 show the cumulative damage of both the released and trailing blades, it can be observed that the most likely FBO event correspond to the same combinations of phase 1 damage and release speed as it was the case in Figure 52 and Figure 53 (56% - 65% blade damage and 95% release speed before the update as well as 68% - 75% blade damage and 90% release speed after the update). This posterior updating needs to be taken into account for the purposes of the dynamics analysis of the fan which is going to benefit from a more accurate representation of the events which are considered to be the most likely.

The contour plots for the joint distribution of the phase 3 damage and release speed are shown below in Figure 56 and Figure 57. The main difference between the two is similar to the one seen in Figure 54 and Figure 55: while before the update the only region with a high probability corresponds to 50% - 65% phase 3 damage and 95% redline speed, after adding the data points, FBO events having 55% - 70% phase 3 damage and 90% redline speed become more probable than before, meaning that any subsequent vibration analysis of the fan should also be focused in the latter region.

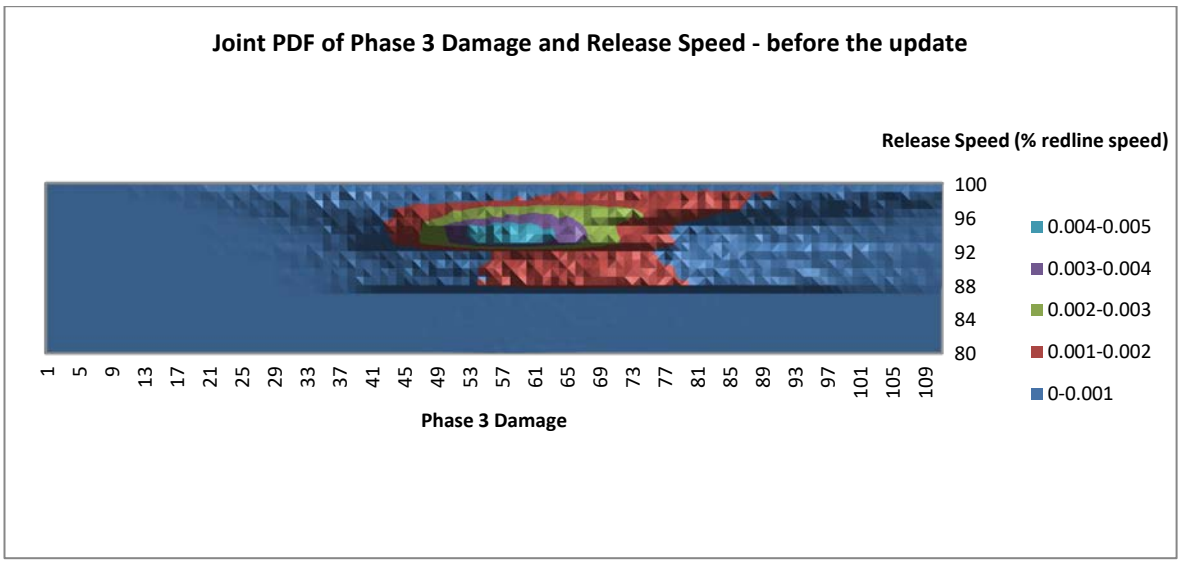


Figure 56: Prior predictive distribution of phase 3 damage and release speed

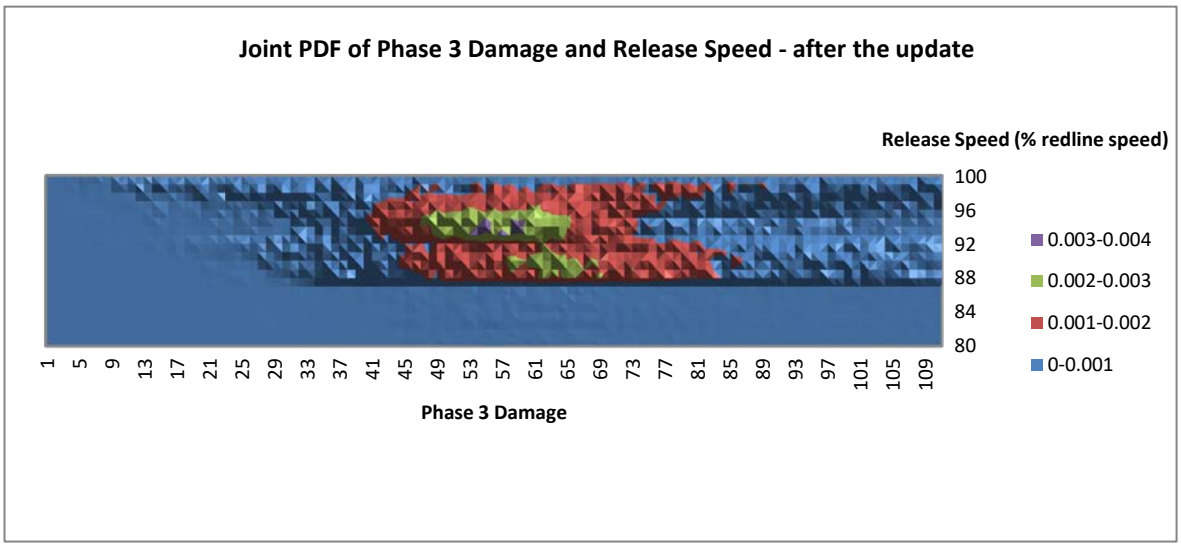


Figure 57: Posterior predictive distribution of phase 3 damage and release speed

In order to understand why this additional knowledge is beneficial, it is worth reconsidering the schematic from Figure 41. Namely, without using the proposed framework, the total number of FEA simulations which needs to be done can be calculated by examining Figure 57. Without knowing which FBO events are more likely, all possible combinations of damage and release speed are assigned an equal probability. In this case, one FEA simulation needs to be done for each pair of damage and release speed (dmg_i, v_j), where $i \in \{1,2,3,\dots,110\}$ and $j \in \{80,81,82,\dots,100\}$. Therefore, the total number of simulations is given as the product between the total number of elements from each set: $110 * 21 = 2310$ simulations. On the other hand, if only the most likely FBO events are considered, (for instance

the ones from Figure 57 for which the likelihood is 0.002 and above), then the number of (dmg_i, v_j) pairs reduces drastically: $i \in \{47,48,,49,\ldots,69\}$, and $j \in \{88,89,90,\ldots,96\}$, which gives a total number of simulations of approximately 115. Although those numbers can be reduced even further by employed a design of experiments methodology (other than the proposed full factorial presented above), it is already obvious that the framework reduces the number of simulations and hence the computational time by a factor of approximately 20.

The initial motivation for applying the Bayesian-elicitation framework for this case study was to facilitate further FEA undertaken either during or after the preliminary engine design stage. This analysis involves studying the forces and moments occurring in the fan subsequent to an FBO event. It would be extremely time consuming and impractical to test all the possible combinations of damage and speed of release, which is why it is necessary to only look at the ones which are the most likely to occur during an FBO event. The model presented here makes use of an expert system in order to achieve that, and it also uses data in order to make it dynamic in light of real observations.

As the first two case studies focused on the evolution of the framework during this EngD project and created the right environment for creating its two main components, the third case study is the one where the whole framework has been applied in order to tackle a problem that is of high interest. The final engineering problem looked into during this project is generally associated to low levels of knowledge regarding the physical parameters, especially during the preliminary stage of the design process, as it refers to quantifying the fatigue failure of the turbine disc due to Non-metallic particle inclusions. After a discussion about the available literature on the topic, it will be explained in detail how the framework will allow the designer to obtain useful information during the early stages of the aero-engine design process.

4.3 Case Study 3: Uncertainty Quantification of Fatigue Failure within a Turbine Disk

In order to comprehend how to effectively manage ageing turbine components, it is crucial to understand their fatigue behaviour. To be more precise, there are two main types of fatigue failure: crystallographic failure, as well as failure due to Non-metallic particle (NMP) inclusions. The latter is least understood and the goal of this section is to investigate it via the Bayesian-elicitation framework. More specifically, the goal is twofold: to be able to predict the life of the turbine disc under working conditions, as well as predict the cause of failure itself before doing any detailed analysis on the failed component. Understanding this phenomenon can also prevent the premature removal from service of the turbine disc in order to meet the safety standards. This in turn reduces costs, as it allows the turbine disc to potentially operate for a longer period of time without being in danger of approaching failure.

In the first part, some literature on fatigue failure will be explored in order to put this problem further into context. Afterwards, it will be shown why applying the current framework to the problem can give useful results for the aero-engine designer.

4.3.1 Theoretical Background

Generally, the variability in cycles to failure increases as the operating temperature increases and as operating stress level decreases. The former has been observed by Huron and Roth [91] who concluded that crack nucleation occurs exclusively from intrinsic non-metallic inclusions located under the surface of the material for Rene' 88 DT fatigue specimens being tested at 649°C. Conversely, at a lower temperature of around 204°C, a shift in mechanisms seemed to occur, where crack initiation occurred only from crystallographic facets. In addition to that, Caton et al. [92] illustrated the life variability of the same Rene' 88 DT material which can be seen in Figure 58. It seems that the divergence in the lives of the two mechanisms was caused by an increase in the difference between their crack nucleation lives as the stress level decreases.

The same behaviour has been identified for other materials such as an $\alpha + \beta$ titanium alloy as well as a nickel-based superalloy [93]. Moreover, there have been several other studies made on titanium and γ-TiAl [94], [95] for which mechanism based prediction methodology was proposed and was able to reduce the uncertainty in the total fatigue lifetime.

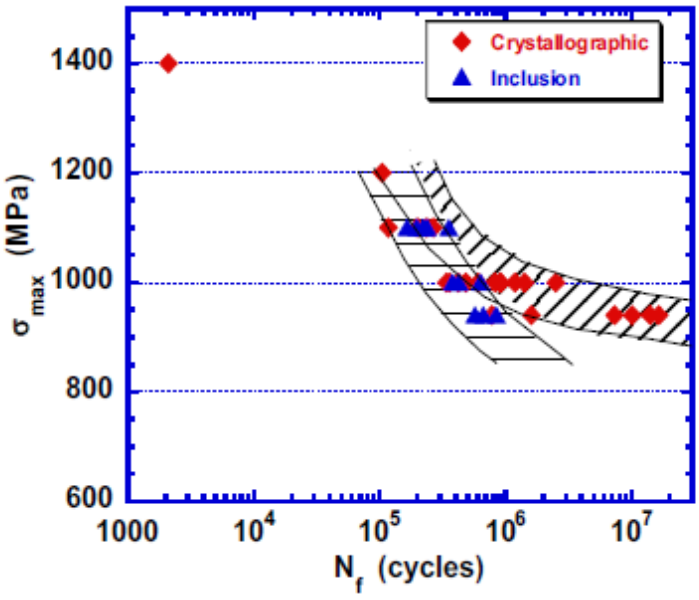


Figure 58: Divergence of fatigue failure mechanisms [92].

