An applied uncertainty analysis on the techno-economic valuation of engine wash procedures

MSc Thesis

Bram Asselman

TU Delft

An applied uncertainty analysis on the techno-economic valuation of engine wash procedures

MSc Thesis

by

Bram Asselman

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on February 24th, 2022.

Student number:	4432029	
Project duration:	February 2021 – January 2022	
Thesis committee:	Dr.ir. B.F. Lopes dos Santos	Chair & supervisor
	Dr. A. Bombelli	Examiner
	Dr.ir. M. Pini	Examiner
	A. A. Pohya, M.Sc.	DLR, Supervisor

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Acknowledgements

Dear reader,

This thesis describes the research carried out at the German Aerospace Center to obtain the degree of Master of Science in Aerospace Engineering at the Delft University of Technology.

Working in this setting has been a unique and enriching experience, which has allowed me to grow professionally and personally. For this, I would especially like to thank my supervisor at DLR, Ali Pohya. His thoughtful and generous guidance, as well as his expertise and mentorship have been invaluable in setting up and conducting this thesis research. My sincere thanks also go to my TU Delft supervisors, Marcia Baptista and Bruno Santos, for sharing their knowledge with me and providing continuous support throughout the thesis process.

Next, I would like to thank my family, and in particular my parents and sisters, who have always supported me during my time in Delft. Finally, my greatest appreciation goes to my girlfriend, Rüveyda. Thank you for your constant support and encouragement, and for the nice moments while working from home.

This thesis concludes my studies at the TU Delft. I look back on a journey of hard work and a lot of fun.

Bram Asselman
Berlin, January 2022

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	Analysis of Variance
C-MAPPS	Commercial Modular Aero-Propulsion System Simulation
CFD	Computational Fluid Dynamics
CLT	Central Limit Theorem
DES	Discrete-event Simulation
DLR	Deutsches Zentrum für Luft- und Raumfahrt
DMC	Direct Maintenance Cost
DMIM	Delta Moment Independent Method
EGT	Exhaust Gas Temperature
EGTM	Exhaust Gas Temperature Margin
EIA	US Energy Information Administration
EOT	Equivalent Operating Time
ESV	Engine Shop Visit
FOD	Foreign Object Damage
FORM	First Order Reliability Method
GSA	Global Sensitivity Analysis
HPC	High Pressure Compressor
HPT	High Pressure Turbine
LLP	Life Limited Part
LPC	Low Pressure Compressor
LPT	Low Pressure Turbine
MCS	Monte Carlo Simulation
MDO	Multi-disciplinary Design Optimisation
MLE	Maximum Likelihood Estimate
MVDO	Mixed Random and Fuzzy Variable Design Optimisation
NPV	Net Present Value
OAT	Outside Air Temperature
OoI	Object of Interest
PBDO	Probability-based Design Optimisation
PCE	Polynomial Chaos Expansion

PHM	Prognostics and Health Management
QoI	Quantity of Interest
RBD	Random Balance Design
RBDO	Reliability-based Design Optimisation
RBMDO	Reliability-based Multidisciplinary Design Optimisation
RS-HDMR	Random Sampling - High Dimensional Model Representation
RUL	Remaining Useful Life
SA	Sensitivity Analysis
SC	Stochastic Collocation
SCSA	Structural and Correlative Sensitivity Analysis
SFC	Specific Fuel Consumption
SORM	Second Order Reliability Method
TOT	Turbine Outlet Temperature
TSFC	Thrust Specific Fuel Consumption
UMDO	Uncertainty-based Multidisciplinary Design Optimisation
XAI	Explainable Artificial Intelligence

Introduction

Airlines nowadays find themselves in a competitive market and are under increasing economic pressure. Strategic decisions therefore have a crucial impact on their short- and long-term competitiveness. A priori overall economic assessments are needed to get a clear picture of the economic feasibility of future investments, for example in new technologies. Inputs to these complex assessments, generally interdisciplinary in nature, often introduce uncertainty. Combining these uncertainties can be difficult due to organisational and computational complexity. Systematic sensitivity analysis can be used to identify which inputs have a high impact on the output uncertainty and which do not, and thus can be helpful in directing model development resources. The aim of this project was to perform an uncertainty analysis, focusing on global sensitivities, for the techno-economic assessment of engine washing procedures, for which holistic uncertainty assessments are rare in the literature. At the same time, the research aims to support practitioners performing uncertainty assessments and, in particular, global sensitivity analysis. Although global sensitivity analysis techniques are well established in mathematics, their application to complex models often remains underexposed. This issue is addressed by conducting a comparative study of different global sensitivity analysis techniques and by applying a selected method to an interdisciplinary model, with the aim of increasing the comprehension of the available techniques and their suitability to different types of problems.

This research project was conducted at the German Aerospace Centre (DLR), Institute for Maintenance, Repair and Overhaul. The study was carried out without project specific data.

This thesis report is organized as follows: in Part I, the scientific paper is presented. Part II contains the Literature Study, followed by a description of the Research Methodologies in Part III. It should be noted that both the Literature Review section and the Research Methodologies section were produced at an earlier time. While they can be relevant as background information, they may not reflect the final objectives and methodology of this study. Finally, Part IV contains supporting work, including a comparative side study of global sensitivity analysis methods and the verification of the model implementation.

I

Scientific Paper

An Applied Uncertainty Analysis on the Techno-economic Valuation of Engine Wash Procedures

Bram Asselman,*

Delft University of Technology, Delft, The Netherlands
German Aerospace Center, Institute of Maintenance, Repair and Overhaul, Germany

Abstract

Overall economic assessments (OEAs) can provide a sound basis for decision-making in the areas of investments in new technologies and the application of existent technologies or operating practices. However, due to their long time horizons and complex nature, OEAs often contain many uncertain inputs, making a deterministic simulation insufficient to reflect the true value of the output. In order to incorporate these uncertainties, a systematic and efficient approach for uncertainty analysis is required. This paper sets out such a process, which consists of an iterative Uncertainty Quantification (UQ) based on importance measures for each uncertainty obtained from a Global Sensitivity Analysis (GSA). Methods for UQ and GSA are generally actively researched and well established in theory, but are infrequently applied on actual problems due to the computational and organisational complexity associated with integrative uncertainty assessments. To address this issue, the process is demonstrated on an interdisciplinary problem, namely the economic valuation of Engine Wash (EW) procedures using the cost-benefit tool LYFE. It is concluded that with this iterative uncertainty quantification procedure, the total uncertainty in the output distribution, measured using the 2.5th and 97.5th percentiles and expressed in terms of the Delta Net Present Value, is reduced from \$45K - \$983K to \$78K - \$584K. To achieve this reduction, additional modelling was carried out for only the two most important of the six uncertainties, determined using the GSA results, which illustrates the efficient allocation of modelling resources.

1 Introduction

The commercial aviation industry is highly complex and competitive, and airlines are under increasing economic pressure [Pohya et al., 2021a]. Strategic decisions (including investments in technologies, the implementation of a certain maintenance strategy, etc.) are therefore crucial for the company's competitiveness in the short and long term [Altavilla et al., 2017]. This calls for Overall Economic Assessments (OEAs) to evaluate the long-term economic effects of these decisions. Such OEAs often combine information, which is not always precisely formulated or known, from different disciplines to generate as output a value for a chosen economic metric. Including uncertainties in the assessment can provide a significant asset to the decision-maker [Uusitalo et al., 2014]. In fact, if these uncertainties are not properly taken into account, the result of the assessment will not reflect the true value (the integrative answer to the question originally posed) [Booker and Ross, 2011]. Despite the evident merits, Uncertainty Quantification (UQ) is not even common in complex assessments due to the computational and organisational complexity [Yao et al., 2011, Pohya et al., 2021b].

Sensitivity Analysis (SA) can play an important role in the efficient assessment of uncertainty, mitigating the presented burden. In the most general sense, SA can be defined as "the study of how the 'outputs' of a 'system' are related to, and are influenced by its 'inputs'" [Razavi et al., 2021]. A review on applied uncertainty and sensitivity analyses in literature found that up to "65% of the reviewed (highly cited) papers are based on inadequate methods" [Saltelli et al., 2019]. One example of bad practise often seen is to apply Local Sensitivity Analysis (LSA) to non-linear models. To explain how the output varies with a change in a given input, LSA relies on the calculation of the partial derivative of the output with respect to a given input variable, and is therefore invalid when applied to non-linear models. This can result in wrong sensitivity results and an underestimation of the output uncertainty [Saltelli et al., 2019]. In this paper Global Sensitivity Analysis (GSA) methods are studied (exploring the entire parameter space, unlike LSA), which can be used to attribute the uncertainty of the output to the individual uncertainties of the inputs and thus focus the allocation of modelling resources on the high-impact uncertainties [Roelofs and Vos, 2018]. This paper presents a process for efficient uncertainty analysis, consisting of the typical UQ procedure (including the identification, quantification and propagation of uncertainty [Booker and Ross, 2011]) and the GSA to guide further modeling effort based on the uncertainty apportionment outcome.

*Msc Student, Air Transport and Operations, Faculty of Aerospace Engineering, Delft University of Technology

The framework is applied for demonstrative purposes to the techno-economic assessment of Engine Wash (EW) procedures. Wear and tear, as well as dirt accumulation cause an aircraft to become less efficient with every flight [Fentaye et al., 2019]. This efficiency reduction leads to an increase in Exhaust Gas Temperature (EGT), thereby also increasing the fuel consumption. As this EGT has an upper limit specified by the manufacturer, the Exhaust Gas Temperature Margin (EGTM) decreases as EGT increases [Ackert, 2011]. At a certain EGTM threshold, an expensive Engine Shop Visit (ESV) is required, where a significant potion of the EGTM is restored. Due to the high cost of ESVs and engine maintenance in general (engine maintenance accounts for around 35%-40% of the Direct Maintenance Cost (DMC) [Ackert, 2011]), countermeasures are brought up. One of these is to perform on-wing engine cleaning during turnaround or overnight stop, which aims to mitigate engine deterioration. This EW procedure typically involves injecting hot water into the engine with the aim of removing accumulated dirt, resulting in fuel savings of up to 1.3% [Hutter, 2006]. However, the overall economic feasibility of EW depends on several uncertain factors, including economic factors (e.g. fuel price, price of an engine wash), physical factors (including the EGTM restoration capabilities of EW, the EGT increase, the sensitivity of the Specific Fuel Consumption (SFC) to the the EGT increase) and operational factors (for instance, the engine wash frequency). To represent these uncertainties and simulate their impact on the economic value of engine cleaning procedures, several sub-models need to be developed and integrated into the cost-benefit tool for aeronautical applications, LYFE.

This paper is structured as follows. Relevant literature and fundamentals are reviewed in section 2. The methodology is presented in section 3, followed by the analysis and results in section 4. Finally, section 5 presents the conclusions and future work.

2 Fundamentals and Literature Review

This chapter consists of a brief overview of the fundamentals in the areas of uncertainty quantification (UQ) and sensitivity analysis (SA), as well as a review of the literature on on-wing engine washing (assessments) and some of the underlying processes.

2.1 Uncertainty Assessment

Uncertainty analysis (UA) consists of propagating uncertainty from the inputs to the output, using for instance Monte Carlo simulations (MCS) [Saltelli et al., 1999]. The part where the input uncertainty is modeled, is often referred to as Uncertainty Quantification (UQ). UA should be clearly distinguished from Sensitivity Analysis (SA), where the goal generally is to attribute the output uncertainty to individual uncertainties. Techniques for UQ and SA are briefly reviewed due to their essential role in this thesis.

2.1.1 Uncertainty Quantification Theories

This section discusses UQ theories with a practise-oriented perspective.

Probability theory. A consistent framework for modeling uncertainty is available through the use of probability theory [Bishop, 2007]. Several types of interpretations on this theory have been given in the past [Cooke, 2004], among which the classical or frequentist interpretation and the subjective interpretation or Bayesian view are the most established ones. In the classical interpretation, probabilities are seen as frequencies of random events [Bishop, 2007]. It is the preferred choice when full statistical information is available [Chen et al., 1999], is straightforward to implement [Roelofs and Vos, 2018] and most decision makers and analysts are familiar with it [Booker and Ross, 2011]. However, it produces the least conservative results [Bae et al., 2004b, Roelofs and Vos, 2018], leading to a potentially false sense of exactness [Helton et al., 2004]. The Bayesian view "interprets probability in terms of degree of belief of a subject" [Cooke, 2004], and is suitable to model uncertainty due to natural randomness (aleatory) and due to a lack of knowledge (epistemic) [Yao et al., 2011, Bae et al., 2004b].

Possibility theory. Developed by Zadeh [Zadeh, 1978], this theory builds on the notion of fuzzy set theory. Citing the author, "the theory of possibility ... is related to the theory of fuzzy sets by defining the concept of a possibility distribution as a fuzzy restriction which acts as an elastic constraint on the values that may be assigned to a variable". Fuzzy sets rely on set membership, which Zadeh suggested is key to representing nonrandom and linguistic uncertainty [Ross et al., 2002]. In classical sets, the membership of an object is precise (also referred to as binary membership, i.e. the object either belongs to the set or does not). In fuzzy sets, an object can have a degree of set membership, with values on the interval [0, 1]. The bounds of this interval correspond to the binary logic used in classical sets. To represent the degree of membership in fuzzy sets, as opposed to simple binary membership for crisp or classical set, many different membership functions can be used. Compared to probability theory, it has a higher applicability to modeling rare events due to less restrictive axioms [Booker and Ross, 2011], and yields more conservative results [Chen et al., 1999]. However, the theory lacks an operational definition [Cooke, 2004] and is less understood by decision makers [Booker and

Ross, 2011].

Evidence theory. Also referred to as Dempster-Shafer theory, evidence theory, is another theory which can be used to represent uncertainty. Compared to probability theory (with its probability distributions), the axioms for basic belief assignments functions in evidence theory are less stringent [Bae et al., 2004a]. Thus, this theory is better suited for modelling imprecise knowledge [Agarwal et al., 2004, J.C.Helton et al., 2007]. However, evidence theory appears to be unreliable for highly inconsistent data [Yao et al., 2011] and, compared to probability theory, worse for decision-making in the long run [Soundappan et al., 2004] and computationally expensive [J.C.Helton et al., 2007].

2.1.2 Sensitivity Analysis

According to a common categorisation, two groups of SA are defined. In Local Sensitivity Analysis (LSA), the effect of making small changes to an input variable is found (in its simplest form) by computing the partial derivative near the instance of interest. In Global Sensitivity Analysis (GSA), the effect of larger input changes (in fact over the entire expected range of input values [Roelofs and Vos, 2018]) is considered to explain which variables strongly influence the model output [Owen, 2014].

Three groups of GSA methods can be distinguished [Iooss and Lemaître, 2015]: screening methods (which involves coarse ranking of the input variables depending on how much influence they have), measures of importance (which provide quantitative information on how influential input variables are) and deep exploration of the model behavior (i.e. going beyond scalar sensitivity indices using graphical techniques and introducing metamodel-based methods to decrease computational cost). Screening techniques can be used if one wants to identify inputs with little influence on the model output while keeping computational costs low. This is useful in many applications since a model usually only contains a small amount of influential parameters [Saltelli et al., 2008]. To understand the fractional contribution of each uncertainty to the total uncertainty, quantitative importance measures can be used, which include regression-based and distribution-based methods, among others. The most widely used distribution-based method, variance-based GSA considers the variance in the output to apportion the output uncertainty to the individual inputs. In analysis of variance (ANOVA), a function is decomposed into different components and the effect and variance of these contributing components are computed [Iooss and Lemaître, 2015]. More formally, given a square integrable function f over Ω^k, the k-dimensional unit hypercube, f can be expanded into terms of increasing dimensionality [Saltelli et al., 2008]:

$$f = f_0 + \sum_i f_i + \sum_i \sum_{j>i} f_{ij} + \ldots + f_{12\ldots k} \tag{1}$$

This expansion, also called High-dimensional Model Representation (HDMR) is unique if each term in Equation 1 has zero mean, making all terms pairwise orthogonal. It then follows that:

$$f_0 = E(Y) \tag{2}$$
$$f_i = E(Y|X_i) - f_0 \tag{3}$$
$$f_{ij} = E(Y|X_i, X_j) - f_i - f_j - f_0 \tag{4}$$

Taking the variance of the decomposed terms, V_i can be considered a measure of sensitivity and, when divided by the unconditional variance $V(y)$ yields the first-order sensitivity index S_i [Saltelli et al., 2008], i.e. the contribution of input x_i to the output variance:

$$S_i = \frac{V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i))}{V(Y)} \tag{5}$$

Higher-order sensitivity indices which represent interaction effects can be computed in a similar fashion. For a complete discussion and derivation, the reader is referred to [Saltelli et al., 2008]. It should be noted that these sensitivity measurements can be calculated using different decompositions. For example, the Fourier Amplitude Sensitivity Test (FAST), which enables faster convergence, performs a decomposition into a Fourier basis.

2.2 Engine Performance Deterioration and On-wing Engine Washing

Several engine operating parameters exist, some of which can be used as health indicators providing information about the performance deterioration level of an engine. The most important engine operating parameters are the N1-speed and the Exhaust Gas Temperature (EGT) [Ackert, 2011]. The former is mostly used to indicate the amount of thrust the engine is producing, whereas the latter can be used as a performance deterioration indicator. The EGT is the temperature measured in the exhaust of the engine. The higher this temperature, the lower the engine efficiency at producing its design trust [Ackert, 2011]. That is, lower engine efficiency (mainly caused by lower compressor efficiency) means that in order to produce equal thrust, more fuel is required [Balicki et al., 2014]. Throughout one flight cycle, the highest EGT is usually reached during take-off or initial climb. Furthermore, the EGT increases linearly with increasing Outside Air Temperature (OAT) until the corner point temperature is reached. A general increase in EGT can be observed as the engine ages and can indicate that engine hardware deterioration has occurred [Seemann, 2010, Justin and Mavris, 2015]. Some of the underlying reasons for the loss of engine efficiency are erosive wear of turbine and compressor blades, increased tip clearance of blade tips and fouling (particles deposited on blade surfaces) [Dunn et al., 1987, Lakshminarasimha et al., 1994]. As an EGT upper limit is imposed by the manufacturer to avoid damage to engine parts, the EGTM can be introduced as the difference between the EGT upper limit and the actual EGT [Ackert, 2011]. An Engine Shop Visit (ESV) is required if the EGTM or the Remaining Useful Life (RUL) of Life Limited Parts (LLPs) approaches zero. During such engine removal, a significant portion of the EGTM can be restored.

On-wing engine cleaning procedures can be employed to restore some of the lost EGTM. During engine washing, water and cleaning additives are sprayed into the intake to clean the surfaces of the compressor and potentially turbine stages, while the engine is running. Hence this maintenance task can be used to partially revert deterioration due to fouling. An illustration of the effect of regular on-wing engine washes on the Thrust-specific Fuel Consumption (TSFC) is shown in Figure 1.

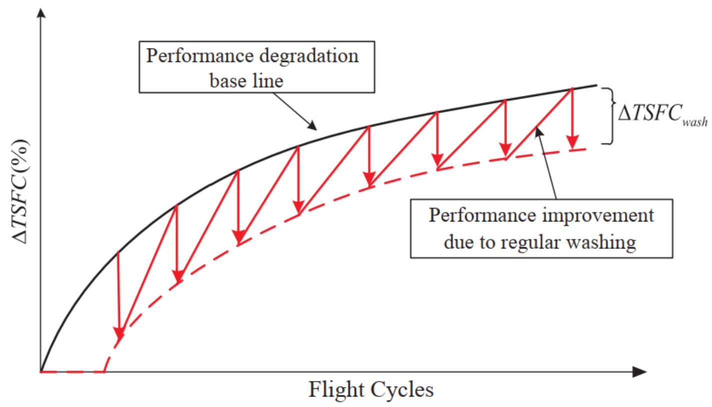


Figure 1: Effect of regular engine washing on engine performance deterioration (from [Chen and Sun, 2018]).

A critical component in simulating the effect of an EW or ESV on the engine health is the engine performance deterioration model. Multiple approaches can be taken, that differ not only in modeling complexity but also in the input and output parameters provided. Literature-based equations can be used that relate the Engine Flight Cycle (EFC) to the EGT [Justin and Mavris, 2015]. Alternatively, data points for this relation can be collected and a suitable statistical technique can be applied to model the rate of deterioration (in this case EGT increase) and the uncertainty in the results. Operation severity curves could be used to include factors like the de-rate, flight hour to flight cycle ratio and environmental factors [Seemann et al., 2011] but its output, the engine maintenance cost, could complicate the integration with other factors such as the effect of an EW. Actual physics-based, data-driven and hybrid models are developed extensively for the purpose of fault diagnostics and RUL prediction. Physics-based models were generally found too complex for this research with the EW assessment being a use-case, or were simply not publicly available. Available data sets (e.g. based on the C-MAPSS software [Chao et al., 2021]) used for training data-driven models [Alozie et al., 2019], while promising, seem not to solve the main problem due to the use of accelerated ageing. In conclusion, a simpler model, thus potentially less accurate, should be developed to represent the deterioration of engine performance so that the model can be integrated with LYFE and other uncertainties that are part of the assessment.

2.2.1 Engine Wash Assessments

Several (economic) assessments of EW procedures for gas turbines with both industrial (stationary) and aerospace applications were found in literature. Giesecke and Igie [Giesecke and Igie, 2012], for instance, highlight the economic value of compressor washing in their techno-economic study on engine compressor washes for short-range aircraft. To model the engine performance deterioration, the software TurboMatch by Cranfield University was used. The authors present the effect of fouled blades in terms of a reduction in compressor efficiency. Based on the compressor efficiency, the non-dimensional mass flow and the pressure ratio, the scaling factors are determined that simulate the deteriorated engine performance. The engine performance deterioration is assumed to be caused only by fouling. This is considered quite a radical assumption, given the numerous papers describing other relevant physical faults leading to degraded engine performance (e.g. Refs. [Diakunchak, 1992, Lakshminarasimha et al., 1994, Naeem, 2008, Döring et al., 2016, Kurz and Brun, 2001]). Other studies investigate on-wing engine cleaning from a different perspective. Boyce and Gonzalez [Boyce and Gonzalez, 2007] developed several tests in a controlled environment to determine the efficacy of engine washing with varying washing frequencies and dissolving agents used for the washing process. A washing program was then developed for a fleet of 36 industrial turbines that maximizes the engine efficiency and minimizes maintenance labor. It should be noted however that due to different operating and environmental conditions for aircraft gas turbines compared to gas turbines used for industrial applications, cost-benefit analyses of engine washes for gas turbines with industrial applications may have limited applicability to the analysis of turbofan engine wash assessments. Chen and Sun [Chen and Sun, 2018] presented an estimation method of the fuel consumption savings due to EWs, taking into account the economic cost of engine washing procedures and the fleet-wide fuel consumption savings. The assessment was applied to a case-study of a fleet of 200 CFM56 engines, and the optimal washing frequency out of six scenarios was determined. Monte Carlo simulations were used to propagate the input uncertainty to a final distribution of the fuel consumption savings. However, the consideration of uncertainty is limited to the fuel consumption. Other important uncertain factors, such as the fuel price and the effect of an EW on the engine deterioration level are not included in the analysis. In summary, uncertainties such as operational and economic factors are not sufficiently considered in EW assessments. One of the goals of this study is to include the uncertainties in a holistic way. In this respect, the proprietary cost-benefit tool LYFE, due to its interdisciplinary nature and discrete-event simulation, can offer considerable added value.

3 Methodology

As part of the systematic and efficient framework for uncertainty assessment, the methodology presented in Figure 2 is used. This framework, inspired by a commonly used approach for uncertainty analysis in Ref. [Booker and Ross, 2011] and to a larger extent by Ref. [Pohya et al., 2021b], integrates the GSA (Global Sensitivity Analysis) into the uncertainty assessment. Each step is further discussed in this section.

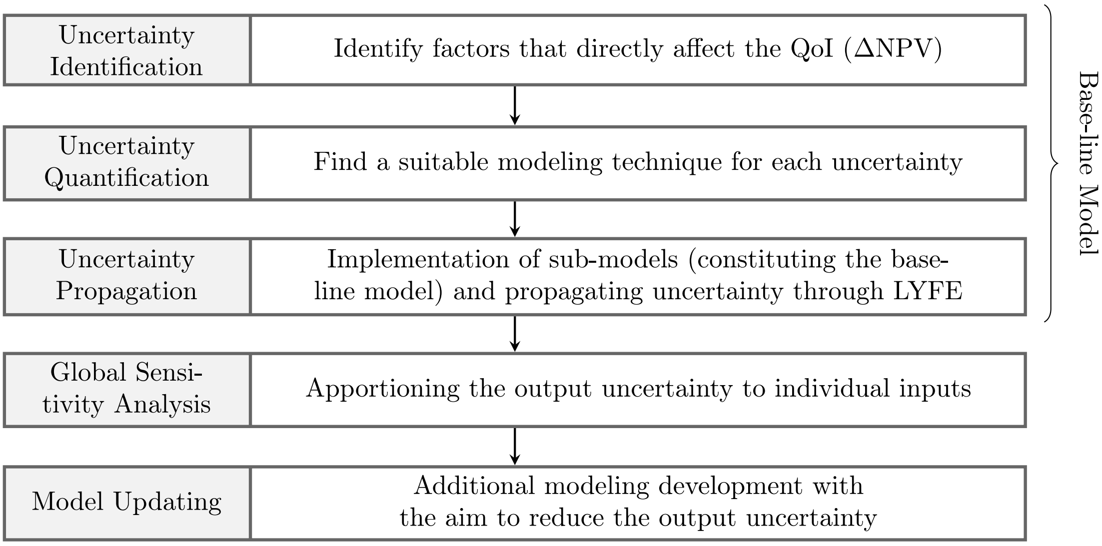


Figure 2: Steps part of the applied framework for uncertainty analysis.

3.1 Uncertainty Identification

"Uncertainty is a fact of life" [Walker et al., 2003]; identifying every uncertainty present in a complex system is an impracticable task. Assumptions allow us to restrict the considered uncertainty. In practical terms, elements (e.g. model parameters) can then be considered uncertain if they directly affect the Quantity of Interest (QoI) or are effected by the QoI, which in this study is the ΔNPV (i.e. the difference in NPV between two scenarios). Uncertainties are propagated, to the best of the assessor's ability, through an environment by considering, for example, the range of values that a parameter can take. Assumptions, much like uncertainty are documented in this work but, unlike uncertain parameters, are treated in a deterministic way. The uncertainty identification is often followed by the uncertainty classification. Here, uncertainties can be categorized, for instance, based on location (uncertainty in input data versus parameters of the model), level (how uncertain is the variable, e.g. can we list possible scenarios and assign probabilities to them?) and type (aleatory uncertainty due to natural randomness versus epistemic uncertainty). With a view to applications in decision-making, the present study contains both aleatory and epistemic uncertainties.

3.2 Uncertainty Quantification

In this phase, uncertainties are quantified using roughly equally distributed modeling and time resources, and in a conservative manner. This is to avoid spending valuable resources on modeling a potentially unimportant parameter (i.e. with low contribution to the output uncertainty). Important parameters will be later identified through the GSA, after which more effort can be put into accurately modeling the uncertainty. Based on data or knowledge from the literature, a suitable distribution is sought for each of the uncertainties identified. While different quantification theories for uncertainty are considered, probability theory was chosen to represent the input uncertainty since most decision makers and analysts are familiar with the theory [Booker and Ross, 2011] and it has a proper operational definition facilitating its interpretation [Cooke, 2004]. Alternatives such as the possibility theory and the evidence theory may be more appropriate in the case of imprecise or limited information, but they complicate effective communication and interpretation of the results for this type of assessment, as stakeholders and non-experts are generally less familiar with them [Helton et al., 2004].

3.3 Uncertainty Propagation

With the uncertainties quantified, samples are drawn from the input distributions and propagated through LYFE as shown in Figure 3.

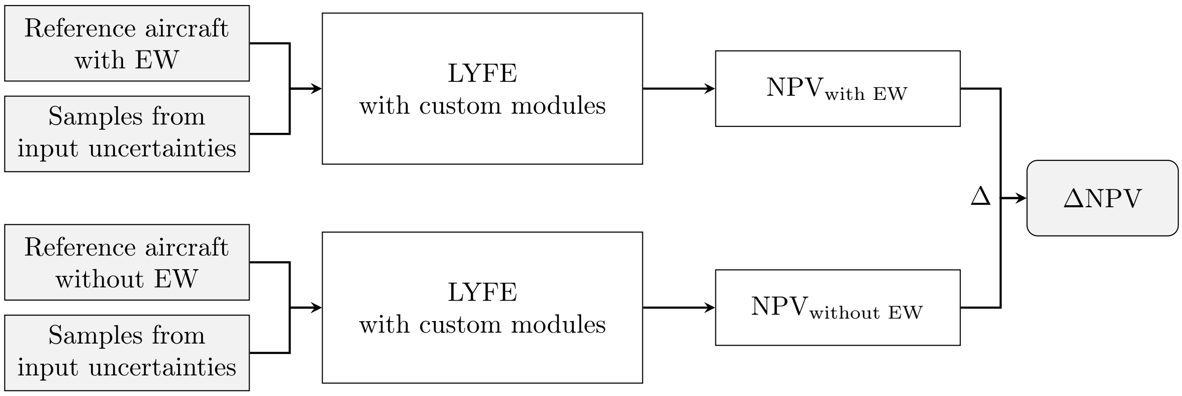


Figure 3: From input to output: the flowchart displays one full simulation (involving two calls of LYFE, one with EW and one without) to obtain the ΔNPV for a sample set. This is repeated N times to obtain the output distribution.

LYFE is a software tool developed by DLR to provide a generic (i.e. not case specific) environment for performing cost-benefit analyses of aeronautical technologies [Pohya et al., 2021a]. Its core module AirLYFE allows to evaluate these technologies, operational procedures and maintenance strategies from the operators point of view. The software uses discrete-event simulation in order to capture the primary and secondary (downstream) effects of economically relevant events over the entire life cycle. In contrast to equation-based costing tools, LYFE can therefore analyze temporal effects, e.g. delays due to unforeseen circumstances. To model the interactions between new and existing parameters, several custom modules are created such that changes to the source code are not required. A comprehensive representation of the modular program structure can be found in [Pohya et al., 2021a].

Several economic metrics are provided by the output of the simulation. In the present study the Net Present Value (NPV), which is measured as the discounted value of a project's or product's cash flows [Sobel et al., 2009], is examined. The NPV is the prevailing economic measure when making investment decisions [Graham and Harvey, 2001], and can be more formally described as [Pohya et al., 2021a]:

$$\mathrm{NPV} = \sum_{t=1}^{T} \frac{R_t - C_t}{(1+r)^t}, \tag{6}$$

where R_t and C_t represent the annual revenues and cost respectively, and r is the discount rate.

The reference aircraft used for the assessment is the Airbus A321-231, equipped with two CFM56-5B turbofan engines which have an initial EGTM of 66°C. The flight schedule used for the discrete event simulation is based on a real route network that was used in [Pohya et al., 2021a], and was originally retrieved by the author by collecting a year's worth of flight data from *flightradar.com* for a Finnair A321 aircraft. Some of the most relevant input assumptions are summarised in Table 1. Other inputs in LYFE have been set to their default value, which is intended to reflect average conditions. This is a reasonable assumption since those parameters generally do not directly affect the ΔNPV.

Table 1: Assumptions in the assessment.

Category	Parameter	Value
Aircraft	Type	Airbus A321-231
	Number of seats	209
	Engine type	CFM56-5B
	Entry into service	01/01/2025
	Termination criterion	25,000 flight cycles
Economics	Discount rate	0.10
	Inflation rate	0.02

3.4 Global Sensitivity Analysis

An introduction to GSA and derivation of variance-based sensitivity indices was provided in section 2. GSA is an active field of research and new techniques or improvements to existing techniques are constantly being developed. Due to differences in the assumptions they make and the sensitivity measure they produce (or at least the way it is computed), there is no GSA technique that is the most appropriate choice for every modelling application. This can seriously complicate the selection for a suitable technique given a particular problem. With the following brief description, a practise-oriented overview in terms of a qualitative comparison of GSA methods is given. The Python library SAlib [Herman and Usher, 2017], which is considered the most comprehensive software for SA in Python [Douglas-Smith et al., 2020], is used for this purpose mainly because it contains a wide range of different methods and can be easily integrated with the simulation framework LYFE.

Table 2 shows a selection of methods evaluated against several criteria deemed important for practitioners. Consider for instance a so-called given-data situation (where data is observed and the original distribution is not known). Generally, methods with a dedicated sampler can not be applied to such problems since they expect samples to be created using a specific sampling scheme. However, more and more methods are being developed that can be applied directly to existing data [Razavi et al., 2021], e.g. Delta Moment Independent Method (DMIM) [Plischke et al., 2013, Borgonovo, 2007] or Random Balance Design - Fourier Amplitude Sensitivity Testing (RBD-FAST) [Tarantola et al., 2006]. Dependency between inputs is another critical point (as ignoring correlation effects biases SA results [Do and Razavi, 2015]) that is receiving increasing attention in literature [Razavi et al., 2021]. Note that "the correlation effect is different from the 'interaction effect' which refers to the presence of non-additivity of the effects of individual inputs on the system output" [Razavi et al., 2021, Razavi, 2015]. Kucherenko et al. for instance realise a generalization of Sobol' indices for dependent inputs using copulae [Kucherenko et al., 2012, Razavi et al., 2021]. Li et al. [Li et al., 2010] generalise Sobol' indices using a surrogate approach called Random Sampling-High Dimensional Model Representation (RS-HDMR) to estimate the sensitivity indices. This approach is referred to as Structural and Correlative Sensitivity Analysis (SCSA) in Table 2.

Table 2: Feature comparison of GSA techniques as implemented in SALib [Herman and Usher, 2017], partly based on [Pohya et al., 2021b]. N and D represent the number of samples and the number of parameters respectively. Refs.: Sobol' [Sobol, 2001], eFAST [Saltelli et al., 1999], RBD-FAST [Tarantola et al., 2006], SCSA [Li et al., 2010], DMIM [Borgonovo, 2007], DBGSM [Sobol and Kucherenko, 2009], Morris [Morris, 1991]

	Method	**Required model evaluations**	**Given-data compatible**	**SIs provided**			**Correlated inputs**	**Supports grouping**
				S_I	S_{II}	S_T		
Distribution-based	Sobol'1	$N \cdot (D+2)$	-	✓	-	✓	-	✓
	Sobol'2	$N \cdot (2D+2)$	-	✓	✓	✓	-	✓
	eFAST	$N \cdot D$	-	✓	-	✓	-	-
	RBD-FAST	N	✓	✓	-	-	-	-
	SCSA	N	✓	✓	✓	✓	✓	-
	DMIM	N	✓		N.A.		✓	-
Derivative-based	DBGSM	$N \cdot (D+1)$	-		N.A.		-	-
	Morris	$N \cdot (D+1)$	-		N.A.		-	✓

Another important aspect to consider when selecting a GSA technique is the type of importance measure returned by the method. In models with high interaction effects between the input variables, first order sensitivity indices will only be able to explain the main effects of the inputs on the output variance. In these cases, methods providing higher-order and/or total order sensitivity indices, which represent the interaction between variables, could be a necessity. Other distribution-based methods produce different indices to explain the importance of variables. DMIM, for example, is a moment-independent method as "it measures the difference between the unconditional distribution of the output and its conditional counterparts" [Razavi et al., 2021]. Instead of only looking at the variance in the output distribution, the new measure considers the complete input/output distribution [Borgonovo, 2007].

The integration of the GSA methodology with the simulation framework is rather straightforward. The cost-benefit analysis tool LYFE is now considered a black-box model. The functionality for carrying out the variance-based GSA was implemented in a top-level script that is responsible for:

(a) generating samples according to defined probability distributions (listed in Appendix A). For GSA methods that rely on a specific sampling scheme (e.g. Sobol' and eFAST), the respective SALib sampling module is used which generates samples from the uniform distribution. To obtain samples from the desired distribution, the percentile function (which is basically the inverse cumulative distribution function) associated with that distribution is applied.