The resistance to fatigue of high strength alloys has been observed to reduce if NMPs are present. Fatigue crack initiation at inclusions depends on several factors such as the matrix material, the types of inclusions and also on the properties of the interface between the inclusion itself and the matrix. As a result, controlling those factors is crucial for an alloy designer. In order to clarify a few micro-mechanisms related to crack initiation, several microscopic observations have been conducted in various literature sources. For instance, in high-stress steels, the fatigue limit is much lower than the yield strength value, while the yield strength itself is reduced by the presence of inclusions [96], [97]. Also, Lankford and Kusenberger [98] quantified a series of stages which occur during fatigue crack initiation at inclusions. The first stage is related to the detaching of the inclusion from the matrix, and it is at the interface with the matrix that cracks start initiating. Also, together with debonding, the greatest part of crack initiation life is consumed. Morrow studied the influence of inclusion size on the fatigue limit on high strength steels [99], and his analytical results appear to agree with the experimental results of Cummings [97].

There have been plenty of studies which are able to provide a comprehensive outline of all the materials-related mechanisms which govern fatigue properties of superalloys, however there is very little discussion on the variability in fatigue behaviour due to the two main failure mechanisms. By applying the Bayesian-elicitation framework, it is possible to quantify uncertainties and make predictions regarding the life and failure cause of a turbine disc by looking into the behaviour of coarse grain (CG) RR1000 samples (RR1000 is a Rolls-Royce plc proprietary alloy). The following section will therefore give more details regarding the physics of the problem.

4.3.2 Outline of the Problem

Being able to design turbine components requires a solid understanding of the inevitable fatigue behaviour. The relationship between cycles to failure and the Walker strain (explained below) is given by the following equation [100]:

$$N = N_0 \left(\frac{\varepsilon + U}{\varepsilon - U}\right)^{\frac{1}{2A}} \tag{66}$$

Here N is the number of cycles to failure, ε is the Walker strain, while N_0, U and A are calibration parameters. The Walker strain is defined by Equation (67):

$$\varepsilon_{\text{Walker}} = \frac{\sigma_{\max}}{E}\left(\frac{\Delta\varepsilon_{\text{eq}} E}{\sigma_{\max}}\right)^{m} \tag{67}$$

Here, $\sigma_{\max}$ is the maximum stress used during the fatigue testing, E is Young's modulus, $\Delta\varepsilon_{\text{eq}}$ is the strain range, and m is the Walker strain exponent. The reason the Walker strain is used here instead of the usual strain is because the experimental data was obtained for different R values, where $R = \frac{\sigma_{\min}}{\sigma_{\max}}$. In other words, when plotting cycles to failure against the regular strain, multiple curves have to be drawn, each corresponding to one particular R value. The benefit of using Walker strain becomes apparent because it can consolidate all the fatigue data for various R values into a single curve [100].

The probabilistic framework was first and foremost used to draw judgements on the calibration parameters N_0, U and A from Equation (66). An expert elicitation was organised for this purpose (for which ethical approval was obtained), and outputs of the elicitation were subsequently passed on to the second component

of the framework in order to update them with real data. As a result, the following section is going to discuss the elicitation itself, the theory behind which was put forward in section 2.4.

4.3.3 Obtaining the Priors

Prior data has been elicited regarding the calibration parameters N_0, U and A from Equation (66). The elicitation workshop used the SHELF framework [54], and for 9 different temperatures, the quartiles of the prior distributions for the calibration parameters have been provided by the expert. The alloy which has been under consideration was the Rolls-Royce plc specific CG RR1000. Assuming there is no available fatigue data on this specific alloy, the judgements provided were extrapolated by the expert from a similar Rolls-Royce plc alloy called FG (fine grain) RR1000. Before discussing the data, it would be worth showing how the theory from section 2.4 was applied for this specific problem.

Most of the pre-elicitation phase followed the indications given in the Sheffield Elicitation Framework [54]. Namely, in order to assess the suitability of potential experts, an Expertise Questionnaire was handed out to them, similar to the one from Appendix A. Also, the manager of the experts was also asked to complete a General Expertise Measure (see Appendix B), in order to indirectly assess the experts asked to participate in the workshop. After the results from those two surveys were deemed acceptable, a time and a place have been established with the expert in order to hold the elicitation.

During the first part of the workshop, a "Context" form has been filled in which referred to the participant's expertise, how he gained this expertise, what is his interest regarding the outcome of the elicitation as well as whether he possesses any prior knowledge on probability and statistical knowledge. Because he proved a sound statistical background as he had taken statistical classes at university, the probability briefing part of the elicitation was skipped. Subsequently, an actual "Pre-elicitation" form was filled in containing details about the problem, after which the actual elicitation procedure took place. That had been conducted using the quantile elicitation technique presented in detail in SHELF [54]. The reason why the elicitation was conducted with a single expert was due to the extremely specific nature of the knowledge required and the fact that one person had been identified as possessing it at that particular time. As already mentioned, the elicited quantities are the three calibration parameters from Equation (67). The main results from the elicitation procedure are outlined as follows:

- The probability distributions for each of the three calibration parameters have only been elicited at the following temperature values: 20 °C, 300 °C, 400 °C, 500 °C, 550 °C, 600 °C, 650 °C, 700 °C, 725 °C, 750 °C, 775 °C. The reason for this was simply because those were the only temperatures that the fatigue testing had been performed at.
- Each individual PDF for each parameter at each temperature was considered to have a uniform distribution, and the initial results from that are shown in Table 11 below (multiplied by a random constant):

Table 11: Elicited quantiles for the calibration parameters (LB – Lower Bound, UB – Upper Bound)

	U		N_0		A	
Temperature (Celsius)	LB	UB	LB	UB	LB	UB
20	0.0003	0.0005	11	12	0.05	0.053
300	0.0003	0.0005	11	12	0.05	0.053
400	0.0003	0.0005	1	3	0.05	0.053
500	0.0003	0.0005	1	3	0.05	0.053
550	0.0025	0.003	1	3	0.05	0.053
600	0.0025	0.003	1	3	0.05	0.053
650	0.0025	0.003	1	3	0.04	0.045
700	0.0025	0.003	1	3	0.04	0.045
725	0.0003	0.0005	11	12	0.04	0.045
750	0.0003	0.0005	11	12	0.04	0.045
775	0.0003	0.0005	11	12	0.05	0.053

- In order to check how the preliminary elicited results compare to the experimental data, the new feedback method from section 3.3 has been applied. For each temperature, all Walker strain values were used as the arguments of the function in Equation (66), and together with the lower and upper bounds of the three coefficients from Table 11, two different sets of cycles to failure were then obtained. Those sets of data were afterwards compared to the actual values of failure cycles which were found from experiments. In order to better grasp this idea, the failure cycles against Walker strain is plotted at 650 °C.

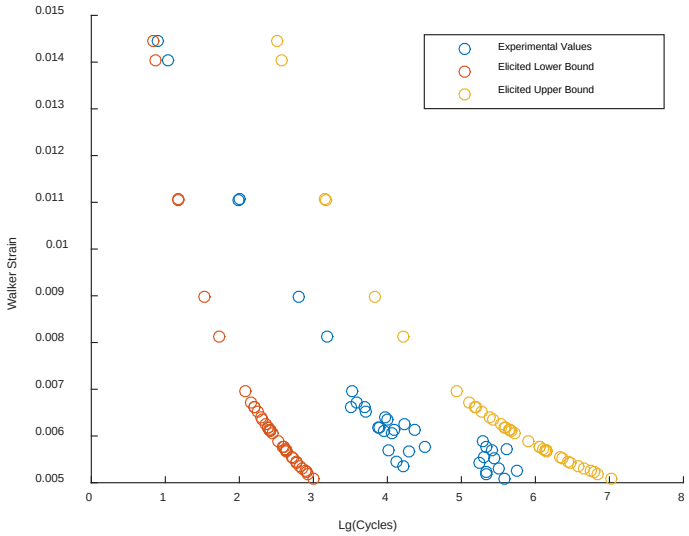


Figure 59: Experimental and elicited Walker strain at elevated temperature.

It is important to note at this point that the plot in Figure 59 shows the elicitation results *after* the feedback process (the actual values have been multiplied by a random number in order to obscure the data itself, although the trends are representative of the actual ones). In other words, the original elicited values did not seem to give proper upper and lower bounds of the failure cycles which contained the experimental values, as in Figure 59. This is in-line with the alternative feedback method: instead of gaining feedback about the distributions from Table 11, it is deemed more useful to convert the distributions into physical quantities which the expert can find easier to trace. After the results in Figure 59 were considered suitable, the elicitation process was stopped.

Although the priors of the calibration parameters N_0, U and A were found, their temperature dependence should be modelled as well. Therefore, a functional dependence between the three coefficients N_0, U and A as well as temperature had to be found. In order to achieve this, a curve fitting step had to be taken first. For each calibration parameters, the mean of each distribution was plotted against temperature, and together with the expert, it was decided that the temperature dependence can be modelled as a 2nd degree polynomial (at least for the temperature range that is relevant for the problem). The equations below illustrate this idea:

$$U(T^*) = a_2 T^{*2} + a_1 T^* + a_0$$

$$N_0(T^*) = b_2 T^{*2} + b_1 T^* + b_0 \quad (68)$$

$$A(T^*) = c_2 T^{*2} + c_1 T^* + c_0$$

Instead of using the physical temperature, a non-dimensional parameter T^* has been used. Its relationship with respect to the actual temperature is the following:

$$T^* = \frac{T - 300}{775 - 300} \quad (69)$$

As it can be deduced from Equation (69), the 9 temperatures at which the elicitation has been performed range between 300 ^{o}C and 775 ^{o}C. In addition to that, the a_i, b_i and c_i terms (with i ranging between 1 and 3) are the quadratic polynomial coefficients. By considering the mean and the 95% confidence intervals of the polynomial coefficients after the fitting process, individual prior distributions for all a_i, b_i and c_i have been created. It is those prior distributions that are then used further for the Bayesian updating procedure. Note that, due to the flexibility of the Bayesian framework, it is possible for the priors to have various distribution types (such as uniform, normal, lognormal, beta/scaled beta and so on). However, for all further analysis, it will be assumed that all 9 prior distributions corresponding to the coefficients from Equation (68) are normal:

$$a_i \sim N\left(\mu_{a,i}, \sigma_{a,i}^2\right)$$

$$b_i \sim N\left(\mu_{b,i}, \sigma_{b,i}^2\right) \quad (70)$$

$$c_i \sim N\left(\mu_{c,i}, \sigma_{c,i}^2\right)$$

As the priors have been set, the next step would be to use the second component of the framework in order to perform the update which ultimately offers the possibility of performing predictions.

4.3.4 Updating the Priors

In order to visualize how it is possible for the experimental data to update the prior distributions, it would be useful to understand how the Bayesian component is applied to this particular problem:

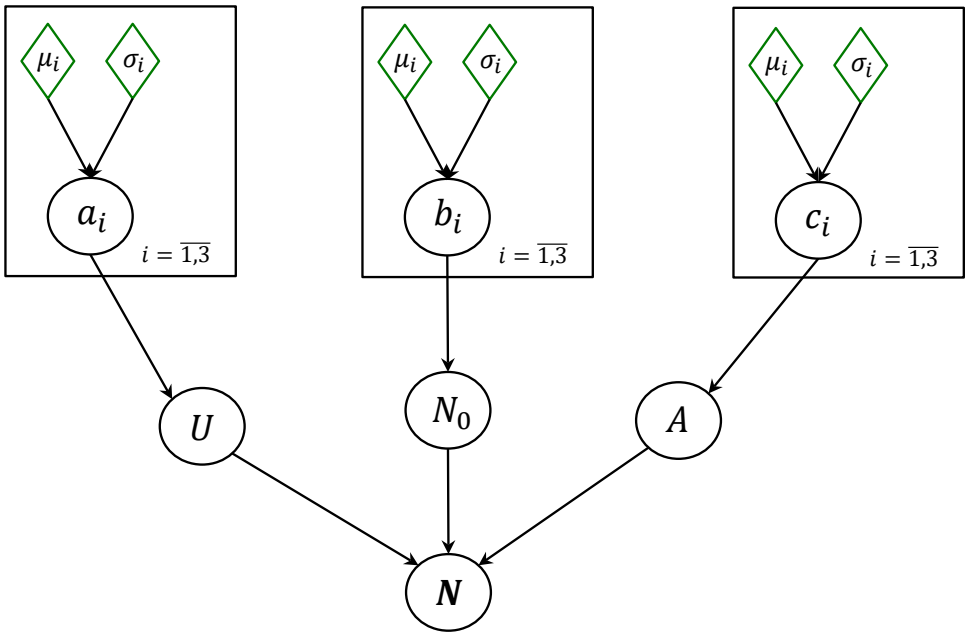


Figure 60: Bayesian Network for the failure cycles case study; the arrows start upstream from the prior distributions and propagate downstream into the likelihood function given by N (number of cycles).

Originally, the values of the likelihood function representing failure cycles are solely based upon the prior distribution of the three calibration parameters N_0, U and A (as well as the Walker strain), which in turn depend upon the nine polynomial coefficients which have been assigned normal distributions. An important feature to recall at this point is that the likelihood function shows how the model thinks the data should look like based on the prior beliefs. Afterwards, the data is being added (in terms of the experimental cycles to failure), and that has the effect of updating the prior normal distributions to posteriors (and hence propagate the new information "upstream"). This process is illustrated in Figure 61:

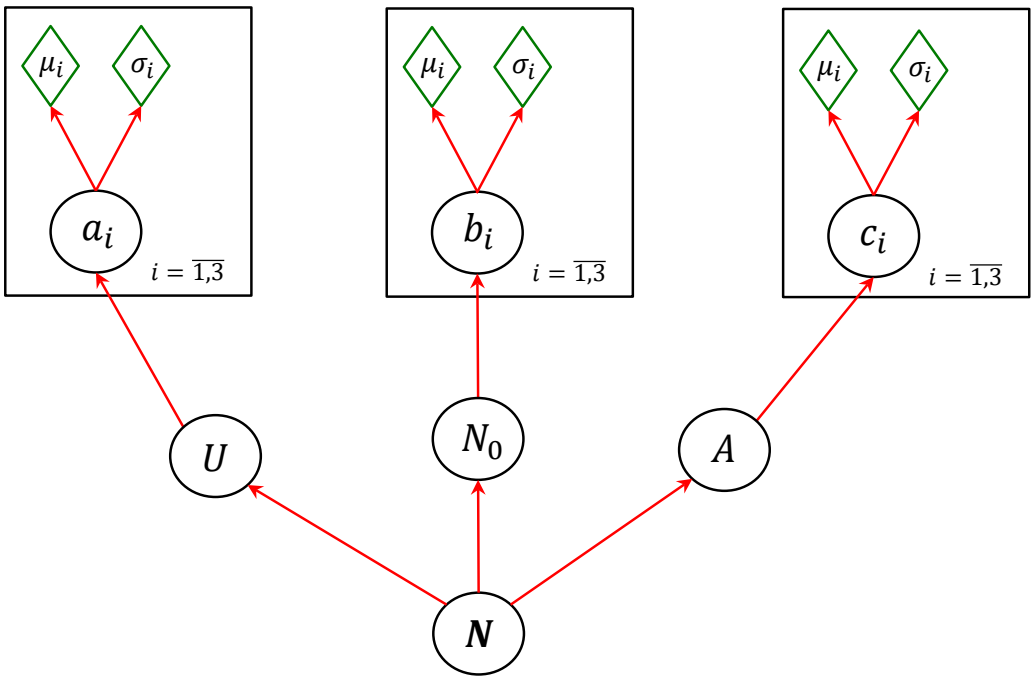


Figure 61: Bayesian update using experimental data. Here, the experimental data is propagated upstream in order to update the original prior distributions to posterior values.

Finally, the new posterior distributions can propagate downstream one more time (Figure 62) in order to create the posterior predictive distribution for the failure cycles (i.e. how the new model thinks the data should look like). This process is repeated as long as there is still unused new data, and ultimately it can be applied in order to give a prediction regarding the expected cycles to failure whether or not the temperature and Walker strain are specified.

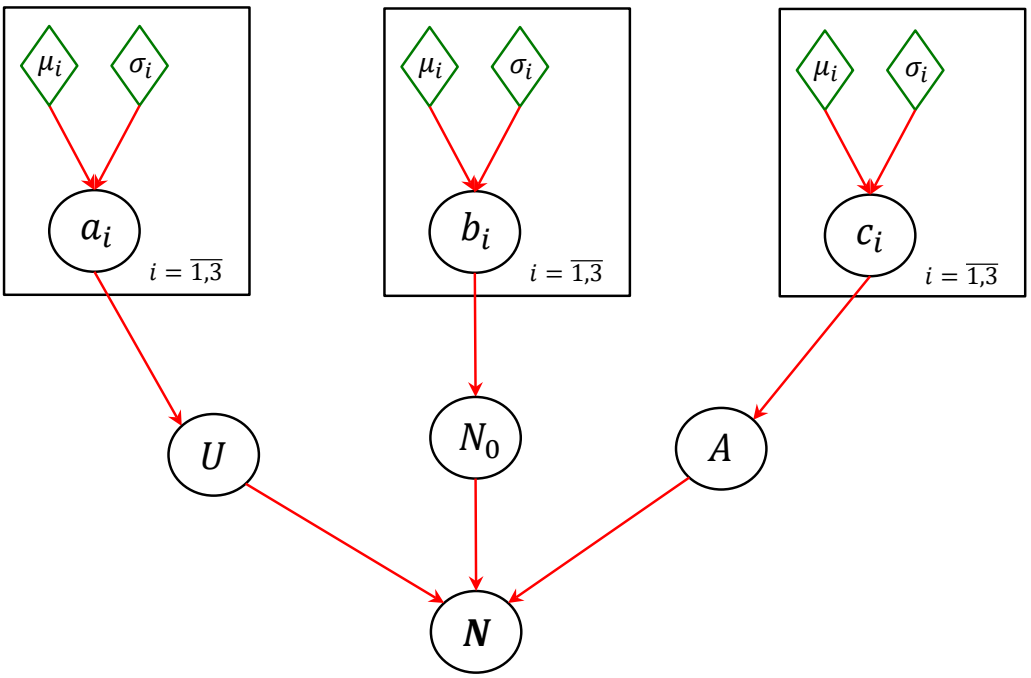


Figure 62: The newly updated posterior distributions propagate downstream once more in order to generate the posterior predictive distributions

After the elicitation workshop described in the previous section, three sets of polynomial coefficients have been assigned joint prior distributions, which have the following values (which once again have been multiplied by a random constant):

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \sim N\left(\begin{pmatrix} -0.0017 \\ 0.002 \\ 0.0008 \end{pmatrix}, \begin{pmatrix} 0.0037^2 & 0 & 0 \\ 0 & 0.004^2 & 0 \\ 0 & 0 & 0.001 \end{pmatrix}\right)$$

$$\begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \sim N\left(\begin{pmatrix} 25.1992 \\ -27.1965 \\ 8.4946 \end{pmatrix}, \begin{pmatrix} 28.4347^2 & 0 & 0 \\ 0 & 31.1826^2 & 0 \\ 0 & 0 & 7.8^2 \end{pmatrix}\right) \quad (71)$$

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \sim N\left(\begin{pmatrix} -0.008 \\ 0.003 \\ 0.0336 \end{pmatrix}, \begin{pmatrix} 0.0334^2 & 0 & 0 \\ 0 & 0.0367^2 & 0 \\ 0 & 0 & 0.0092^2 \end{pmatrix}\right)$$

As it can be deduced from the null non-diagonal elements, an independence assumption has been made regarding the coefficients. Knowing this, the following step is related to discussing the experimental data.

4.3.5 The Experimental Data

The priors have been updated by using high fidelity experimental data. Almost 200 individual fatigue tests have been performed on coarse grain RR1000, and the results have been gathered in one spreadsheet which contains information regarding the temperature, the Walker strain at which the experiments have been performed, as well as the number of cycles to failure and the failure cause. In order to make the data easier to understand, the spreadsheet was then split into several others which contain information about samples that failed due to the same cause. An example of how the data looks like for inclusion failure specimens can be seen in the table below:

Table 12: Data related to fatigue failure

Temperature (Celsius)	Walker Strain	*Cycles*
650	0.00576	13279
700	0.005718	28241
725	0.005449	16463
750	0.005071	29626

It is important to note that due to the sensitivity of the information presented here, the data in Table 12 has been obfuscated by being multiplied by a certain constant. By investigating all the data, it can be concluded that the component life reduces with increasing temperature and decreasing strain. As Figure 28 suggests, the data is used within the likelihood function in order to perform the Bayesian updating, which is going to be further discussed in the following section.

4.3.6 The Bayesian Updating

Before generating the posterior predictive distribution, which is used to predict life, the posterior distribution itself needs to be computed. As described in section 2.3, a single component Metropolis-Hastings algorithm is used in this regard. As a reminder, the joint posterior distribution is formed of the 9 polynomial coefficients which are updated sequentially.

Let $\boldsymbol{x} = (a_1, a_2, a_3, b_1, b_2, b_3, c_1, c_2, c_3)$ be the vector of parameters at each MCMC iteration. The updating was performed in the same order suggested by the order

of coefficients from vector x. The original starting point for the chain corresponds to the mean of the joint prior distribution. Several other starting points have been tried out, and every time the chain eventually converges to the same values, independent of the starting point. The proposal distribution for each variable was chosen to be normal (although it can very easily be assigned other distributions as well). Its mean was equal to the current state of the chain, while the standard deviation was chosen to be equal to a fraction of it. The standard deviation of the proposal distribution can also be considered an input (by controlling the value of the fraction), and by varying it, it is possible to check for convergence. As a general rule, this standard deviation should be large enough so as to explore the space, but not too large so that convergence becomes a problem due to low acceptance rate.

The likelihood function can once again be assigned various distribution types, although for now, it will be assumed that it has a normal distribution with mean equal to the life obtained from the physical model, and a standard deviation equal to a fraction of the experimental life value. This fraction corresponds to the level of uncertainty (i.e.: the level of confidence in the accuracy of the experimental results). In the beginning, this level of uncertainty has been set to 10%, although the simulation has been carried out for other values, as it will be shown in the results section. Due to the nature of the simultaneous updating, the overall likelihood function looks exactly like in Equation (49): a product of the likelihood functions for each data point. In the following section, the results will be presented and also their significance will be discussed.