(b) calling LYFE for each sample in the sample set. This happens twice (with and without EW) per sample as depicted in Figure 3.

(c) collecting the results from all simulations.

(d) running the actual GSA. Given-data methods (which do not rely on a specific sampling sequence) rely on the input and corresponding output for the computation of the sensitivity indices. Other variance-based methods expect model outputs to be presented in a specific scheme (the Sobol' method, for instance, expects samples to be created according to the Sobol' sequence). Depending on the method used, results may include first-order, second-order and total-order sensitivity indices.

Based on a qualitative comparison of different distribution-based GSA methods presented in section 3.4, as well as a quantitative analysis of these methods on a mechanistic surrogate of LYFE[1], the Sobol' method was selected. This variance-based technique produces first- and total-order sensitivity indices and has consistently showed adequate performance in the quantitative comparison. It should be noted that for this application, several other of the methods considered meet these conditions, since there are no complicating factors such as the availability of given-data only or the presence of dependent inputs. The selected number of samples was 2048 (power of two to preserve the balance properties of the Sobol' sequence [Owen, 2020]), which results in a total of 32,768 evaluations of LYFE. These simulations are run in parallel on a workstation with an Intel Xeon Gold 6146 processor (24 cores) and 768GB RAM, and take about a day to complete.

3.5 Model Updating

Variance-based GSA allows the identification of the uncertainties with the largest contribution to the output variance. As part of an efficient uncertainty assessment, more effort can then be put into modelling the high-impact uncertainties. Literature will be revisited with the aim of making the uncertainty representation more specific or advanced (for instance by including more uncertain parameters) and thereby potentially reducing the uncertainty. This is by no means always the case, for when the complexity of the model is increased, deep uncertainty may be revealed. In that case, the state after model improvement may contain more uncertainty than before, due to previously unrecognised ignorance of the magnitude of uncertainty. To obtain an estimate for the output distribution, Monte Carlo simulations will be performed. The subsequent analysis of the changes in the overall uncertainty includes a comparison of the output distributions for the different model versions.

In conclusion, the proposed process contributes to a systematic and efficient uncertainty analysis, as the majority of development resources are spent on uncertainties that matter most (i.e. contribute the greatest amount to the uncertainty of the output).

4 Analysis and Results

The research steps from Figure 2 can be categorized into three phases which represent the high-level model development process: the base-line model development, the global sensitivity analysis and the model rework. This section is structured accordingly.

[1] A quantitative comparison of distribution-based and variance-based GSA methods, including a convergence analysis and a side study on interaction and correlation effects is available as part of the Supporting Work.

4.1 Base-line Model

Identified uncertainties are shown in Figure 4. In addition to the Engine Wash (EW) price and the EW interval, the fuel cost has a direct influence on the ΔNPV. The fuel cost is influenced by, indeed, the fuel price and the sensitivity of the Specific Fuel Consumption (SFC) to the Exhaust Gas Temperature (EGT) (which is used to translate a change in EGT into a change in fuel consumption). A more rapid deterioration in performance (or an increase in EGT, which is affect by the operational severity) will also affect fuel consumption and thus fuel cost, as will the EW effect, where a higher value is assumed to further improve fuel efficiency by increasing the EGTM. The quantification of these uncertainties is addressed next.

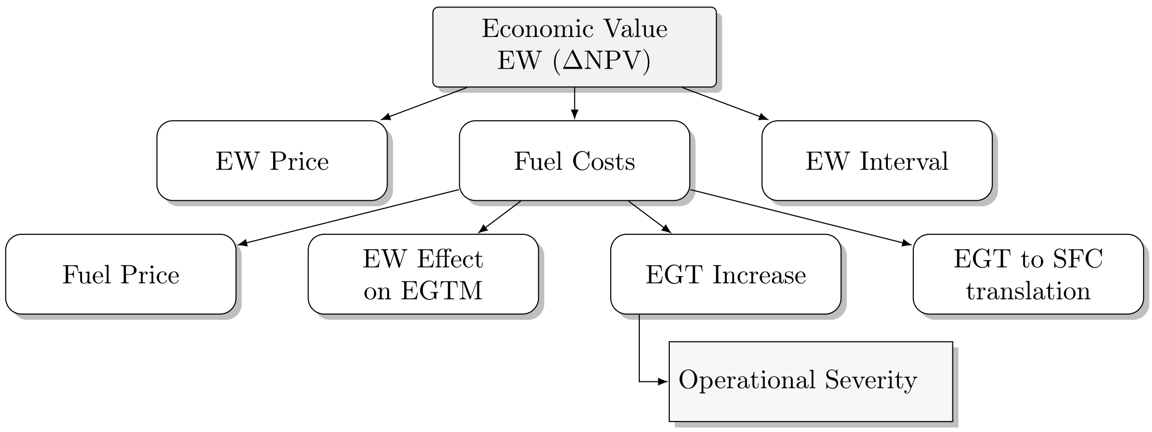


Figure 4: Breakdown of factors affecting the economic value (ΔNPV) of engine wash procedures.

4.1.1 EGT Increase

Various modeling techniques for the engine performance deterioration have been discussed in section 2. Due to the demonstrative purpose of the use-case and the need for a measure of performance deterioration per Engine Flight Cycle (EFC) and in terms of the EGT increase, data points retrieved from interviews with MRO providers [ac2, 2007a] were used as a basis to quantify the uncertainty in the deterioration profile of engine performance. This data set is very limited, both in terms of size and diversification of operators. Nevertheless, the data points can prove useful to get a general idea of the behavior of the EGT increase. Engine degradation is higher during the first few thousand EFCs and reduces as the number of EFCs increases. To present some uncertainty around these observations, a Bayesian linear regression with non-informative priors was performed. The Bayesian regression provides a more desirable interpretation than its frequentist counterpart for this use-case with a very limited sample size and trains a fully probabilistic model. 95% prediction intervals were obtained from the predictive posterior distribution and based on these a lower and upper bound were assumed for the EGT

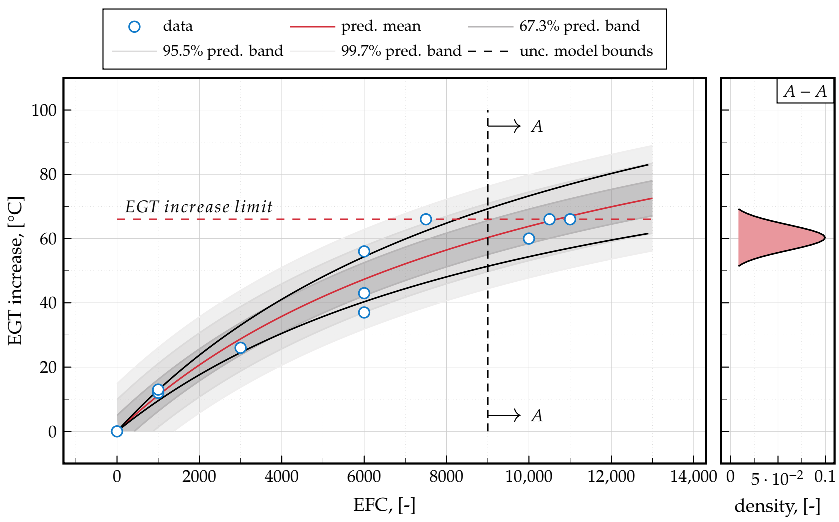


Figure 5: Left: predictions of Bayesian regression based on data points obtained from interviews with MRO providers [ac2, 2007a], along with the assumed bounds used to model the uncertainty. Right: probability density of the truncated normal distribution of EGT values at a particular EFC.

increase. Spaced vectors are then generated between the upper and lower limit. A vector is picked based on an interpolation coefficient sampled from truncated normal distribution with $\mu = 0.5$ and $\sigma = 0.2551$ and double truncation 0.025, 0.975. This corresponds to Figure 5 (right), where the eventual distribution corresponds to an EFC of 9,000.

4.1.2 EW Effect on EGTM

The variability in the EW effect is modeled using a truncated normal distribution. This assumption is based on indications in literature [Ackert, 2011, Bonnet, 2017], stating a maximum EGTM restoration value of 15°C and mean values around 7°C. However, as the engine considered in this use-case has a particularly low EGTM (66°C), the distribution was *re-scaled* based on the proportion of EGTM values of the reported engine against the CFM56-B3 of this use-case. The respective moments are shown in Figure 6.

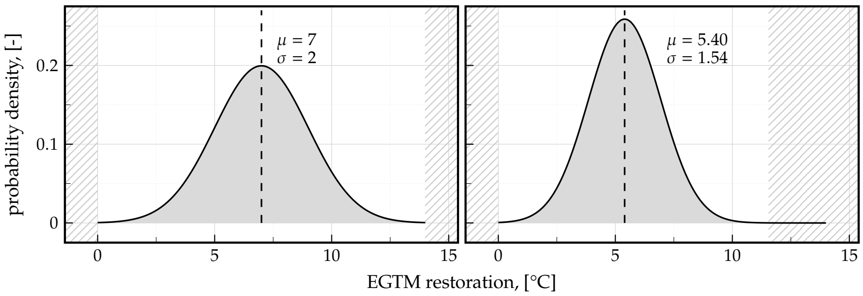


Figure 6: The effect of an EW on the EGTM is modeled using a truncated normal distribution based on [Ackert, 2011, Bonnet, 2017] (left), which was then re-scaled (right) due to the comparably low EGTM of the CFM56-B3 engine.

The EGTM restoration effect due to EW is currently constant over the entire degradation profile of the engine. This means that performing an EW right at the moment a brand new engine leaves the factory would have the same positive effect on the EGTM as washing a mature-run engine with heavily deteriorated EGTM. To overcome this unrealistic behavior, the EW effect was modeled to scale with the EGT increase as shown in Figure 7.

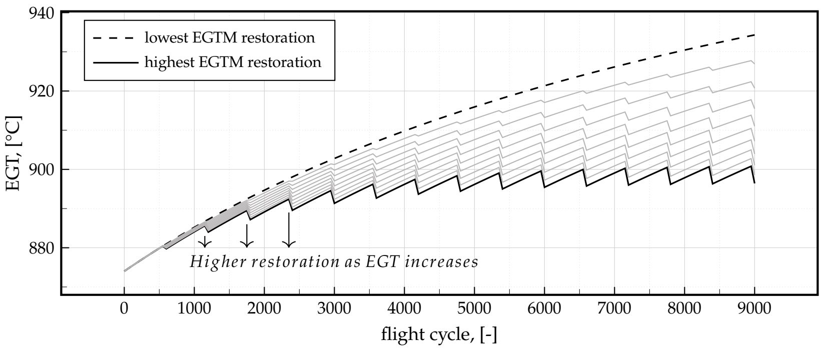


Figure 7: Varying EGTM restoration effect due to EW. For this visualization, mean EGT increase is assumed and ESVs are excluded.

4.1.3 Fuel Price

The annual energy outlook by the EIA [eia, 2021] provides various fuel price development scenarios depending on macro-economic factors. From these cases, the upper and lower extremes (referring to high and low oil prices respectively) were used as bounds for developing a set of spaced vectors, partly based on the work of Pohya et al. [Pohya et al., 2021b]. An interpolation coefficient, which determines the actual fuel price profile, is sampled from the standard uniform distribution. The uniform distribution is selected to sample the interpolation coefficient from to represent the uncertainty in a conservative way. When the left bound of the interval is sampled, i.e. zero, the fuel price development equals the lowest scenario from the EIA outlook. When the sampled value is one, the highest pricing scenario is selected. The result from this procedure together with the historical data is shown in Figure 8.

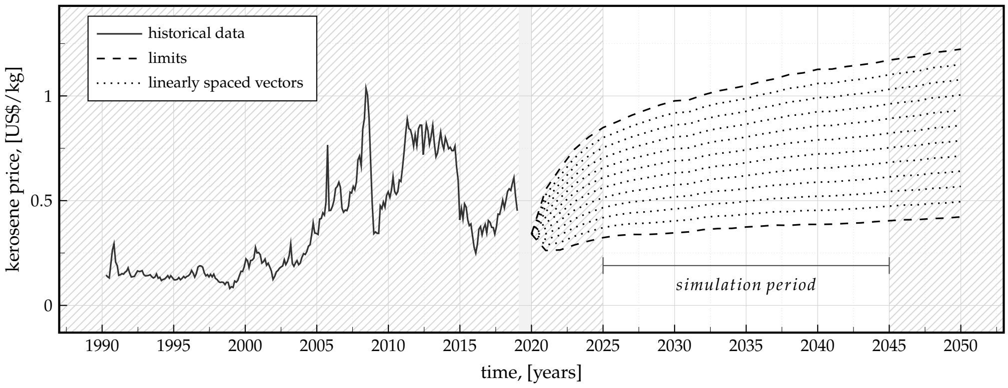


Figure 8: Historical fuel price data and linearly spaced fuel price scenarios between two extreme cases defined by the EIA [eia, 2021]. A sample drawn from a standard uniform distribution determines the pricing scenario selected for a particular simulation of LYFE.

4.1.4 Other Uncertainties

The EW interval, EW price and translation of EGT into SFC were directly modeled using uniform distributions and are thus described in brief. According to Lufthansa Technik[2], on-wing engine cleaning is performed with a defined frequency and the interval recommended by the OEM seems to be around three to eight months [Chen and Sun, 2018], which are used as bounds for the uniform distribution. This interval, however, may depend on certain constraints (e.g. environmental) faced by the operator. The EW price interval used to represent uncertainty is based on two sources (a Lufthansa Technik document[3] and an EW assessment [Chen and Sun, 2018]), which indicate the cost of engine cleaning for one engine, later corrected for the time value of money. The quantified sensitivity of SFC with respect to EGT is assumed to be between 0.0833 and 0.1 % change of SFC per °C change in EGT, based on simple equations from literature relating both parameters ([Justin and Mavris, 2015, Chen and Sun, 2018]). The statistical moments of all uncertainties as part of the base-line model are listed in Appendix A.

The modeling effort goes further than quantifying each identified uncertainty. To allow for the propagation of uncertainty, the interaction between the uncertain variables and the LYFE framework is modelled. The effect of an ESV, for example, plays an important role in the assessment, due to its ability to restore a large portion of the EGTM (between 60-80% of the original EGTM level [ac2, 2007b]). An ESV can be triggered by EGTM erosion but also by the expiry of LLPs. In the latter, the EGTM may be anywhere between original levels and zero. A fixed EGTM restoration rate per ESV, independent of the current EGT levels, would therefore largely eliminate the effect of ESV. The effect was modelled so that it varied linearly with the EGT increase, and the maximum effect corresponded to the restoration factor assigned to the particular ESV multiplied by the initial EGTM. This is visualized in Figure 9, where RF denotes the EGTM restoration factor.

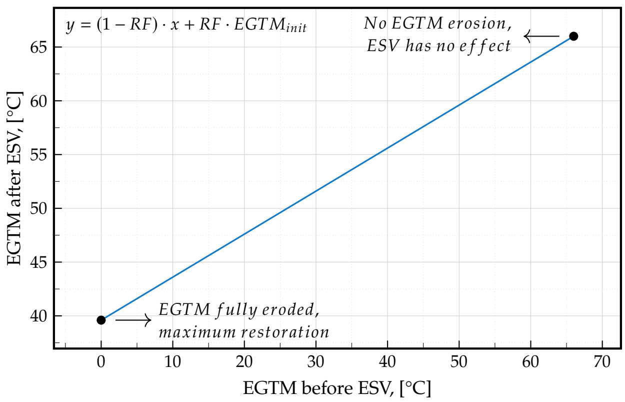


Figure 9: Modeled effect of the ESV on the EGTM for varying performance deterioration levels.

[2] *Product sheet: Cyclean Engine Wash.* Lufthansa Technik. 2019
[3] *Price List Line Maintenance Services.* Lufthansa Technik, 2012

4.1.5 Deterministic Case

With the uncertainties quantified and the interactions with LYFE implemented, which involved adding several custom modules to the simulation framework, a deterministic case is simulated with and without EW. For this reference case, the mean of the distribution for each uncertain variable was taken, which can be found in Appendix A. Therefore, these two deterministic simulations give an idea of the economic value of EW (ΔNPV) under *average conditions*, which was found to be $328K. More importantly, however, the analysis of a simulation with and without EW can give further insight into the developed custom module and the simulation framework as a whole. Figure 10 (top) shows the evolution of the EGT over the number of flight cycles. For the case without EW, the engine operates at a higher EGT for extended periods. Due to the relationship of the EGT with the SFC, this leads to an increase in fuel consumption and therefore in the fuel burned, as can be seen in the lower graph that shows the fuel savings by regularly performing EWs. This effect becomes stronger with increasing EGT difference and decreases strongly with each ESV, as during an ESV a large part of the EGTM is restored for both scenarios.

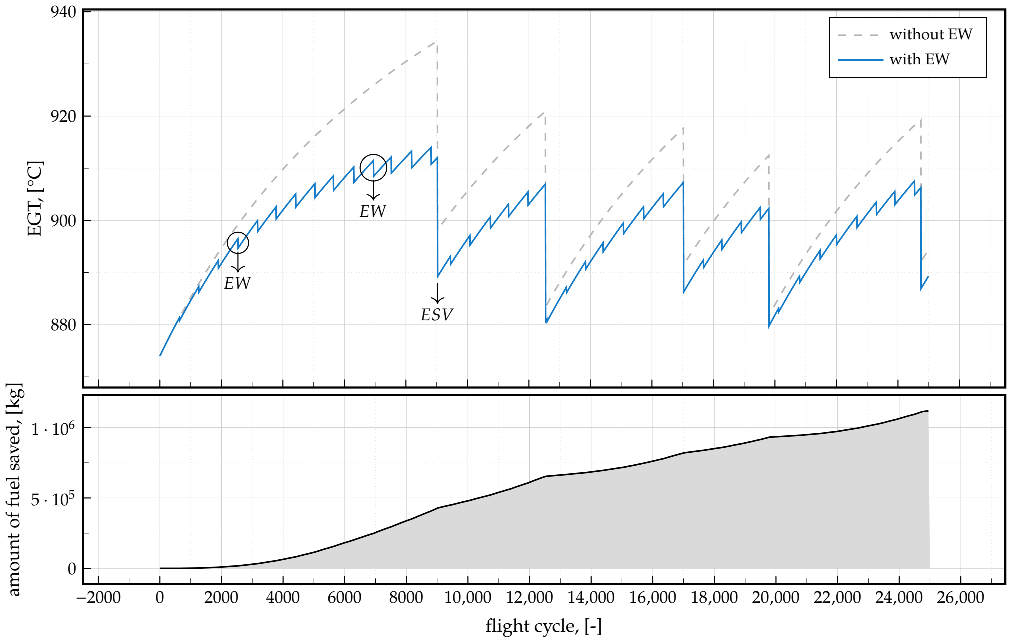


Figure 10: EGT (top) evolution for reference case with and without EW. Difference in fuel burn between the washed and unwashed case is depicted on the bottom. Visualisation inspired by [Pohya et al., 2021a].

4.2 Global Sensitivity Analysis

The resulting first and total-order sensitivity indices produced by the Sobol' method for GSA applied to the base-line model are shown in Figure 12 (the raw data can be found in Appendix B), along with the distribution of simulation outputs in Figure 11. The estimated variance in the distribution is 56.4×10^9. Although variance and mean would not be sufficient to describe the skewed distribution, variance can still be an informative measure of the degree of spread in the distribution. The Sobol' GSA method is based on the variance decomposition, which means that first order indices can be interpreted as the fractional contribution of each input to this output variance. The first and total order sensitivity of x_1: *EGT to SFC relationship* and x_2: *EW price* is below 0.01, indicating a very low contribution of these variables to the output variability. Further effort to model these parameters is not needed, as their effect on the system output as well as the potential reduction in uncertainty is minimal. x_3: *fuel price* and x_4: *EGT increase* seem to have the highest fractional contribution, with a first order sensitivity index of 0.27 and 0.38 respectively. The further development of uncertainty modelling can therefore be focused on these high-impact uncertainties in an attempt to maximise the information or precision in the output distribution for the NPV in the most efficient way. It can be concluded on the basis of the 95% confidence intervals that S_{T_i} is larger than S_i for x_5: *EW effect* and x_6: *EW interval* (and to a lesser extent for x_3: *fuel price*). This indicates the presence of interaction effects (i.e. multiple variables together affect the output through interaction). However, in this case, the order of parameters ranked by sensitivity index is identical with and without interaction effects, and inputs with a low contribution to the output uncertainty are easily identified. Therefore, no higher-order sensitivity indices are calculated.

The results of the individual simulations were analysed to gain a better understanding of the model output. Most of the conclusions below are drawn based on simulations for the lower and upper bound of each distribution, which can be accessed in full in Appendix C. Of all the uncertainties, x_4: *EGT increase* has the greatest effect, and therefore it is investigated in particular. When the EGT rises fast enough, the EGT upper limit of the engine is reached before the first planned ESV, so that an additional ESV is required. This phenomenon, caused by the simplified maintenance processes in the model, occurs less for the washed engine than for the unwashed engine (since the washed engine does not reach the EGT limit as quickly). This results in ΔNPVs on the order of a million USD due to the EGT increase alone. x_3: *fuel price*, as second most important variable, has a very large effect on the NPV, compared to the other uncertainties, with a negative NPV for the highest fuel price scenario (upper bound). The economic value of EW is five times larger for the highest fuel price than for the lowest fuel price scenario (lower bound). This is in accordance with the realization that for higher fuel prices, the positive effect of EW on the fuel consumption (and hence fuel burn) reduction becomes increasingly favorable, thereby increasing the ΔNPV. x_1: *EGT to SFC relationship*, like the fuel price, influences the ΔNPV through the total fuel costs, but contains a much smaller uncertainty range. Its effect on the ΔNPV is therefore comparably small. x_2: *EW price* also appears to have a small effect, as the EW price (even when multiplied by the number of EWs performed) is orders of magnitude lower than typical values for the NPV.

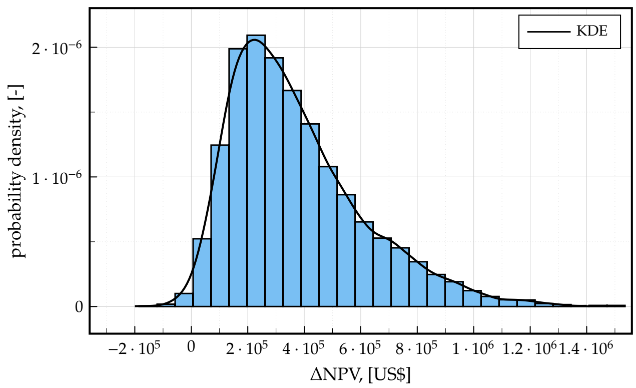


Figure 11: Output distribution for model evaluations run as part of the GSA.

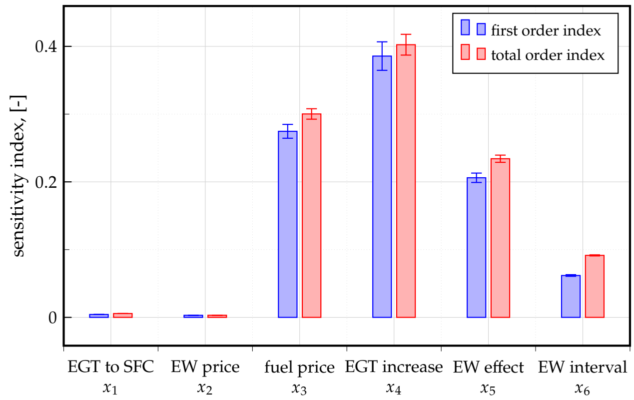


Figure 12: First and total order sensitivity indices with corresponding 95% confidence intervals obtained using the Sobol' method for GSA.

4.3 Updated Models

The uncertainty assessment so far treated the development of base-line models and the apportionment of the output uncertainty to input variables to identify high-impact variables. Based on the GSA results, the two variables with the largest first and total-order sensitivity index, the EGT increase and fuel price, are selected for *model updating*, with the goal of reducing the total uncertainty by using all available resources, thus demonstrating the process for efficient uncertainty analysis. The reworked models are then implemented in the simulation framework and the results are analysed.

4.3.1 EGT Increase Model

The operational severity strongly affects the engine maintenance costs, due its effect on the performance deterioration [Ackert, 2011]. Four important factors in determining the operational severity are: the flight length, take-off de-rate, ambient temperature and environment (a dusty or sandy environment causes higher blade distress and thus a greater decline in performance). As part of this demonstrative model update, the ambient temperature was chosen as additional modeling parameter. The main reason is that weather data from over 6000 airports around the world is available in LYFE, from which samples can be created to generate random but representative weather data [Pohya et al., 2021a]. Furthermore, the ambient temperature plays an important role in the engine degradation as a higher temperature during take-off leads to a greater deterioration in performance because the engine will operate at a lower EGTM.

To make the EGT increase dependent on the temperature, a new model was developed. As the use-case concerns a specific airline flying a specific flight schedule and a pre-determined set of airports is visited, the input to the model can be further specified. For each flight, a temperature sample was generated using the weather information for the departure airport. Based on the combined data for day and night-time, a kernel density estimation was performed to help decide on meaningful bounds for determining the EGT increase, as visualized in Figure 13.

From the estimated cumulative distribution function, the temperature values corresponding to the 2.5th and 97.5th percentile were taken as lower and upper bound for the temperature range translated to some EGT

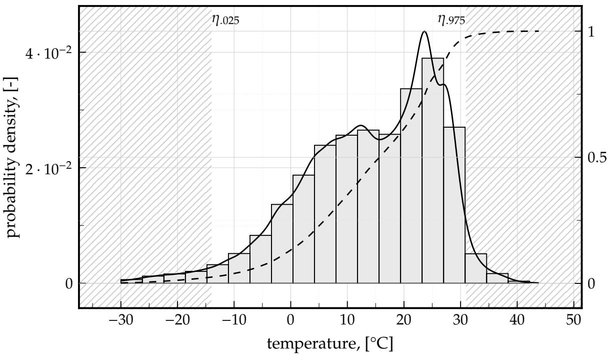


Figure 13: Estimated probability density and cumulative distribution function for temperature data of all airports available in LYFE [Pohya et al., 2021a]. Temperature values for the 2.5% (-14°C) and 97.5% (31°C) percentiles are used as bounds to represent the range of most common temperatures.

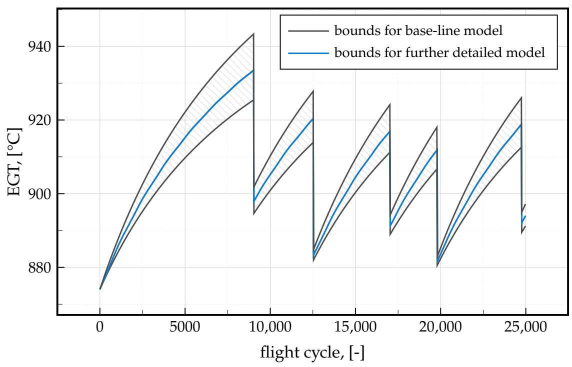


Figure 14: Uncertainty comparison between base-line (dark gray) and further specified EGT increase model (blue).

value. More precisely, the temperature scenario leading to the highest EGT increase (hence greatest performance deterioration), i.e. 31°C or above, causes an EGT increase corresponding to the upper bound step increase in the base-line EGT model at a particular flight cycle. Conversely, temperature values below -14°C are assumed to result in the lower bound EGT step increase per flight cycle. In this way, it is ensured that in the most adverse scenario (leading to the highest engine performance deterioration, i.e. temperature during departure equal to or higher than 31°C), the EGT increase is limited to the upper bound from the base-line model. In the most favorable conditions, the engine performance deterioration will be equivalent to the lower bound scenario from the base-line model.

In usual operations however, exceptional conditions (i.e. close to the temperature bounds) are interspersed with less severe operating conditions. For this use-case, knowledge of the flight schedule allows for further specification of the problem, thereby reducing the uncertainty. The flight schedule for the Finnair A321 is pre-determined, and the flight schedule is flown in exactly the same way in each simulation. The only element of uncertainty is the random component in the determination of the temperature at the airport of departure. Compared to the base-line model, the uncertainty is greatly reduced after having updated (further detailed) the model, as can be observed in Figure 14. The large uncertainty reduction can be explained as follows:

> Before updating the model, the uncertainty can be interpreted as: *EGT increase profile for any airline flying according to a random schedule with a short/medium-haul aircraft equipped with the CFM56-5B engine.*
>
> After updating the model, the scope becomes limited to: *the EGT increase for a Finnair Airbus A321 equipped with the CFM56-5B engine, flying a specific flight schedule.* With the uncertainty only originating from the temperature sampled from the normal distribution for each departure airport, it stands to reason that the new uncertainty is a fraction of what it was for the base-line model.

4.3.2 Fuel Price Model

The base-line fuel price model made use of the outlook under different macro-economic assumptions published by the EIA [eia, 2021]. The lowest and highest jet fuel price scenarios were used as lower and upper bounds for developing a set of linearly (equidistantly) spaced vectors. However, the volatility in the historical data (see Figure 15, top) is not represented by the spaced vectors that simulate the price for the following 30 years. Furthermore, this model assumes that the extreme price scenarios are as likely as the moderate scenarios, an unrealistic assumption. To address both of these model limitations, a bounded random walk model is introduced.

The historical data along with a plot of the autocorrelation function (ACF) of its first discrete difference is displayed in Figure 15. No systematic pattern can be observed in the AFC plot and almost all of the autocorrelations are within the 95% limits used to test their statistical significance (i.e. the weekly changes appear to be statistically independent), making the assumption of a drift-less random walk reasonable [Nau, 2014]. The 1-step forecast error $SE_{forecast,1}$ then equals the root-mean-square of the first discrete difference [Nau, 2014]. This results in the following random walk process:

$$y_{i+1} = y_i + \mathcal{N}(0, SE_{forecast,1}) \tag{7}$$

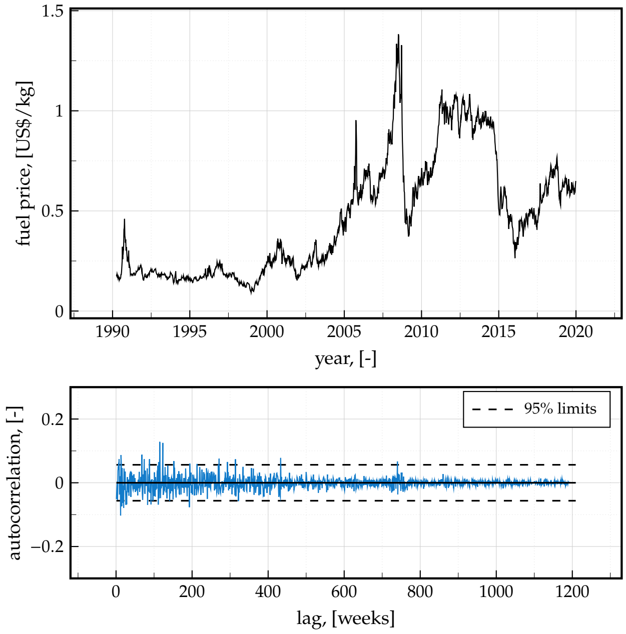


Figure 15: Historical data of kerosene prices retrieved from [eia, 2021] (top) along with the autocorrelation of its first discrete difference (bottom).

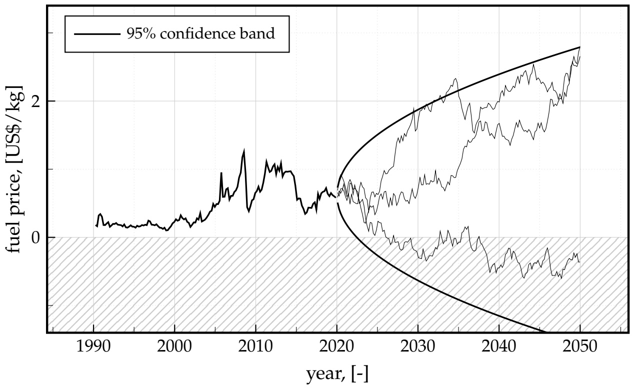


Figure 16: Historical data and random walk simulations for future fuel prices.

An example of several simulations up to the year 2050 yields the result shown in Figure 16. Note that a large part of the simulations would have negative fuel prices at some point. Also, this simple model does not take into account expert knowledge from the EIA energy outlook used in the base-line model. In order to combine both modeling methods, it was decided to bound the random simulations by the lower and upper scenario of the base-line model. The working principle is summarised in Figure 17. In brief, a random walk simulation is performed until adding the next sampled value to the current price would would exceed either one of the price limits that are based on the expert predictions. In that case, the next sample is subtracted instead of added so that the price limits are not exceeded. The process is then continued.

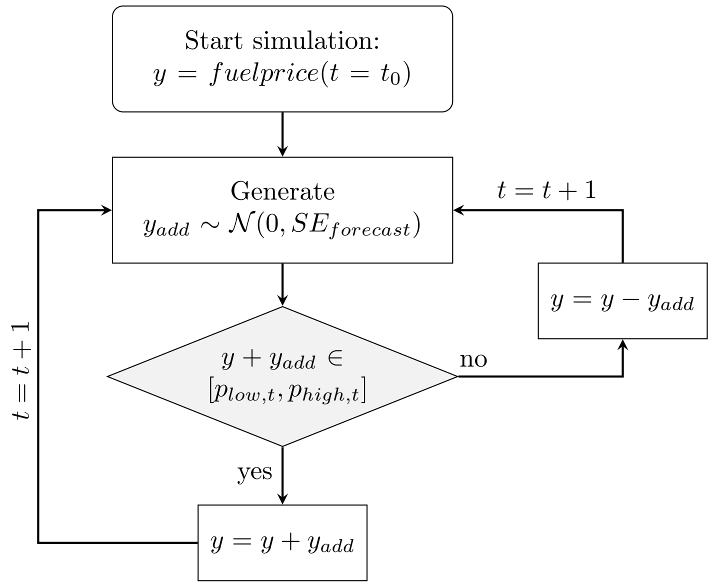


Figure 17: Fuel price simulation process. $p_{low,t}$ and $p_{high,t}$ denote the EIA [eia, 2021] predicted lower and upper fuel price scenarios at time t. The iteration starts at year 2020 and is performed for each week (which is also the resolution of the historical data).

The outlook provided by the EIA, in contrast with the historical data, has only yearly resolution. To obtain weekly resolution just like the past data, an interpolation using cubic splines was performed on the upper and lower bound scenario. This interpolation method, which involves piece-wise cubic polynomials was preferred over regular polynomial interpolation. Polynomial interpolation, in particular in this case, would be susceptible to the Runge's phenomenon (where oscillations occur at the edges of an interval) due to the high required order of the polynomial and the fact that the data points are evenly spaced across the interval [Runge, 1901]. The bounds and an example simulation of the jet fuel price is shown in Figure 18. Interestingly, although the random simulation algorithm is based on the assumption that no drift occurs, the average of a sufficiently large number of simulations at each time step appears to follow the average between the lower and upper bound (except for the first few years), creating a kind of drift. This is further investigated next with a view to the overall uncertainty.