4.3.7 Results and Discussion

The number of iterations used for the simulation was 300,000 due to the fact that beyond this number, there was minimal increase in the fidelity of the results, while the computational time kept increasing. The posteriors of the polynomial coefficients were subsequently used in order to create predictions about life. Due to the fact that there are 9 coefficients and 6 different failure causes, the graphs below will only depict 2 of those coefficients and 2 causes (one corresponding to crystallographic failure, and the other to inclusion failure). This is done in order to just show 4 plots instead of 54. In addition to that, in the first instance, the

uncertainty level used for the likelihood function (as already suggested in the previous section) is 10%. The comparison between the prior and posterior plots for b_3 and c_3 can be seen below:

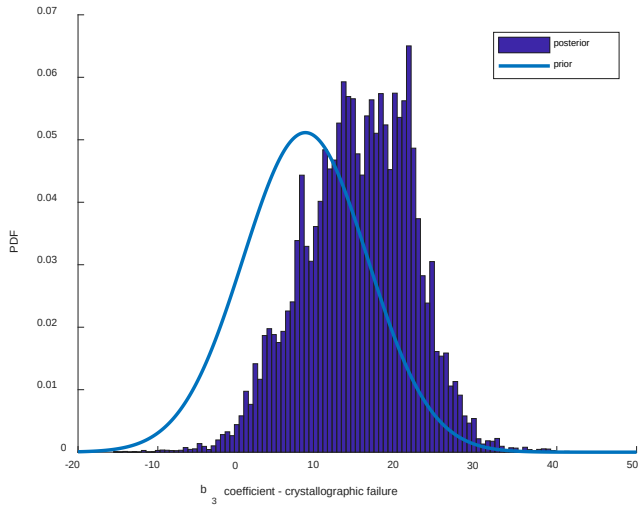


Figure 63: Comparison between prior and posterior of the b3 coefficient (crystallographic failure).

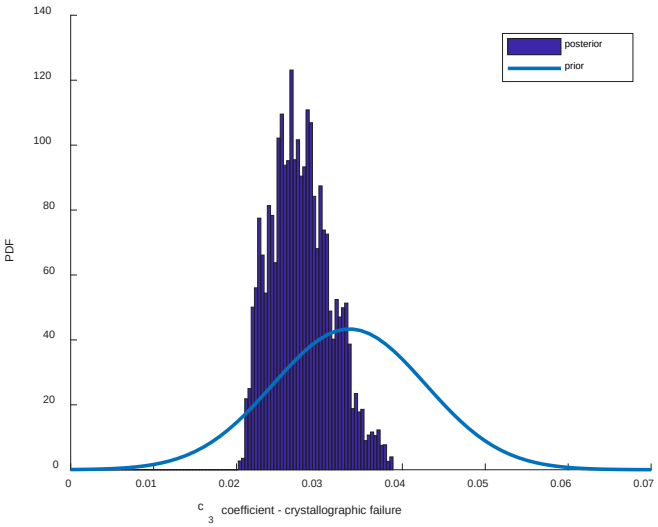


Figure 64: Comparison between prior and posterior of the c3 coefficient (crystallographic failure).

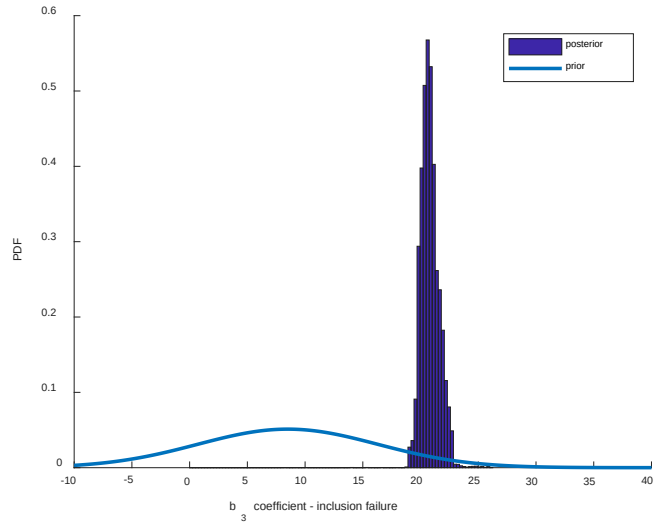


Figure 65: Comparison between prior and posterior of the b3 coefficient (inclusion failure)

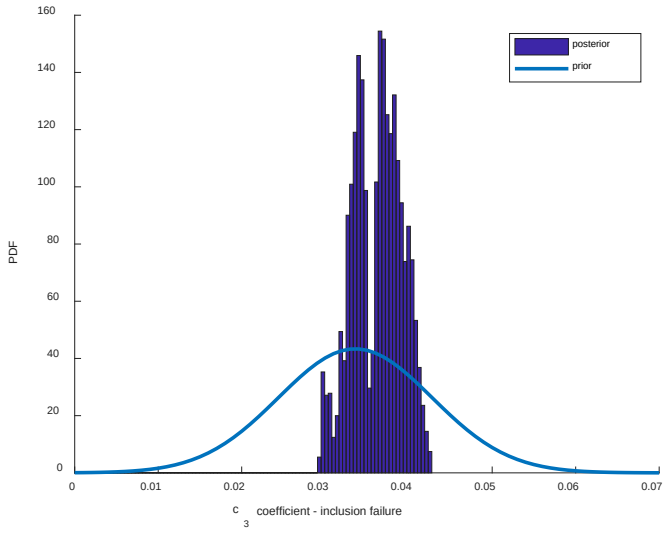


Figure 66: Comparison between prior and posterior of the c3 coefficient (inclusion failure)

In the crystallographic failure case, the joint posterior distribution of the 9 coefficients was used to predict the life for new test conditions: first at 775 ^{o}C and 0.005 Walker strain, and then at 775 ^{o}C and unknown Walker strain. The unknown Walker strain is taken into account simply by taking samples from its distribution (assumed to be uniform) with lower and upper bound equal to 0.0049 and 0.0056 respectively. The posterior predictive distributions are therefore shown in the following plots.

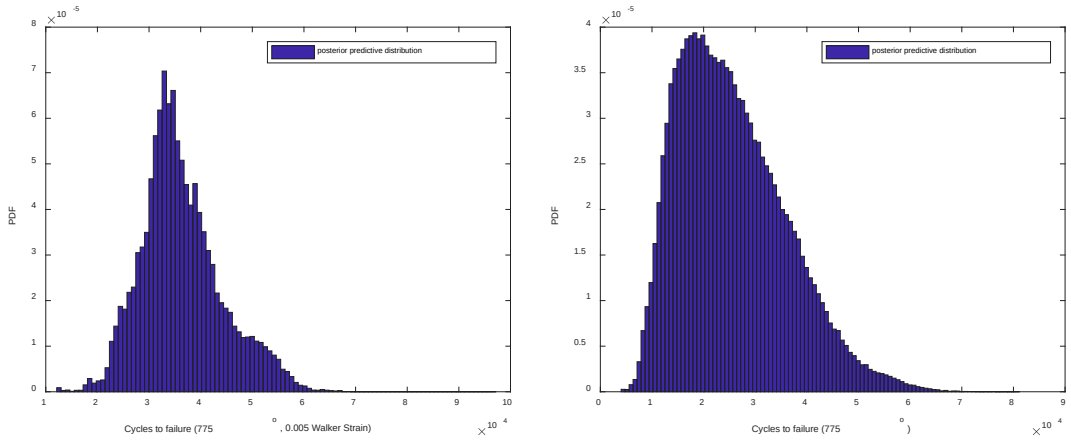


Figure 67: Predicted life for crystallographic failure at 10% uncertainty.

For the inclusion failure case, the uniform distribution for the Walker strain has upper and lower bounds equal to 0.004 and 0.006. The predicted life has been calculated at 800 ^{o}C (as there is no data at this temperature) as well as 0.005 strain (and once again at 800 ^{o}C, while the strain was considered unknown).

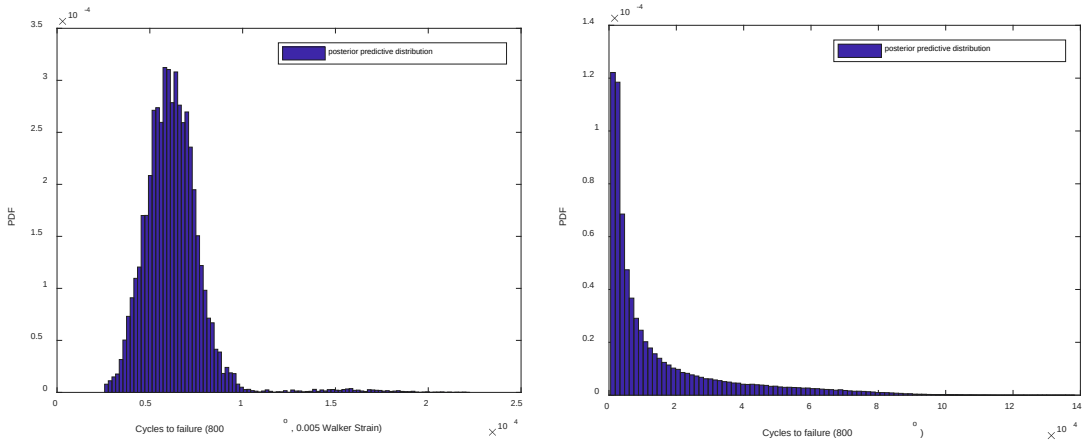


Figure 68: Predicted life for inclusion failure at 10% uncertainty.

One of the common features from Figure 67 and Figure 68 is that the variance of the prediction increases when the strain is unknown. This is also intuitive, as the strain values are sampled from a distribution. In addition to that, it can be observed that the variance difference between the two plots from Figure 68 is much higher

than the one from Figure 67 due to the different strain range; a wider Walker strain distribution causes the predicted life distribution to be wider as well.

A change in the standard deviation can also be observed as the uncertainty of the likelihood function is increased; this behaviour can be seen in the following graphs.

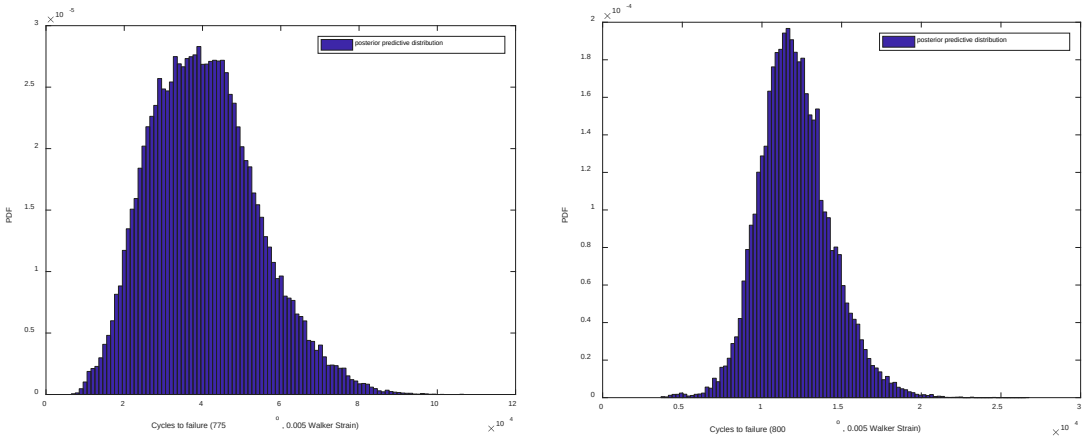


Figure 69: Predicted life (crystallographic failure on the left, inclusion failure on the right) at 30% uncertainty.