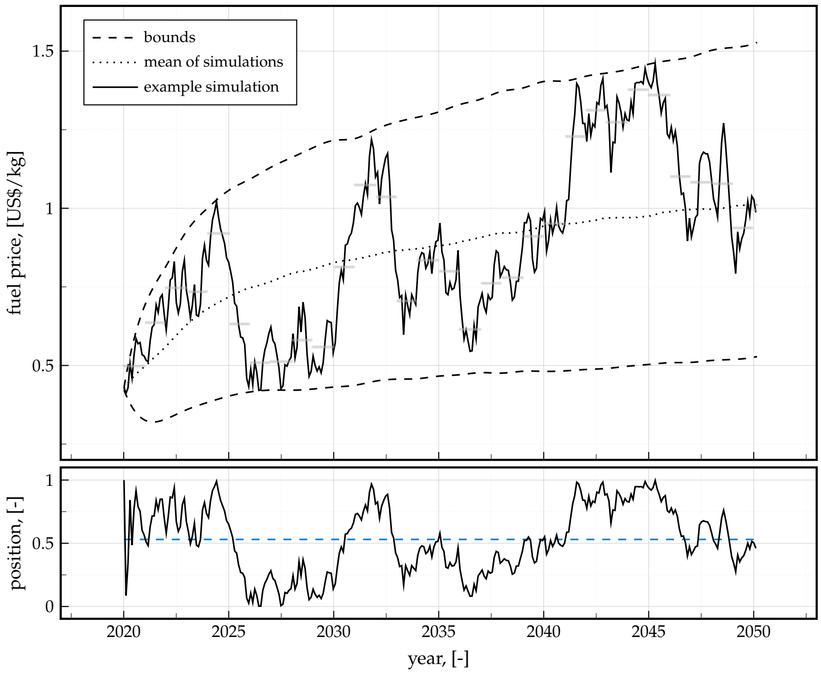


Figure 18: *Top:* bounded version of the random walk. Upper and lower bound are obtained using cubic spline interpolation on yearly data provided by the EIA [eia, 2021]. The mean curve is computed based on 5000 simulations and illustrates the drift created by the bounds. One price simulation is drawn, along with its yearly averages (light-gray horizontal bars) that are used as input into LYFE. *Bottom:* position of the fuel price relative to the lower and upper bound, used for uncertainty comparison.

As the fuel price base-line model and improved model make use of the same bounds, the uncertainty can be directly quantified and compared. The insight from this comparison helps to verify the model implementation. In the base-line model, linearly spaced vectors between the lower and upper bound represent the fuel pricing. That is, each vector maintains the same ratio between the lower and upper bound. In the improved model, due its random nature, this is not the case. Therefore, for each simulation, the average position between upper and lower bound is computed. This is shown for the example simulation in the bottom plot of Figure 18. A kernel density estimation was applied to an array consisting of this particular 'positional mean' for a large number of simulations (200,000) and is shown in Figure 19 next to the uniform distribution used in the base-line model. Since the density is now more concentrated around the mean value, the total uncertainty after propagation is expected to decrease.

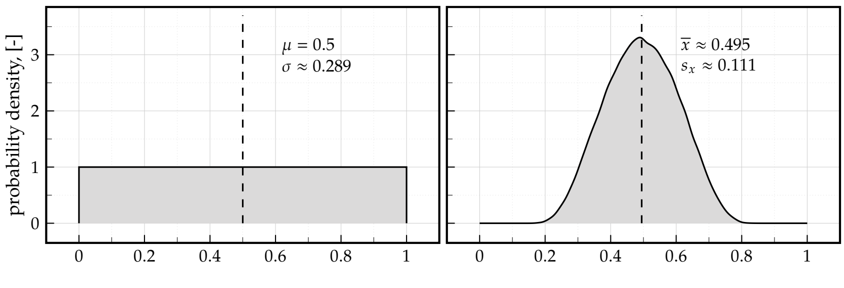


Figure 19: Probability density function for the mean position with respect to lower and upper bound of the fuel price, for base-line model (left) and improved model (obtained using kernel density estimation, right).

4.3.3 Simulation Results

Both updated models were implemented in the discrete-event simulation and can be enabled or disabled (in which case the previous version is used) to be able to assess the results for different versions of the model. Although it is not the intention to perform a GSA at this stage, it is worth noting that due to the random nature of the fuel price model (i.e. the price evolves in a random manner), GSA should not be run with this updated version of the models as the analysis relies on the input and output to determine the contribution of each input uncertainty to the total uncertainty. Instead, Monte Carlo simulations are performed to estimate the output distribution. Figure 20 depicts the output distribution the base-line model, and the updated versions of the fuel price and EGT increase model. 95% confidence intervals of the three parameters under consideration, $\eta_{.025}$ (2.5th percentile), μ (mean), $\eta_{.975}$ (97.5th percentile), are computed using bootstrapping due to the rather limited amount of samples (N = 2000) used to estimate the output distributions. Approximate confidence intervals were obtained using the *percentile method*, which uses the shape of the bootstrap distribution and has been reported to require bootstrap sample sizes in the order of N = 2000 [Efron and Hastie, 2016].

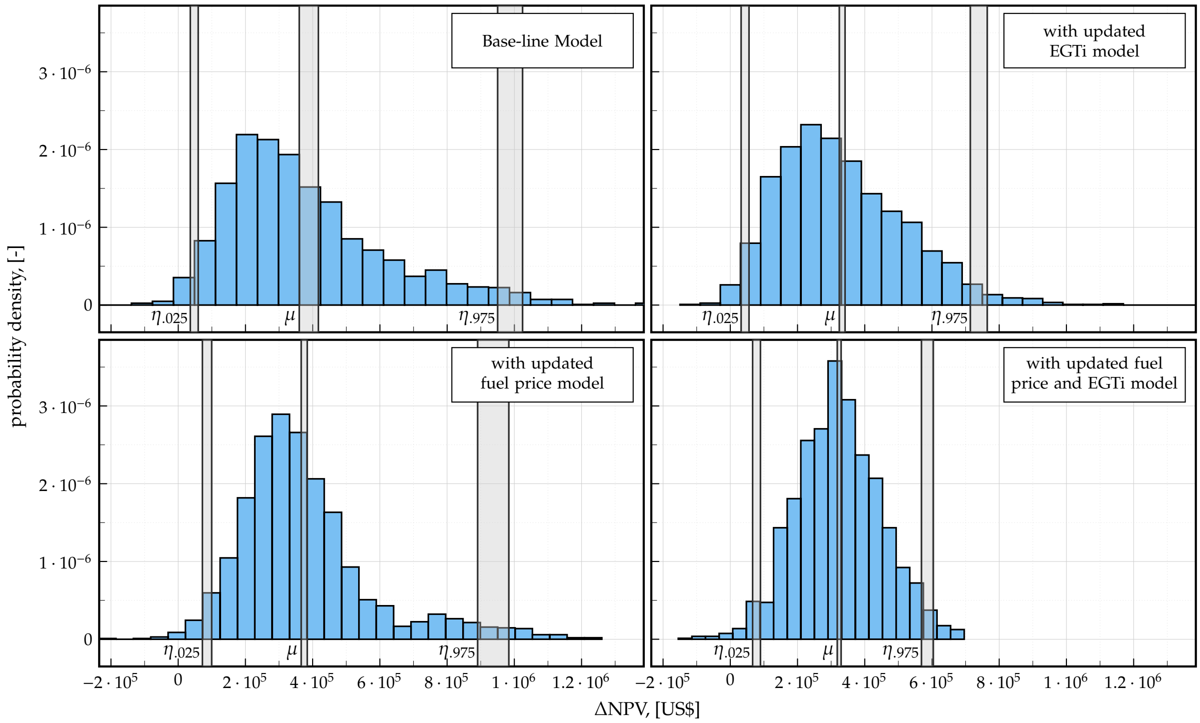


Figure 20: Comparison of estimated output distributions for different model versions. Number of samples N = 2000; the 2.5th and 97.5th percentile are shown with their 95% confidence bounds obtained by bootstrapping.

Using the updated EGT increase model, values for ΔNPV above 6×10^5 become significantly less frequent than for the base-line model. This is also concluded by the comparison of $\eta_{.975}$, which goes from around $1,000K to $730K. In the updated model, the vertical dispersion of the EGT increase simulations is indeed only a fraction of that in the base-line version, as visualized in Figure 14. One of the reasons why the estimated output distribution for the base-line EGT increase model is more fat-tailed than for the updated model, is the fact that the EGT exceeds its limit value before the first scheduled ESV more frequently (caused by a faster engine deterioration rate), requiring one additional expensive ESV. When this happens for the unwashed engine and not for the washed engine, the ΔNPV indeed rises.

The estimated distribution with the updated fuel price model shows a similar range of values as for the base-line model (especially the shape of the distribution seems well preserved), as indicated by the small difference in $\eta_{.025}$ between the two versions, and $\eta_{.975}$ not being statistically different. However, the distribution shows that some of the density in the tails (especially the right tail) of the distribution has shifted towards the mode.

When both the updated fuel price and EGT increase model are used the long right tail of the distributions, which was present in any of the previous model versions considered, disappears. Quantitatively, we observe that $\eta_{.975}$ decreases from around $1,000K to $600K and no values for ΔNPV exist above $700K. Negative ΔNPV values are rare, as is the case for all four versions of the model, which is indicated by the confidence intervals for $\eta_{.025}$ that are always positive. In this case, the uncertainty reduction provides a higher level of information for decision makers, compared to the output distribution for the base-line model. It should be noted, that when the base-line distribution is less decisive towards either positive or negative ΔNPVs, the uncertainty reduction can have a more crucial impact on the decision making process.

In conclusion, estimated percentiles were used to analyse the output of Monte Carlo simulations for different model versions and determine which of the model updates had the greatest impact on the uncertainty reduction. The biggest contribution to this is made by the updated EGT increase model and can be attributed to two main reasons. Firstly, the GSA showed that the uncertainty in the EGT increase contributes to almost 40% of the output variance, with the result that the uncertainty reduction in it has a large impact on the overall uncertainty. Secondly and probably more importantly, it was already noticed that the EGT increase uncertainty reduction is more drastic than for the updated fuel price model model, compared to their respective base-line model.

5 Conclusion and Discussion

Techniques for Uncertainty Analysis and Global Sensitivity Analysis (GSA) are well established in math and are an active field of research. However, their application on actual problems is less frequently looked at, in part due to the difficulties that arise when using these tools in uncertainty management with complex models or a large number of uncertain variables. This work sets out a process for systematic and efficient uncertainty analysis to address this issue, and applies the process on a use-case with a complex simulation environment. The procedure consists of three main steps. First, uncertainties are identified and then quantified using roughly equally distributed modeling and time resources. Next, variance-based GSA is used to apportion the output uncertainty to the each of the input uncertainties and thereby identify high-impact inputs. Finally, further model development is focused on these important variables with the aim to reduce the overall uncertainty.

The framework for uncertainty analysis was applied to the techno-economic assessment of engine cleaning procedures, where the output is measured in terms of ΔNPV. Six uncertainties were identified and quantified using simple probability distributions. It was estimated that for the base-line model the 2.5th and 97.5th percentile for the ΔNPV equal \$45K and \$983K respectively. Of the considered uncertainties, the Exhaust Gas Temperature (EGT) increase and the fuel price were identified by the GSA as having the highest impact on the output uncertainty, with total-order sensitivity indices of 0.402 and 0.274 respectively. Their uncertainty representation and interaction with the simulation framework were then reworked. For the fuel price, a random walk model was combined with a forecast made by the EIA to simulate realistic fuel price behaviour. The EGT increase model was made temperature-dependent and further specified by realizing a pre-determined flight schedule is flown and by incorporating airport weather data in the model. With both updated models integrated into the simulation model, the estimated 2.5th and 97.5th percentile of the output distribution are now \$78K and \$584K, with bootstrapping used to indicate statistical significance of the simulation results. Due to the iterative process for uncertainty quantification, the total uncertainty is greatly reduced while additional modeling was only carried out for two out of six uncertainties based on the GSA results.

Due to the demonstrative purpose of the use-case, the focus of the analysis did not lie on the accuracy of the models. However, especially considering the lack of integrative uncertainty assessments on the subject of engine cleaning in literature, this study can help to promote further research in this field. Next steps could be to improve the overall accuracy of the model by collecting more data for each of the uncertainties considered, possibly in combination with the use of a physics-based approach (e.g. to model the degradation of engine performance). This would then permit a greater focus on the actual outcome of the economic value of the engine cleaning procedures in terms of the ΔNPV.

This methodology presented is generic in nature and can therefore be applied to any uncertainty assessment as long as input data to the model is available or input distributions are known and the model output is clearly defined (i.e. it answers a specific question). The framework can help save costly modeling resources by identifying parameters with low importance early in the uncertainty analysis process and focus model development on parameters that actually affect the system output. In addition to the applied nature of this research, which should encourage systematic uncertainty analysis and global sensitivity analysis in complex simulation environments, it is also addressed to practitioners in the field for example through the qualitative and quantitative (see Supporting Work) comparison of methods for GSA. Many of the problems and insights described in applying this framework are also applicable to other use cases with a potentially complex character, and thereby shall lead to a better overall comprehension of the value of GSA and its major role in an efficient uncertainty management process.

Further work regarding the uncertainty modeling may involve more extensive use of Bayesian inference or non-probabilistic methods such as possibility theory or evidence theory (assuming an operational definition is well defined) to deal with limited or imprecise information. Another limitation of this study is that independence between inputs was assumed due to the limited data available. Dependence between inputs is an additional challenge for the GSA, as most methods assume independence between inputs and the interpretation of the resulting indices can become more complicated. It appears that although several methods have been developed for correlated inputs, their operating principles vary widely and further work is needed to make them more accessible to practitioners.

References

[ac2, 2007a] (2007a). CFM56-5A/-5B maintenance analysis & budget. *Aircraft Commerce*, (50):15–27.

[ac2, 2007b] (2007b). CFM56-5A/-5B series specifications. *Aircraft Commerce*, (50):6–9.

[eia, 2021] (2021). Petroleum and other liquids. U.S. Energy Information Administration. https://www.eia.gov/dnav/pet/PET_PRI_SPT_S1_D.htm. Accessed: 2021-09-30.

[Ackert, 2011] Ackert, S. (2011). Engine maintenance concepts for financiers - elements of turbofan shop maintenance costs. Technical Report. https://www.icao.int/Meetings/EnvironmentalWorkshops/Documents/ICAO-TransportCanada-2006/Hutter.pdf. Accessed: 2021-09-10.

[Agarwal et al., 2004] Agarwal, H., Renaud, J. E., Preston, E. L., and Padmanabhan, D. (2004). Uncertainty quantification using evidence theory in multidisciplinary design optimization. *Reliability Engineering and System Safety*, 85(1):281–294.

[Alozie et al., 2019] Alozie, O., Li, Y.-G., Wu, X., Shong, X., and Ren, W. (2019). An adaptive model-based framework for prognostics of gas pathfaults in aircraft gas turbine engines. *International Journal of Health Management*, 10(2).

[Altavilla et al., 2017] Altavilla, S., Montagna, F., and Cantamessa, M. (2017). A Multilayer Taxonomy of Cost Estimation Techniques, Looking at the Whole Product Lifecycle. *Journal of Manufacturing Science and Engineering*, 140(3).

[Bae et al., 2004a] Bae, H.-R., Grandhi, R. V., and Canfield, R. A. (2004a). An approximation approach for uncertainty quantification using evidence theory. *Reliability Engineering System Safety*, 86(3):215–225.

[Bae et al., 2004b] Bae, H.-R., Grandhi, R. V., and Canfield, R. A. (2004b). Epistemic uncertainty quantification techniques including evidence theory for large-scale structures. *Computers and Structures*, 82(13-14):1101–1112.

[Balicki et al., 2014] Balicki, W., Glowacki, P., Szczecinski, S., Chachurski, R., and Szczecinski, J. (2014). Effect of the atmosphere on the performances of aviation turbine engines. *acta mechanica et automatica*, 8(2):70–73.

[Bishop, 2007] Bishop, C. M. (2007). *Pattern Recognition and Machine Learning*. Springer.

[Bonnet, 2017] Bonnet, A. (2017). Avoiding high speed rejected take-offs due to egt limit exceedance. https://safetyfirst.airbus.com/avoiding-high-speed-rejected-takeoffs-due-to-egt-limit-exceedance/. Accessed: 2021-10-05.

[Booker and Ross, 2011] Booker, J. M. and Ross, T. J. (2011). An evolution of uncertainty assessment and quantification. *Scientia Iranica*, 18(3):669–676.

[Borgonovo, 2007] Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering and System Safety*, 92(6):771–784.

[Boyce and Gonzalez, 2007] Boyce, M. P. and Gonzalez, F. (2007). A study of on-line and off-line turbine washing to optimize theoperation of a gas turbine. *Journal of Engineering for Gas Turbines and Power*, 129(1).

[Chao et al., 2021] Chao, M. A., Kulkarni, C. S., Goebel, K., and Fink, O. (2021). Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics. *Prognostic Data Sets*, 6(1).

[Chen and Sun, 2018] Chen, D. and Sun, J. (2018). Fuel and emission reduction assessment for civil aircraft enginefleet on-wing washing. *Transportation Research Part D*, 65:324–331.

[Chen et al., 1999] Chen, S., Nikolaidis, E., and Cudney, H. H. (1999). Comparison of probabilistic and fuzzy set methods for designing under uncertainty. In *40th Structures, Structural Dynamics, and Materials Conference and Exhibit*.

[Cooke, 2004] Cooke, R. (2004). The anatomy of the squizzel: The role of operational definitions in representing uncertainty. *Reliability Engineering System Safety*, 85(1-3):313–319.

[Diakunchak, 1992] Diakunchak, I. S. (1992). Performance deterioration in industrial gas turbines. *Journal of Engineering for Gas Turbines and Power*, 114(2):161–168.

[Do and Razavi, 2015] Do, N. C. and Razavi, S. (2015). Correlation effects? a major but often neglected component in sensitivity and uncertainty analysis. *Water Resources Research*, 56(3).

[Douglas-Smith et al., 2020] Douglas-Smith, D., Iwanaga, T., Croke, B. F. W., and Jakeman, A. J. (2020). Certain trends in uncertainty and sensitivity analysis: An overview of software tools and techniques. *Environmental Modelling and Software*, 124(19).

[Dunn et al., 1987] Dunn, M. G., Padova, C., Moller, J. E., and Adams, R. M. (1987). Performance deterioration of a turbofan and a turbojet engine upon exposure to a dust environment. *Journal of Engineering for Gas Turbines and Power*, 109(3):336–343.

[Döring et al., 2016] Döring, F., Staudacher, S., Koch, C., and Weißschuh, M. (2016). Modeling particle deposition effects in aircraft engine compressors. *Journal of Turbomachinery*, 139(5).

[Efron and Hastie, 2016] Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge University Press, USA.

[Fentaye et al., 2019] Fentaye, A. D., Baheta, A. T., Gilani, S. I., and Kyprianidis, K. G. (2019). A review on gas turbine gas-path diagnostics:state-of-the-art methods, challenges and opportunities. *Aerospace*, 6(7).

[Giesecke and Igie, 2012] Giesecke, D. and Igie, U. (2012). Performance and techno-economic investigation of on-wing compressor wash for a short-range aero engine. In *ASME Turbo Expo 2012: Turbine Technical Conference and Exposition*.

[Graham and Harvey, 2001] Graham, J. R. and Harvey, C. R. (2001). The theory and practice of corporate finance: evidence from the field. *Journal of Financial Economics*, 60(2):187–243.

[Helton et al., 2004] Helton, J. C., Johnson, J. D., and Oberkampf, W. L. (2004). An exploration of alternative approaches to the representation of uncertainty in model predictions. *Reliability Engineering and System Safety*, 85(1-3):39–71.

[Herman and Usher, 2017] Herman, J. and Usher, W. (2017). Salib: An open-source python library for sensitivity analysis. *Journal of Open Source Software*, 2(9):97.

[Hutter, 2006] Hutter, I. (2006). Engine deterioration and maintenance actions. `https://www.icao.int/Meetings/EnvironmentalWorkshops/Documents/ICAO-TransportCanada-2006/Hutter.pdf`. Presentation. Accessed: 2021-08-15.

[Iooss and Lemaître, 2015] Iooss, B. and Lemaître, P. (2015). A review on global sensitivity analysis methods. *Operations Research and Computer Science Interfaces Series, Springer*.

[J.C.Helton et al., 2007] J.C.Helton, Oberkampf, J. D. J. W. L., and Storlie, C. B. (2007). A sampling-based computational strategy for the representation of epistemic uncertainty in model predictions with evidence theory. *Computer Methods in Applied Mechanics and Engineering*, (37-40):3980–3998.

[Justin and Mavris, 2015] Justin, C. Y. and Mavris, D. N. (2015). Aircraft and engine economic evaluation for fleet renewal decision-making and maintenance contract valuation. *Proceedings of the Institution of Mechanical Engineers Part G Journal of Aerospace Engineering*, 229(11).

[Kucherenko et al., 2012] Kucherenko, S., Tarantola, S., and Annoni, P. (2012). Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, 183(4):937–946.

[Kurz and Brun, 2001] Kurz, R. and Brun, K. (2001). Degradation in gas turbine. *ASME Journal of Engineering for Gas Turbines and Power*, 123(1):70–77.

[Lakshminarasimha et al., 1994] Lakshminarasimha, A. N., Boyce, M. P., and Meher-Homji, C. B. (1994). Modeling and analysis of gasturbine performance deterioration. *Journal of Engineering for Gas Turbines and Power*, 116(1).

[Li et al., 2010] Li, G., Rabitz, H., Yelvington, P. E., Oluwole, O. O., Bacon, F., Kolb, C. E., and Schoendorf, J. (2010). Global sensitivity analysis for systems with independent and/or correlated inputs. *The Journal of Physical Chemistry A*, 114:6022–6032.

[Morris, 1991] Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174.

[Naeem, 2008] Naeem, M. (2008). Impacts of low-pressure (lp) compressors' fouling of aturbofan upon operational-effectiveness of a military aircraft. *Applied Energy*, 85(4):243–270.

[Nau, 2014] Nau, R. (2014). Notes on the random walk model. Technical report, Fuqua School of Business, Duke University. Accessed: 2021-11-14.

[Owen, 2014] Owen, A. B. (2014). Sobol' indices and shapley value. *Journal of Uncertainty Quantification*, 2:245–251.

[Owen, 2020] Owen, A. B. (2020). On dropping the first sobol' point.

[Plischke et al., 2013] Plischke, E., Borgonovo, E., and Smith, C. L. (2013). Global sensitivity measures from given data. *European Journal of Operational Research*, 226(3):536–550.

[Pohya et al., 2021a] Pohya, A. A., Wehrspohn, J., Meissner, R., and Wicke, K. (2021a). A modular and lifecycle based discrete event simulation framework for quantifying the true value of aircraft technologies, maintenance strategies, and operational procedures. *Advances in Engineering Software.*

[Pohya et al., 2021b] Pohya, A. A., Wicke, K., and Kilian, T. (2021b). Introducing variance-based global sensitivity analysis for uncertainty enabled operational and economic aircraft technology assessment. *Aerospace Science and Technology.* submitted.

[Razavi, 2015] Razavi, S. (2015). What do we mean by sensitivity analysis? the need for comprehensive characterization of global sensitivity in earth and environmental systems models. *Water Resources Research*, 51(5).

[Razavi et al., 2021] Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Piano, S. L., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J. H. A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., Smith, S., Sheikholeslami, R., Hosseini, N., Asadzadeh, M., Puy, A., Kucherenko, S., and Maier, H. R. (2021). The future of sensitivity analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling Software*, 137.

[Roelofs and Vos, 2018] Roelofs, M. and Vos, R. (2018). Technology evaluation and uncertainty-based design optimization: A review. *American Institute of Aeronautics and Astronautics Inc.*

[Ross et al., 2002] Ross, T. J., Booker, J. M., and Parkinson, W. J. (2002). *Fuzzy Logic and Probability Applications: Bridging the Gap.*

[Runge, 1901] Runge, C. T. (1901). Über empirische funktionen und die interpolation zwischen äquidistanten ordinaten. *Zeitschrift für Mathematik und Physik*, 46:224–243.

[Saltelli et al., 2019] Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., and Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling and Software*, 114:29–39.

[Saltelli et al., 2008] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Caribonia, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer.*

[Saltelli et al., 1999] Saltelli, A., Tarantola, S., and Chan, K. P. S. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1):39–56.

[Seemann, 2010] Seemann, R. (2010). Modeling the life cycle costof jet engine maintenance. M.Sc. Thesis.

[Seemann et al., 2011] Seemann, R., Schilling, T., Langhans, S., and Gollnick, V. (2011). Modeling the life cycle cost of jet engine maintenance. In *60. Deutscher Luft- und Raumfahrtkongress (DLRK).*

[Sobel et al., 2009] Sobel, M. J., Szmerekovsky, J. G., and Tilson, V. (2009). Scheduling projects with stochastic activity duration to maximize expected net present value. *European Journal of Operational Research*, 198(3):697–705.

[Sobol, 2001] Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280.

[Sobol and Kucherenko, 2009] Sobol, I. M. and Kucherenko, S. (2009). Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulations*, 79(19):3009–3017.

[Soundappan et al., 2004] Soundappan, P., Nikolaidis, E., Haftka, R. T., Grandhi, R., and Canfiel, R. (2004). Comparison of evidence theory and bayesian theory for uncertainty modeling. *Reliability Engineering System Safety*, 85(1-3):295–311.

[Tarantola et al., 2006] Tarantola, S., Gatelli, D., and Mara, T. A. (2006). Random balance designs for the estimation of first order global sensitivity indices. *Reliability Engineering and System Safety*, 91(6):717–727.

[Uusitalo et al., 2014] Uusitalo, L., Lehikoinen, A., Helle, I., and Myrberg, K. (2014). An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environmental Modelling and Software*, 63:24–31.

[Walker et al., 2003] Walker, W. E., Harremoes, P., Rotmans, J., Sluijs, J. P. V. D., Asselt, M. B. A. V., Janssen, P., and Krauss, M. P. K. V. (2003). Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1):5–17.

[Yao et al., 2011] Yao, W., Chen, X., Luo, W., van Tooren, M., and Guo, J. (2011). Review of uncertainty-based multidisciplinary design optimization methodsfor aerospace vehicles. *Progress in Aerospace Sciences*, 47(6):450–479.

[Zadeh, 1978] Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1):3–28.

Appendices

A Appendix A: Base-line model uncertainties

Table 3: Statistical parameters and interpretation summary for the uncertainties part of the base-line model.

	Probability distribution		**Moments**		
Parameter	Type	Bounds	Mean	Std dev	Interpretation
EGT to SFC	uniform	[8.33e-4, 1e-3]	9.16e-4	4.82e-5	% increase in SFC per °C EGT increase
EW price	uniform	[6638, 9216]	7927	744	Economic cost in US$ of performing EW for both engines
fuel price	uniform	[0, 1]	0.5	0.288	Interpolation coefficient used to select fuel price scenario (highest coefficient leads to higher fuel price)
EGT increase	truncated Gaussian	[0, 1]	0.5	0.255	Interpolation coefficient used to select performance degradation profile (higher coefficient leads to steeper EGT increase)
EW effect	truncated Gaussian	[0, 11.56]	5.398	1.542	Performance restoration measured in °C of EGT reduction as a result of EW
EW interval	uniform	[90, 240]	165	43.3	Interval in days between two EWs

B Appendix B: Global sensitivity analysis results

Table 4: First and total order Sobol' indices.

Variable	$S_i(\pm 2\sigma)$	$S_{T_i}(\pm 2\sigma)$
x_1: EGT to SFC	0.0042 (± 0.0040)	0.0056 (± 0.0007)
x_2: EW price	0.0031 (± 0.0028)	0.0031 (± 0.0002)
x_3: fuel price	0.2746 (± 0.0372)	0.3002 (± 0.0257)
x_4: EGT increase	0.3857 (± 0.0546)	0.4024 (± 0.0380)
x_5: EW effect	0.2060 (± 0.0335)	0.2341 (± 0.0226)
x_6: EW interval	0.0617 (± 0.0198)	0.0915 (± 0.0085)

C Appendix C: Deterministic simulations for base-line model

Table 5: Individual runs for extreme values of the uncertain variables. The left and right bound refer to the lower and upper extremes of each probability distribution. Notation: NPV_w (washed, NPV for scenario with engine wash), NPV_u (unwashed, NPV for scenario without engine wash)

Variable	Left bound of distribution, all values in US$	Right bound of distribution, all values in US$	Absolute difference of ΔNPV [US$]
x_1: EGT to SFC	NPV_w: 22,928,037 NPV_u: 22,628,498 ΔNPV: 299,539	NPV_w: 22,672,232 NPV_u: 22,320,576 ΔNPV: 351,656	52,117
x_2: EW price	NPV_w: 22,820,524 NPV_u: 22,471,293 ΔNPV: 349,231	NPV_w: 22,779,797 NPV_u: 22,471,293 ΔNPV: 308,504	40,727
x_3: Fuel price	NPV_w: 45,862,922 NPV_u: 45,749,143 ΔNPV: 113,779	NPV_w: -262,602 NPV_u: -806,558 ΔNPV: 543,956	430,177
x_4: EGT increase	NPV_w: 23,008,536 NPV_u: 22,729,837 ΔNPV: 278,699	NPV_w: 22,591,826 NPV_u: 21,617,368 ΔNPV: 974,458	695,759
x_5: EW effect	NPV_w: 22,725,102 NPV_u: 22,478,048 ΔNPV: 247,054	NPV_w: 23,141,687 NPV_u: 22,471,293 ΔNPV: 670,394	423,340
x_6: EW interval	NPV_w: 22,941,088 NPV_u: 22,473,584 ΔNPV: 469,504	NPV_w: 22,725,102 NPV_u: 22,478,048 ΔNPV: 247,054	222,450

II

Literature Study
previously graded under AE4020

1
Introduction

Wear and tear, as well as dirt accumulation, cause aircraft engines to become less efficient with every flight. This efficiency reduction leads to an increase in Exhaust Gas Temperature (EGT), thereby also increasing the fuel consumption. As this EGT has an upper limit, the Exhaust Gas Temperature Margin (EGTM) decreases as EGT increases. At a certain EGTM threshold, an expensive Engine Shop Visit (ESV) is required, where a significant portion of EGTM is restored.

Due to the high cost of ESVs and engine maintenance in general (engine maintenance accounts for around 35% - 40% of the dmc [4]) countermeasures are brought up. One of these is to perform on-wing engine cleaning during turnaround or overnight stop, which aims to mitigate engine deterioration. This engine wash (EW) procedure typically involves injecting hot water into the engine with the aim of removing accumulated dirt, resulting in fuel savings of up to 1.3% [55]. However, the overall economic feasibility of on-wing engine cleaning depends on several economic factors (eg. fuel price, price of an engine wash), physical factors (eg. the EGTM restoration capabilities of an engine wash, the effect of EGT increase on the SFC) and operational factors (eg. the engine wash frequency). A model that integrates these factors is needed to simulate their impact on the economic value of engine cleaning procedures.

This model, its parameters and inputs introduce uncertainties with different locations, levels and types [151]. To address these uncertainties, a systematic process for uncertainty assessment is suggested to be integrated into the use-case. This framework should guide the practitioner through the process of uncertainty assessment (including identifying, classifying, quantifying, propagating and combining uncertainty [16]) in an interdisciplinary environment with the aim to aid in decision making. As a consequence, differentiating between different types of uncertainty (aleatory vs epistemic) is important, as it enables the recipient to determine whether the output uncertainty can be reduced by attaining more knowledge or the uncertainty is largely due to natural randomness. Global sensitivity analysis (GSA) can support the uncertainty modeling process by ranking input uncertainties based on their effect on the output uncertainty, and thereby enabling efficient management of modeling resources. The engine wash assessment under uncertainty, making use of the in-house framework for life cycle based evaluation of aircraft technologies named LYFE, serves as a suitable use-case to demonstrate the developed methods due to the complex simulation environment and multi-source uncertainties.

1.1. Research Objective

The following research objective was defined:

The research objective is to perform a systematic uncertainty assessment in an interdisciplinary simulation environment that aids in decision making by analyzing available methods for uncertainty quantification and global sensitivity analysis, and demonstrating and verifying the framework on the use-case of engine cleaning procedures.

The research objective can be broken down into two main parts, namely the external goal and the internal goal. These parts are clarified below.

External Goal

- Systematic uncertainty assessment: a framework is required that enables systematic consideration of the relevant uncertainties.
- Interdisciplinary simulation environment: relevant uncertainties originate, due to the interdisciplinary use-case of engine wash, from different fields. The simulation environment LYFE (further explained in subsection 4.1.3) encapsulates numerous inputs that affect the cash flow during the aircraft lifecycle such as sfc, maintenance costs, engine wash schedule, etc.
- That aids in decision making: a framework that allows decision makers to decide whether it is worthwhile to gather more knowledge, this refers to the differentiation between aleatory and epistemic uncertainty. The uncertainty apportionment provided by global sensitivity analysis aids in decision making too by identifying high-impact uncertainties, enabling decision makers to direct further modeling effort accordingly.

Internal Goal

- Analyzing available methods for uncertainty quantification and global sensitivity analysis: the systematic framework for uncertainty assessment on a high level will at least consist of following activities: uncertainty identification, uncertainty classification, uncertainty quantification and uncertainty propagation (including global sensitivity analysis).
- Demonstrating and verifying the method on the use-case of engine cleaning procedures: the use-case involves the quantification of the economic value of engine cleaning procedures. This use-case was chosen with the sole purpose of demonstrating the methodology for uncertainty assessment. The main objective therefore remains the systematic integration of uncertainties in a complex environment. The results of the economic evaluation, and thus also the accuracy of models developed for this purpose, are not of primary interest.

1.2. Research Framework

The research framework outlines the steps needed to be taken in order to achieve the research objective. These steps also provide a useful basis for constructing the research question(s), as it becomes clear which data and knowledge need to be gathered in order to achieve the research objective. A schematic of the research framework is shown in Figure 1.1.

The top three boxes primarily comprise the knowledge to be acquired. Theory on engine performance deterioration as well as other factors affecting the engine wash effectiveness will be thoroughly studied. Uncertain variables will then be identified and quantified using the available knowledge. Next, the uncertainty can be propagated through the cost-benefit simulation tool LYFE (subsection 4.1.3) and a global sensitivity analysis is performed to apportion the output uncertainty to individual input uncertainties. Finally, the developed models will be verified and the results will be analyzed.

1.3. Research Question

From the research framework, one or multiple research questions are drawn up. It was decided to go with one overarching research question, which is then broken down into several sub-questions. Each of them define (a part of) the knowledge that can be used or needs to be gathered in order to achieve the research objective.

Which methodology needs to be developed to perform a systematic and efficient assessment of the uncertainties present in the process of quantifying the economic value of engine wash procedures under uncertainty?

Sub-models

1. Which factors are affecting the economic value of engine wash procedures and what are appropriate models to simulate their behavior and impact on the engine wash economic value?

Uncertainty Analysis

2.a. Which theory is suited to model the uncertain parameters defined in the identification phase?

2.b. Which method for global sensitivity analysis can be used to apportion the combined uncertainty to the input parameters?

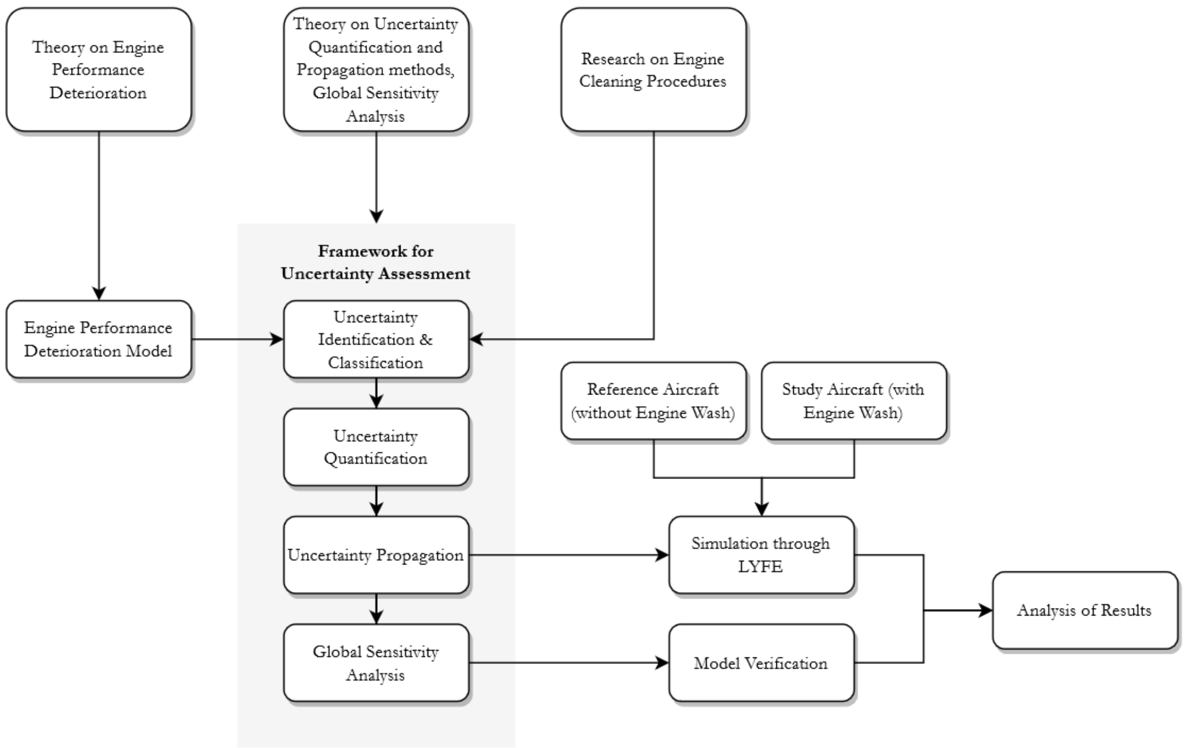


Figure 1.1: Research Framework.