As the likelihood uncertainty is increased even further, the predicted life becomes more conservative ("wider") as it can be seen below:

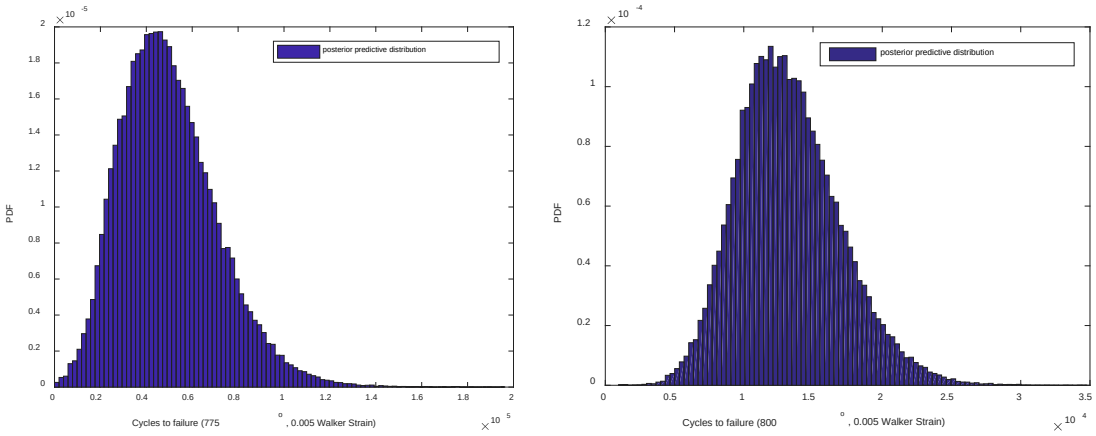


Figure 70: Predicted life (crystallographic failure on the left, inclusion failure on the right) at 50% uncertainty

In order to better quantify the effect of the likelihood uncertainty on the predicted life, multiple plots depicting various uncertainty levels have are shown below in the same graph. It can also be noted that the normalized histograms have been fitted to lognormal distribution via the MATLAB function "allfitdist.m". One of the main reasons lognormal distributions are suitable (as opposed to normal ones for instance) is because physically the life cannot take negative values.

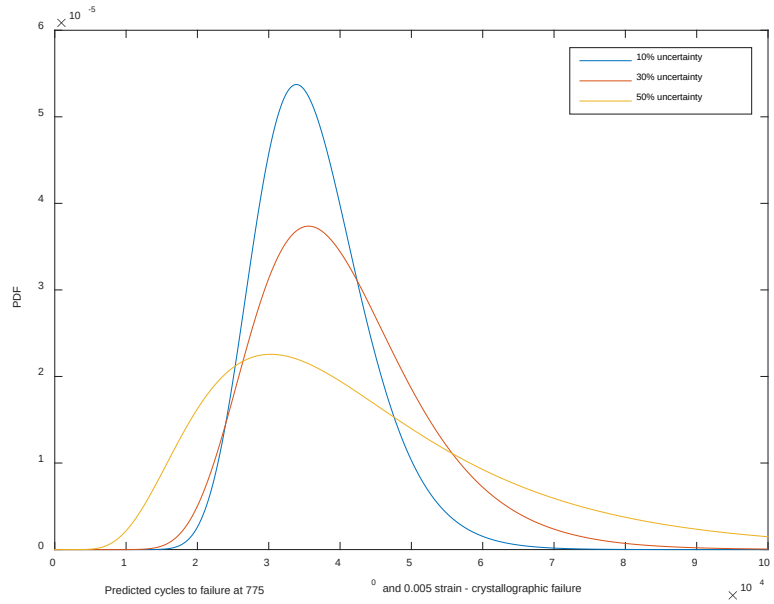


Figure 71: Predicted life at various uncertainty levels (crystallographic failure)

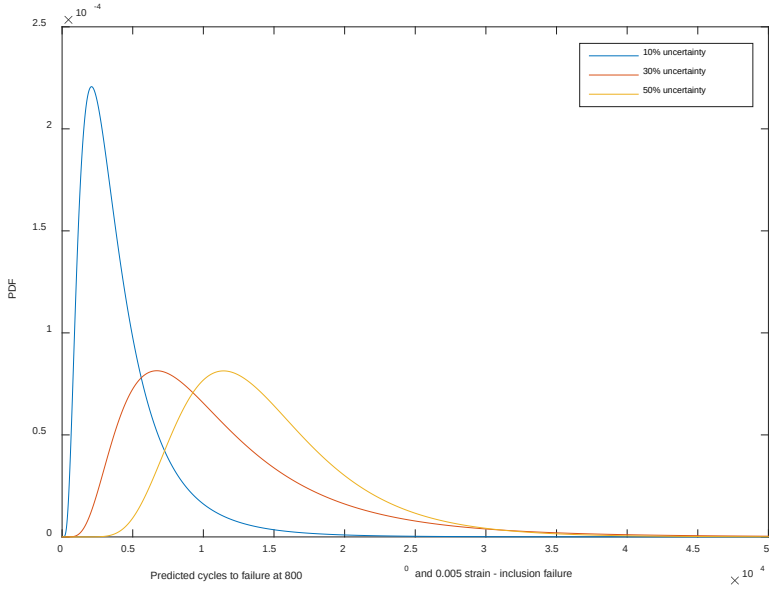


Figure 72: Predicted life at various uncertainty levels (inclusion failure)

Another feature of interest is to be able to predict which failure causes are the most probable given a certain life and working temperature. The strain is considered to have a probabilistic nature: it is a uniform distribution between 0.005 and 0.006. In addition to that, the likelihood uncertainty picked in this example is 10%. Consequently, the predicted life at 600^0 for three different causes looks as follows:

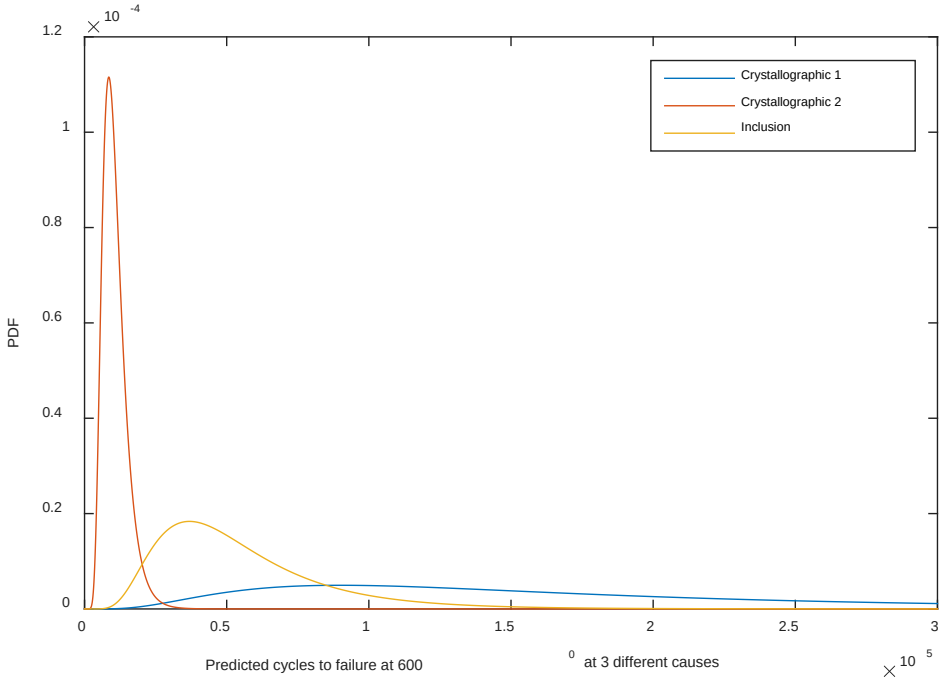


Figure 73: Predicted life at 600^0 for 3 different causes

In order to assess which cause is the most responsible for a certain failure, it is only necessary to examine the graph and compare the relative PDF values. For instance, for lives of 150,000 or higher, it can be seen that “Crystallographic 1” is the principal cause of failure, whereas for lives lower than 25,000 the failure is mainly due to “Crystallographic 2”. Most inclusion failures can thus be expected to occur for lives in the range of 40,000 to 60,000. The same kind of graph can be plotted for other temperatures as well, and it can also be done while considering temperature to be probabilistic.

Similar to the previous two case studies, this one also stands to show the versatility of the Bayesian-elicitation framework presented in Chapter 3. Hopefully, it is obvious by this point that similarly to how the framework was applied in order to study FBO events, grain growth and fatigue failure, there could be plenty other aero-engine related problems where its capabilities can shine. As described in Chapter 1, the main reason for developing this was to give the designer useful information during the preliminary design process in situations when data is scarce or expensive to come across.

Chapter 5 Summary and Conclusions

The journey that is this EngD project commenced from examining the aero-engine market, and that was the principal motivation which lead to describing the

importance of the preliminary aero-engine design process. Following that, the need to develop techniques that produce fast and reliable results during this stage has been identified. Later in Chapter 2, the entire theory behind the probabilistic framework proposed at the end of Chapter 1 was laid out. This included a general discussion regarding uncertainty, after which the focus was directed on more specific topics such as Bayesian inference, expert systems as well as judgement elicitation. All those topics have been discussed in detail as they form the foundation of the Bayesian-elicitation framework which was described in detail in Chapter 3. Not only was it outlined how all the theory is interconnected, but the validation process was emphasized too. As the power of the framework can only become apparent when applied to real-life case studies, the next natural step was to use it in order to solve several case studies related to the aero-engine preliminary design process; that was presented in Chapter 4.

5.1 Summary

The first case study was concerned with analysing in service FBO events, more specifically with the loss of balance of the fan during the windmilling phase which succeeds the actual event. The Bayesian-elicitation framework was applied in order to identify the characteristics of the most likely FBO events. This was done in order to allow the designer to obtain critical knowledge during the preliminary stage of the aero-design process, which is when some of the most important design decisions need to be taken (as mentioned in Chapter 1).

The second case study dealt with studying the grain growth of a turbine disk during the manufacturing process which comprises of two phases: isothermal forging as well as heat treatment. As it was mentioned several times, investigating this case study had two objectives: developing the updating algorithm as well as quantifying the grain size distribution given several manufacturing inputs. The reason is that the grain size directly impacts the fatigue life of the component, which is in turn related to cost. As it was highlighted in Chapter 1, by the end of the preliminary design stage, most of the committed cost is allocated to the project, so doing this efficiently is paramount.