These research questions form the basis for conducting a literature review. Factors that influence the economic value / effectiveness of engine cleaning will be studied. A model for engine performance deterioration that can be integrated with LYFE is likely to be one of the key sub-models. Literature will be looked at from a general perspective, but special attention will be paid to modeling techniques that yield a parameter that can be operationalized (eg. a change in EGT, SFC or efficiency) and integrated with LYFE.

Literature on uncertainty analysis will be looked at from a technological assessment perspective. That is, without diving into higher-order mathematics, techniques and methods used for classifying, quantifying and propagating uncertainty are to be demystified and their applicability to this research problem is to be analyzed. Global sensitivity analysis methods will also be studied and compared.

The literature review on engine performance deterioration modeling and engine wash assessments can be found in chapter 2. Literature on uncertainty analysis and sensitivity analysis is discussed in chapter 3. The thesis project plan is presented in chapter 4 and a conclusion with regard to the research questions is provided in chapter 5.

2

Engine Wash Assessment

The use-case for the systematic uncertainty analysis is the assessment of the economic value of engine wash procedures. The engine cleaning considered in this study is carried out on turbofan engines. A general discussion of basic principles of gas turbine engines and its deterioration mechanisms, as well as an introduction to on-wing engine cleaning, are therefore provided in section 2.1. This will provide a foundation for further consideration of literature on engine deterioration and engine cleaning in section 2.2. From the surveyed literature, a research gap is identified and summarized in section 2.3.

2.1. Fundamentals

This section briefly introduces the development of the gas turbine and explains its basic principles. This is relevant to this review since literature on deterioration mechanisms and its modeling techniques are clearly not restricted to modern turbofan engines, but instead find discussions and applications in the broader field of jet propulsion as well as land-based and sea-based turbine systems. An excursion to modern turbofan engines is eventually made, as this type of jet engine will be used in the use-case of the thesis project. Once the relevant characteristics of the (aircraft) gas turbine have been described in subsection 2.1.1, performance deterioration is discussed in subsection 2.1.2. Engine cleaning operations will afterwards be presented in subsection 2.1.3 as a way to partially recover performance degradation.

2.1.1. Basic Principles of the Gas Turbine Engine

The turbine is in many aspects the desired choice for generating mechanical power used for various applications [126]. Gas turbines tend to be potentially highly reliable and feature low oil consumption due to the absence of reciprocating and rubbing members [126]. The gas turbine was initially developed for electricity generation at the start of the twentieth century. Around the 1950s, shortly after it was introduced as a means of aircraft propulsion in the form of the jet engine, the turbine started to gain popularity in various other fields.

The simplest way to represent a gas turbine engine is shown in Figure 2.1. From this schematic, one can identify three essential components. The compressor, the combustion chamber and the turbine. If one were to simply compress the working fluid (in this case air) and allow direct expansion in the turbine, the net power would be zero since, assuming zero efficiency loss, the turbine would be just able to drive the compressor. To increase the power output of the turbine, the air temperature in between the compressor and turbine can be increased by means of fuel combustion in the combustion chamber. The net power in this case will be larger than zero, meaning the turbine can also provide more power than the amount needed to drive the compressor. It should be noted that the amount of fuel that can be added to the system is limited. A maximum fuel/air mixture that can be used exists, which is limited by the working temperature of the turbine blades [126].

Until now, we have been assuming zero losses in any of the components involved. In practice however, losses in the compressor and turbine occur, leading to the efficiency reduction of the gas turbine. One can now identify two factors that influence the system performance: the efficiency of the individual components and the maximum turbine temperature [126]. The former in particular will play an essential role when reviewing gas turbine deterioration mechanisms.

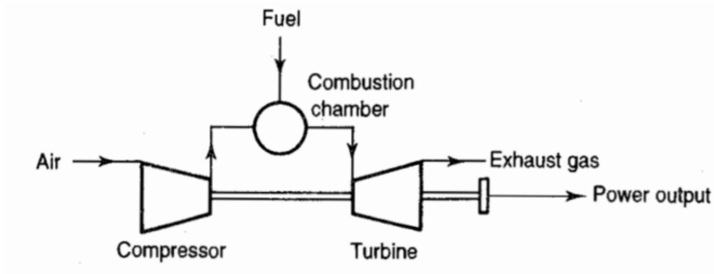


Figure 2.1: A simple representation of a gas turbine (from [126]).

The simple gas turbine design depicted in Figure 2.1 can be further extended, i.e. components can be added and a large variety of combinations of different components exists since the most suited configuration for specific applications tends to differ. One common variation often seen is a multi-spool arrangement. This arrangement is required when the desired pressure ratio can not be generated by a single compressor. The twin-spool arrangement, as depicted in Figure 2.2, then consists of a low-pressure and high-pressure variant of both the compressor and turbine, where the the low-pressure compressor is driven by the low-pressure turbine and the high-pressure compressor by the high-pressure turbine. This configuration was introduced for jet engines but aero-derivate engines using this arrangement have been developed for other applications as well [126].

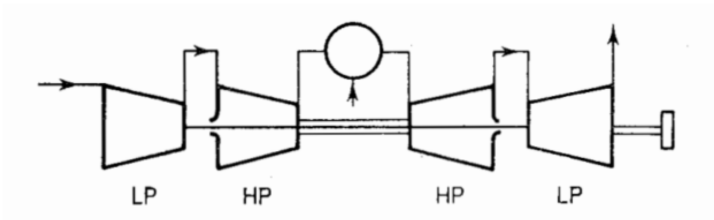


Figure 2.2: Schematic of a twin-spool gas turbine arrangement (from [126]).

Gas turbines have generated a great impact on the aerospace propulsion field. In the 1950s a gas turbine, called the turbojet, was introduced in civil aviation. The setup is similar to the one in Figure 2.1, except that the shaft at the right it replaced by a propelling nozzle to create high velocity jet [126]. This was followed by the development of the turboprop (for lower speed aircraft) and afterwards the turbofan engine (for high subsonic speeds), the latter of which soon replaced the turbojet due to its higher propulsive efficiency and lower levels of exhaust noise caused by the bypass flow. Both twin-spool and triple-spool are possible arrangements for conventional turbofan engines [4]. An example of a typical turbofan engine seen today is shown in Figure 2.3. Most of the air does indeed not go through the core of the engine, creating a bypass airflow. The fan, which has a similar function as the propeller in a turboprop engine, generates most of the thrust (between 50% and 85% of the total thrust). Generally speaking, a higher bypass ratio (proportion of the air that bypasses the core relative to the air traveling through the core) results in lower noise and higher fuel efficiency [4].

2.1.2. Engine Performance Deterioration

Several engine operating parameters exist, some of which can be used as health indicators providing information about the performance deterioration level of an engine. The most important engine operating parameters are the N1-speed end the EGT [4]. The former is mostly used to indicate the amount of thrust the engine is producing, whereas the latter can be used as a performance deterioration indicator.

The EGT is the temperature measured in the exhaust of the engine. The higher this temperature, the lower the engine efficiency at producing its design trust [4]. That is, lower engine efficiency (mainly caused by lower compressor efficiency) means that in order to produce equal thrust, more fuel is required [12]. Throughout one flight cycle, the highest EGT is usually reached during take-off or initial climb. A general increase in EGT can be observed as the engine ages (the engine burns more fuel to deliver a certain amount of thrust) and can indicate that engine hardware deterioration has occurred [63, 129]. Some of the underlying reasons for the loss of engine efficiency are erosive wear of turbine and compressor

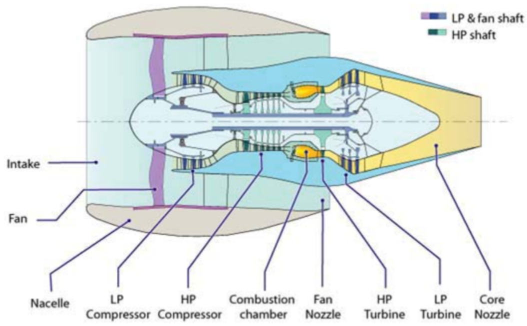


Figure 2.3: Example of a modern turbofan engine (from [4]).

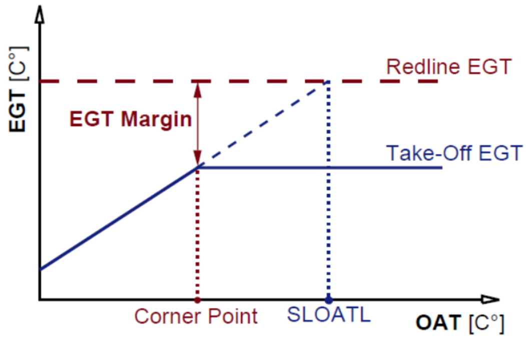


Figure 2.4: Relationship between EGT and OAT (from [130]).

blades, increased tip clearance of blade tips and fouling (particles deposited on blade surfaces) [34, 73]. As an EGT upper limit is imposed by the manufacturer to avoid damage to engine parts, the EGT margin can be introduced as the difference between the EGT upper limit and the actual EGT [4]. An ESV is required if the EGT margin approaches zero. During such engine removal, a portion of the EGT Margin can be restored.

Furthermore, the EGT increases with the Outside Air Temperature (OAT), as can be observed in Figure 2.4. EGT increases linearly with increasing OAT until the corner point temperature is reached corresponding to the EGT redline. As the constant line indicates however, the aircraft can be operated beyond the corner point temperature, albeit with reduced thrust [4]. EGT margin deterioration is one of the primary causes for Engine Shop Visits, and is reported to be the main cause in the case of first-run engines (before the first refurbishment) used for short-haul operation, as well as mature-run engines (after the first refurbishment) in long-haul operation [4]. Other important causes for engine removals are hardware deterioration and expiry of Life Limited Parts (LLPs) [62]. The latter are parts that can not be contained if they fail and are replaced after a predefined number of flight cycles [4].

One of the main takeaways from this brief introduction to engine performance deterioration linked to engine maintenance is the importance of the EGT as condition monitoring parameter. The EGT (and directly related EGT margin) also makes for a suitable engine health indicator within the scope of this research project since the EGT increase can be conveniently used as input for the in-house economic assessment framework LYFE.

2.1.3. On-wing Engine Cleaning

On-wing engine cleaning procedures can be employed to restore some of the lost EGT margin. During engine washing, water and cleaning additives are sprayed into the intake to clean the surfaces of both the compressor and turbine stages, while the engine is running. Hence this maintenance task can be used to partially revert deterioration due to fouling. An illustration of the effect of regular on-wing engine washes on the Thrust Specific Fuel Consumption (TSFC) is shown in Figure 2.5.

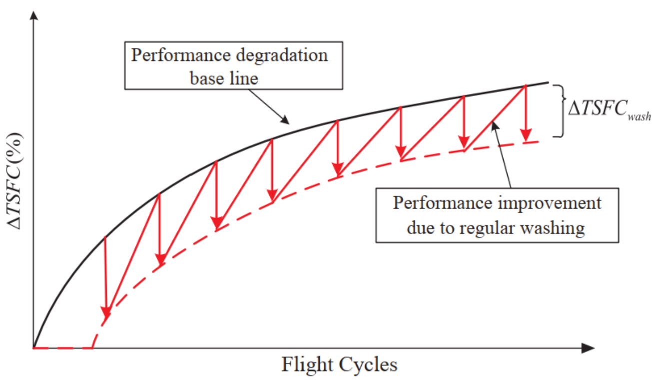


Figure 2.5: Effect of regular engine washing on engine performance deterioration (from [22]).

While the exact washing process depends on the aircraft and engine, a general sequence holds for most cases. First, a solution of water and cleaning additives is sprayed in the engine to clean component surfaces. The engine is then rinsed with demineralized water (usually several times) until the waste water is clear [56]. Generally, on-wing engine cleaning is performed at intervals recommended by the manufacturer (yet this remains at the discretion of the airline) [12, 56]. Besides being used as a means of restoring some portion of the lost EGT margin by keeping the gas path clean, engine washes may also be performed after Foreign Object Damange (FOD) (eg. a bird strike).

2.2. Literature Consideration

The research questions defined during the conceptual research design revealed that there is a need for knowledge on what factors play a role in determining the economic value of engine wash procedures and how this system of relations can be modeled in an integrative way. Literature is considered to provide a knowledge basis for further research in the modeling phase of this research project. Factors that are found to affect the economic feasibility of engine wash are discussed in subsection 2.2.1. Methods for engine performance deterioration are described in subsection 1.2.1. Lastly, subsection 2.2.3 treats engine wash assessments found in literature.

2.2.1. Factors Affecting Economic Value of Engine Wash

To build a model that simulates the effect of factors affecting the Quantity of Interest (QoI), they must first be identified and their relationship to the QoI described. To this end, Figure 2.6 provies a breakdown of the factors that are considered to directly affect the QoI. That is, these factors directly affect the difference in economic value between the use-case with engine wash and the same use-case without engine wash. Many other factors (eg. load factor, type of aircraft and engine), while affecting the economic value of these two scenarios individually, are considered to have only a minor effect on the difference in economic value between the scenarios, since no direct effect on the QoI is observed. Only the factors in Figure 2.6 will be treated in a non-deterministic (i.e. as an uncertainty). Other inputs and parameters will be set to a certain value in the assessment. More on the motivation behind this differentiation can be found in subsection 3.3.1.

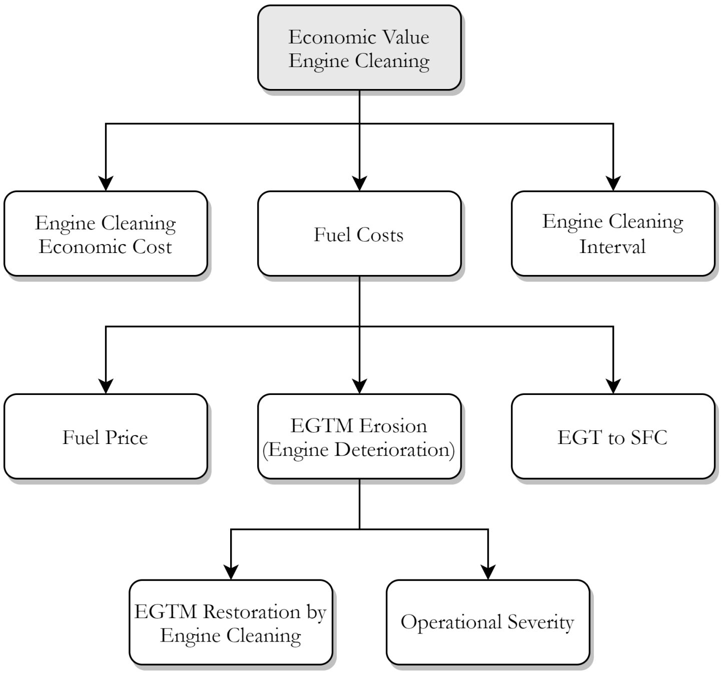


Figure 2.6: Factors affecting the economic value of on-wing engine cleaning.

Engine cleaning economic cost: only two sources that provide a value for the price of an engine wash were found. One is from Chen and Sun [22], stating "the cost of one on-wing washing for a narrow-body aircraft with two engines is estimated to be 3,017 US$ taking into account the technician's laber hours and materials". They add that the rent for the equipment is approximately 180,000 US$. According to a price list from Lufthansa Technik published in 2012 [144], one engine water wash for narrow body aircraft was priced at 3,600 EUR per engine. These charges do not include towing, run up and hangar usage.

Engine cleaning interval: frequent on-wing engine cleaning can prevent dirt from building up to such an extent that causes non-recoverable degradation. According to Lufthansa Technik [2], engine washes should be performed with a defined frequency. Chen and Sun [22] cite the interval recommended by the OEM to be three to eight months. This interval depends however on certain constraints (e.g. environmental) faced by the operator.

Fuel price: a largely unpredictable and therefore highly uncertain parameter. The Annual Energy Outlook published by the Energy Information Administration [5] provides an outlook for the fuel price (till 2050) visualized in Figure 2.7 which includes a set of scenarios depending on several economic factors.

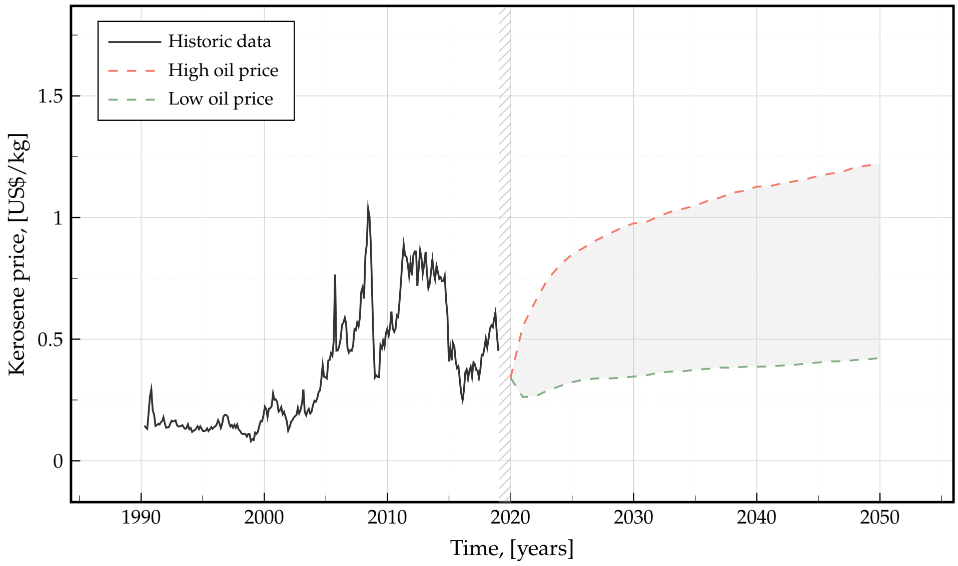


Figure 2.7: Historic data and outlook of the fuel price under different scenarios (data retrieved from the U.S. Energy Information Administration [5]), split by a gray bar since they do not connect.

EGT to SFC: as engine performance deterioration may be expressed in terms of EGT, translating a change in EGT to a change in SFC becomes essential to facilitate the changed fuel burn calculation. Literature presents several simple equations relating both parameters (eg. Ref. [22, 63]).

Operational severity: includes operational factors like take-off de-rate and the flight hour to flight cycle ratio, as well as environmental factors like the Outside Air Temperature (OAT) and the environment (erosive-corrosive environments, or eg. places with high dust concentration) [4, 63]. One way to include the effects of multiple operational and/or environmental effects is through so-called severity curves, an example of which is shown in Figure 2.8. The severity factor, which in this example depends on the derate and flight hour to flight cycle ratio, is then used to alter the engine maintenance cost per flight hour [130]. Ackert [4] presents a severity curve that includes the effect of three possible environmental conditions. Although these figures seem promising at first, it is not immediately clear if and how data from these curves could be related to EGT, which is a desired output parameter from the deterioration model. More advanced techniques for engine performance modeling are described in subsection 1.2.1.

EGTM restoration by engine cleaning: publicly available data on the EGTM restoration capabilities of an engine water wash is scarce. Some indication is provided by Airbus [15] in a report on EGT limit exceedance. Average EGTM recovery is cited to be 7°C, with maximum values up to 15°C, the latter in agreement with Ackert [4]. Lufthansa Technik claims an EGTM improvement of up to 25°C for its CyClean engine wash technology [2].

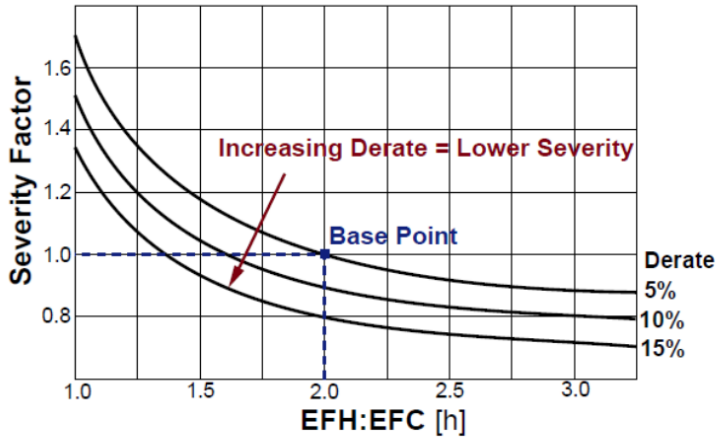


Figure 2.8: An example operational severity curve (from [130]).

2.2.2. Modeling Engine Deterioration

Factors affecting the QoI have been demystified in Figure 2.6 from an operational perspective. Severity curves can account for operational and environmental effects in modeling the engine performance deterioration but cannot be easily quantified in terms of EGT. Literature-based equations do provide a relation between the EGT increase and the number of flight cycles (eg Ref. [63]). Alternatively, interviews with MRO providers are available in Ref. [1] from which EGT values can be retrieved for various numbers of flight cycles. However, as these representations for engine performance deterioration are very simplified, more advanced techniques should also be reviewed. To this end, a more detailed study of the physical processes behind engine performance deterioration and a review on data-driven and physics-based modeling techniques is performed.

Types of engine degradation can be broadly classified into [28, 93]:

- Recoverable performance degradation: can be recovered by cleaning or washing.
- Nonrecoverable performance degradation: cannot be recovered by cleaning or washing but can be recovered through eg. replacement [35].
- Permanent performance degradation: cannot be recovered even after an engine shop visit and replacing/repairing components.

Fentaye et al. [40] reviewed literature on physical faults in gas turbines and provides an extensive description of each fault including its effects and performance change indications. A brief description is given below.

Fouling: performance deterioration due to adherence of particles on the blade surfaces [73]. These particles can be dust, dirt, oil droplets, insects etc. [40]. According to Lakshminarasimha [73], the fouling increase can be assumed linear with time and causes between 70 to 85 % of the gas turbine engine performance loss.

Erosion: concerns the material loss on blade surfaces caused by ingestion of particulate matter [40, 73]. This causes non-recoverable performance degradation [35].

Corrosion: also non-recoverable, caused by for example salts and acid [40].

Blade tip clearance: one of the causes for blade tip increase is thermal expansion, causing the blade tips to rub against the casing and resulting in the loss of material, which is not recoverable through washing [40].

Foreign Object Damage: besides bird ingestion, also gravel or hailstones can damage the internals and cause performance deterioration which can not be recovered with engine washes [40].

By realizing that engine washes can only reduce fouling of the faults listed above (a large proportion of performance deterioration due to fouling is recoverable [73]), the degradation of the flow capacity and more importantly efficiency (of eg. the compressor) are interesting parameters that could be used to model engine deterioration and the impact of engine cleaning.

The considered literature on engine performance degradation modeling has been divided in two groups: physics-based models and data-driven models. Physics-based models as defined here are models that simulate the thermodynamic processes and can therefore are expected to more accurately model the deterioration process. Data-driven approaches are reviewed with the perspective of finding data that could be used for deducing relationships from between parameters of interest.

Physics-Based Modeling Approaches and Tools

A physics-based model is defined by Alozie et al. [9] as "a performance model that describes the behavior of the gas turbine components based on the thermodynamics of the working fluid to provide information about the configuration and operation of the real engine.". Considering physical parameters and modeling the degrading engine based on such parameters makes the assessment significantly more complicated, since these parameters usually vary dynamically with the wide range of operating and environmental conditions specified in this assignment and since the required knowledge on the physical modeling techniques for engines needs to be acquired [165]. Therefore, only a brief review of relevant studies in the field is performed.

Lakshimarasimha et al. present a model to simulate the effect of fouling and erosion [73]. They developed a procedure to simulate how these two forms of deterioration affect the engine performance parameters. Giesecke and Igie [42] employ the tool TURBOMATCH developed at Cranfield University to simulate the jet engine performance. Using this software, one can specific an engine model with a particular set of components, as well as the engine inlet conditions and the design point. To simulate the degraded performance, a reduction in isentropic compressor efficiency as well as the non-dimensional mass flow is specified, allowing to incorporate the effect of engine washes as explained in subsection 2.2.3. Wensky et al. [155] model the environmental influences on engine performance degradation. Zaita et al. [168] developed a performance deterioration model for rotating components. The authors present some simple equations that relate the degree of fouling and erosion to the location inside several gas turbine components. The model outputs ΔTSFC and takes among other parameters the mission profile, engine characteristics and geographical location as input. Alozie et al. [8] present a degradation model for the Equivalent Operating Time (EOT) derived from first principles and empirical data correlations. Interesting input parameters are the internal power settings, ambient and environmental conditions. Kurz and Brun [69] describe the engine behavior using governing equations, as well as a set of deviation factors using which the degree of deterioration could be studied. The most important inputs to the model are the change in compressor efficiency and reduction in airflow, while the output consists of the reduction in power and overall efficiency.

Most of these models are not publicly available and some require a high level of understanding of gas turbine modeling. The commercial software GasTurb is available for this thesis and could be used to model performance deterioration. The main disadvantage is that due to its complexity, a large time investment would be required to make effective use of this simulation tool.

Data-driven Modeling Approaches

Data-driven modeling approaches rely on statistical methods and machine learning models to understand patterns in performance deterioration data [9]. This means these approaches do not require insight into the underlying physical behavior of the system, which can be helpful in case building models to simulate physical characteristics is considered too complicated [165]. On the other hand, in order to make data-driven approaches successful, historical data is needed that is representative for the pursued application.

However, particular in the commercial aviation industry, where gas turbine engines are the most common means of propulsion, there is a lack of run-to-failure data sets [127]. Real fault progression data is expensive to acquire and collecting relevant data is hard, making publicly available data sets scarce. Saxena et al. [127] noticed that the lack of public data sets hinders progress in engine prognostics, motivating them to construct a data set to be used for a Prognostics and Health Management (PHM) data challenge, where the goal is the predict the Remaining Useful Life (RUL) of the aircraft engine based solely on historical data. This means that the physical process behind the performance deterioration is bypassed, reducing the analysis to a purely data-driven one.

To generate data to be used in the PHM data challange the authors used the cmapss software, which can be used to simulate a large commercial turbofan engine [74, 127]. The inputs consist of the the fuel flow and health parameters (e.g. component efficiencies) that can be varied in order to simulate performance degradation.

Five simulation data sets are available, each with different operating conditions and fault modes. Although the prediction of RUL has no direct added value for this research (the data set was generated mainly for benchmarking prognostic and fault diagnostic approaches), the simulation data consists of several parameters (e.g. component efficiencies) with corresponding simulated output data (eg. temperatures at various stages). The data set has been used in many studies related to prognostics and fault diagnosis. Lei et al. categorized some of these studies [74], and stated that several researchers trained machine learning models on the available training data to predict the RUL (Refs. [47, 58, 102, 157, 160]). The data is also used to develop other prognostic approaches that employ for instance unit-to-unit similarity (Refs. [37, 109]) or multi-feature fusion (Refs. [77, 78]). Others have employed the available data to come up with new prognostic approaches (Refs. [53, 87, 142]).

Recently, a new dataset [21] has become available based on the same C-MAPSS tool. This new dataset features real flight data from take-off to landing, from which the flight conditions are extracted. Furthermore, the previous dataset contained only two failure modes, while the new one features continuous degradation in the fan, Low Pressure Compressor (LPC), High Pressure Compressor (HPC), High Pressure Turbine (HPT), Low Pressure Turbine (LPT). These data sets do have some critical limitations. Accelerated aging was used instead of a realistic scenario for a real engine (with a full lifespan in the order of thousands of flight cycles). This means that no meaningful conclusions can be drawn from the rate of deterioration. Furthermore, the EGT is not directly provided by the output. Further research would be required to translate the tot (which is provided) into EGT.

Some studies use real flight sensor data to analyze the behavior of EGT with varying environmental and operating factors. Yildirim and Kurt [162] for instance present in their work a multiple linear regression model relating various input parameters to EGT, albeit an analysis without the inclusion of the time parameter. This means no relation between EGT and either flight hours or flight cycle can be drawn from the data, discounting its value for this research. The same issue is the case for the work of Yilmaz, who obtained a relationship between EGT and engine operational parameters at isolated time instances for the CFM56-7B turbofan engine [163].

Modeling techniques that combine physics-based and data-driven modeling are often referred to as hybrid modeling techniques (applications in prognostics can be found in Refs. [76, 158, 159]). This approach was considered beyond the scope of this research as literature states it is not mature yet [10].

Summary of Modeling Techniques

Engine performance deterioration can be modeled using data-driven, physics-based as well as hybrid techniques (which are beyond the scope of this study). Within each category, a range of complexity levels exist. A brief note: complexity here refers to both the inherent complexity of the model and the effort required by the practitioner to apply a certain technique or use a model. The latter may require a complete understanding of the model, if the model has to be build by the practitioner (eg. training and testing deep neural network, developing a physical model from a set of governing equations, etc.). In other cases the complexity might be due to the need to become familiar with a commercial software package.

Deriving an equation relating flight cycles to an increase of EGT, for instance based on a handful of data points retrieved from interviews with MRO providers, is considered a data-driven model. The same is true for a complex neural network which takes as input multiple parameters indicating the health condition of a component and outputs a rul prediction. The biggest challenge seems to be acquiring sufficient data with the desired inputs and outputs.

Physics-based models also have varying degrees of complexity, ranging from relatively simple equations to self-contained commercial software. Kurz and Brun [69] for instance model the gas turbine using a set of governing equations. On the other side of the complexity spectrum are full-fledged engine performance models which can be used to model engine performance deterioration (eg. Gasturb [41]). More advanced software, for instance using Computational Fluid Dynamics (CFD) is deemed beyond the scope of this research. A graphical summary of data-driven and physics-based models with their limitations and advantages with respect to the current study, and ranked in terms of complexity, is shown in Figure 2.9.

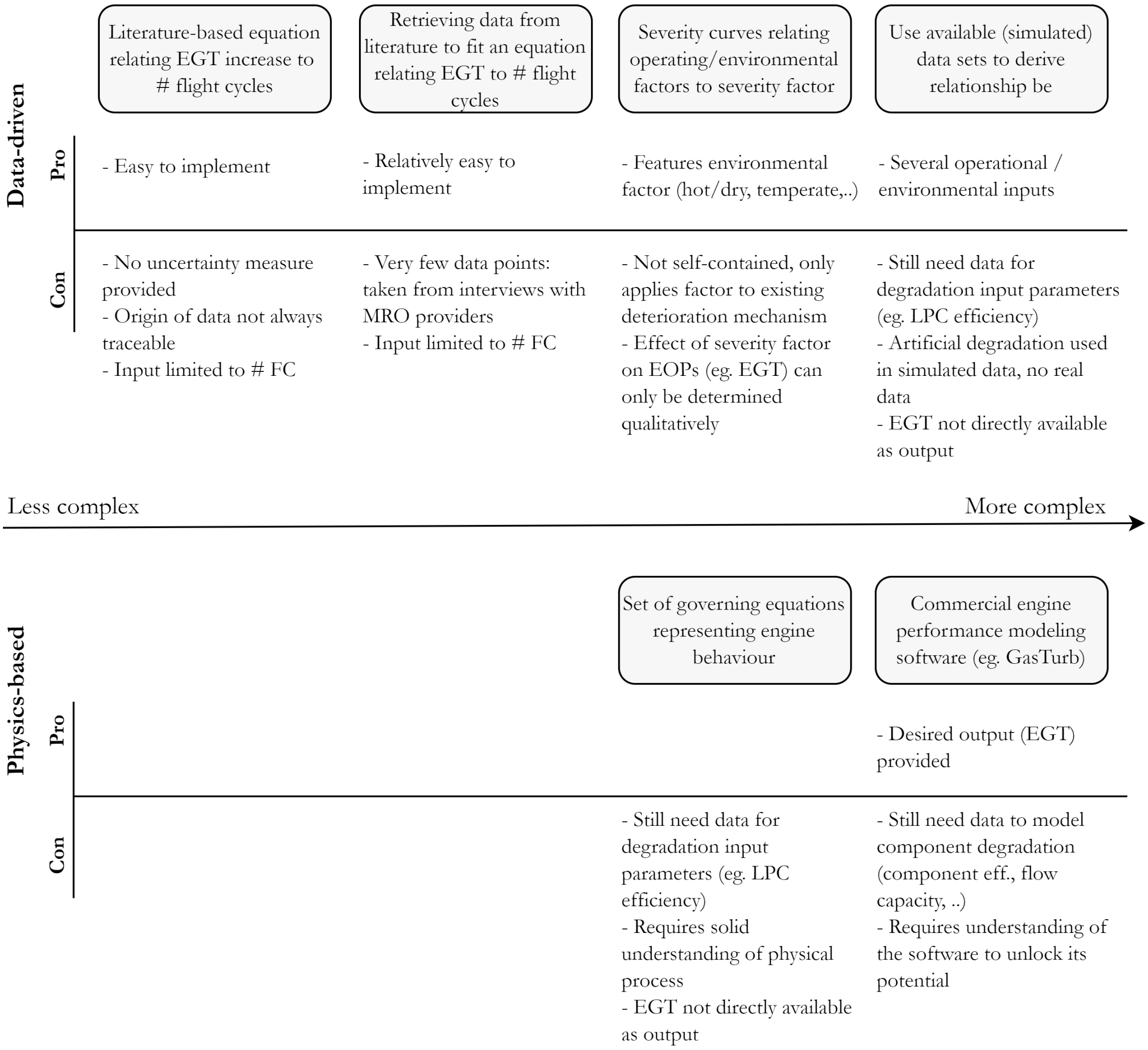


Figure 2.9: Comparison of engine performance deterioration modeling techniques considered in this study.

2.2.3. Engine Wash Assessments

Several engine wash assessments were found in literature, both for gas turbines with industrial and aerospace applications. Giesecke and Igie [42] highlight the economic value of compressor washing in their techno-economic study on engine compressor washes for short-range aircraft. To model the engine performance deterioration, the software Turbomatch by Cranfield University was used. The authors present the effect of fouled blades in terms of a reduction in compressor efficiency. Based on this parameter, the non-dimensional mass flow and the pressure ratio, the scaling factors are determined that simulate the deteriorated engine performance. The engine performance deterioration is assumed to be caused only by fouling. This is considered quite a radical assumption, given the numerous papers describing other relevant physical faults leading to degraded engine performance (eg. Refs. [28, 35, 69, 73, 93]).

Other studies investigate on-wing engine cleaning from a different perspective. Boyce and Gonzalez [18] developed several tests in a controlled environment to determine the efficacy of engine washing with varying washing frequencies and dissolving agents used for the washing process. A washing program was then developed for a fleet of 36 industrial turbines that maximizes the engine efficiency and minimizes maintenance labor. It should be noted however that due to different operating and environmental conditions for aircraft gas turbines compared to gas turbines used for industrial applications, cost-benefit analyses of engine washes and washing interval optimization for gas turbines with industrial applications may have limited relevance to the analysis of turbofan engine wash assessments.