Finally, the third case study investigates fatigue failure of a turbine disc due to non-metallic particle inclusions. This is where the framework was used to its full potential, and this was achieved by applying the elicitation and the updating

components, both of which were described in Chapter 3. The case study itself is of interest because the available knowledge on inclusion failure is limited. As a direct effect of that, there are cases when aero-engine components (such as the turbine disc) are taken out of service much earlier than they should be. This is a somewhat crude "safety factor approach", and it is believed that it can be refined further given more available knowledge on the topic. This knowledge has the form of expert judgements and real-life data which can be used to predict both the life and the failure cause of the turbine disc (although the same methodology can be applied to any other component subjected to similar working conditions).

The Bayesian-elicitation framework is the main selling point of this project, and throughout the sections above it was shown how useful it can be during the preliminary stages of the design process when data is generally hard to come by. The number of applications where this framework shines is not limited just to the three case studies presented here, but it can be used for any aero-engine problem where cutting costs via knowledge gathering and uncertainty quantification is one of the top priorities.

5.2 Conclusions and Summary of Original Contributions

In order to understand the impact of this project on the field of aero-engine design and how it can contribute to the preliminary design process, it would be worth discussing the outcomes of this research in the context of the research hypothesis presented in section 1.3. The hypothesis stated as follows:

"Using probability-based methods during the preliminary stage of the aero-engine design process can allow fast and accurate investigation of the design space in order to aid the generation of an optimum design which can afterwards be passed to the detailed design phase. Ultimately, this can have the effect of making the entire design process faster and less expensive."

The non-deterministic approach to this project was the factor that formed the foundation of this EngD. In Chapter 1 the importance of the preliminary design stage was established and from that, the need for efficient decision making arises. Together with the literature findings from Chapter 2, the Bayesian-elicitation framework was proposed to meet the requirements of the research hypothesis. This was done because also in Chapter 2, it was underlined how

unifying the two pillars: expert elicitation and Bayesian inference into a flexible framework can meet the desired objective.

After careful consideration of mathematical and psychological theory, the framework was built in MATLAB and was intended to be as flexible as possible so that parts of it can be changed for others (if desired by the user), while keeping its overall philosophy intact. After the framework was validated, it has been applied to a variety of aero-engine specific case studies in a way in which would satisfy the goal set within the research hypothesis.

It would also be worth stating (or re-stating) the novel contributions that were developed during this EngD. The principal contribution is the framework itself. Its building blocks are put together in a manner unlike what was found in the literature and it is also able to solve preliminary aero-engine related problems that would allow the designer to take decisions during this stage that would ultimately reduce cost. Secondly, a new method of obtaining feedback subsequent to an expert elicitation was put forward. That is, instead of asking on feedback with regards the PDFs themselves, the feedback can instead be on the prior predictive distribution as this converts the calibration parameters into physical parameters which for the expert are potentially easier to grasp.

A final interesting aspect is that the concepts outlined in the title of the project have all been dealt with in one form or another:

- Model Validation: as shown in Chapter 3, especially in section 3.6, the model (or the framework in the context of this project) has been validated by being applied to a case study found in the literature. It is also mentioned that the validation of the elicitation component is taken care of during the feedback process.
- Uncertainty Quantification: the entire EngD revolved around this idea, and theory on Bayesian Inference and Expert Systems, which are uncertainty quantification methods, is described throughout Chapter 2.
- The Preliminary Aero-Engine Design Process: the framework was applied to several case studies related to making this stage of the design process more efficient.

5.3 Future Work

Like any research work carried out in general, there are areas where this project could see improvements. Firstly, the expert elicitation component based on SHELF [54] could be replaced with something else. One option would be to use a Gaussian process approach instead of fitting the judgements to set PDFs [48], as this can potentially capture the uncertainty without constraining them beyond the fundamental definition that its integral has to be 1. In addition to that, it is also possible to consider the elicited coefficients as dependent. However, that would add a high amount of complexity to the problem as covariances between various parameters have to be investigated and this can prove to be a difficult task especially if the case study itself is niche. This could mean that the data can be hard to obtain, or it could be a difficult task to find experts with deep understanding on the topic.

The framework itself could be developed by adding several other possibilities for the distribution types (or by including the Gaussian process-based PDF mentioned earlier). If the methodology described in Chapter 3 is followed, it is expected that adding more such possibilities into the framework should be straightforward. Additionally, other machine learning algorithms could be used as an alternative to Bayesian Inference, although that is expected to modify this framework quite significantly.

From the point of view of the case studies, it is possible to look into more aero-engine related problems. The field of research need not be constrained strictly to aircraft engines, as there are a number of other aerospace-related areas where uncertainty quantification is a top priority. Of course, the physical models behind the existing case studies could be improved. Unfortunately, there is a limit to this as beyond a certain point they could become too expensive from a computational perspective. It is also important to keep in mind that no matter how accurate the model becomes, at the end of the day it is just an approximation. This is an idea that was neatly captured in 1976 in a publication by Box and Draper who said that “All models are wrong, but some are useful” [101].

Appendix A Expertise Questionnaire

This questionnaire is intended to find out about the nature of your job, and the type of judgements that you make while performing it. These answers will be used to prepare for the upcoming elicitation workshop on

Inclusion Fatigue Failure of a Turbine Disc

In particular, we are interested in whether or not your job requires you to make probabilistic judgements, and how you make such judgements. In addition, we are interested to find out what sort of aids you use in making judgements, whether you received any relevant training, and whether you receive feedback about the quality of your judgements.

Part A: General description of your job

1. What is the title of your job?

   _____

   _____

2. How would you describe your area of expertise?

   _____

   _____

3. How many years of experience would you say you had in your area of expertise?

---

4. Would you describe that experience to be:
   Please tick ONE circle.

   - o Wholly practical and/or field-based
   - o Mostly practical and/or field-based but some theoretical and/or lab-based
   - o About equally practical/theoretical and field/lab-based
   - o Mostly theoretical and/or lab-based
   - o Wholly theoretical and/or lab-based

Part B: The judgements you make

5. Describe the most important judgements you make on a regular basis in your job.

---

6. When you have to make work judgements, to what extent do you rely on your judgements alone, and to what extent do you rely on other information sources (such as manuals of statistics, computer databases or programs, etc.)?
   Please tick ONE circle.

   - o I always use my judgement alone.
   - o I mostly use my own judgement.
   - o I use partly my own judgement, and partly other sources.
   - o I mostly use other sources.
   - o I always use other sources alone (not personal judgement)

7. If you use other information sources, please describe them below.

_____

_____

8. In making your work judgements, do you receive any feedback on their accuracy?
   Please tick ONE circle.

   - o Always
   - o Often
   - o Sometimes
   - o Rarely
   - o Never

9. If you receive feedback, what form does this take?

_____

_____

10. How often after a judgement, on average, do you receive feedback?
    Please tick ONE circle.

- o The same day
- o Within a week
- o Within a month
- o Within a year
- o Longer than a year
- o I do not receive feedback

11. How would you rate the ease of making good judgements in your work? Please tick a box that represents your opinion.

Very difficult

Very easy

1	2	3	4	5	6	7

12. Do you make use of a formal model for making your work judgements? Please tick a box that best represents your opinion.

Never

Always

1	2	3	4	5	6	7

13. How would you rate the availability of data you use for your work judgements?
Please tick a box that best represents your opinion.

Poor Availability

Plentiful

1	2	3	4	5	6	7

14. How would you rate the quality of data that you use for your work judgements?
Please tick a box that best represents your opinion.

Very poor

Very good

1	2	3	4	5	6	7

15. Did you receive any training to make judgements? If so, please describe below.

---

16. Do you ever make any of the following types of judgements at work (numerically, verbally, or by any other means)? Please tick and fill in as many as are relevant.

- o I estimate the likelihood/probability of
  _____
- o I estimate the chances of
  _____
- o I estimate confidence in
  _____

17. How often, on average, are you called upon to make judgements of risk or uncertainty?
Please tick ONE circle.

- o At least once a day
- o At least once a week
- o At least once a month
- o Less than once a month

18. When you make judgements of risk or uncertainty, what forms do they take?
    Please tick as many circles as are relevant.

    - o Numerical estimates (e.g. 0.5, 50%, 1 in 2)
    - o Verbal estimates (e.g. likely, infrequent)
    - o Comparative (e.g. “the risk is similar to another risk”)

19. If you do make numerical judgements, what forms do these take?
    Please tick as many circles as are relevant.

    - o Percentages (e.g. 50% chance)
    - o Point probabilities (e.g. 0.5 chance)
    - o Confidence intervals (e.g. range within which you are 95% confident the true value falls)
    - o Probability distributions (as previous but more than one range assessed for each quantity)
    - o Frequencies (e.g. 3 out of 10 chances of occurring)
    - o Odds (e.g. odds of 2 to 1 chance of occurring)
    - o Ratings on scales (e.g. point 2 on a 7-point scale of likelihood)
    - o Other types of numerical judgements: please provide details below

    _____

20. Please give an example of a type of judgement of risk or uncertainty you typically make (if you do make such judgements).

    _____

    _____

21. Did you receive any training to make judgements of risk and uncertainty? If so please describe below.

    _____

22. When you have to make judgements of risk and uncertainty do you rely on your own judgement alone or do you also use other information sources (such as manuals of statistics, computer databases or programs, etc.)? Please tick ONE circle.

- o I always use my judgement alone
- o I mostly use my own judgement
- o I use partly my own judgement, and partly other sources
- o I mostly use other sources
- o I always use other sources alone (not personal judgement)

23. If you use other information sources, please describe them below.

THANK YOU FOR YOUR TIME AND EFFORT.

Appendix B Generalized Expertise Measure

Please rate the proposed candidate on the characteristics below using the scale:

Question no./rating	*1*	*2*	*3*	*4*	*5*
Q1					
Q2					
Q3					
Q4					
Q5					
Q6					
Q7					
Q8					
Q9					
Q10					
Q11					
Q12					
Q13					
Q14					
Q15					

1 represents “Not true at all”

5 represents “Definitely true”

This person:

Q1. Has knowledge specific to a field of work.

Q2. Shows they have the education necessary to be an expert in the field.

Q3. Has the qualifications required to be an expert in the field.

Q4. Has been trained in their area of expertise.

Q5. Is ambitious about their work in the company.

Q6. Can assess whether a work-related situation is important or not.

Q7. Is capable of improving themselves.

Q8. Is charismatic.

Q9. Can deduce things from work-related situations easily.

Q10. Is intuitive in the job.

Q11. Is able to judge what things are important in their job.

Q12. Has the drive to become what they are capable of becoming in their field.

Q13. Is self-assured.

Q14. Has self-confidence.

Q15. Is outgoing.

Appendix C Creating a Dirichlet Distribution

Let us assume that there is a set of k uncertain quantities which form the vector $\mathrm{x} = (x_1, x_2, x_3, \dots, x_k)^T$ which has the properties that $x_i \geq 0$ for all $i = 1,2, \dots, k$ and also $\sum_{i=1}^{k} x_i = 1$. Consequently, x_i can be thought of as a proportion of members from a population which belong to category i [89]. For explaining purposes, it will be assumed that the elicited judgements come from a single expert, although the method is flexible enough to make it suitable for a group of experts as well. The results obtained from the elicitation are assumed to have a Dirichlet distribution unless the expert believes this is not an accurate representation of the knowledge possessed.