Cheng and Sun [22] presented an estimation method of the fuel consumption savings due to engine washes, taking into account the economic cost of engine washing procedures and the fleet-wide fuel consumption savings. The assessment was applied to a case-study of a fleet of 200 CFM56 engines, and the optimal washing frequency out of six scenarios was determined. Monte Carlo simulations were used to propagate the input uncertainty to a final distribution of the fuel consumption savings.

Summarizing, the engine wash assessments carried out in available literature are often not integrative with respect to the operational and economic factors considered. This research aims at a more holistic analysis, to which the proprietary cost-benefit tool LYFE adds great value. Furthermore, uncertainties originating from the inputs or models are rarely included in the reviewed assessments. This thesis project, with its focus on analyzing the uncertainty, can provide significant added value in this regard.

2.3. Research Gap

The literature review found that integrative techno-economic engine wash assessments are rare and not integrative with respect to the factors considered to affect the engine wash economic value. To close this gap, a model for engine performance deterioration is required that can be integrated with engine wash assessments. This approach should be compatible with the framework LYFE (eg. by expressing performance deterioration in terms of for instance EGT). Current data-driven approaches do not provide a solution to this problem as they are mostly developed for fault diagnostics and RUL prediction, which differs from the perspective and focus of the current study, and the available data is not (directly) useful for this research project. Developed physics-based models are often either not publicly available or too complex for this thesis, with the engine wash assessment being just a use-case. A simpler, hence less accurate, model may be developed that relies on relations between parameters of interest, to represent engine performance deterioration. In addition, the effect of an engine wash on the EGTM needs to be further investigated.

3
Uncertainty Analysis

In order to improve decision making in cases involving high uncertainty, it is important to understand well the concept of uncertainty and how it can be dealt with [120]. In a model-based decision support context, uncertainty can be found in a multitude of locations and take different forms and levels [151]. In this study, uncertainties relevant in the use-case of an engine wash assessment will be analyzed. From a high-level perspective, the study therefore involves identifying, classifying, quantifying and propagating uncertainties relevant in the system in order to systematically combine the uncertainties present in the system [16]. However, uncertainty quantification as a field has not reached the maturity of a field like linear algebra [141]. Reasons include the recent emergence of the field compared to well-established ones and the fact that the field seems to be closely linked with applications. A large set of theories and methods for uncertainty quantification have been developed over the past decades (often tailored for a specific application), yet no over-arching theory of uncertainty quantification has been established [141]. This has an important consequence for this literature survey. As relevant literature can be dispersed over different fields and applications that employ uncertainty analysis, literature outside the field of aerospace engineering is also considered.

Section 3.1 states the motivation for considering uncertainties and performing an uncertainty analysis, followed by a description of some of the main building blocks for uncertainty analysis methods and theories in section 3.2. Subsequently, different phases of the uncertainty assessment process and the respective literature are described in section 3.3. Previous work on these concepts is discussed from an engineer's perspective, i.e. relevant literature and its link to the problem at hand are addressed with a focus on how and why a certain theory or technique can (or cannot) help with solving this research problem, rather than detailing the involved higher-order mathematics. Finally, the research gap is presented in section 3.4.

3.1. Motivation

Teng et al. state: "uncertainty analysis is an indispensable part of model prediction for non-idealised environmental systems" [145]. While their research focuses on uncertainty quantification in flood inundation modeling, the relevance of this quote on the research conducted in this thesis, with a use-case involving uncertainty quantification in a complex environment, cannot be ignored. Uusitalo et al. claim that models that incorporate uncertainties may be a substantial asset for the decision maker [147]. They also add that one should represent the uncertainties in the most honest possible way. Deliberately selecting only a few of a total set of relevant uncertainties could mislead the decision maker and would thus run counter to one of the main purposes of the uncertainty analysis, i.e. communicating the uncertainty present in a research outcome in a transparent way.

In this thesis, uncertainty analysis will be applied to the engine wash assessment use-case by systematically and efficiently considering the relevant uncertainties. This requires a holistic analysis of the available methods, frameworks and techniques for doing so, which is further discussed in section 3.3.

3.2. Fundamentals

This section briefly describes several topics that are fundamental to uncertainty analysis. The main theories with strong applicability to uncertainty quantification are introduced. The goal is to provide a brief introduction to concepts frequently used in analyzing uncertainty, treated with a practice-oriented perspective.

3.2.1. Probability Theory

A consistent framework for modeling uncertainty is available through the use of probability theory [14]. Several types of interpretations on this theory have been given in the past [25]. Two types of interpretations that can be of use in this thesis are: the classical or frequentist interpretation and the subjective interpretation or Bayesian view. The Bayesian view "interprets probability in terms of degree of belief of a subject" [25] (introduction is given in Ref. [148]). In the classical interpretation, probabilities are seen as frequencies of random events [14]. That is, the probability of eg. a particular scenario A can be defined as the fraction of the number of times scenario A occurs out of the total number of trials, with the number of trials approaching infinity [14]. In reality however, performing an infinite amount of trials is not possible. Instead, a sample set of out of the population can be randomly selected. When this sample set is sufficiently large, one can then infer parameters from the sample set about the total population [70]. The clt tells us that as the sample size approaches infinity, the sample means are normally distributed, with the mean equal to the population mean. A general introduction to the clt is given in [70].

3.2.2. Possibility Theory

A deliberate excursion is made towards fuzzy set theory and its applications in uncertainty quantification due to the limitations of probability theory to represent nonrandom (or epistemic) uncertainty [119]. Zadeh introduced the notion of fuzzy sets, which he claimed is a more appropriate way of dealing with uncertainty where the imprecision does not originate from random variables, but instead from the "absence of sharply defined criteria of class membership" [166]. These fuzzy sets indeed rely on set membership, which Zadeh suggested is key to representing nonrandom and linguistic uncertainty [119]. In classical sets, the membership of an object is precise (also referred to as binary membership, i.e. the object either belongs to the set or does not. In fuzzy sets, an object can have a degree of set membership, with values on the interval [0, 1]. The bounds of this interval correspond to the binary logic used in classical sets. To represent the degree of membership in fuzzy sets, as opposed to simple binary membership for crisp or classical set, many different membership functions can be used. An example of such membership function for both types of sets is shown in Figure 3.1. In this example a triangular membership function is used.

Fuzzy logic builds on the notion of fuzzy sets and can be used to build fuzzy models [89]. This technique might be useful for modeling (a part of) the engine performance deterioration under different operating/environmental parameters, and is therefore briefly described. Moraga [89] gives a comprehensive introduction to fuzzy logic and fuzzy modeling, and presents several applications. A slightly adapted and simple linguistic example of a fuzzy logic model is:
"Rule 1: If outside is freezing and the window does not close properly then the room will become very cold."
Here the two conditions as well as the conclusion are given by fuzzy sets. Possibility theory as developed by Zadeh [167] also builds on the fuzzy set theory and can be used to represent uncertainty. Citing the author, "the theory of possibility ... is related to the theory of fuzzy sets by defining the concept of a possibility distribution as a fuzzy restriction which acts as an elastic constraint on the values that may be assigned to a variable" [167]. Very roughly, a possibility distribution is obtained from the membership function, and the theory presents two likelihood measures: possibility and necessity. A general introduction is provided by Ref. [167]. The use of this theory for uncertainty quantification is discussed in Table 1.2.2.

3.2.3. Evidence Theory

Evidence theory, also referred to as Dempster-Shafer theory or belief functions, can also be used to represent uncertainty. Unlike probability theory (with its probability distributions) and possibility theory (with its possibility distributions), evidence theory has no function that represents the information

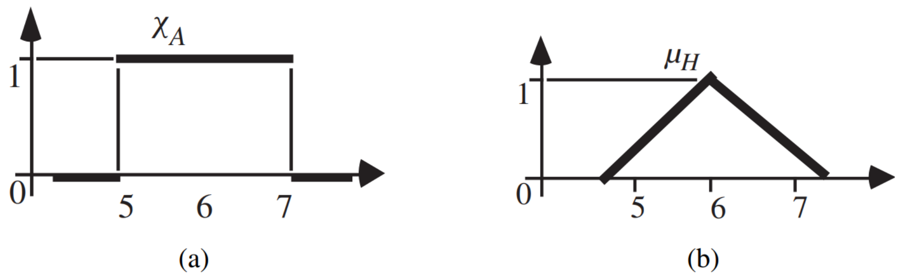


Figure 3.1: An example membership function for a crisp set (a) and fuzzy set (b) (from [119]).

about the evidence, making this theory suitable for modeling imprecise knowledge [11]. Uncertainty is measured with belief and plausibility, which define the lower and upper bounds of a probability interval [11, 161]. The interested reader is referred to the work of Shafer [131].

3.3. Literature Consideration

The need for a framework for systematic uncertainty assessment in complex environments was made clear in the conceptual research design. To gather the knowledge required to answer the research questions, a literature study was conducted on the topic of uncertainty analysis and sensitivity analysis. The literature is presented in a step wise manner, to a certain extent building on the basic steps for a systematic uncertainty assessment outlined by Booker and Ross [16].

3.3.1. Identification

Identifying uncertainty involves selecting the uncertainties present in a system. It is noticed by Meijer [84] that different definitions of uncertainty are given in various studies and that there exists no general agreement on the classification of uncertainty either (more on this in subsection 1.2.2). In this study, uncertainty is defined as: "any deviation from the unachievable ideal of completely deterministic knowledge of the relevant system", adopted from Walker et al. [151]. Uncertainty therefore not only occurs when there is a lack of knowledge. In fact, gathering more knowledge could either reduce or increase the uncertainty.

It may be useful to distinguish uncertainties from assumptions. "Uncertainty is a fact of life" [151] and identifying every uncertainty present in a complex system is an impracticable task. Assumptions allow us to restrict the considered uncertainty. In practical terms, elements (eg. model parameters) can then be considered uncertain if they affect the Object of Interest (OoI) or are effected by the OoI, which is indeed problem specific. Uncertainties are propagated, to the best of the assessor's ability, through an environment by considering, for example, the range of values that a parameter can take. An example of uncertainty is the amount of rain that will fall tomorrow in a particular place. Assumptions, much like uncertainty will be documented throughout this the research project, but unlike uncertain parameters, will be handled in a deterministic way. An assumption could be for example discarding third- and higher-order terms in a Taylor series, or for the previous example, assuming 20 mm rainfall.

3.3.2. Classification

The past decades, objective and perceived uncertainty have given food for discussion, given the debate among researches [60, 66]. According to proponents of the perceived view on uncertainty, the environment is not certain or uncertain [84]. They argue uncertainty depends on the individual, more specifically on how the environment is perceived by that individual [26, 86]. The objective view on uncertainty assumes that uncertainty originates from the environment and can be measured objectively [27, 84]. The perceptive view on uncertainty may be adopted in research where the behavior of individuals or actors is important. In the current study, actors could be stakeholders and non-experts whom uncertainty results are presented to. Although one could argue this induces perceived uncertainties, the relevance of the individual in this research is lower than in typical socio-technical studies (where the perceptive view on uncertainty might be a more appropriate choice). The objective view on uncertainty is therefore adopted.

Various ways of uncertainty characterization (or classification or categorization) have been proposed or adopted in literature, depending on the type of application [114, 115, 135, 147, 151]. Walker et al. categorized uncertainties in a holistic way in the context of model-based decision support [151]. The authors argued uncertainties can be categorized based on three dimensions, the nature of uncertainty; the level of uncertainty; the location of uncertainty. The authors provide a rigorous analysis of the terminology and topology in uncertainty analysis as a means of harmonizing different contributions in existing literature [151]. Kwakkel et al. reviewed the literature that extends on the uncertainty matrix constructed by Walker et al [72]. They presented an updated, synthesized uncertainty matrix with the aim of harmonizing the emerged variants of the original framework. The streamlined topology description as well as its general applicability in model-based decision support activities led to its partial adoption in this framework for uncertainty assessment. Several modifications to the framework were deemed appropriate due to the papers' decision-oriented focus and the more technical perspective of this thesis. Some changes were inspired by other literature. As part of the methodological framework proposed by Walker et al., the authors presented a so-called uncertainty matrix. They argue that systematically using this matrix for uncertainty identification, characterization and communication can increase the quality of model-based decision-support. Furthermore, it is suggested to combine this matrix with sensitivity analysis, which can be used at a later stage to screen out variables that have marginal contribution to the output uncertainty. This uncertainty matrix, or a derivative of it, can also be presented in reporting as indicated by the authors to make the uncertainty characterization process transparent towards stakeholders [151].

First, the three dimensions of uncertainty initially proposed by Walker et al. are described and a careful analysis was made of the applicability of the proposed definitions and subdivisions to the research problem at hand. Adjustments are described and finally the complete matrix for classification is presented.

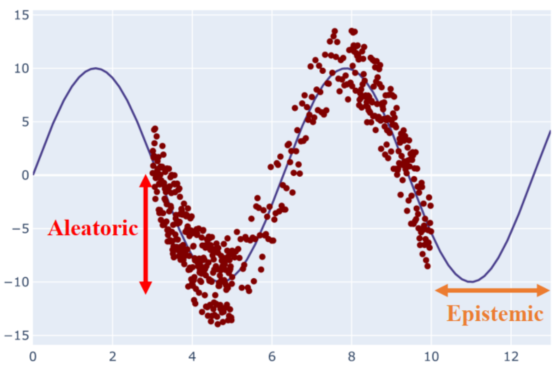


Figure 3.2: Representation of aleatory vs epistemic uncertainty (from [3].

Nature of uncertainty

The nature of uncertainty (also referred to as form or type) has been treated using various taxonomies and terminology in literature [161]. Since different classification bases exist, the context and scope of the research usually determines the scheme used for uncertainty classification [147]. The uncertainty type is commonly classified into one of the following categories: aleatory and epistemic and uncertainty [64], defined as shown below. This classification scheme seems to be widely accepted by the engineering community and finds application in model-based decision support [3, 72, 145, 147, 151, 152, 161].

> **Aleatory:** uncertainty caused by natural randomness of a process or system. This type of uncertainty is therefore also referred to as type A, random or stochastic uncertainty. Aleatory uncertainty can not be reduced but, generally speaking, can be quantified [52, 118, 161]. An example of aleatory uncertainty is the occurrence of turbulence at high altitudes.

> **Epistemic:** also called reducible, type B or cognitive uncertainty, is caused by the imperfection of the existing knowledge. This means that, contrary to aleatory uncertainty, epistemic uncertainty can be reduced by increasing the level of knowledge [52, 151, 161]. An example of epistemic uncertainty is the effect of the take-off derate on the EGT margin erosion, since the uncertainty can be reduced by acquiring more knowledge.

Differentiating between aleatory and epistemic uncertainty in the output uncertainty also holds a distinct advantage to decision making. Stakeholders can then decide whether investing more time and money in acquiring this knowledge is justified or the uncertainty is largely due to natural randomness (aleatory) and increasing knowledge will not be effective at reducing the uncertainty.

Another categorization scheme worth mentioning uses the notion of inference to classify uncertainty types [16]. Inference in this context is described as the deviation between the measured and desired quantity. Several examples of inference types in engineering applications are: predictive (eg. forecasting based on historical data, thereby inferring future behavior from knowledge about the past), statistical (eg. inferring behavior of population from a limited sample), etc. Uncertainty from making inferences is often neglected, although its presence is likely [16]. However, such a categorization seems less appropriate for use in decision making than, for instance, differentiating between aleatory and epistemic uncertainty. Furthermore, uncertainty from inference is rarely considered in papers, and no generally accepted framework for its considerations has been found. Therefore, the differentiation between different types of inference is beyond the scope of this study.

Level of uncertainty

One of the dimensions of uncertainty identified by Walker et al. is the level of uncertainty [151]. In their original framework, three levels were distinguished: statistical uncertainty, scenario uncertainty and recognised ignorance. After revision of the framework by Kwakkel et al. [72], these were replaced the following levels: shallow uncertainty, medium uncertainty, deep uncertainty and recognised ignorance. A short description based on Kwakkel et al. [72] and its application to this thesis provided below. It should be noted that the intended purpose of specifying the uncertainty level it not to assign a degree of uncertainty to specific parameters in the most 'correct' way. Instead, as noticed by the authors, this approach is indeed subject to interpretation of the assessor, making its main goal promoting effective communication by handling them in a systematic and transparent way [72].

Level 1 (shallow uncertainty): an uncertainty is considered shallow if alternative model structures can be enumerated and probabilities to these different model structures can be assigned. This can be seen as a situation in which sufficient knowledge/data is available to model the uncertainty using probability theory.

Level 2 (medium uncertainty): for this level of uncertainty, alternatives can be ranked and a likelihood can be assigned to each scenario. Consider following fictitious example. Engine performance deterioration due to a particular type of FOD is either unlikely, likely or very likely. The three scenarios can be ranked based on likelihood of occurring but how much more likely one is than the other is not specified. In this case, probabilities should not be the primary choice for modeling uncertainty.

Level 3 (deep uncertainty): (model structure) alternatives can be listed but cannot be ranked. Going back to the previous example, those three scenarios could be listed but it can not be specified which one is more likely. Deep uncertainty, also referred to as severe uncertainty [71], seems to be infrequently addressed in literature. According to Popper et al. [107], deep uncertainty arises when one faces problems for which the understanding is very limited. The authors have developed methods to systematically deal with deep uncertainty. Schawabe et al. [128] phrase deep uncertainty as "A decision-making situation where Knightian uncertainty, conflicting divergent paradigms and emergent decision making are relevant". Hallegatte et al. render the term more operational for policy analysis [7]: "Deep uncertainty is a situation in which analysts do not know or cannot agree on (1) models that relate key forces that shape the future, (2) probability distributions of key variables and parameters in these models, and/or (3) the value of alternative outcomes" [44]. A clear link with predictions about the future indeed appears. Given the complex and interdisciplinary environment of this research problem with input parameters carrying large uncertainty, this advocates considering deep uncertainties in this analysis.

Level 4 (recognized ignorance): concerns uncertainty for which no (model) alternatives can be listed. As for the interpretation of uncertainty due to ignorance in this study, it is seen as uncertainty that one recognizes the existence of, but no further knowledge or information is available that could be used to represent the uncertainty. Uncertainty can be grouped into this level category when one for example accepts the possibility of being surprised.

In the context of this thesis, assigning the level of uncertainty in the uncertainty classification step is essential. Given the multi-level uncertainties in the simulation environment, defining the uncertainty level in a systematic way can benefit the selection of suitable uncertainty quantification techniques in further stages of the uncertainty assessment.

Location of uncertainty

The location (often called source) of uncertainty refers to where the uncertainty originates from within the entire system [151]. Similar to the categorization of uncertainty types, many different ways to group uncertainty locations can be found in literature. Kwakkel et al. provides a rather detailed breakdown with a decision-oriented focus [72]. For this thesis however, an approach from a more general simulation model perspective is deemed more appropriate. Du and Chen categorize uncertainty locations into two groups: external and internal uncertainty [33]. This classification is based on whether or not the uncertainty originates from within the simulation environment [161]. External uncertainty is the input data uncertainty (or input uncertainty), and internal uncertainty can be broken down into model parameter uncertainty and model structure uncertainty. Model parameter uncertainty is caused by limited information when fitting model parameters for an assumed model form. Uncertainty in the model structure is due to the validity of the chosen model and its underlying assumptions [33].

It was also mentioned that errors may be present due to the computer implementation of the model [33]. Events leading to model error include for instance programming errors and round-off errors, the effect of which can be estimated by performing model verification [161]. These errors are technically not a type of uncertainty, but they influence the model accuracy [118]. Kwakkel et al. and Walker et al. refer to this as model implementation and computer implementation respectively [72, 151]. The latter description is deemed least ambiguous and has therefore been adopted in this study.

Uncertainty Classification Matrix

Based on available studies and an interpretation and modification of past work in the context of this thesis project, a matrix for systematically classifying identified uncertainty is presented in Table 1.1.

Table 3.1: Matrix for uncertainty classification (partly based on [72]).

	Level				**Nature**	
Location	Level 1: shallow uncertainty	Level 2: medium uncertainty	Level 3: deep uncertainty	Level 4: recognised ignorance	Aleatory	Epistemic
Input data						
Model parameters						
Model structure						
Computer implementation						

3.3.3. Quantification

With the uncertainties classified according to the framework outlined in subsection 1.2.2, an appropriate representation of each uncertainty using one of many available techniques needs to be found. Schwabe et al. [128] state that "a suitable metric is defined as being one which avoids the need for data normalization in order to achieve statistically significant results". For instance, if full statistical information is not given, a more conservative theory (eg. possibility theory) might be a more appropriate choice than a more restrictive one (eg. probability theory) [16].

Various general theories for representing uncertainty have been presented in section 3.2. Since not every theory is equally suitable for every situation, this section aims to provide some guidelines for uncertainty quantification (also called uncertainty modeling). To this end, an overview of appropriate methods is provided in Table 1.2. Based on literature, frequently addressed advantages and limitations relevant to this research problem are provided, as well as a set of applicative studies found in literature. It should be noted that this list of theories and methods for uncertainty modeling is not exhaustive. The aim is to provide a top-level overview of techniques used in literature with a practice-oriented perspective, that can be used when selecting a quantification method.

Some final remarks are given with respect to the different theories and their applicability to the different types, levels and/or location of uncertainty. Probability theory is the least conservative theory (out of the theories considered in this study) and may be used if full statistical information is available [11, 23]. That is, modeling uncertainty using probability theory involves assuming a probability distribution and fitting the model parameters through for instance Maximum Likelihood Estimate (MLE) methods [118]. Generally speaking and employing the taxonomy presented in subsection 1.2.2, probability theory can be useful for modeling shallow uncertainty originating from input data or model parameters, of the aleatory type. Approaches relying on the subjective interpretation of probability, i.e. Bayesian approaches, can be used to represent model structure uncertainty [31], and is reported to suited for both aleatory and epistemic uncertainty [11, 161]. Possibility theory is a more general theory giving more conservative results than probability theory and thus is better suited to modeling medium uncertainty, i.e. when limited information is available [24, 149]. Evidence theory appears to be able to deal with a large range of uncertainty levels (from shallow to deep uncertainty). Both possibility theory and evidence theory are reported to work well for aleatory and epistemic uncertainty [161]. Finally, interval analysis is considered the simplest way (conceptually) to represent epistemic uncertainty [38, 161]. As only a minimum and maximum value are provided, and it is assumed that nothing is known about the uncertainty (except these bounds) [38], it seems a suitable technique for modeling deep uncertainty.

It should be noted however that the presented analysis is merely based on general practice found in literature. The way an uncertainty occurs, what its location is and how uncertain it is should be evaluated on a separate basis such that a suitable modeling technique can be selected.

3.3.4. Propagation

In previous steps, uncertainty has been identified, classified and techniques to represent uncertainty have been considered. The next step is to combine quantified uncertainties and propagate them for instance through the black-box model. Many techniques for uncertainty propagation are available. These can be broadly categorized into intrusive and non-intrusive methods [161].

The first practically used methods for uncertainty propagation were Monte Carlo Simulation (MCS) methods [85]. These are non-intrusive, meaning that no changes to the original simulation model (which handles the deterministic propagation) are required [161]. MCS is therefore particularly popular for uncertainty propagation through a black-box system [118]. Another advantage is that the estimation accuracy is insensitive the the problem dimensionality [161]. The idea is to sample repeatedly from a probability distribution and propagate the value through the model (perform a simulation), and analyze the samples. One of the main problems with MCS it that, because of the Central Limit Theorem (CLT), a large number of samples is needed in order to get accurate results [118]. Different sampling strategies can be used to reduce the amount of samples needed. Helton et al. [50] reviewed sampling-based methods for uncertainty, including different sampling strategies such as Latin hypercube sampling [48, 83], random sampling and importance sampling. They conclude Latin hypercube sampling can be used instead of random sampling in case the computational cost is to be reduced. Yao et al. [161] in their review on uncertainty propagation methods for Uncertainty-Based Multidisciplinary Design Optimization (UMDO) list MCS methods for other uncertainty quantification theories, namely possibility theory [133], evidence theory [61] and interval analysis [65].

Intrusive methods do require modification of the simulation model. That is, the simulation model is reformulated (into a so-called surrogate model) to include uncertainty directly in the system [161]. A commonly used example is Polynomial Chaos Expansion (PCE) (note that also non-intrusive methods based on PCE exist, eg. Ref. [39]), which uses differential equations to analyze the uncertainty in a system [118]. PCE-based methods are reported to be less computationally expensive than MCS-based methods [169]. According to Roelofs and Vos [118], PCE-based methods can potentially replace MCS-based methods, which are still the most popular method for uncertainty propagation. Another surrogate model (or meta-model) is Stochastic Collocation (SC) which appears to be even more efficient than PCE [39, 103]. A comparison of PCE and SC for uncertainty analysis can be found in [39].

Other methods for uncertainty propagation exist, for instance methods for reliability analysis (eg. First Order Reliability Method (FORM) and Second Order Reliability Method (SORM)) [161]. Roelofs and Vos argue however that few studies use these methods and thus their applicability may be low [118].

Table 3.2: Overview of theories and methods for uncertainty modeling.

	Advantages	Disadvantages/Limitations	Applications
Probability Theory	Freq.: Relatively straightforward to implement [118] Freq.: Most decision makers and analysts are familiar with it [16, 118] Freq.: Preferred choice when full statistical information is available [23] Bayes.: Suitable for aleatory and epistemic uncertainty [11, 161]	Freq.: Produces least conservative results [11, 118] Freq.: May give false appearance of exactness [49] (caused by previous point) Freq.: Questionable for modeling epistemic uncertainty [6] Bayes: relies on 'Principle of Insufficient Reason' when constructing Bayesian belief (explained in Ref. [13])	Freq.: [22, 106, 136] Bayes.: [68, 140, 153, 154]
Possibility Theory	Higher applicability to rare events due to less restrictive axioms (compared to probability theory [16]) Can be used with limited information about the uncertainty [167], more conservative results [23] Suitable for aleatory and epistemic uncertainty [161]	Less understood by decision makers [16] Lacks operational definition [25]	[19, 45, 91, 164]
Evidence Theory	Suitable when limited information [6, 61], can deal with well characterized uncertainty as well as near-total ignorance [97] Useful when conflicting evidence is present. If not the case, possibility theory is more appropriate. [91, 118] No assumptions required from analyst [118] With increasing information, results approach results obtained through probability theory [61, 92, 118] (advantageous for industry [161]) Can be used to represent model form uncertainty [101]	Evidence rule potentially unreliable for highly inconsistent data [161] Tough to make decision in case of wide bounds [118, 140] Worse for decision making than probability theory in long run (but can be combined with Bayesian Theory) [140] Propagation computationally expensive compared to probability theory [61]	[6, 92, 96]
Interval Analysis	Straightforward communication to stakeholders due to simplicity [49]	Effective propagation might be challenging [38], despite simple uncertainty representation Computational cost prohibitive when number of inputs is large and output range can be incorrect if function contains local extrema [91]	[82, 110, 111]

Combining Uncertainty Types

So far the description has not dealt with propagating uncertainties of different types. This might occur when epistemic uncertainty is not modeled using some probability distribution, but aleatory uncertainty is. Several approaches were reviewed for handling such situations.

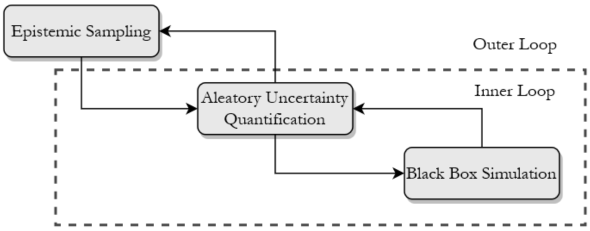


Figure 3.3: Schematic representation of second-order probability based on [20].

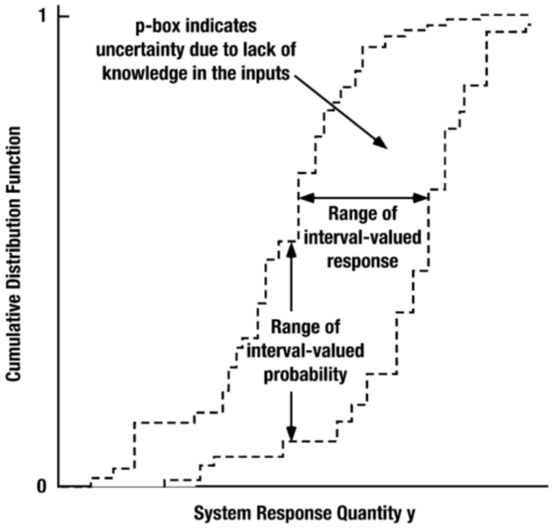


Figure 3.4: Example of p-box visualization (from [98]).

Ideally, aleatory and epistemic uncertainty should be propagated simultaneously, for instance by segregating the propagation of both types of uncertainty using a nested loop (sampling-based) method involving second-order probability [38]. This way, the effect of aleatory and epistemic uncertainty is dealt with at the same time [20]. A schematic representation adopted from Brune et al. [20] of this approach is shown in Figure 3.3. In the outer loop, epistemic variables are sampled. These samples are then passed to the inner loop where for each of these samples aleatory variables are sampled and propagated through the model. The amount of samples after completing this iterative procedure can indeed be obtained by multiplying the amount of epistemic samples by the amount of aleatory samples [20].

Roy and Oberkampf [121] employed this approach in their framework for uncertainty quantification in scientific computing. They quantified aleatory uncertainty using probability distributions, while epistemic uncertainty was treated using interval analysis. A cumulative distribution function of the system response was then computed based on the samples from the aleatory uncertainty. As the epistemic variables (outer loop) simply contain a range of values bound by a minimum and maximum value, a CDF of the system response was computed for each outer loop. From this ensemble of CDFs, a probability box (or p-box) was built, an example of which is shown in Figure 3.4. A p-box incorporates the degree to which knowledge is available in the representation of the model output. The model output is indeed not represented by an exact probabilistic value, but instead using an interval-valued probability [121]. The advantage of being able to treat aleatoric and epistemic uncertainty separately comes at the expense of higher computational complexity. To decrease this computational effort, Eldred and Swiler [38] proposed an approach that employs a stochastic expansion method for the inner loop, while the outer loop bounds are determined using interval optimization.

Huang et al. have studied combining aleatory and epistemic uncertainty in Reliability-Based Multidisciplinary Design Optimization (RBMDO), by modeling them using random and fuzzy variables respectively [54]. Du and Choi proposed a Mixed Random and Fuzzy Variable Design Optimization (MVDO) method to fill the gap between Possibility-based Design Optimization (PBDO) and Reliability-based Design Optimization (RBDO), as in the former the input uncertainty needs to treated as random variables whereas in the latter only fuzzy variables can be used to represent uncertainty [32]. Further research would be necessary to understand the basis used in these papers for combining both uncertainty quantification techniques and understand the applicability of proposed methods in the field of Multi-disciplinary design optimization (MDO) to this thesis. Singpurwalla and Booker [134] incorporate fuzzy sets within the framework of probability theory. The foundation for their argument is based on, inter alia, the subjective interpretation of probability theory.

Although some of these studies seem to address the propagation of mixed uncertainty using different techniques, the problem remains challenging [46]. Segregation of aleatory and epistemic uncertainties has been a common thread in the approaches presented. However, the implementation tends to be application specific, depending on the model complexity, the selected quantification techniques for different types of uncertainty the problem contains and the amount of samples generated. Integrative (mixed) uncertainty assessment are rare, especially with a scale comparable to the multi-source and multi-level environment in the current research.

3.3.5. Visualization and Communication

Presenting results from uncertainty assessments to stakeholders or non-experts remains challenging [49]. However, effective communication of the output from the analysis is an important element of the uncertainty assessment. More specifically, the scenario in which results are interpreted poorly or even wrongly by stakeholders needs to be avoided. The notion of perceived uncertainty, as treated earlier, describes this well. Although perceived uncertainty will not be considered here, it is deemed advisable to consider the effect of modeling choices on the communication of uncertainty, and thus the decision making process.

Arguably one of the essential aspects that determine how well a presentation of results is conceived is the underlying uncertainty modeling techniques used. Helton et al. performed an exploration of alternative uncertainty modeling approaches and considered the uncertainty visualization aspect in their discussion [49]. They state that interval analysis provides the most convenient basis for communicating uncertainty due to its simplicity, i.e. only a lower and upper bound are provided. Probability theory is due to its popularity in many fields known to most technical people, although it might be understood worse than generally thought. When evidence theory and possibility theory are used in analyses, effective communication and interpretation of results becomes harder since among stakeholders and non-experts, low familiarity with these techniques exists. Presentation of results obtained in this research project should therefore be well thought-through. While uncertainty visualization and communication (also referred to as uncertainty management) might not dictate the actual theory or method used for uncertainty modeling, it should receive sufficient attention in the final part of the uncertainty assessment.

3.3.6. Sensitivity Analysis

Sensitivity Analysis (SA) is concerned with studying the effect of the uncertainty in model inputs to on the model output uncertainty [48]. It is closely related to uncertainty analysis, which instead focuses on quantifying the uncertainty in the model output [124]. Hence, it is advised to perform uncertainty analysis and SA in tandem. SA methods are commonly classified into either Local Sensitivity Analyses or Global Sensitivity Analyses, depending on the scope of the model input variation. Several other techniques, whether or not borrowed from other fields, can be used to explain the model outcome. As they share the same objective of giving model insight, they are discussed in this section together with SA methods despite not necessarily being a type of SA.

Local Sensitivity Analysis

Local Sensitivity Analysis (LSA) can provide insight as to how the output behaves as one or more input parameters are varied [118]. This is especially useful for deterministic problems, but can also be used in non-deterministic problems for instance to verify individual parameters. The sensitivity of the model output with respect to the varied input can be found by computing its partial derivative. These derivatives can be computed in multiple ways, including symbolic differentiation, numerical differentiation and automatic differentiation. Specific limitations and the computational cost of each technique may be considered when selecting a technique for a particular application.

Global Sensitivity Analysis

In LSA, the effect of making small changes to an input variable was found by computing the partial derivative near the instance of interest. In Global Sensitivity Analysis (GSA), the effect of larger input changes (in fact over the entire range of input values [118]) is considered to explain which variables strongly influence the model output [100]. An example workflow for GSA is shown in Figure 3.5.

Iooss and Lemaitre [57] distinguish three groups of methods: screening methods (which involves coarse ranking of the input variables depending on how much influence they have), measures of importance (which provide quantitative information on how influential input variables are) and deep exploration of the model behavior (i.e. going beyond scalar sensitivity indices using graphical techniques and introducing metamodel-based methods to decrease computational cost). The latter is only briefly touched upon as an extension to quantitative importance measures.

Screening techniques can be used if one wants to identify inputs with little influence on the model output while keeping computational costs low. This is useful in many applications since a model usually only contains a small amount of influential parameters [124]. Screening parameters is generally achieved by discretizing the inputs into different levels [57]. Several techniques exist for sampling parameters, such as one-at-a-time sampling and fractional factorial sampling. Despite Saltelli et al. [124] stating that one-at-a-time sampling is not efficient when the number of parameters is large compared to the number of parameters that influence the model output, most of the popular screening methods used in engineering applications rely on this sampling technique. The way one-at-a-time sampling works is by varying the value of only one input across multiple simulations. An effective method for screening is the Elementary Effects (EE) method, the idea of which was introduced by Morris. This method is able to group the effect of each input has into three categories: negligible, linear and additive, or nonlinear or involved in interactions with other methods [124]. Further details on this method can be found in Refs. [57, 90, 124].

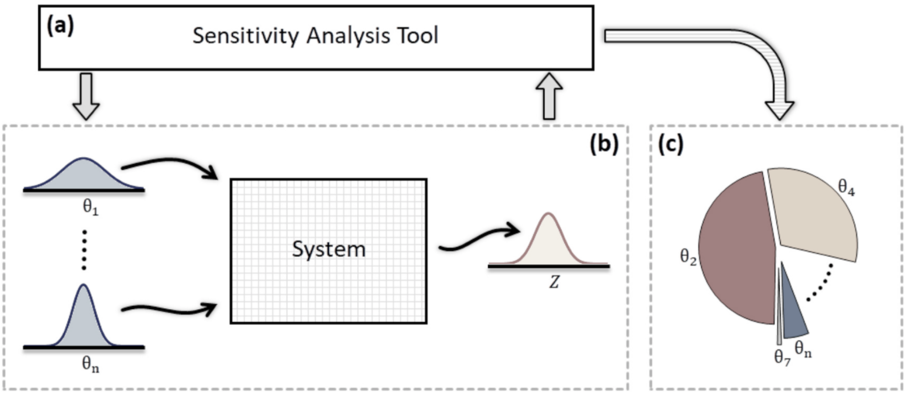


Figure 3.5: Top-level overview of how GSA is often performed. Samples from input distributions are propagated through a system. The output and input are used by a GSA method to compute sensitivity indices, and apportioned uncertainty is visualized in (c). From Ref. [113].

Various categorizations of quantitative importance measures exist. In this brief review, several regression-based approaches and distribution-based (including variance-based) techniques are described. Regression-based approaches rely on linear models to explain the change in the model output given the input values [57]. The coefficients of the linear model that fit the original model best then explain the global effect of each variable [108]. Several global sensitivity measures can be derived from such fitted model, including Pearson Correlation Coefficient, Standard Regression Coefficient and Partial Correlation Coefficient.

Another group of methods are variance-based. Analysis of Variance (ANOVA) is a popular method in which a function is decomposed into different components and the effect and variance of these contributing components are computed [57]. The variance of these components can then be used to calculate sensitivity indices. The first-order index for instance S_i defines the contribution of a variable i to the total variance. Higher-order sensitivity indices (representing interactions between inputs) can be calculated too, as well as total sensitivity index. Calculating these indices is expensive due to the presence of integrals in the variance calculation, which is why often simulation-based methods are used [108]. The calculation of Sobol' indices [138] for instance relies on Monte Carlo simulations. To get accurate results for the sensitivity indices however, a large number of simulations are required [57]. One can use the Quasi-Monte Carlo methods which has been reported to reduce the computational cost in some cases by a factor of 10 [124]. Common Quasi-Monte Carlo sequences include Sobol sequences and Latin Hypercube [108].

Fourier Amplitude Sensitivity Sampling (FAST) is a method that decomposes a function into a Fourier basis. In a study by Tarantola et al. [143], FAST has been combined with Random Balance Design (RBD). However, FAST might become unstable and biased when the number of inputs is higher than 10 [57, 146]. Given the potentially large amount of uncertainties in the use-case of this thesis, this limitation might become prohibitive for using FAST.

The Python library SAlib [51] features most of the methods described above, including the Elementary Effects for screening and quantitative importance measures such as Sobol' and FAST. Furthermore, SALib is considered by Douglas-Smith [30] the most comprehensive software for UA/SA in Python.

In Table 3.3, various techniques for GSA featured in SALib are evaluated against several criteria considered important for engineers performing sensitivity analysis. Consider for instance a so-called given-data situation (where data is observed and the original distribution is not known). Generally, methods with a dedicated sampler can not be applied to such problems since they expect samples to be created using a specific sampling scheme. Recently however, methods have emerged that can directly be applied to existing data [113], eg. Ref. [104] (DMIM) or [143] (RBD-FAST).

Dependency between inputs is another critical point (as ignoring correlation effects biases SA results [29]) that is receiving increasing attention in literature [113]. Note the difference between the dependency or correlation effect and interaction effects: "the correlation effect is different from the 'interaction effect' which refers to the presence of non-additivity of the effects of individual inputs on the system output" [112, 113]. Kucherenko et al. for instance realize a generalization of Sobol' indices for dependent inputs using copulae [67, 113]. Li et al. [75] generalize Sobol' indices using a surrogate approach called Random Sampling-high Dimensional Model Representation to estimate the sensitivity indices. However, according to Wiederkehr [156], the method "fails to detect even strong interaction effect in many cases".

Another important aspect to consider when selecting a GSA technique is the type of importance measure returned by the method. In models with high interaction effects, first order sensitivity indices will only be able to explain a portion of the output variance. In these cases, methods providing higher-order and/or total order sensitivity indices could be a necessity. Other distribution-based methods produce different indices to explain the importance of variables. DMIM [17], for example, is a moment-independent method as "it measures the difference between the unconditional distribution of the output and its conditional counterparts" [113]. Instead of only looking at the variance in the output distribution, the new measure considers the complete input/output distribution [17].

By comparing GSA methods on the basis of these criteria, specific recommendations (depending on the type of model) could be made to practitioners applying GSA. The defined criteria are very relevant in the context of this thesis, due to the interdisciplinary use-case with potentially dependent variables, observed-data and other model-specific requirements.

Table 3.3: Feature comparison of GSA techniques as implemented in SALib [51], partly based on [105]. N and D represent the number of samples and the number of parameters respectively.

	Method	**Total number of simulations**	**Given-data compatible**	**SIs provided**			**Correlated inputs**	**Supports grouping**
				S_I	S_{II}	S_T		
Distribution-based	Sobol'1 [138]	$N \cdot (D+2)$	-	✓	-	✓	-	✓
	Sobol'2 [138]	$N \cdot (2D+2)$	-	✓	✓	✓	-	✓
	eFAST [122]	$N \cdot D$	-	✓	-	✓	-	-
	RBD-FAST [143]	N	✓	✓	-	-	-	-
	HDMR [75]	N	✓	✓	✓	✓	✓	-
	DMIM [17]	N	✓	N.A.			✓	-
Derivative-based	DBGSM [139]	$N \cdot (D+1)$	-	N.A.			-	-
	Morris [90]	$N \cdot (D+1)$	-	N.A.			-	✓

Other Methods for Explainability

Several other techniques from other fields could be used to provide insight on which input parameters are most affecting the behavior of the output of the model. In machine learning for instance, the popularity of XAI (Explainable Artificial Intelligence) has increased significantly over the past years. This is caused by the overall increased application of machine learning and in particular its sub-field deep learning [150]. Models have become increasingly accurate but, at the same time, many of those lack explainability and interpretability. In the context of this research project, two methods for explainability that create visual explanations are considered: Local Interpretable Model-Agnostic Explanations [116] and SHapley Additive exPlanations (SHAP) [80]. Since these methods are model-agnostic, they are applicable to the interdisciplinary model in this thesis. Practical advantages and disadvantages of each method are described in the work of Molnar [88].

The idea behind Local Interpretable Model-Agnostic Explanation (LIME) is to train local surrogate models on the outcome/predictions of a black-box model in order to explain individual model outcomes [88]. To do so, the input data is perturbed and propagated through the black-box model, resulting in different model outcomes. The input and output data for these variations are then used to train a weighted interpretable model (which is a simple model like for instance linear regression), which has similar performance to the original model locally, but not necessarily globally. The weights depend on the distance from the new samples to the instance being explained.

SHapley Additive exPlanations (SHAP) construct an interpretable model that approximates the original model by using additive feature attribution methods (which can be seen as linear combinations of the features) [150]. The method builds on Shapley values, a technique from coalitional game theory that was introduced in 1953 for assigning the payouts to each player in a coalition depending on their respective contribution to the total payout [88, 132]. Unlike Sobol' indices in GSA, SHAP does not give global explanations, i.e. it does not provide global feature importance [79]. Lundberg et al. developed a set of methods that can globally explain the model structure for tree-based machine learning models by combining local explanations from many samples "while retaining local faithfulness to the original model [117], which produces detailed and accurate representations of model behaviour" [81]. Another difference mentioned by the author of the SHAP software package [79] is that Sobol' indices use regular conditional expectation, whereas SHAP uses mostly interventional expectations. Janzig et al. [59] argue that the latter is conceptually correct when attributing features. A more extensive comparison between Sobol' indices and Shapley values has been conducted by Owen [100]. Owen explains that Shapley values for variable importance do not match the Sobol' indices, but instead Sobol' indices bracket the Shapley value. The author concludes that, since Sobol' indices are less computationally expensive, they might be used as bounds for the Shapley value.

3.4. Research Gap

In the literature survey, a framework for uncertainty classification has been presented, along with an overview of uncertainty quantification, uncertainty propagation and global sensitivity analysis methods and theories with their typical uses. Several problems and potential challenges have been identified. It is understood that most challenges originate from the uncertainty complexity, i.e. the presence of uncertainties of different type, level and source complicates the uncertainty analysis in various ways. These include findings ways for combining epistemic and aleatory uncertainty; dealing with the computational expense of combining uncertainties; modeling uncertainty originating from model choice and parameter selection. Finding a method for GSA suited to a specific model and project requirements can be challenging too, since a wide variety of methods exist with distinct properties. Important criteria, depending on the application could be given-data capability, computational cost, type of sensitivity measure provided, suitability to dependent inputs. These considerations seem to be underrepresented in available literature and form a research gap to be addressed by this thesis. Once these issues have been understood and addressed, one can provide value to practitioners by developing a framework for performing systematic and efficient uncertainty analysis in interdisciplinary environments.

4
Thesis Project Plan

With the conceptual research design introduced and the literature discussed, the project plan for this Master's thesis is defined. To some extent, the research methodology has been touched upon in the literature study, albeit from a more technical perspective (for instance the steps for uncertainty assessment). In this chapter, a practical approach is used to further clarify which steps need to be taken to answer the research questions, how this will be done, and in which order they occur. This is discussed in section 4.1. Subsequently, the project planning is provided in section 4.2 by means of a Gantt chart.

4.1. Research Breakdown

Several topics have been discussed in the literature overview. To make sure the interrelations are clear and to provide a concrete set of steps that can be used as a guide throughout the research phase, a high-level framework has been proposed for the development of the sub-models, as well as for the uncertainty assessment. These have also been translated into a set of work packages. Finally, the discrete event simulation framework LYFE and its role in this thesis are explained.

4.1.1. Development Methodology

The literature view treated various methods to model engine performance deterioration. The flowchart in Figure 1.1 depicts the process of selecting a modeling technique to simulate the parameter impact. Such techniques can range from a simple linear relationship found in literature between two parameters, to an advanced software tool relying on a mathematical model to replicate a thermodynamic process. To bring structure to this selection process, the flowchart outlines steps to be taken. Relationships between parameters can be grouped into sub-models. Several hard requirements apply to the the models. First, the model should be compatible and integratable with the simulation environment (more about this in subsection 4.1.3). That is, output parameters from the developed model should be fed into the in-house cost benefits tool. Furthermore, the model should be feasible to be developed within the thesis time frame. Recalling that the engine wash assessment merely serves as a use-case, modeling tools that require a large development time should be avoided. Finally, albeit rather a criterion for model selection than a hard requirement, the research quality ideally should surpass that of existing engine wash assessments (elaborated in subsection 2.2.3).

Similar to the sub-model development, a flowchart has been made of the methodology for uncertainty assessment to aid in conducting efficient research. Section 1.4 shows the process which builds on the methodology for sub-model development shown in Figure 1.1. Once uncertainties have been identified, relevant uncertainties are classified using the categorization scheme described in subsection 1.2.2.

At this point in the framework, it is not known which variables explain most of the uncertainty. In order to avoid the loss of time and computational resources on propagating uncertainties with little effect on the output, the model complexity can be reduced (by reducing the dependencies for a specific uncertainty, and representing it with for instance a simple distribution bounded by the minimum and maximum value the variable is assumed to take). A good basis for deciding where to reduce complexity can be the level dimension of uncertainty, which has been determined for each uncertainty in the classification framework. In a further stage, its influence on the output uncertainty can be determined

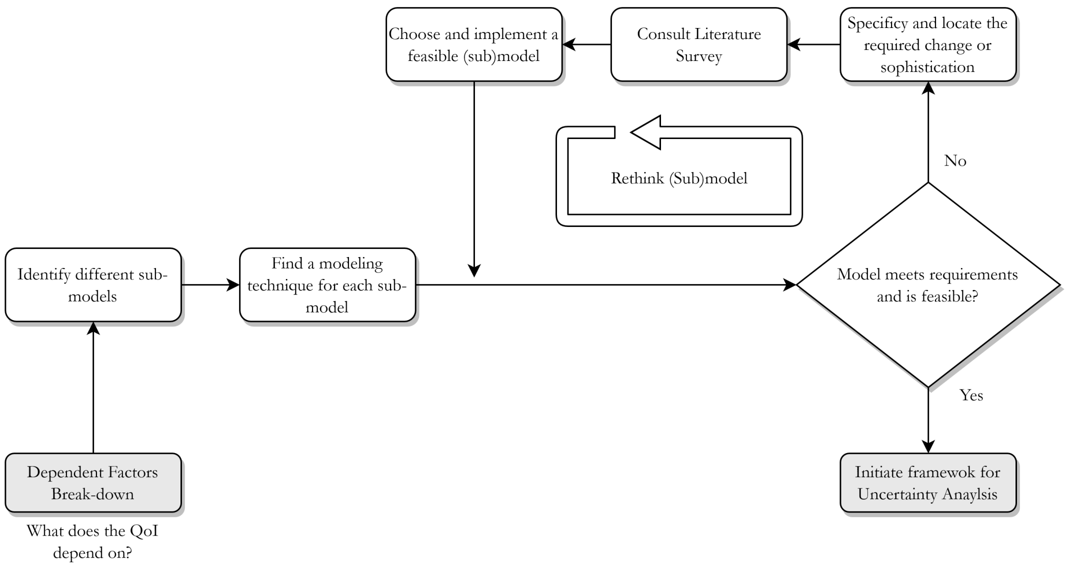


Figure 4.1: Flowchart depicting sub-models development process.

through a GSA method. A large influence then justifies the need for a more sophisticated model, i.e. increasing the complexity. This procedures is represented in section 1.4 by the left-most feedback loop.

In uncertainty quantification, each uncertainty is modeled by means of an appropriate technique (Table 1.2 can be used for this purpose). Uncertainty propagation refers to combining the uncertainties by propagating them through the discrete event simulation framework LYFE. The large grey box in section 1.4 represents activities related to sensitivity analyses, which can be used for multiple purposes. This part in particular is subject to changes and is merely included to give an overview of the potential uses for sensitivity analyses in this thesis. The main purposes of applying sensitivity analysis in this study are model verification, potentially screening out variables that have a small influence in order to reduce the uncertainty problem scale [161], and apportioning the output uncertainty to individual uncertainties. Verification of parameters can be performed by applying local sensitivity analysis (varying parameters one-at-a-time and observing the resulting changes in the output). The developed model can be verified by interchanging sub-models with a less/more sophisticated version (i.e. using a different model, we change the uncertainty in the system by altering the bounds of a variable) and observe the influence on the output uncertainty. For instance, when the bounds get smaller, the total uncertainty should reduce. One must be cautious however when interpreting these results, as deep uncertainty is always around the corner when making a model more complex. Next, GSA techniques can be used to apportion the output uncertainty to individual input uncertainties. This by itself also serves to understand the model, besides making it possible to differentiate between aleatory and epistemic uncertainty.

It should be noted that this research will be conducted without any industrial partner. Currently, no experimental data is available that could be used to validate the sub-models. However, since this research project is aimed at developing an integrative uncertainty analysis framework that can be used in aerospace economic assessments, the lack of experimental data is not an issue. There are no accuracy and precision requirements of the performance deterioration models, which renders the validation of this model beyond the scope of this research. If experimental data would become available during the course of the thesis, model validation could still be performed, yet it is not a primary aim of the research.

4.1.2. Work Packages

The work to be done during this thesis project can be divided into following work packages.

- Development of the model to simulate the degrading environment and engine deterioration. Inputs for this model can be selected based on the literature consideration in section 2.2 while the output(s) holds a parameters which can be readily linked to a change in e.g. EGT or SFC, and used as an input to the in-house economic assessment framework named LYFE.

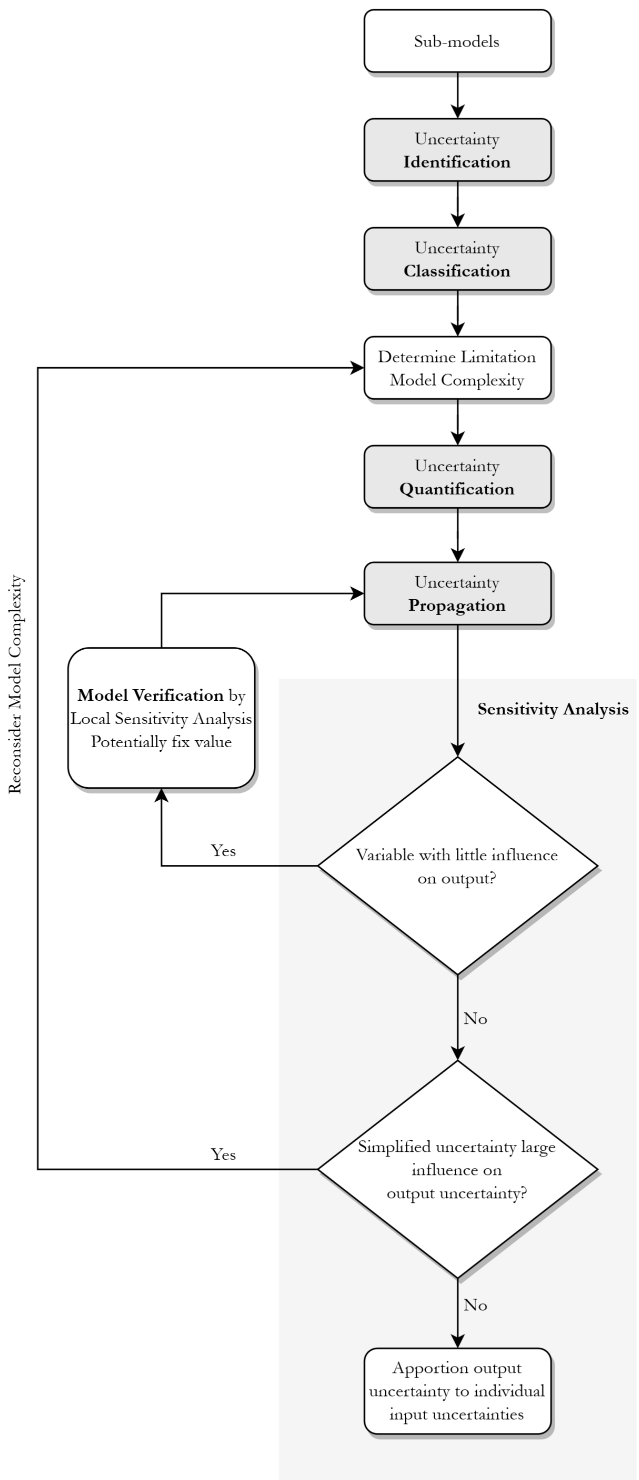


Figure 4.2: Flowchart depicting uncertainty assessment process.

- Integration of sub-models (degrading environment) into LYFE (coding effort).
- Creation of a deterministic simulation of the economic efficiency of engine cleaning procedures as a maintenance application.
- Identification and classification of uncertainty using the categorization scheme presented in this

review.

- Select quantification and propagation techniques for characterized uncertainties.
- Implement quantification and propagation of characterized uncertainties (coding effort).
- Verification of parameters and sub-models by means of local sensitivity analysis and strategic replacement of sub-models.
- Global sensitivity analysis to apportion the output uncertainty to individual input uncertainties.
- Visualization of the non-deterministic economic value of engine cleaning procedures (eg. probability distribution).
- Critical discussion and documentation of results.

4.1.3. Discrete Event Simulation Framework

The cost-benefit tool named LYFE, developed by DLR, has been mentioned several times throughout this report. LYFE is a software tool developed to provide a generic (i.e. not case specific) environment for performing cost-benefit analyses of aeronautical technologies by researchers and analysts alike [106]. Its core module called AirLYFE allows to evaluate these technologies, including aircraft, operational procedures as well as maintenance strategies, from the operator's point of view. This is the simulation environment which will be used throughout this thesis project and will be simply referred to as LYFE.

The software uses discrete event simulation in order to capture the primary (i.e. direct) and secondary (i.e. downstream) effects of economically relevant events. In contrast to equation based costing tools, LYFE can therefore analyze temporal effects, e.g. delays due to unforeseen circumstances. The amount of effects can reach up to 300,000 for a typical aircraft lifespan, resulting in a running time of the simulation of around one to two minutes. An important aspect of LYFE is the modular structure of the framework. The main advantage of the modularity is the high customizability. It is expected that no modifications to the source code of LYFE will be necessary for this study. A custom module (which is run by the core module) will be written to integrate the developed model in LYFE.

To perform an economic assessment of an aircraft (technology) with LYFE, two simulations are performed: the reference case and the study case. The aircraft with desirable properties for this research project will be chosen as the reference aircraft. 35 aircraft are available in LYFE per default, including popular commercial airliners such as several variants of the Boeing 737 family and Airbus A320 family, as well as several long-haul aircraft types. The study aircraft will be essentially the same aircraft, apart from the added engine cleaning maintenance task.

A schematic of the inputs that go into LYFE and the output from the model is shown in Figure 1.2. However in reality, a single run of LYFE per default handles the two separate simulations and the subsequent calculation of the economic metrics and generation of the standardized report. The NPV, measured as the discounted value of a project's or product's cash flows [137], is one of the parameters outputted by the model and arguably the most essential in the determination of the economic value of engine wash procedures. The overall results can then conveniently be presented as ΔNPV, representing economic values. The bell-shaped curve depicted next to these parameters in the figure indicate we are looking for a distribution of ΔNPV, rather than a deterministic result. This is achieved by propagating uncertain parameters through the black-box model.

4.2. Project Planning

A planning is made for the thesis project using a Gantt chart shown in Figure 3.

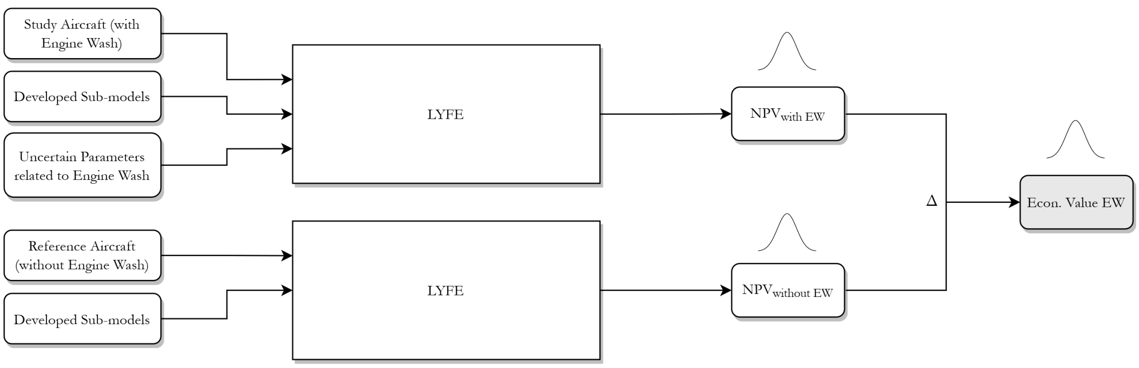


Figure 4.3: Visualization of the inputs and outputs to the cost-benefit tool LYFE for this use-case.

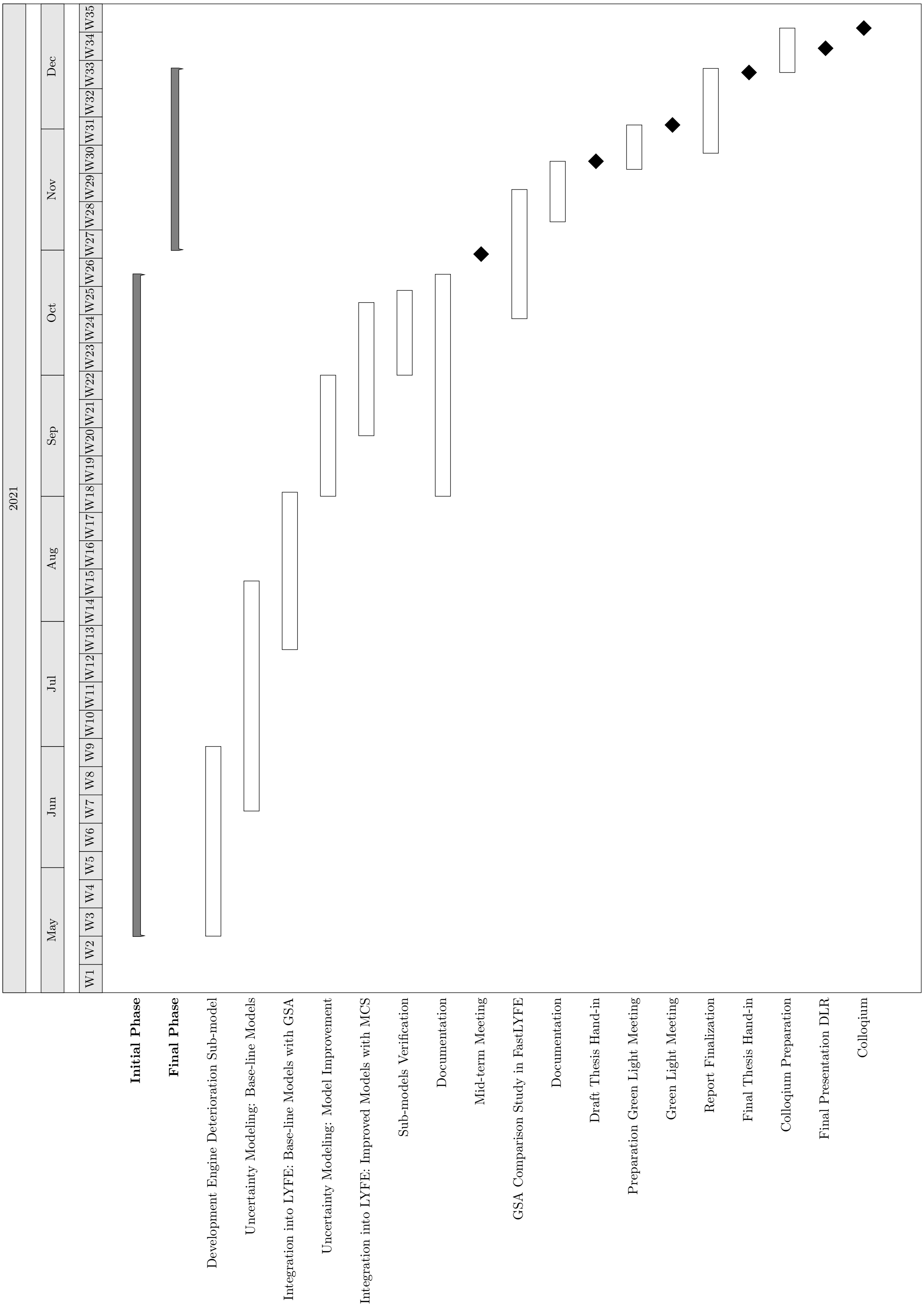


Figure 4.4: Thesis project Gantt chart

5
Conclusion

In the foregoing review on engine wash assessments and uncertainty analysis, knowledge and data have been gathered with the aim of providing a basis that can be used to answer the research questions.

1. Which factors are affecting the economic value of engine wash procedures and what are appropriate models to simulate their behavior and impact on the engine wash economic value?

Factors affecting the economic value of engine wash procedures are broken down in subsection 2.2.1. Data and/or knowledge has been acquired about each factor to support subsequent modeling. Engine performance deterioration is found to be an important factor in this assessment. Therefore, different modeling techniques were reviewed and evaluated on the basis of requirements related to this project (e.g. input/output, complexity).

2.a. Which theory is suited to model the uncertain parameters defined in the identification phase?

2.b. Which method for global sensitivity analysis can be used to apportion the combined uncertainty to the input parameters?

Theories for representing the identified uncertainties are reviewed in Table 1.2.2 and propagation methods were studied in Table 1.2.2. Methods for GSA are discussed Table 1.2.2. These theories and methods are discussed and compared from an engineering perspective and with the aim to aid in developing a systematic and integrative uncertainty assessment.

To conclude, the literature review revealed that integrative techno-economic engine wash assessments are rare. Furthermore, partly due to complexity, uncertainty quantification and sensitivity analysis are often not undertaken in such interdisciplinary studies, or the analysis remains limited to Monte Carlo simulations [105, 161]. This is a problem because:

a) integrative assessments are required to evaluate the economic feasibility of engine wash procedures in a realistic way. To close this gap, a model for engine performance deterioration is particularly required that can be integrated with engine wash assessments. This approach should be compatible with the framework LYFE (e.g. by expressing performance deterioration in terms of for instance EGT). Current data-driven approaches do not solve the problem as they are mostly developed for fault diagnostics and RUL prediction, which differs from the perspective and focus of the current study, and the available data is not (directly) useful for this research project. Developed physics-based models are often either not publicly available or too complex for this thesis with the engine wash assessment being just a use-case. A simpler, hence less accurate, model may be developed that relies on relations between parameters of interest, to represent engine performance deterioration.

b) several factors affecting the economic value of engine wash are highly uncertain (e.g. fuel price). To capture the true value of economic feasibility, the integrated answer must be based on uncertainties, that are modeled using all available knowledge [16]. Most challenges for the uncertainty analysis stem from the complexity of uncertainty in this assessment, i.e. the presence of uncertainties of different type, level and source. These challenges include findings ways for combining epistemic and aleatory uncertainty; dealing with the computational expense of combining uncertainties; modeling uncertainty originating from model choice and parameter selection. Finding a method for GSA suited to a specific problem and project requirements can be challenging too, since a wide variety of methods exist with distinct properties. Once these issues have been understood and addressed, this thesis could provide value to practitioners by developing a framework for performing systematic and efficient uncertainty and sensitivity analysis in interdisciplinary environments.

The methodology for the development of models and uncertainty analysis framework is discussed in section 4.1 and shows in which way the identified research gap is to be filled. Finally, a set of work packages summarizes the work to be done in a chronological order.

III

Research Methodologies

previously graded under AE4010

Executive Summary

This project plan is created for a MSc. thesis project dealing with uncertainty quantification applied to the evaluation of on-wing engine wash procedures. To assess the value of engine wash procedures, the engine performance deterioration rate under different conditions is considered, as well as the level of degradation restoration by engine cleaning. These aspects depend on a variety of operative and environmental factors, which bring uncertainty of different levels and sources into the system. This requires a method for systematic uncertainty analysis in a complex environment to be developed. The uncertainty analysis plays an important role in the research, partly because of the differentiation between epistemic and aleatory uncertainty which aids in decision making. Methods for developing a framework for systematic uncertainty analysis as well as the development of a model for engine performance deterioration are used/developed after an extensive literature review. As part of the framework for uncertainty analysis, uncertainties are identified, classified and finally quantified and propagated through the simulation using one of the techniques presented in an extensive literature study. A global sensitivity analysis then allows for apportionment of the total uncertainty to the individual uncertainties, revealing if and how much uncertainty can be reduced.

This research aims to develop a method for systematic uncertainty analysis in a complex simulation environment with uncertainties of various sources and of different types (aleatory and epistemic). Such integrative assessments are rare, presumably due to the increased difficulty caused by the uncertainty complexity (e.g. the need for combining epistemic and aleatory uncertainty and dealing with the associated computational cost of considering numerous uncertainties; finding a basis for combining values obtained through different uncertainty quantification techniques). This research contributes to integrative uncertainty analyses in complex environments by demonstrating the developed method on the use-case of engine wash assessments.

1.1. Introduction

Wear and tear, as well as dirt accumulation, cause aircraft engines to become less efficient with every flight. This efficiency reduction leads to an increase in Exhaust Gas Temperature (EGT), thereby also increasing the fuel consumption. As this EGT has an upper limit, the Exhaust Gas Temperature Margin (EGTM) decreases as EGT increases. At a certain EGTM threshold, an expensive Engine Shop Visit (ESV) is required, where a significant portion of EGTM is restored.

Due to the high cost of ESVs and engine maintenance in general (engine maintenance accounts for around 35% - 40% of the Direct Maintenance Cost (DMC) [4]) countermeasures are brought up. One of these is to perform on-wing engine cleaning during turnaround or overnight stop, which aims to mitigate engine deterioration (and therefore delaying ESVs). This Engine Wash (EW) procedure typically involves injecting hot water into the engine with the aim of removing accumulated dirt, resulting in fuel savings of up to 1.3% [55]. This value depends on a variety of factors, including the airline's de-rating policy, the outside air temperature, the flight hour to flight cycle ratio, etc. To simulate the impact of these factors on the economic value of engine cleaning procedures, a model that integrates relevant operational and environmental factors on the engine performance deterioration needs to be developed.

This model, and the input variables, introduce uncertainty. Depending on the amount and sort of input variables, choice of model and model parameters, uncertainties can have different locations, types and levels [151]. A framework for multi-source and multi-level uncertainty assessment is therefore needed to systematically address the uncertainties. This framework should guide the practitioner through the process of uncertainty assessment (including identifying, classifying, quantifying, propagating and combining uncertainty) in an interdisciplinary environment with the aim to aid in decision making. As a consequence, differentiating between different types of uncertainty (aleatory vs epistemic) is important, as it enables the recipient to determine whether the output uncertainty can be reduced by attaining more knowledge or the uncertainty is largely due to natural randomness. The engine wash assessment under uncertainty serves as a suitable use-case to demonstrate the developed methods since many uncertain factors affect the economic value of engine cleaning procedures, hence this use-case allows for a multi-source and multi-level uncertainty assessment.

1.2. State-of-the-art/Literature Review

Two major topics in this project can be distinguished requiring an extensive literature study. Engine performance deterioration and engine wash assessments are reviewed in subsection 1.2.1, techniques and theories for uncertainty assessment are dealt with in subsection 1.2.2.

1.2.1. Engine Performance Deterioration

Several engine operating parameters exist, some of which can be used as health indicators providing information about the performance deterioration level of an engine. The most important engine operating parameters are the N1-speed end the EGT [4]. The former is mostly used to indicate the amount of thrust the engine is producing, whereas the latter can be used as a performance deterioration indicator.

The EGT is the temperature measured in the exhaust of the engine. The higher this temperature, the lower the engine efficiency at producing its design trust [4]. That is, lower engine efficiency (mainly caused by lower compressor efficiency) means that in order to produce equal thrust, more fuel is required [12]. Throughout one flight cycle, the highest EGT is usually reached during take-off or initial climb. A general increase in EGT can be observed as the engine ages (the engine burns more fuel to deliver a certain amount of thrust) and can indicate that engine hardware deterioration has occurred [63, 129]. Some of the underlying reasons for the loss of engine efficiency are erosive wear of turbine and compressor blades, increased tip clearance of blade tips and fouling (particles deposited on blade surfaces) [34, 73]. EGT margin deterioration is one of the primary causes for ESV, and is reported to be the main cause in the case of first-run engines (before the first refurbishment) used for short-haul operation, as well as mature-run engines (after the first refurbishment) in long-haul operation [4]. Other important causes for engine removals are hardware deterioration and expiry of Life Limited Parts (LLP) [62].

On-wing engine cleaning procedures can be employed to restore some of the lost EGT Margin. During engine washing, water and cleaning additives are sprayed into the intake to clean the surfaces of both the compressor and turbine stages, while the engine is running. Hence this maintenance task can be used to partially revert deterioration due to fouling. This has been reported to increase the EGT Margin by up to 15 °C [4, 22].

Multiple assessments of engine cleaning procedures have been performed in literature, both for gas turbines with industrial and aerospace applications.

Giesecke and Igie [42] highlight the economic value of compressor washing in their techno-economic study on engine compressor washes for short-range aircraft. They also presented a breakdown of washing costs of compressor washing for a short-range aircraft engine. To model the engine performance deterioration, the software TURBOMATCH by Cranfield University was used. The authors present the effect of fouled blades in terms of a reduction in compressor efficiency. Boyce and Gonzalez [18] developed several tests in a controlled environment to determine the efficacy of engine washing with varying washing frequencies and dissolving agents used for the washing process. Cheng and Sun [22] presented an estimation method of the reduction in engine performance deterioration due to engine washes, taking into account the economic cost of engine washing procedures and the fleet-wide fuel consumption savings.