The vector $\mathrm{x} = (x_1, x_2, x_3, \dots, x_k)^T$ is said to have a Dirichlet distribution with the parameter vector $\mathrm{d} = (d_1, d_2, d_3, \dots, d_k)^T$ if the PDF is given by the following expression:

$$f(\mathrm{x}|\mathrm{d}) = c(\mathrm{d}) \prod_{i=1}^{k} {x_i}^{d_i - 1} \tag{72}$$

The normalising constant from Equation (73) is given by:

$$c(\mathrm{d}) = \frac{\Gamma\left(\sum_{i=1}^{k} d_i\right)}{\prod_{i=1}^{k} \Gamma(d_i)} \tag{73}$$

The gamma function $\Gamma(n)$ has its usual expression given by:

$$\Gamma(n) = (n-1)!, \text{ if } n \in \mathbb{N}^* \tag{74}$$

The means and variances of each x_i are given by Equation (75) below:

$$E(x_i|\mathrm{d}) = \frac{d_i}{n}, \quad \mathrm{Var}(x_i|\mathrm{d}) = \frac{d_i(n-d_i)}{n^2(n+1)}, \quad \text{where } n = \sum_{i=1}^{k} d_i \tag{75}$$

It is inevitable for the x_i variables to be correlated because of the relation $\sum_{i=1}^{k} x_i = 1$. The expression for covariance is given in Equation (76) below:

$$\mathrm{Cov}(x_i, x_j) = -\frac{d_i d_j}{n^2(n+1)}, \quad \text{for } i \neq j \tag{76}$$

A special property of the Dirichlet distribution is that each marginal distribution of each x_i is a beta distribution with parameters d_i and $n - d_i$: $x_i \sim \mathrm{Be}(d_i, n - d_i)$.

The steps taken in order to perform an elicitation using the Dirichlet distribution, which were also briefly presented in section 4.2.2, are as follows:

- Prepare and train the experts in order for them to get familiarized with the ideas of probability judgements as well as the elicitation method which will be used during the workshop.

- By using O'Hagan's SHELF software, several quantiles are elicited for each x_i and a beta distribution which has similar quantile values is built. It is important to note at this point that more often than not, a compromise has to be made in order to construct the beta distributions; realistically no beta distribution will exactly fit the elicited quantiles [89]. This is indeed reflected by the way in which the SHELF software operates; it chooses a beta distribution which fits the judgements as close as possible. The next step consists of the elicitor showing the beta distributions to the expert so that feedback can potentially be given and the expert is then given an opportunity to update the initial estimates. Alternately, the expert can insist that his beliefs are not represented accurately by beta distribution, and at this point the procedure stops, the conclusion being that no Dirichlet distribution is suitable for the problem.

- The means from all beta distributions will most probably need to be adjusted. The constraint $\sum_{i=1}^{k} x_i = 1$ always holds for a Dirichlet distribution, however because each x_i has been elicited separately, it is unlikely to satisfy this equation in practice. If x_i has a beta distribution of

the form $x_i \sim \mathrm{Be}(d_i, e_i)$ and the expected value is $E(x_i) = \frac{d_i}{d_i + e_i}$, the constraint can be written in an equivalent form presented in Equation (77) below:

$$r = \sum_{i=1}^{k} \frac{d_i}{d_i + e_i} = 1 \tag{77}$$

Realistically, the parameter r above will most likely not be equal to 1. This discrepancy can be accounted for by applying a correction to d_i and e_i in the following way:

$$d_i{}^* = \frac{d_i}{r}, \qquad e_i{}^* = d_i + e_i - d_i{}^* \tag{78}$$

- Although Equation (77) is satisfied after the previous step, it is highly likely that the sum of beta parameters for each x_i (which is the n value) is different for each of the marginal distributions. The joint Dirichlet distribution has only a single n value associated to it, meaning that a compromise value for n should be used. The question therefore becomes: *"Which value of n would produce a distribution that reflects the expert's knowledge in the best way possible?"*. There are several suggestions by Vazquez [89] in this respect:
  - The compromise value for n could for instance be given by the middle value $n_{mid} = \frac{n_{\min} + n_{\max}}{2}$, where $n_{\min}$ and $n_{\max}$ correspond to the lowest and highest values of n for all k beta distributions.
  - Another option would be to use a conservative value $n = n_{\min}$ as in this way the elicitor does not assume more knowledge about the x_i values than it had already been elicited.
  - An optimization technique could be applied, and Vazquez [89] showed the full solution on how to obtain the optimum value of n which is:

$$n_{opt} = \left(\frac{\sum_{i=1}^{k} v_i^* (n_i + 1)}{\sum_{i=1}^{k} v_i^* \sqrt{(n+1)}} \right)^2 - 1, \qquad v_i^* = \frac{d_i{}^* (n_i - d_i{}^*)}{n_i{}^2 (n_i + 1)} \tag{79}$$

After the value of n is decided upon, the final values for each pair of beta parameters can also be found:

$$d_i(n) = n\frac{{d_i}^*}{{d_i}^* + {e_i}^*} \tag{80}$$

- The final step consists of showing the expert each beta distribution as obtained from the joint Dirichlet distribution in order to obtain feedback as to how well each beta distribution fits the expert's original beliefs. Once again, the expert may believe that the final Dirichlet distribution does not match his judgements in any way, meaning that the elicitation process terminates with the conclusion that no Dirichlet distribution can represent the expert's uncertainty about the problem. Otherwise, it can be stated that the vector x has the distribution $\mathrm{x}\sim\mathrm{Di}\big(\mathrm{d}(n)\big)$.

Bibliography

1. Profir, B., Eres, M.H., Moss, M., Bates, R., *Uncertainty Quantification via Elicitation of Expert Judgements*, in *AIAA Forum and Exposition 2016*2016: Washington D.C.

2. Profir, B., Eres, M.H., Moss, M., Bates, R., *Updating Elicited Expert Judgements using a Bayesian Network*, in *AIAA 2017 Aviation*2017, American Institute of Aeronautics and Astronautics: Denver, Colorado.

3. Profir, B., Eres, M.H., Moss, M., Bates, R., *Quantifying Uncertainties during the Early Design Stage of a gas Turbine Disc by Utilizing a Bayesian Framework*, 2017, University of Southampton.

4. Profir, B., Eres, M.H., Moss, M., Bates, R., *Investigation of Fan Blade Off Events Using a Bayesian Framework*, in *ASME Turbo Expo 2017: Turbomachinery Technical Conference and Exposition*2017: Charlotte, North Carolina.

5. Profir, B., *Markov Chain Monte Carlo Analysis of a Case Study on Grain Growth within a Turbine Disk*, 2016. p. 29.

6. *IATA Fact Sheet - Industry Statistics*. 2019.

7. *IATA Forecasts Passenger Demand to Double Over 20 Years*. 2016.

8. *IATA Fact Sheets - Industy Facts and Statistics*, 2018.

9. *Commercial Aero-Engine Market Analysis and Update*. 2018. London, UK.

10. Karl, A. and Bates, R., *Uncertainty management during the design of advanced aero engines*, 2017.

11. Hryshchenko, T. *Integrated Design and Analysis of Turbofan Engines*.

12. ABET, *Criteria for accrediting engineering programs 2015-2016*, 2015: Engineering Accreditation Commission, 2015.

13. Keane, A.J. and Nair, P.B., *Computational Approaches for Aerospace Design the Pursuit of Excellence*, 2005, Wiley: Chichester. p. xx, 582 p., [4] p. of plates.

14. Verhagen, W.J.C., et al., *A Critical Review of Knowledge-Based Engineering: An Identification of Research Challenges.* Advanced Engineering Informatics, 2012. **26**(1): p. 5-15.

15. Roy, C.J., Oberkampf, W.L., *A Comprehensive Framework for Verification, Validation and Uncertainty Quantification in Scientific Computing.* Computer Methods in Applied Mechanics and Engineering, 2011. **200**: p. 2131-2144.

16. Yao, W., Chen, X.Q., Luo, W.C., Van Tooren, M., Guo, J., *Review of Uncertainty-based Multidisciplinary Design Optimization Methods for*

Aerospace Vehicles. Progress in Aerospace Sciences, 2011. **47**(6): p. 450-479.

17. Thunnissen, D.P., *Uncertainty classification for the design and development of complex systems*, in *The third annual predictive methods conference*2003: Newport Beach, CA.

18. Walton, M., *Managing Uncertainty in Space Systems Conceptual Design using Portfolio Theory*, 2002, Massachusetts Institute of Technology.

19. Padmanabhan, D., *A Probabilistic Approach to Aircraft Design Emphasizing Guidance and Stability and Control Uncertainties*, 1998, Georgia Institute of Technology.

20. Stone, J.V., *Bayes' Rule - A Tutorial Introduction to Bayesian Analysis*. 1 ed. 2013: Sebtel Press.

21. Lindley, D.V., *Why Isnt Everyone a Bayesian - Comment.* American Statistician, 1986. **40**(1): p. 6-7.

22. Larsen, R.J. and Marx, M.L., *An Introduction to Mathematical Statistics and its Applications*. 4th ed. 2006, Upper Saddle River, N.J.: Perason/Prentice Hall. viii, 920 p.

23. Cook, J.D. *Conjugate Prior Relationships*. Available from: http://www.johndcook.com/blog/conjugate_prior_diagram/#geometric.

24. Smith, A.F.M., et al., *Progress with Numerical and Graphical Methods for Practical Bayesian Statistics.* Statistician, 1987. **36**(2-3): p. 75-82.

25. Naylor, J.C. and Smith, A.F.M., *Applications of a Method for the Efficient Computation of Posterior Distributions.* Applied Statistics-Journal of the Royal Statistical Society Series C, 1982. **31**(3): p. 214-225.

26. Gamerman, D. and Lopes, H.F., *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*. 2006: Chapman & Hall/CRC.

27. Muller, P., *Metropolis Based Posterior Integration Schemes.*, 1991: Purdue University.

28. Hart, A., *Knowledge Acquisition for Expert Systems*. 1986, 120 Pentonville Road, London N1 9JN: Kogan Page Ltd.

29. Welbank, M., *A Review of Knowledge Acquisition Techniques for Expert Systems*, 1983: Martlesham Heath, Ipswich, England.

30. Cowell, R.G., Dawid, A.P., Lauritzen, S.L., Spiegelhalter, D.J., *Probabilistic Networks and Expert Systems*. 1999, New York: Springer-Verlag.

31. Cooper, G.F., *Current Research Directions in the Development of Expert Systems Based on Belief Networks.* Applied Stochastic Models and Data Analysis, 1989. **5**: p. 39-52.

32. Rich, E., Knight, K., and Nair, S., *Artificial Intelligence*. 3 ed. 2009, New Delhi: McGraw-Hill.

33. Sakellaropoulos, G.C. and Nikiforidis, G.C., *Prognostic performance of two expert systems based on Bayesian belief networks.* Elsevier, 2000: p. 431 - 442.

34. Pearl, J., *Causal Diagrams for Empirical Research*, 1995, University of California: Great Britain.

35. Guo, B., *Knowledge Representation and Uncertainty Management: Applying Bayesian Belief Networks to a Safety Assessment Expert System.* 2003.

36. Pitchforth, J. and Mengersen, K., *A Proposed Validation Framework of Expert Elicited Bayesian Networks.* Elsevier, 2013.

37. Jongsawat, N. and Premchaiswadi, W., *Bayesian Network Inferene with Qualitative Expert Knowledge for Decision Support Systems.* 2010.

38. Karandikar, J.M., Abbas, A.E., and Schmitz, T.L., *Tool Life Prediction using Bayesian Updating. Part 2: Turning tool life using a Markov Chain Monte Carlo Approach.* Elsevier, 2014.

39. Karandikar, J.M., Kim, N.H., and Schmitz, T.L., *Prediction of Remaining Useful Life for Fatigue-Damaged Structures using Bayesian Inference.* Elsevier, 2012.

40. Zhu, S.P., et al., *Bayesian Framework for Probabilistic Low Cycle Fatigue Life Prediction and Uncertainty Modelling of Aircraft Turbine Disk Alloys.* Elsevier, 2013.

41. Yeratapally, S.R., Glavicic, M.G., Argyrakis, C., Sangid, M.D., *Bayesian Uncertainty Quantification and Propagation for Validation of a Microstructure Sensitive Model for Prediction of Fatigue Crack Formation.* Reliability Engineering & System Safety, 2017.

42. Franklin, R.C.G., Spiegelhalter, D.J., Macartney, F.J., Bull, K., et al., *Evaluation of a Diagnostic Algorithm for Heart-Disease in Neonates.* British Medical Journal, 1991. **302**(6782): p. 935-939.

43. Winston, P.H., *Artificial Intelligence*. 1993.

44. Zadeh, L.A., *Fuzzy Sets as a Basis for a Theory of Possibility.* Fuzzy Sets and Systems, 1978.

45. Shafer, G.R., *A Mathematical Theory of Evidence*. 1976, Princeton, New Jersey: Princeton University Press.

46. Barber, D., *Bayesian Reasoning and Machine Learning*, 2010.

47. Steyvers, M., *Computational Statistics with Matlab*. 2011.

48. O'Hagan, A., *Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment.* 2013.

49. Kadane, J.B. and Wolfson, L.J., *Experiences in elicitation.* Journal of the Royal Statistical Society Series D-the Statistician, 1998. **47**(1): p. 3-19.

50. Meyer, M.A., Booker, J.M., *Eliciting and Analyzing Expert Judgement: A Practical Guide*. 2001, Philadelphia, PA: ASA-Society of Industrial and Applied Mathematics.

51. O'Hagan, A., *Eliciting Expert Beliefs in Substantial Practical Applications.* Journal of the Royal Statistical Society Series D-the Statistician, 1998. **47**(1): p. 21-35.

52. Garthwaite, P.H., Kadane, J.B., and O'Hagan, A., *Statistical Methods for Eliciting Probability Distributions.* Journal of the American Statistical Association, 2005. **100**(470): p. 680-700.

53. Landquist, H., Norman, J., Lindhe, A., Norberg, T., Hassellov, I.M., Lindgren, J.F., Rosen, L., *Expert Elicitation for Deriving Input Data for Probabilistic Risk Assessment of Shipwrecks.* Elsevier, 2017.

54. O'Hagan, A., and Oakley, J. E. *SHELF: the Sheffield Elicitation Framework*. 2010.

55. Johnson, F.A., et al., *Expert Elicitatation, Uncertainty and the Value of Information in Controlling Invasive Species.* Elsevier, 2017.

56. Soll, J.B. and Klayman, J., *Overconfidence in Interval Estimates.* Journal of Experimental Psychology-Learning Memory and Cognition, 2004. **30**(2): p. 299-314.

57. Hogarth, R.M., *Judgement and Choice*. 1987, Chichester: John Wiley.

58. Kahneman, D. and Tversky, A., *Psychology of Prediction.* Psychological Review, 1973. **80**(4): p. 237-251.

59. Slovic, P., *From Shakespeare to Simon: Speculations - and some Evidence - about Man's Ability to Process Information.* Oregon Research Institute Monograph, 1972. **12**(2).

60. Tversky, A. and Kahneman, D., *Judgment under Uncertainty - Heuristics and Biases.* Science, 1974. **185**(4157): p. 1124-1131.

61. Tversky, A. and Kahneman D., *Extensional Versus Intuitive Reasoning - the Conjunction Fallacy in Probability Judgment.* Psychological Review, 1983. **90**(4): p. 293-315.

62. Wilson, A.G., *Cognitive Factors Affecting Subjective Probability Assessment.* 1994.

63. Kadane, J.B., et al., *Interactive Elicitation of Opinion for a Normal Linear-Model.* Journal of the American Statistical Association, 1980. **75**(372): p. 845-854.

64. Murphy, A.H. and Winkler, R.L., *Probability Forecasting in Meteorology.* Journal of the American Statistical Association, 1984. **79**(387): p. 489-500.

65. Erlick, D.E., *Absolute judgement of discrete quantities randomly distributed over time.* Journal of Experimental Psychology, 1964. **67**: p. 472-482.

66. Nash, H., *The Judgements of Linear Proportions.* American Journal of Psychology, 1964. **77**: p. 480-484.

67. Pitz, G.F., *Response Variables in the Estimation of Relative Frequency.* Perceptual and Motor Skills, 1965. **21**: p. 867-873.

68. Shuford, E.H., *Percentage Estimation of Proportion as a Function of Element, Types, Exposure Type, and Task.* Journal of Experimental Psychology, 1961. **61**: p. 430-436.

69. Beach, L.R., Swenson, R.G., *Intuitive Estimation of Means.* Psychonomic Science, 1966. **5**: p. 161-162.

70. Spencer, J., *Estimating averages.* Ergonomics, 1961. **4**: p. 317-328.

71. Peterson, C.R.M., A., *Mode, Median and Mean as Optimal Strategies.* Journal of Experimental Psychology, 1964. **68**: p. 363-367.

72. Hofstatter, P.R., *Uber die Schatzung von Gruppeneigenschaften.* Zeitschrift fur Psychologie, 1939. **145**: p. 1-44.

73. Clemen, R.T., Fischer, G.W., and Winkler, R.L., *Assessing Dependence: Some Experimental Results.* Management Science, 2000. **46**(8): p. 1100-1115.

74. Cooke, R.M., *Experts in Uncertainty: Opinion and Subjective Probability in Science*. 1991, Oxford: Oxford University Press.

75. Pill, J., *The Delphi Method: Substance, Context, a Critique, and an Annotated Bibliography*. 1971.

76. Aspinall, W., *A Route to a more Tractable Expert Advice.* Nature, 2010: p. 463, 294-295.

77. Cook, J.D. *Determining Distribution Parameters from Quantiles*. 2010.

78. Kerman, J., *A Closed-form Approximation for the Median of the Beta Distribution.* 2011.

79. van Dorp, J.R. and Mazzuchi, T.A., *Solving for the Parameters of a Beta Distribution under two Quantile Constraints.* Journal of Statistical Computation and Simulation, 2000. **67**(2): p. 189-201.

80. Bhat, H.S. and Kumar, N. *On the derivation of the Bayesian Information Criterion*.

81. Humphreys, F.J. and Hatherly, M., *Recrystallization and Related Annealing Phenomena*. Elsevier, 2004.

82. Profir, B., Eres M.H., Bates, R., Argyrakis, C., *Quantifying Uncertainties During the Early Design Stage by Utilizing a Probabilistic Bayesian Framework*, in *21st International Conference on Engineering Design*2017: Vancouver, Canada.

83. Parr, I., et al., *Inhomogenous Grain Coarsening Behavior in Supersolvus Heat Treated Nickel-Based Superalloy RR1000*, in *Superalloys 2016*2016, TMS (The Minerals, Metals and Materials Society): Warrendale, PA, USA. p. 447-456.

84. Collins, D., *Modelling and characterisation of the microstructure in a polycrystalline nickel-based superalloy*, 2011, Cambridge University, Pembroke College.

85. Galindo-Nava, E.I. and Rivera-Díaz-Del-Castillo, P.E.J., *Grain Size Evolution During Discontinuous Dynamic Recrystallization.* Scripta Materialia, 2014. **72-73**.

86. Tu, S., *The Dirichlet-Multinomial and Dirichlet-Categorical models for Bayesian inference.*

87. Groll, G.V., *Windmilling in Aero-Engines*, 2000, University of London. p. 151.

88. O'Hagan, A. and Oakley, J.E., *Probability is perfect, but we can't elicit it perfectly.* Reliability Engineering & System Safety, 2004. **85**(1-3): p. 239-248.

89. Zapata-Vazquez, R.E., O'Hagan, A., and Bastos, L.S., *Eliciting Expert Judgements about a Set of Proportions.* Journal of Applied Statistics, 2014. **41**(9): p. 1919-1933.

90. Johnson, M. *Bayesian Inference of Dirichlet-Multinomials*.

91. Huron, E.S. and Roth, P.G., *The Influence of Inclusions on Low Cycle Fatigue Life in a P/M Nickel-Base Disk Superalloy.* Superalloys, 1996.

92. Caton, M.J., Jha, S.K., Rosenberger, A.H., Larsen, J.M., *Divergence of Mechanisms and the Effect on the Fatigue Life Variability of Rene' 88 DT.* Superalloys, 2004.

93. Jha, S.K., Caton, M.J., and Larsen, J.M., *A New Paradigm of Fatigue Variability Behavior and Implications for Life Prediction.* Materials Science and Engineering, 2007.

94. Jha, M.K., Larsen, J.M., and Rosenberger, A.H., *Tracking Fatigue Life Variability of a Nearly Fully Lamellar γ-Ti-Al Based Alloy.* Materials and Metallurgical Transactions, 2004.

95. Jha, S.K., Larsen, J.M., *Dual Fatigue Failure Modes in Ti-6Al-2Sn-4Zr-6Mo and Consequences on Probabilistic Life Prediction.* Scripta Materialia, 2003. **48**.

96. Stulen, F.B. and Cummings, H.N., *Relation of Inclusions to the Fatigue Properties of High-Strength Steels*, in *Fatigue Metals*1956. p. 429-44.

97. Cummings, H.N., Stulen, F.B. and Schulte, W.C., *Fatigue Strength Reduction for Inclusions in High Strength Steels.* ASTM, 1958. **58**: p. 505-14.

98. Lankford, J. and Kusenberger, F.N., *Initiation of Fatigue Cracks in 4340 Steel.* Metall. Trans., 1973. **4**.

99. Tanaka, K. and Mura, T., *A Theory of Fatigue Crack Initiation at Inclusions.*

100. Nicholas, T., *High Cycle Fatigue - A Mechanics of Materials Perspective*. 2006, Air Force Institute of Technology, Department of Aeronautics and Astronautics, Wright Patterson AFB, Ohio, USA.

101. Box, G.E.P. and Draper, N.R., *Empirical model-building and response surface*. (Wiley series in probability and mathematical statistics. Applied probability and statistics). 1987.