The engine wash assessments performed in available literature are often not integrative with respect to the operational and environmental factors considered. Furthermore, in most investigations on engine cleaning procedures in literature, the uncertainties originating from the inputs or models are not included in the analysis. This thesis project, with its focus on analyzing the uncertainty, can provide significant added value in this regard.

Engine Performance Deterioration Modeling

The considered literature on engine performance degradation modeling has been divided in two groups: physics-based models and data-driven models.

Physics-Based Modeling Approaches and Tools

A physics-based model is defined by Alozie et al. [9] as a performance model that describes the behavior of the gas turbine components based on the thermodynamics of the working fluid to provide information about the configuration and operation of the real engine.. Lakshimarasimha et al. present a model to simulate the effect of fouling and erosion [73]. They developed a procedure to simulate how these two forms of deterioration affect the engine performance parameters. Giesecke and Igie [42] employ the tool TURBOMATCH developed at Cranfield University to simulate the jet engine performance. Using this software, one can specific an engine model with a particular set of components, as well as the engine inlet conditions and the design point. To simulate the degraded performance, a reduction in isentropic compressor efficiency as well as the non-dimensional mass flow is specified. Kurz and Brun [69] describe the engine behavior using governing equations, as well as a set of deviation factors using which the degree of deterioration could be studied. The most important inputs to the model are the change in compressor efficiency and reduction in airflow, while the output consists of the reduction in power and overall efficiency.

Most of these models are not publicly available and some require a high level of understanding of gas turbine modeling. The commercial software GasTurb is available for this thesis and could be used to model performance deterioration. The main disadvantage is that due to its complexity, a large time investment would be required to make effective use of this simulation tool.

Data-driven Modeling Approaches

Data-driven modeling approaches rely on statistical methods and machine learning models to understand patterns in performance deterioration data [9]. This means these approaches do not require insight into the underlying physical behavior of the system, which can be helpful in case building models to simulate physical characteristics is considered too complicated [165]. On the other hand, in order to make data-driven approaches successful, historic data is needed that is representative for the pursued application.

Especially in the commercial aviation industry however, there is a lack of run-to-failure data sets [127]. Saxena et al. [127] noticed that the lack of public data sets hinders progress in engine prognostics, motivating them to construct a data set to be used for a Prognostics and Health Management (PHM) data challenge, in which the goal is the predict the Remaining Useful Life (RUL) of the aircraft engine provided only with historical data.

To generate data to be used in the PHM data challange the authors used the Commercial Modular Aero-Propulsion System Simulation software (C-MAPPS), which can be used to simulate a large commercial turbofan engine [74, 127]. Recently, a new dataset [21] has been made available based on the same C-MAPSS tool. This new dataset features real flight data from take-off till landing, which the flight conditions are retrieved from. Accelerated aging was used however instead of a realistic scenario for a real engine. This means that no meaningful conclusions can be drawn from the changes in output parameters with increasing flight cycles.

1.2.2. Uncertainty Analysis

The need for a framework for systematic uncertainty assessment in complex environments is made clear in the conceptual research design. To gather the knowledge required to answer the research questions, a literature study was conducted on the topic of uncertainty analysis and sensitivity analysis. The literature is presented in a step wise manner, to a certain extent building on the basic steps for uncertainty assessment outlined by Booker and Ross [16].

Identification

In this study, uncertainty is defined as: "any deviation from the unachievable ideal of completely deterministic knowledge of the relevant system", adopted from Walker et al. [151]. Uncertainty therefore not only occurs when there is a lack of knowledge. In fact, gathering more knowledge could either reduce or increase the uncertainty.

Classification

Various ways of uncertainty characterization (or classification or categorization) have been proposed in literature [114, 115, 135, 147, 151]. Walker et al. categorized uncertainties in a holistic way in the context of model-based decision support [151]. The authors categorized based on three dimensions, the nature; the level; the location. Kwakkel et al. reviewed the literature that extends on the uncertainty matrix constructed by Walker et al [72] and presented an updated matrix. Its streamlined topology description as well as its general applicability in model-based decision support led to its (partial) adoption in this project.

Nature of uncertainty

The most common way of classifying the uncertainty type is by using two categories: aleatory and epistemic and uncertainty [64]. This classification scheme seems to be widely accepted by the engineering community, as well as for model-based decision support [3, 72, 145, 147, 151, 152, 161]. The following definition of aleatory and epistemic uncertainty is considered.

Aleatory: uncertainty caused by natural randomness of a process or system. Aleatory uncertainty can not be reduced but, generally speaking, can be quantified [52, 118, 161]. An example of aleatory uncertainty is the occurrence of turbulence at high altitudes.

Epistemic: also called reducible, type B or cognitive uncertainty, is caused by the imperfection of the existing knowledge. This means that, contrary to aleatory uncertainty, epistemic uncertainty can be reduced by increasing the level of knowledge [52, 151, 161]. An example of epistemic uncertainty is the effect of the take-off derate on the EGT margin erosion, since the uncertainty can be reduced by acquiring more knowledge.

Level of uncertainty A short description and application-specific (in the context of this thesis) interpretation of each level is provided below, based on Kwakkel et al. [72].

Level 1 (shallow uncertainty): an uncertainty is considered shallow if alternative model structures can be enumerated and probabilities to these different model structures can be assigned. This can be seen as a situation in which sufficient knowledge/data is available to model the uncertainty using probability theory.

Level 2 (medium uncertainty): for this level of uncertainty, alternatives can be ranked and a likelihood can be assigned to each scenario. Consider following fictitious example. Engine performance deterioration due to a particular type of FOD is either unlikely, likely or very likely. The three scenarios can be ranked based on likelihood of occurring but how much more likely one is than the other is not specified. In this case, probabilities should not be the primary choice for modeling uncertainty.

Level 3 (deep uncertainty): alternatives can be listed but cannot be ranked. Going back to the previous example, those three scenarios could be listed but it can not be specified which is more likely. Hallegatte et al. render the term more operational for policy analysis [7]: "Deep uncertainty is a situation in which analysts do not know or cannot agree on (1) models that relate key forces that shape the future, (2) probability distributions of key variables and parameters in these models, and/or (3) the value of alternative outcomes" [44].

Level 4 (recognized ignorance): concerns uncertainty for which no (model) alternatives can be listed.

In the context of this thesis, assigning the level of uncertainty in the uncertainty classification step is essential. Given the multi-level uncertainties in the simulation environment, defining the uncertainty level in a systematic way will benefit the selection of suitable uncertainty quantification techniques in further stages of the uncertainty assessment.

Location of uncertainty

The location (often called source) of uncertainty refers to where the uncertainty originates from within the entire system [151]. Similar to the categorization of uncertainty types, many different ways to group uncertainty locations can be found in literature. Du and Chen categorize uncertainty locations into two groups: external and internal uncertainty [33]. External uncertainty is the input data uncertainty (or input uncertainty), and internal uncertainty can be broken down into model parameter uncertainty and model structure uncertainty. Model parameter uncertainty is caused by limited information when fitting model parameters for an assumed model form. Uncertainty in the model structure is due to the validity of the chosen model and its underlying assumptions [33]. The matrix for systematically classifying identified uncertainty is presented in Table 1.1.

Quantification

With the uncertainties classified according to the framework outlined in subsection 1.2.2, an appropriate representation of each uncertainty using one of many available techniques needs to be found. Schwabe et al. [128] state that "a suitable metric is defined as being one which avoids the need for data normalization in order to achieve statistically significant results". For instance, if full statistical information is not given, a more conservative theory (e.g. possibility theory) might be a more appropriate choice than a more restrictive one (e.g. probability theory) [16].

Various general theories for representing uncertainty exist. Since not every theory is equally suitable

Table 1.1: Matrix for uncertainty classification (partly based on [72]).

	Level				Nature	
Location	Level 1: shallow uncertainty	Level 2: medium uncertainty	Level 3: deep uncertainty	Level 4: recognised ignorance	Aleatory	Epistemic
Input data						
Model parameters						
Model structure						
Implementation						

for every situation, this section aims to provide some guidelines for uncertainty quantification (also called uncertainty modeling). To this end, an overview of appropriate methods is provided in Table 1.2. Based on literature, frequently addressed advantages and limitations relevant to this research problem are provided, as well as a set of applicative studies found in literature.

Propagation

In previous steps, uncertainty has been identified, classified and techniques to represent uncertainty have been considered. The next step is to combine quantified uncertainties and propagate them for instance through the black-box model. Many techniques for uncertainty propagation are available. These can be broadly subdivided in two groups: intrusive and non-intrusive methods [161].

The first practically used methods for uncertainty propagation are Monte Carlo Simulation methods [85]. These are non-intrusive, meaning that no changes to the original simulation model (which handles the deterministic propagation) are required [161]. Intrusive methods do require modification of the simulation model. That is, the simulation model is reformulated to include uncertainty directly in the system [161]. A commonly used example is Polynomial Chaos Expansions, which uses differential equations to analyze the uncertainty in a system [118].

Sensitivity Analysis

Sensitivity Analysis (SA) is concerned with studying the effect of the uncertainty in model inputs to on the model output uncertainty [48]. Sensitivity Analysis is closely related to uncertainty analysis, which instead focuses on quantifying the uncertainty in the model output [124]. Hence, it is advised to perform uncertainty analysis and Sensitivity Analysis in tandem, in this order [124]. SA methods are therefore commonly classified into either Local Sensitivity Analyses or Global Sensitivity Analyses, depending on the scope of the model input variation.

One group of models can be identified that relies on linear models to explain the change in the model output given the input values [57]. The coefficients of the linear model that fits the original model best then explain the global effect of each variable [108]. Another group of methods are variance-based. Analysis of Variance for instance is a popular method in which a function is decomposed into different components and the effect and variance of these contributing components are computed [57]. Fourier Amplitude Sensitivity Sampling (FAST) is a method that decomposes a function into a Fourier basis instead of the linear decomposition which was the case for Sobol [108]. In a study by Tarantola et al. [143], FAST has been combined with Random Balance Design. Selecting an appropriate method for a particular application involves performing trade-off between accuracy and computational cost [118]. Regarding implementation in this thesis, a Python library called SAlib [51] is available that features quantitative importance measures such as Sobol' and FAST.

1.2.3. Research Gap

Engine Performance Deterioration

The literature review revealed that techno-economic engine wash assessments that take into account operational and environmental factors affecting the Object of Interest (OoI) are scarce, and uncertainties are rarely considered. This is a problem because: a) these factors affect the engine performance deterioration, being a key factor to the economic value of engine washes; b) uncertain parameters are generally treated in a deterministic way. To close this gap, a new approach needs to be developed integratable with the factors considered and the framework LYFE. Physics-based models are often either not publicly available or too complex for the scope of thesis. A simpler model that relies on relations between parameters of interest found in literature may be appropriate.

Table 1.2: Overview of theories and methods for uncertainty modeling.

	Advantages	Disadvantages/Limitations	Applications
Probability Theory	Freq.: Relatively straightforward to implement [118] Freq.: Most decision makers and analysts are familiar with it [16, 118] Freq.: Preferred choice when full statistical information is available [23] Bayes.: Suitable for aleatory and epistemic uncertainty [11, 161]	Freq.: Produces least conservative results [11, 118] Freq.: May give false appearance of exactness [49] (caused by previous point) Freq.: Questionable for modeling epistemic uncertainty [6] Bayes: relies on 'Principle of Insufficient Reason' when constructing Bayesian belief (explained in Ref. [13])	Freq.: [22, 106, 136] Bayes.: [68, 140, 153, 154]
Possibility Theory	Higher applicability to rare events due to less restrictive axioms (compared to probability theory [16]) Can be used with limited information about the uncertainty [167], more conservative results [23] Suitable for aleatory and epistemic uncertainty [161]	Less understood by decision makers [16] Lacks operational definition [25]	[19, 45, 91, 164]
Evidence Theory	Suitable when limited information [6, 61], can deal with well characterized uncertainty as well as ... near-total ignorance [97] Useful when conflicting evidence is present. If not the case, possibility theory is more appropriate. [91, 118] No assumptions required from analyst [118] With increasing information, results approach results obtained through probability theory [61, 92, 118] (advantageous for industry [161]) Can be used to represent model form uncertainty [101]	Evidence rule potentially unreliable for highly inconsistent data [161] Tough to make decision in case of wide bounds [118, 140] Worse for decision making than probability theory in long run (but can be combined with Bayesian Theory) [140] Propagation computationally expensive compared to probability theory [61]	[6, 92, 96]
Interval Analysis	Straightforward communication to stakeholders due to simplicity [49]	Effective propagation might be challenging [38], despite simple uncertainty representation Computational cost prohibitive when number of inputs is large and output range can be incorrect if function contains local extrema [91]	[82, 110, 111]

Uncertainty Analysis

A framework for uncertainty classification has been presented, along with an overview of uncertainty quantification, uncertainty propagation and sensitivity analysis techniques. Several potential challenges have been identified. It is understood that most challenges originate from the uncertainty complexity, i.e. the presence of uncertainties of different type, level and source complicates the uncertainty analysis in various ways. These include findings ways for combining epistemic and aleatory uncertainty and dealing with the associated computational cost; finding a basis for combining values obtained through different uncertainty quantification techniques. Other difficulties include dealing with dependencies between input parameters in global sensitivity analysis. These issues seem to be underrepresented in available literature in the context of uncertainty analyses and therefore form a research gap to be addressed by this thesis. Once these issues have been understood and addressed, one can provide value to practitioners by developing a framework for performing systematic and efficient uncertainty analysis in interdisciplinary environments.

1.3. Research Question, Aim/Objectives and Sub-goals

This section consists out of two main parts: the research questions and the research objective.

1.3.1. Research Question(s)

It was decided to go with one overarching research question, which is then broken down into several sub-questions. Each of them define (a part of) the knowledge that can be used or needs to be gathered in order to achieve the research objective.

Which methodology needs to be developed to perform a systematic and efficient assessment of the uncertainties present in the process of quantifying the economic value of engine wash procedures under different types of uncertainty?

Sub-models

1.a. Which factors are affecting the economic value of engine wash procedures and what are appropriate models to simulate their impact?

1.b. What model can be used to simulate the effect of engine wash procedures on engine performance deterioration?

Uncertainty Analysis

2.a. Which theories are appropriate to model the parameters defined in the identification phase and which uncertainty propagation method is most suitable?

2.b. Which method for global sensitivity analysis can be used to apportion the combined uncertainty to the input parameters?

These research questions provide a basis for conducting a literature review. Regarding the development of sub-models, simulating engine performance deterioration under the influence of environmental and operational parameters identified by answering question 1.a deserves the main focus. Literature is looked at from a general perspective, yet special attention goes to modeling techniques which output parameter that can be operationalized (e.g. a change in EGT / SFC / efficiency) and integrated with the lifecycle simulation tool.

1.3.2. Research Objective

"The research objective is to perform a systematic uncertainty assessment in an interdisciplinary simulation environment by analyzing available methods for identifying, differentiating and propagating through this environment, and demonstrating and verifying the method on the use-case of engine cleaning procedures."

The research objective can be broken down into two main parts, namely the external goal and the internal goal. These parts are clarified below.

External Goal

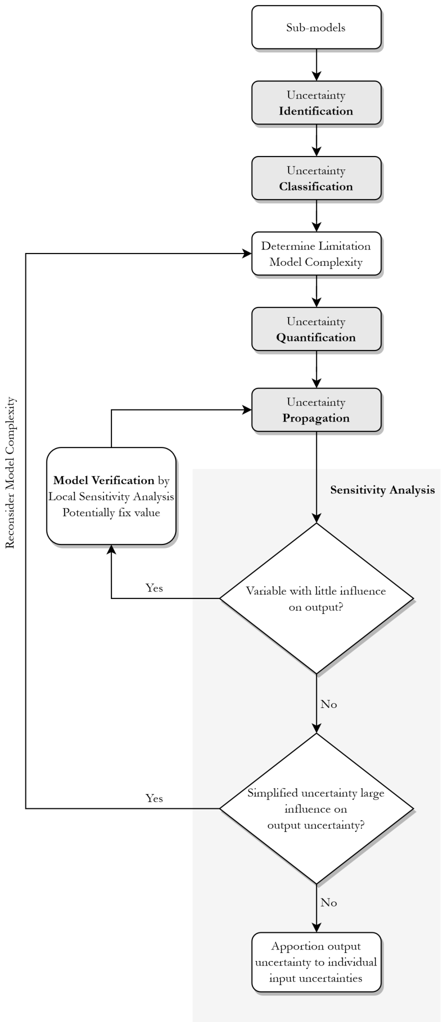
Systematic uncertainty assessment: a framework is required that enables systematic consideration of the relevant uncertainties.

Interdisciplinary simulation environment: relevant uncertainties originate, due to the interdisciplinary use-case of engine wash, from different fields. The simulation environment called LYFE encapsulates numerous inputs that affect the cash flow during the aircraft lifecycle such as SFC, maintenance costs, flight schedule, etc.

Internal Goal

Analyzing available methods for identifying, differentiating and propagating through this environment: the systematic framework for uncertainty assessment on a high level will at least consist of these activities.
Demonstrating and verifying the method on the use-case of engine cleaning procedures: the use-case involves the quantification of the economic value of engine cleaning procedures. This use-case was chosen with the sole purpose of demonstrating the methodology for uncertainty assessment. The main objective therefore remains the systematic integration of uncertainties in a complex environment. The results of the economic evaluation, and thus also the accuracy of models developed for this purpose, are not of primary interest.

1.4. Theoretical Content/Methodology



The literature view treated various methods to model engine performance deterioration. The flowchart in Figure 1.1 depicts the process of selecting a modeling technique to simulate the parameter impact. Such technique can range from a simple linear relationship found in literature between two parameters, to an advanced software tool relying on a mathematical model to replicate a thermodynamic process. To bring structure to this selection process, the flowchart outlines steps to be taken. Relationships between parameters can be grouped into sub-models. Several hard requirements apply to the the models. First, the model should be compatible and integratable with the simulation environment. That is, output parameters from the developed model should be fed into the in-house cost benefits tool. Furthermore, the model should be feasible to be developed within the thesis time frame.

Similar to the sub-model development, a flowchart has been made of the methodology for uncertainty assessment to aid in conducting efficient research. Section 1.4 shows the process which builds on the methodology for sub-model development shown in Figure 1.1. Ones uncertainties have been identified, relevant uncertainties can be categorized, quantified and propagated using the techniques described in the literature review. The large rectangle in section 1.4 represents activities related to sensitivity analyses. The main purposes of applying sensitivity analysis in this study are model verification, screening out variables that have a small influence in order to reduce the uncertainty problem scale [161], and apportioning the output uncertainty to individual uncertainties.

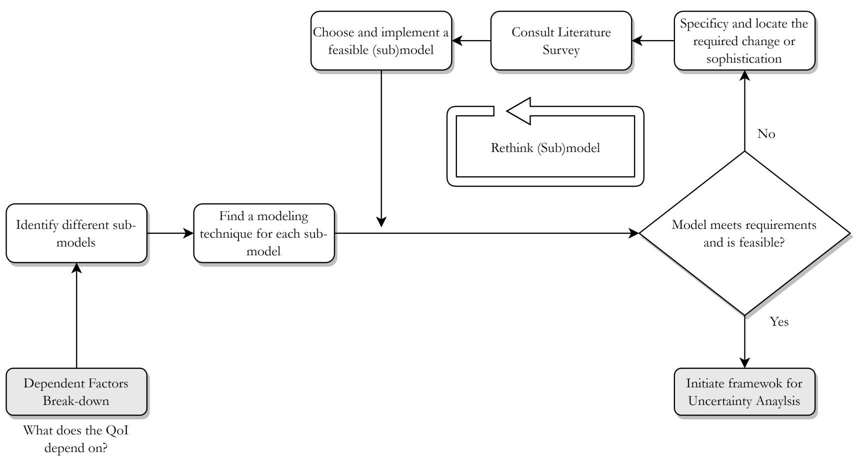


Figure 1.1: Flowchart depicting sub-models development process.

1.5. Experimental Set-up

The experimental set-up for this research project revolves around the discrete event simulation tool LYFE, which is a software tool developed to provide a generic environment for performing cost-benefit analyses of aeronautical technologies [106]. In contrast to equation based costing tools, LYFE can therefore analyze temporal effects, e.g. delays due to unforeseen circumstances. For this study, a custom module will be written to integrate the developed engine performance deterioration model in LYFE. To perform an economic assessment of an aircraft (technology), two simulations are performed: the reference case and the study case. The aircraft with desirable properties for this research project will be chosen as the reference aircraft. The study aircraft will be essentially the same aircraft, apart from the added engine cleaning maintenance task.

LYFE is written primarily in Python hence the custom module to be added during development phase (to perform uncertainty analysis, incorporate the model for engine performance deterioration and postprocess the output for Global Sensitivity Analysis) will be written in this language too. Several (well-documented) libraries will be used, such as SALIb for performing GSA. Depending on the number of required simulations, computational expense might be a limitation requiring special attention to algorithmic efficiency of the code. Besides this, the engine performance modeling tool GasTurb can be used to simulate engine deterioration as a function of, for instance, several operating and environmental conditions. This software is made available by the institute.

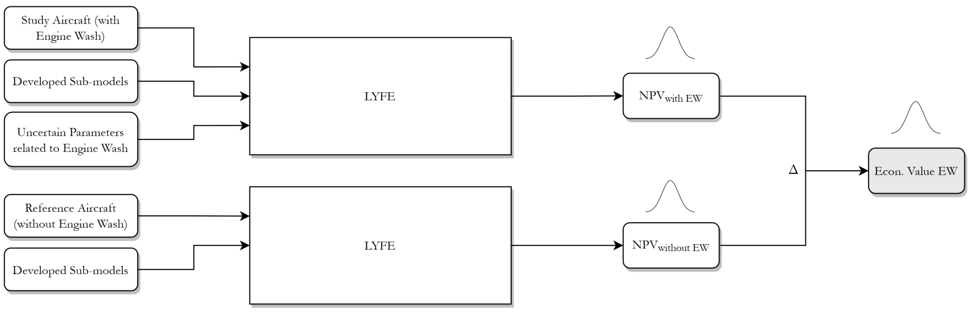


Figure 1.2: Visualization of the inputs and outputs to the cost-benefit tool LYFE for this use-case.

1.6. Results, Outcome and Relevance

For modeling engine performance deterioration, several data sets described in the literature review can be investigated as part of a data-driven modeling approach. In the case of physics-based modeling, an engine performance model is to be supplied with an engine design point and the deterioration of the engine performance can be modeled through modulation of for instance component efficiencies. The desired results show the behavior of engine performance deterioration under various environmental and operating conditions. This enhanced performance degradation model, as well as a model for the effectiveness of engine wash is to be implemented in the cost-benefit analysis framework LYFE. By treating uncertain parameters in a non-deterministic way and sampling from their distributions, a probabilistic distribution for the output is obtained. A schematic of the inputs that go into LYFE and the output from the model is shown in Figure 1.2. The overall results can then conveniently be presented as ΔNPV (Net Present Value), representing economic values. The real value comes from the systematic uncertainty analysis that will be performed on the uncertain parameters. The application on engine performance modeling and the integration in LYFE serves as an ideal use-case for demonstrating the to be developed methodology for systematic uncertainty analysis due to the large number of multi-level and multi-source uncertainties.

Verification of parameters can be performed by applying local sensitivity analysis. The developed model can be verified by interchanging sub-models with a less/more sophisticated version (i.e. using a different model, we change the uncertainty in the system by altering the bounds of a variable) and observe the influence on the output uncertainty. It should be noted that this research will be conducted without any industrial partner. Currently, no experimental data is available that could be used to validate the sub-models. However, since this research project is aimed at developing an integrative uncertainty analysis framework that can be used in aerospace economic assessments, the lack of experimental data is not an issue. There are no accuracy and precision requirements of the performance deterioration models, which renders the validation of this model beyond the scope of this research. If experimental data would become available during the course of the thesis, model validation could still be performed, yet it is not a primary aim of the research.

1.7. Project Planning and Gantt Chart

The work to be done during this thesis project can be divided into following work packages.

- Development of the model to simulate the degrading environment and its effect on engine deterioration. Inputs are operational and environmental factors that affect performance degradation while the output(s) holds a parameters which can be readily linked to a change in e.g. EGT or SFC, and used as an input to the in-house economic assessment framework named LYFE.
- Integration of sub-models (degrading environment) into LYFE (coding effort).
- Creation of a deterministic simulation of the economic efficiency of engine cleaning procedures as a maintenance application.
- Identification and classification of uncertainty using the categorization scheme presented in this review.
- Select quantification and propagation techniques for characterized uncertainties.
- Implement quantification and propagation of characterized uncertainties (coding effort).
- Verification of parameters and sub-models by means of local sensitivity analysis and strategic replacement of sub-models.
- Global sensitivity analysis to apportion the output uncertainty to individual input uncertainties.
- Visualization of the non-deterministic economic value of engine cleaning procedures (e.g. probability distribution).
- Critical discussion and documentation of results.

A Gantt chart has been created to visualize the planning of the major work packages, and can be found in Figure 3. For the initial phase of the thesis, the first task consists of developing a model for engine performance deterioration. Independent of this development, uncertainties can be assessed and quantified, which therefore overlaps to a certain degree in the planning.

1.8. Conclusions

This thesis research project aims to tackle challenges that arise when performing integrative uncertainty assessments in an interdisciplinary environment. Through the use-case of techno-economic engine wash assessments, aleatory and epistemic uncertainties of different levels are introduced and aim to be propagated through the in-house simulation environment. Expected difficulties were identified, both regarding modeling and combining the uncertainties as well as modeling the engine performance deterioration, which plays an important role in the assessment. A framework for developing the required models has been outlined which forms the basis for this research project. This framework for uncertainty analysis will be applied to the research problem, facilitating the development of a method for conducting uncertainty assessments with multi-source and multi-level uncertainties.

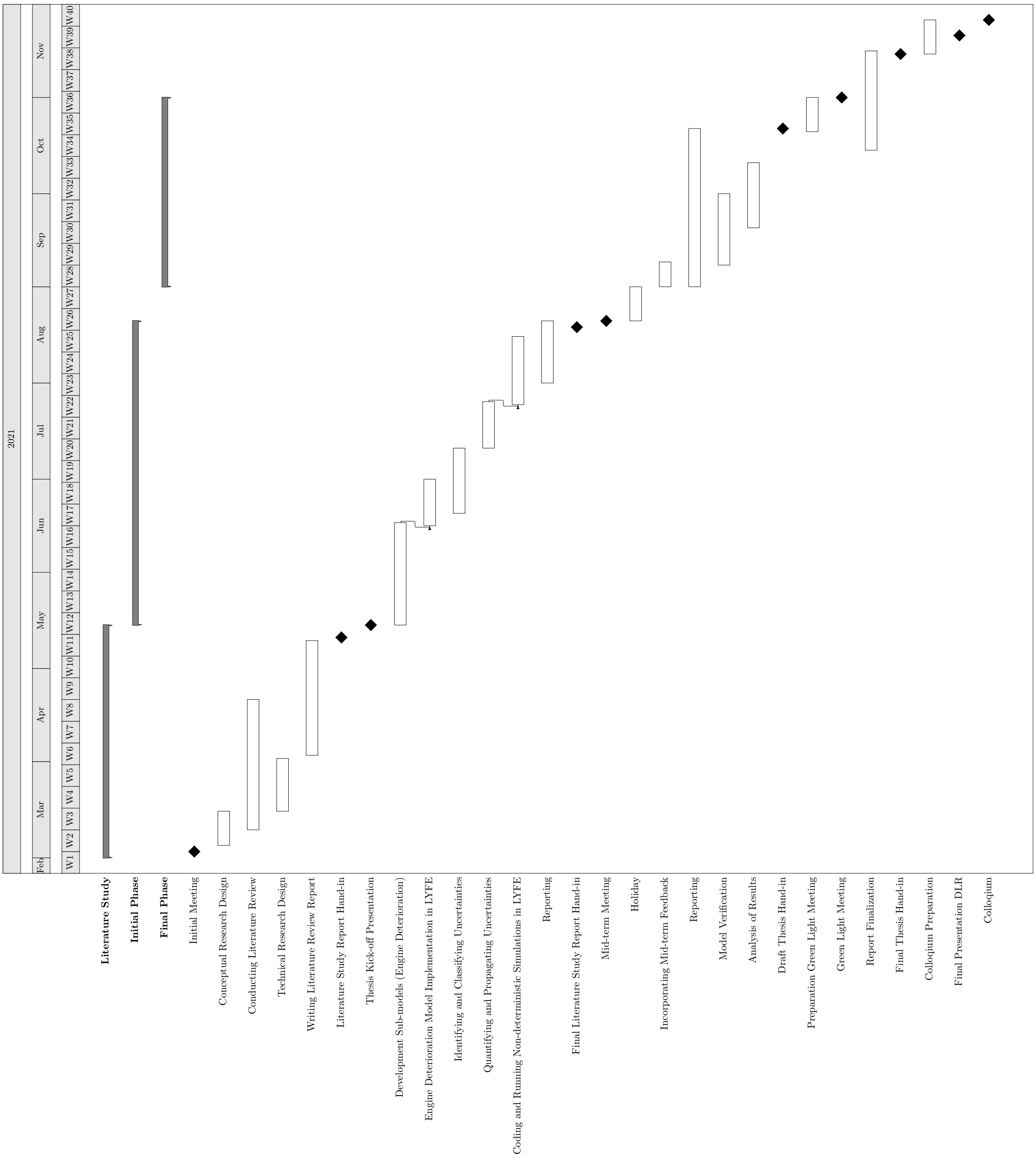


Figure 3: Thesis Project Gantt Chart

IV

Supporting work

A

Appendix 1: Applied GSA comparison study on mechanistic surrogate of LYFE

Global Sensitivity Analysis (GSA) is an active field of research with a large potential for generating better understanding of simulation models and the importance of their input parameters. Furthermore, GSA provides several clear advantages over Local Sensitivity Analysis (LSA), including its validity for non-linear models and the ability to identify interaction effects and correlations effects. Crucially, GSA methods allow effective exploration of the parameter space, even in large dimensions. Despite these clear advantages, LSA is seen more often in practice. One reason for this may be that the techniques rely on different assumptions for the computation of sensitivity measures, which can lead to differences in the interpretation of these measures. Moreover, the fact that a large number of techniques are available with no instructions on when to use which, results in a lack of clarity and thus makes it difficult to choose a suitable GSA technique for a particular problem.

This side study aims to make the application of GSA more accessible, by performing GSA using a variety of methods on a range of different model scenarios and by analysing and discussing the results. This applied study is not intended to provide a comprehensive summary of GSA techniques, nor does it claim to rigorously investigate the causation between the mathematical definitions for each of the theories and the results observed.

In this chapter, first the adapted use-case and simulation framework used for this study are described. Next, the analysis and results are discussed.

A.1. Use-case Description

In the main paper, the Analysis of Variance (ANOVA) was performed to obtain Sobol' indices used to apportion the output uncertainty to the individual inputs. A comprehensive comparison study of GSA techniques on the Lifecycle Cashflow Environment LYFE was deemed computationally intractable due to the large number of simulations required to form a conclusive argument based on statistically significant results. FastLYFE on the other hand, a mechanistic surrogate of LYFE, is considered a viable test-bed for such comparative side study.

As the mechanistic surrogate is not a discrete-event simulation (DES), the model presents a simplified version of the maintenance and operational processes available in LYFE. Because of this, implementing the same use-case presented in the paper is not feasible. Therefore, a hypothetical use-case is developed, with similar characteristics with regard to the uncertainty. The use-case involves an improved engine which has a lower Specific Fuel Consumption (SFC), but also a higher acquisition price. Both the SFC reduction and the extra economic cost associated with the more advanced engine are treated non-deterministically. Other uncertainties taken into account are the load factor, fuel price and flight distance. The output of each simulation is measured in terms of the Net Present Value (NPV). The output used for the sensitivity analysis (SA) is the ΔNPV, obtained by taking the difference between the NPV for the simulation with the improved engine and the one without.

A.2. Analysis and Results

Three individual analyses are performed. First, four GSA methods (variance-based as well as moment-independent distribution-based techniques) from the SALib library [51] for Python are applied on the FastLYFE use-case for different numbers of samples[1], and the corresponding confidence intervals are discussed. Next, a selective pair of parameters is tweaked within the model. That is, minimal changes are made to the use-case implementation in FastLYFE. The goal here is to create strong interaction effects between inputs and compare the resulting first and total-order sensitivities for different GSA methods. Questions that can then be answered include: Do the considered methods agree on the first and total-order sensitivities? Can second-order sensitivities generate further insight into the model? Finally, in the last experiment correlation is introduced between two parameters using a Gaussian copula. GSA is then performed using the correlated input data and the original use-case implementation. Note that here, changes are made to the input distributions. This is in contrast to the experiment on interaction effects, where the actual use-case implementation in FastLYFE was modified. The analysis of this final experiment will look at the results of GSA methods for correlated variables as well as methods that assume independence among inputs, applied to a correlated dataset. The main goal here is to investigate the impact of correlated input data (with gradually increasing correlation coefficient) on the resulting sensitivities and compare the results for different methods. Summarising, this side study entails:

- An applied analysis of Sobol', eFAST, SCSA, RBD-FAST and DMIM on FastLYFE. The goal is to compare the confidence bounds in the sensitivities with the amount of model evaluations and based on this information analyse which methods seem to provide faster convergence rates.
- A brief look at interaction effects in GSA using Sobol', eFAST and SCSA. The use-case implementation in FastLYFE is slightly changed to generate large interactions. Results and potential benefits of GSA methods that provide total and/or second-order indices are discussed.
- An analysis of the results for a selected set of GSA methods (that do assume independence among inputs: RBD-FAST, and do not assume independence among inputs: SCSA and DMIM) applied on correlated input data. The emphasis here lies on the comparison of the resulting sensitivities between methods and for varying correlation strengths.

A.2.1. Base Version

The results of four variance-based (Sobol', eFAST, SCSA, RBD-FAST) and one moment-independent distribution-based (DMIM) GSA techniques are shown in Figure A.1, along with their 95% confidence interval. The analysis is performed for four different number of samples per uncertainty, namely 256, 512, 2048 and 8192. Note that any amount of samples is a power of two, to preserve the balance properties of the Sobol' sequence which is used in the Sobol' method (not doing so may reduce the accuracy and rate of convergence) [99]. To enable easy comparison, the number of required model evaluations for each number of samples per variable is shown in Table A.1. Note that for the sake of clarity it is ignored here that, in fact, two simulations are required to be able to compute the ΔNPV for a certain sample set.

Table A.1: Number of model evaluation required for each considered GSA technique applied to a problem with five uncertain variables, for an increasing number of samples per variable.

Method	N = 256	N = 512	N = 4048	N = 8192
Sobol'	1792	3584	14,336	57,344
eFAST	1280	2560	10,240	40,960
SCSA	256	512	2048	8192
RBD-FAST	256	512	2048	8192
DMIM	256	512	2048	8192

For a low number of samples per variable (N = 256), extended Fourier Amplitude Sensitivity Sampling (eFAST) appears to be unstable (looking at the sensitivities for the SFC decrease and the fuel price) and in this case yields invalid first-order sensitivity indices (summing up to more than one [43]). Despite the narrow confidence interval for these values, the results for double the number of samples seem to have stabilised at a value that does seem consistent with other variance-based methods. While

[1]Please note that a brief introduction to the GSA techniques used in this study is provided as part of the literature review in the scientific paper.

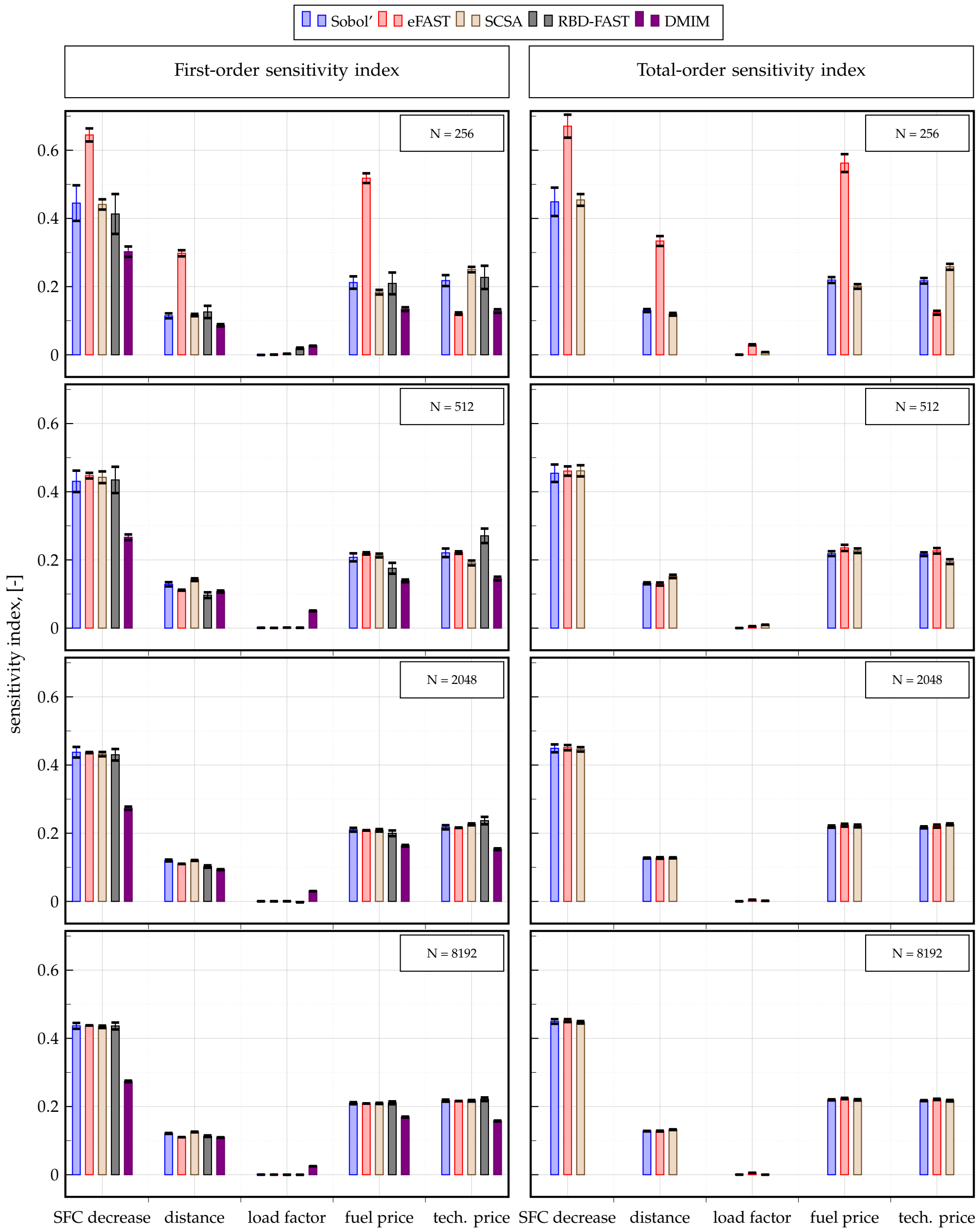


Figure A.1: Results for different GSA techniques applied to the mechanistic surrogate of LYFE along with their 95% confidence intervals. Note that RBD-FAST and DMIM do not provide total-order sensitivity indices.

the exact reason behind this behavior might need to be further investigated (repeated experiments yielded similar results), the practitioner may benefit from exercising caution when using eFAST for such a low number of samples and either increase the amount of samples or perform a comparison with other GSA methods as a benchmark to confirm or refute the observed results.

As a more general remark, the number of model evaluations performed seems not always in proportion with the width of the confidence bounds, across different methods. It is noticed that in particular the Sobol' method leads to results with comparably wide confidence intervals for a small sample size N (256 and 512), especially considering the large amount of required model evaluations (compared to for instance Structural and Correlative Sensitivity Analysis (SCSA)) as shown in Table A.1. SCSA performs well in this test, given that for 1/7 of the amount of required model evaluations for the Sobol' method, results seem precise (narrow confidence interval, even for small number of samples) and accurate. Furthermore, SCSA provides estimates for second-order sensitivity indices (which will be treated later in this study). It should be noted, however, that while bootstrap confidence intervals can provide a useful measure of uncertainty, they are "neither exact nor optimal" [36]. Moreover, as the resampling procedure used in SALib may differ for different methods, a more thorough study of the bootstrapping procedures used would be useful. For larger sample sizes (2048 and 8192), the results of the variance-based methods become more consistent and the confidence intervals more similar. In many real-life scenarios with complex models, however, running that many simulations might be too computationally expensive. In those cases it might be worthwhile to explore alternatives that allow for faster convergence. Based on this brief analysis, eFAST and SCSA have that ability.

Another interesting observation concerns the results for Delta Moment Independent Measure (DMIM), the only moment-independent method in this list. Moment-independence in this case means that for the calculation of the sensitivity index, the entire output distribution is considered, as opposed to merely the first two moments (i.e. mean and variance) as is the case for variance-based methods. As DMIM can be readily compared to Sobol' indices [17] and under the assumption that no dependency exists among the input variables, it can be concluded that factors that have a large effect on the variance do not necessarily have the same effect on the entire distribution. This is illustrated by the sensitivity of the load factor, which shows that DMIM seems to register a significant contribution to the output distribution, while Sobol' indices remain practically zero.

A.2.2. Interaction Effects

Interaction effects between a set of inputs refer to the "presence of non-additivity of the effects of individual inputs on the system output" [112] and can be interpreted as the way in which these inputs collectively (in the model) influence the output. This should not be confused with correlation, which refers to a statistical dependence between input variables [113]. Interaction effects are briefly introduced using a couple of simple examples with three (independent) variables x_1, x_2, x_3 that follow a uniform distribution over $[-\pi, \pi]$.

Starting with Equation A.1, we expect to see equal first-order indices S_i since the random variables have the same bounds and x_1, x_2 and x_3 have the same constant in the equation. The results in Table A.2 confirm this. Also, since the first-order indices are equal to the total indices, no interactions are observed.

$$f(\mathbf{x}) = x_1 + x_2 + x_3 \tag{A.1}$$

Table A.2: Sobol' sensitivity indices (first, second, and total-order) for an introductory example without interaction.

	S_i	S_{T_i}
x_1	0.333	0.333
x_2	0.333	0.333
x_3	0.333	0.333

	S_{ij}
(x_1, x_2)	0.000
(x_1, x_3)	0.000
(x_2, x_3)	0.000

In Table A.3, associated with Equation A.2, we observe total-order indices different from first-order indices for x_1 and x_2, indicating the presence of interaction effects. The interaction effect is caused by the multiplication between x_1 and x_2. That is, the effect of increasing x_1 on the system output $f(\mathbf{x})$ not only varies with constants associated with x_1, but also with the magnitude of x_2. This is in contrast to the previous example, where the effect of increasing x_1 was not affected by the value of x_2.

$$f(\mathbf{x}) = x_1 + x_1 \cdot x_2 + x_3 \tag{A.2}$$

Table A.3: Sobol' sensitivity indices (first, second, and total-order) for an introductory example with modeled interaction.

	S_i	S_{T_i}
x_1	0.189	0.811
x_2	0.000	0.621
x_3	0.189	0.189

	S_{ij}
(x_1, x_2)	0.622
(x_1, x_3)	0.000
(x_2, x_3)	0.000

Finally, in Equation A.3, x_1 and x_2 only affect the system output through interaction with each other. It therefore follows that their first-order indices are zero (i.e. x_1 and x_2 do not separately affect the output, but only together) and the total-order indices, S_{T_1} and S_{T_2}, equal the second-order index S_{12}, as shown in Table A.4. x_3 has no interaction effects so S_3 equals S_{T_3}.

$$f(\mathbf{x}) = x_1 \cdot x_2 + x_3 \tag{A.3}$$

Table A.4: Sobol' sensitivity indices for a second introductory example with modeled interaction.

	S_i	S_{T_i}
x_1	0.000	0.767
x_2	0.000	0.767
x_3	0.233	0.233

	S_{ij}
(x_1, x_2)	0.767
(x_1, x_3)	0.000
(x_2, x_3)	0.000

To demonstrate the effect of interactions on a more realistic use-case, the surrogate of LYFE was used. Note that due to required changes (albeit limited to the simple multiplication of the flight distance and SFC decrease) in the source code, the results are not to be compared with results from the base version. Rather, the value comes from comparing the results among different GSA methods. Only methods that provide total-order sensitivity indices (Sobol', eFAST and SCSA) are considered in the analysis. The results in Figure A.2 show that all three methods agree well on both first and total-order sensitivity indices. The total-order index for the fuel price and flight distance in particular is considerably higher (± 0.10) than the first-order index, indicating the presence of strong interaction effects. This insight into the model would have been missed by the analyst when using GSA techniques that return first-order sensitivity indices only. One can gain even more insight in the interaction effects by computing second-order sensitivity indices, which measure the fractional contribution due to the interaction between pairs of uncertain input variables. These interaction effects are listed in Table A.5. It should be noted that to obtain statistically significant results, the number of samples was increased by a factor of four. Nevertheless, confidence bounds for the Sobol' generated indices are comparably wide, especially when placed next to the SCSA method. In many real-life situations, obtaining statistically significant second-order indices may therefore become computationally prohibitive when using the Sobol' method. The results indicate that most of the higher-order effects is due to interaction between the flight distance and fuel price. It is noteworthy that the adjustment of the model was to deliberately interact the decrease in SFC due to the new technology and the flight distance (for which pair a non-zero but small interaction effect was detected). This shows that even in this relatively simple surrogate of LYFE, it becomes difficult to keep track of how variables affect each other. In more complex models, this task would quickly become impossible. Calculating higher-order sensitivity indices is a proven method of providing this insight in a systematic manner.

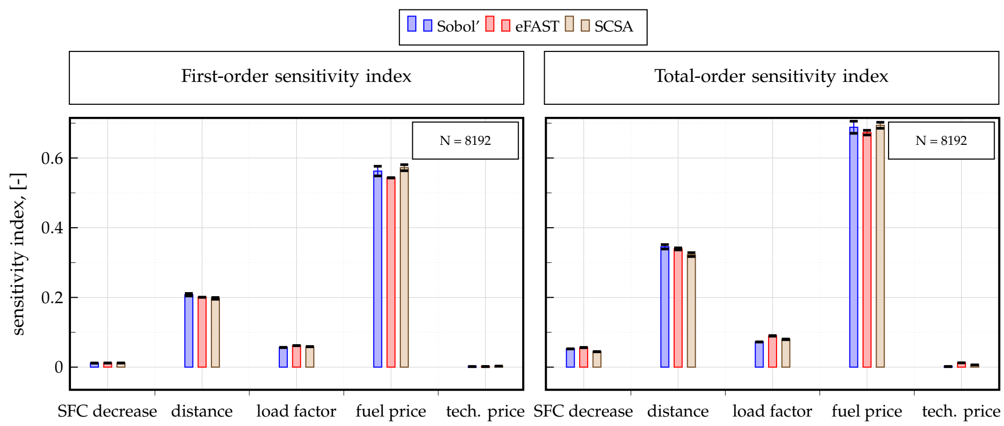


Figure A.2: Results for different GSA techniques applied to the mechanistic surrogate of LYFE with increased interaction effects. Only methods that provide total-order sensitivity indices are displayed.

Table A.5: Second-order sensitivity indices computed using the Sobol' and SCSA method.

Input variable pair	**Sobol'**	**SCSA**
(SFC decrease, distance)	0.015 ± 0.007	0.008 ± 0.001
(SFC decrease, load factor)	-0.002 ± 0.006	0.002 ± 0.000
(SFC decrease, fuel price)	0.021 ± 0.007	0.019 ± 0.001
(SFC decrease, tech. price)	0.000 ± 0.007	0.000 ± 0.000
(distance, load factor)	0.013 ± 0.017	0.013 ± 0.002
(distance, fuel price)	0.102 ± 0.020	0.102 ± 0.003
(distance, tech. price)	0.001 ± 0.015	0.000 ± 0.000
(load factor, fuel price)	0.000 ± 0.006	0.000 ± 0.000
(load factor, tech. price)	0.000 ± 0.005	0.000 ± 0.000
(fuel price, tech. price)	0.000 ± 0.016	0.000 ± 0.000

It was concluded that when relying solely on the first-order and total-order sensitivity indices, it can be difficult to understand the interaction effects between the individual parameters. Second-order effects, as provided, for example, by the Sobol' method (at the cost of more than 1.7 times the required number of model evaluations, for five variables), provide the pairwise interaction effects and thus allow the analyst to gain more insight into the model.

A.2.3. Correlation effects

To demonstrate the application of GSA methods to a use-case with a set of correlated input variables, the original sampling implementation was slightly adapted. Specifically, the fuel price and extra cost of the new technology are, in this hypothetical scenario, assumed to be correlated. The main goal of this analysis is to understand how the results for different GSA methods compare in the case of dependent inputs. Concretely, three methods are considered: DMIM, FAST-RBD and SCSA. FAST-RBD indeed assumes independence of input variables and could therefore give additional insight in the results. Also, the effect of unknowingly applying GSA methods which assume independence to problems with correlated variables can thereby be assessed. A given-data method, FAST-RBD was selected out of all variance-based methods assuming independence among the inputs, since the SALib library does not support the generation of correlated samples for sampling scheme dependent methods.

To introduce dependency between the fuel price and the technology price, both assumed to follow a uniform distribution, the Gaussian copula function was used. Copula functions can be used to "join or 'couple' multivariate distribution functions to their one-dimensional marginal distribution functions" [94]. A particular family of copulas is the Gaussian copula [95], which is given by:

$$\mathbb{C}_\Lambda(\{F_i(y_i)\}_{i=1}^{N}) = \Phi_\Lambda(\{\Phi^{-1}(F_i(y_i))\}_{i=1}^{N}) \tag{A.4}$$

where $\Phi_\Lambda(\cdot)$ is the standard multivariate Gaussian distribution with covariance matrix Λ and $\Phi^{-1}(\cdot)$ is the inverse of the standardized univariate Gaussian CDF.

In the bivariate case, this reduces to:

$$\mathbb{C}_\Lambda(\{F_i(y_i)\}_{i=1}^{N=2}) = \Phi_\Lambda(\Phi^{-1}(F_1(y_1)), \Phi^{-1}(F_2(y_2))) \tag{A.5}$$

The covariance matrix Λ used is the one of a standard bivariate Gaussian (Figure A.3, left) with correlation coefficient ρ. It should be noted that this ρ is going to differ from the ρ of the samples generated by the Gaussian copula. However, since we just want to introduce different levels of correlation (as a basis for comparison of the methods) and thus are not interested in the actual correlation value, this is deemed acceptable for the present study. Samples from the Gaussian copula are then generated which all lie between 0 and 1 (the marginals are obtained by taking the CDF of the standard Gaussian univariates and hence follow the standard uniform distribution). Finally, to obtain correlated samples of the fuel price and technology price, the percentile function associated with the respective desired marginals was applied to the samples from the Gaussian copula. The result is visualized in Figure A.3 (right).

The resulting first and total-order sensitivity indices are shown in Figure A.4 for different values of the correlation coefficient ρ. In contrast to what was the case when observing results for the interaction effects study, here the results can be compared to the base version, since only the input variables (fuel

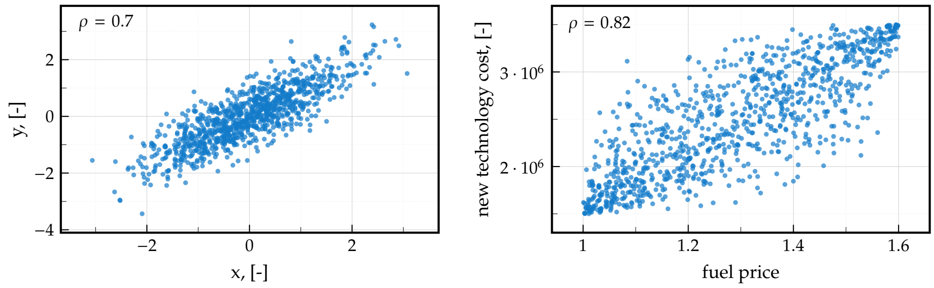


Figure A.3: The bivariate Gaussian (left, 1000 samples) is used for the Gaussion copula. Correlated samples are drawn and scaled to the desired marginal distributions (right).

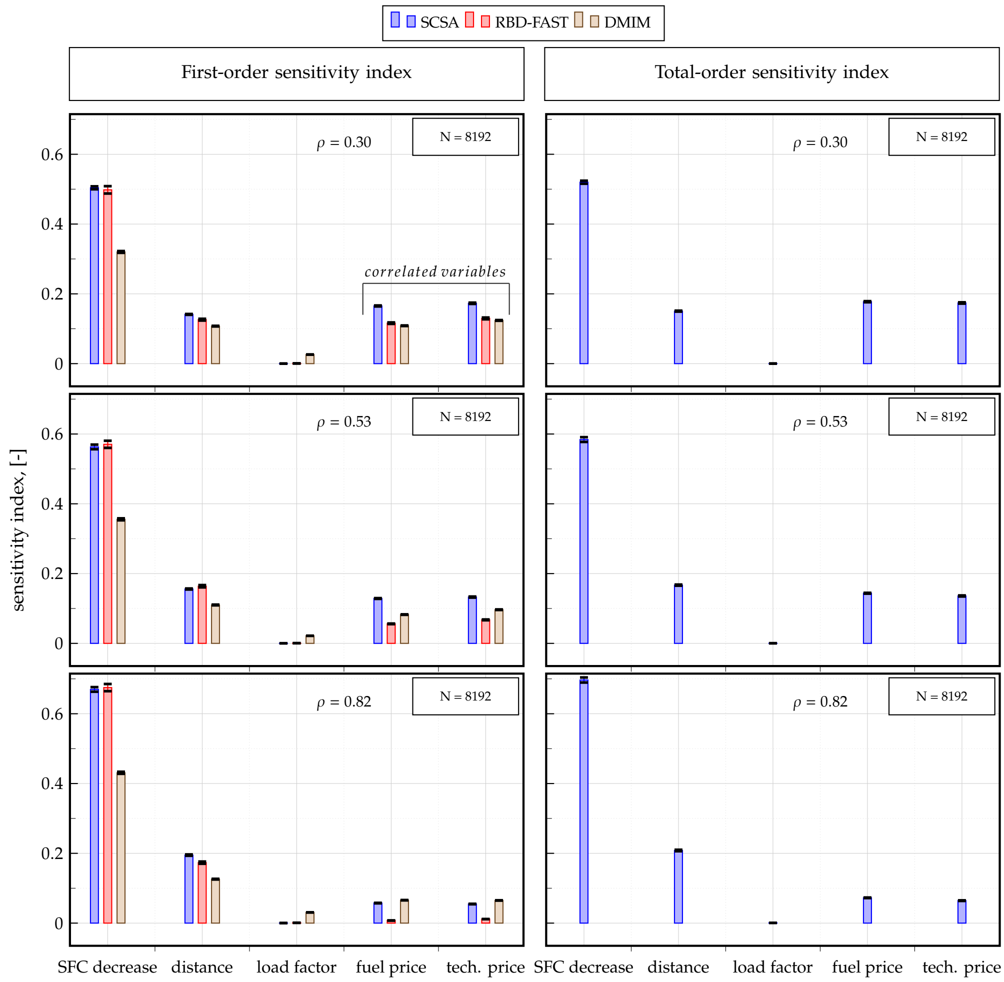


Figure A.4: Results for different GSA techniques applied to the mechanistic surrogate of LYFE with modeled correlation between the fuel price and technology price. Note that RBD-FAST and DMIM do not provide total-order sensitivity indices.

price and technology price) have been correlated and no changes to the actual model have been made. Compared to the results for the base version in Figure A.1, introducing correlation seems to lower the sensitivity of the correlated variables and increase the contribution of variables for which the samples are not correlated. This effect continues as the correlation coefficient increases.

The scenario where an analyst unknowingly applies a variance-based method that is not appropriate for dependent inputs on a dataset with correlated inputs is briefly investigated. First-order sensitivity indices for non-correlated variables obtained using any variance-based method seem to be similar. For correlated variables, however, the underestimation by Random Balance Designs Fourier Amplitude Sensitivity Test (RBD-FAST) of the first-order index S_i increases with the strength of the correlation. For $\rho = 0.82$, the S_i computed by RBD-FAST is close to zero, while the SCSA and DMIM return a first-order index between 0.05 and 0.10. With increasing correlation strength, S_i seems to reduce to its component that is due to the correlation effect, which is indeed not picked up by methods such as RBD-FAST that assume independence among input variables. While results from DMIM and SCSA for $\rho = 0.82$ are close, this is not the case for lower correlation values. However, the same behavior was observed in the uncorrelated case in Figure A.1, suggesting that the cause of this lack of agreement here is also the fact that DMIM considers the entire output distribution as opposed to just the variance. In conclusion, the RBD-FAST results for uncorrelated variables (e.g. decrease SFC, distance) are consistent with SCSA and DMIM, but the first-order indices S_i for correlated variables are continuously underestimated by methods that assume independence. The magnitude of this underestimation increases as the correlation increases. As a result, the analyst may fail to capture part of the sensitivity if he or she is unaware of the correlation in the data set and applies a method that assumes uncorrelated inputs.

B

Appendix 2: Verification

This research was conducted without any industrial partner. No experimental data is available that could be used to validate the sub-models. However, since this research project is aimed at developing an integrative uncertainty analysis framework that can be used in economic assessments, the lack of experimental data is not necessarily a problem. For example, there are no accuracy and precision requirements for the performance deterioration models, which renders the validation of this model beyond the scope of this research. Nevertheless, several parts of the research are actually validated. LYFE, without the custom functionality implemented in this thesis, is a validated product. Also the implementation of methods included in the Python library for Global Sensitivity Analysis (GSA), SALib, is tested [51].

Areas where validation is not feasible are therefore verified. First, it should be noted that GSA in itself is an effective tool for model verification[1] [125]. The GSA is indeed only performed on the base-line model, but the updated model implementation for high-impact uncertainties is to a certain extent verified by the nature of the analysis. That is, for the reworked representation of the fuel price and EGT increase uncertainty, the range of the values the variable takes is either reduced by further specifying the input (EGT increase) or extreme values for the uncertainty are now less likely to occur due to introduced randomization (fuel price). In both cases, the analysis concluded that the overall uncertainty after propagation reduced compared to the the base-line case, in line with the expectations.

The verification of the application of GSA in technology assessments is in part covered by the GSA comparison study on the mechanistic surrogate of LYFE, which indicates that GSA can be applied to technology assessments. This is concluded on the basis of the results of the analysis and comparison of different methods with very different characteristics, whereby the results are consistent for the different methods.

General sanity checks are performed in section B.1 for each of the implemented sub-models. This is done using a simple form of LSA (Local Sensitivity Analysis) and is mainly intended to detect implementation errors. In section B.2, random samples from each of the input distributions are propagated through the model, with the aim to gain further insight into the simulation framework including the implemented sub-models.

B.1. Sanity Checks

To verify the sub-model implementation part of the base-line model, a simple form of LSA is performed and the results are analysed. For each of the uncertain input variables, the left and right-most value of the distribution is taken and used for a simulation of LYFE, while keeping all other variables constant (equal to the mean of their input distribution). This approach ignores the shape of the input probability distribution and is therefore only suitable, to a certain extent, to give a general indication of the order of magnitude of importance of each uncertain variable (measured in terms of the difference in ΔNPV for the extremes of the input distribution). Nevertheless, the main objective here is to determine if a change in the input value of each parameter causes a sensible change according to our judgement (for

[1]During GSA, hundreds or thousands of simulations are run with input values sampled from the entire parameter space. Any mistakes in the implementation are likely going to result in errors during the propagation of these values.

example, asking: "a higher value of the fuel price causes a higher ΔNPV, does this intuitively make sense?"). Hence, this verification step focuses on the sub-models that were integrated into LYFE as part of the base-line model. The results are presented in Table B.1.

Table B.1: Local sensitivity analysis of the uncertain variables. The left and right bound refer to the lower and upper extremes of each probability distribution. The absolute difference between these ΔNPVs gives an indication on the output spread imposed by a particular uncertainty. Notation: NPV_w (washed, NPV for scenario with engine wash), NPV_u (unwashed, NPV for scenario without engine wash)

Variable	Left bound of distribution, all values in US$	Right bound of distribution, all values in US$	Absolute difference of ΔNPV [US$]
EGT to SFC	NPV_w: 22,928,037 NPV_u: 22,628,498 ΔNPV: 299,539	NPV_w: 22,672,232 NPV_u: 22,320,576 ΔNPV: 351,656	52,117
EW price	NPV_w: 22,820,524 NPV_u: 22,471,293 ΔNPV: 349,231	NPV_w: 22,779,797 NPV_u: 22,471,293 ΔNPV: 308,504	40,727
Fuel price	NPV_w: 45,862,922 NPV_u: 45,749,143 ΔNPV: 113,779	NPV_w: -262,602 NPV_u: -806,558 ΔNPV: 543,956	430,177
EGT increase	NPV_w: 23,008,536 NPV_u: 22,729,837 ΔNPV: 278,699	NPV_w: 22,591,826 NPV_u: 21,617,368 ΔNPV: 974,458	695,759
EW effect	NPV_w: 22,725,102 NPV_u: 22,478,048 ΔNPV: 247,054	NPV_w: 23,141,687 NPV_u: 22,471,293 ΔNPV: 670,394	423,340
EW interval	NPV_w: 22,941,088 NPV_u: 22,473,584 ΔNPV: 469,504	NPV_w: 22,725,102 NPV_u: 22,478,048 ΔNPV: 247,054	222,450

Sanity checks are then performed for each uncertain variable, based on the values from Table B.1. This approach aims to verify the sub-model implementation in LYFE.

EGT to SFC: for the right bound (high value), the NPV is lower than for the left bound (low) of the distribution. This is in line with our intuition, which tells that for the same EGT (Exhaust Gas Temperature), a higher SFC (Specific Fuel Consumption) will result in higher fuel burn and hence increased fuel costs, leading to a lower NPV and vice-versa. Its effect on the ΔNPV is comparably small, as shown by the difference between the ΔNPV for both cases.

EW price: the NPV for the highest EW price is lower than for a low EW price. This is indeed as expected. The effect on the ΔNPV seems relatively weak, compared to the other variables. This can be explained by the fact that any value in the range of possible EW prices is several orders of magnitude lower than typical values for the NPV.

Fuel price: has a very large effect on the NPV, compared to the other uncertainties, with a negative NPV for the highest fuel price scenario (right bound). The economic value of EW is five times larger for the highest fuel price than for the lowest fuel price scenario (left bound). This is in accordance with the realization that for higher fuel prices, the positive effect of EW on the fuel consumption (and hence fuel burn) reduction becomes increasingly favorable, thereby increasing the ΔNPV.

EGT increase: high values for this factor should lead to a lower NPV, simply due to the higher fuel consumption caused by a larger EGT increase reducing the NPV. This behavior is observed from the comparison between the high and low EGT increase values. Furthermore, the EW economic value (ΔNPV) varies greatly between both engine degradation scenarios. For high EGT increase values, the economic value of EW is almost four times larger than for the lowest deterioration profile. The observed behavior is in line with our expectations. As the engine deteriorates quicker, the EW becomes more effective at restoring lost EGT margin (due to the way it is implemented in this use-case), making the maintenance task more desirable. A real-world example would be an engine operating in a hot and dusty environment. The unwashed engine will suffer from fast deterioration rates, while the washed engine can benefit from washes more than

it could if the operation took place in more favorable conditions.
EW effect: a higher amount of restored EGT margin due to EW seems to increase the NPV. This is as expected, since a higher EGT margin (hence lower EGT increase) leads to a lower fuel consumption, increasing the NPV. The effect on ΔNPV seems relatively large, compared to other uncertainties. This is expected as EW is only performed for the washed engine scenario, thus any change in its effect will directly impact the ΔNPV.
EW interval: the largest EW interval bears a lower NPV than the smallest interval. Due to the greater deterioration caused by increasingly infrequent washing, the higher fuel consumption leads to higher economic costs. The lower EW costs (caused by performing less EWs) clearly do not off-set this difference, mainly due to the relatively low order of magnitude of the EW price compared to the overall NPV (where big contributors are the expensive ESV (Engine Shop Visit) required when reaching the EGT red-line and increased fuel costs due to more severe engine degradation). The effect on the ΔNPV is looked at in slightly more detail due to the particularly high ΔNPV for the left bound value. A short experiment in Figure B.1 revealed that the optimal fixed EW frequency would in fact be once every 50 days, which is lower than the minimum value used for modeling the EW interval uncertainty (which was based on literature). This deviation is due to the nature of the implementation, in which, for instance, the EW interval does not correlate with the EGT margin restoration capabilities. This is in part compensated for by making the restoration effect dependent on the interval within the simulation framework, but this method proves to be limited.

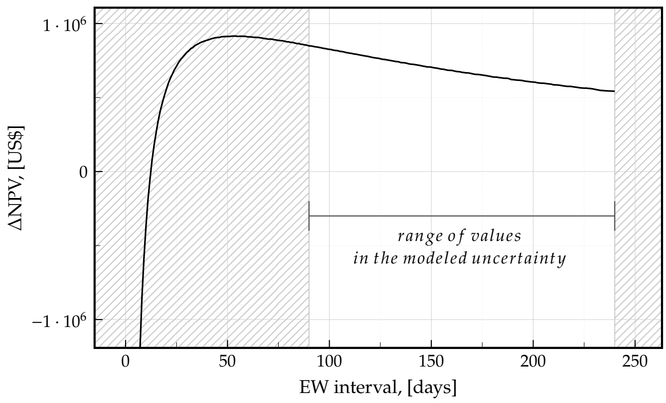


Figure B.1: ΔNPV plotted against the EW interval length.

Furthermore, ΔNPV's absolute differences on an ordinal scale are in line with the GSA results presented in the Scientific Paper. However, one should be cautious in drawing conclusions from this comparison, since the disregard of the input distribution together with the vastly different method of calculating Sobol' sensitivity indices makes comparing the actual GSA results with findings from this simple form of LSA hard.

B.2. Further Insight

To better understand the effect of the simulation framework on the input samples, separate simulations are carried out where all but one of the parameters are kept constant (the mean of their input distribution is taken as the constant). It should be noted that also this is not a verification of the sensitivity indices. Consider again the first order sensitivity index [123]:

$$S_i = \frac{V_{X_i}(E_{\mathbf{X}_{\sim i}}(Y|X_i))}{V(Y)} \tag{B.1}$$

Here, $\mathbf{X}_{\sim i}$ includes all uncertainties except X_i. The expectation is found by fixing X_i and taking the mean of the resulting Y while varying $\mathbf{X}_{\sim i}$. This is done repeatedly for different values of X_i and the variance thereof is then the first order index (not sensitivity index, which first needs to be normalized over $V(Y)$). Sobol' (first order) sensitivity indices are indeed obtained in a more involved way than if we were to simply calculate the variance of the individual output probability distributions shown in Figure B.2. This effort does however give more insight into the simulation framework and also emphasizes the need for global (instead of local) SA.

The shape of some original input distributions is severely changed after the propagation of samples through LYFE. Consider for instance the discontinuities in the output distribution after varying the EGT increase and the EW interval, for which the input distribution followed a normal and uniform distribution respectively, showing the non-linear nature of the simulation framework. This stresses the need for sensitivity measured obtained from an effective exploration of the parameter space, thus GSA, as opposed to the frequently used Local Sensitivity Analysis (LSA).

The remarkable shape of the output distribution when varying the EGT increase input calls for further analysis. When the EGT rises quickly enough, the EGT upper limit of the engine is reached before the first planned ESV, thus requiring an additional ESV. This phenomenon, caused by the simplified maintenance processes in the model, occurs less for the washed engine than for the unwashed engine (since the washed engine will not reach the EGT redline as quickly). This results in values for ΔNPV of the order of a million US$.

In conclusion, these two sections aimed to verify, to the extent possible, the sub-model implementation in the the cost-benefit analysis tool LYFE. Straightforward sanity checks were performed on the base-line model results by evaluating the QoI at extreme points for each input variable. As a next step, sampled values from the individual input distributions were propagated to obtain the distribution of ΔNPVs for each varied input. Findings were generally in line with the outcome from the GSA, and generated further insight into the simulation framework LYFE and the implemented sub-models.

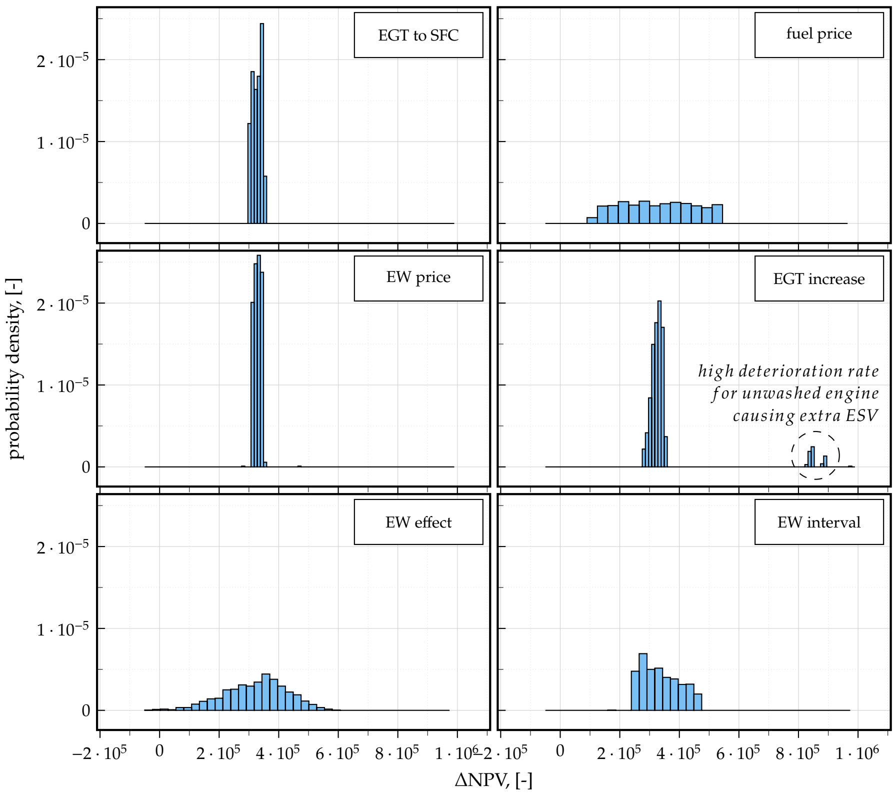


Figure B.2: Output uncertainty distribution when sequentially keeping all but one variable constant, for 1000 model evaluations.

Bibliography

[1] CFM56-5A/-5B maintenance analysis & budget. Aircraft Commerce, (50):15–27, 2007.

[2] Productsheet: Cyclean engine wash. https://www.lufthansa-technik.com/documents/5875801/6257402/Productsheet+Cyclean+Engine+Wash/f7d883a7-d271-4016-9eb8-2c55a03cf9dd, 2019. Accessed: 2021-08-20.

[3] Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. 2020. doi: 10.1016/j.inffus.2021.05.008.

[4] S. Ackert. Engine maintenance concepts for financiers - elements of turbofan shop maintenance costs. Technical Report. https://www.icao.int/Meetings/EnvironmentalWorkshops/Documents/ICAO-TransportCanada-2006/Hutter.pdf, 2011. URL http://www.aircraftmonitor.com/uploads/1/5/9/9/15993320/engine_mx_concepts_for_financiers___v2.pdf. Accessed: 2021-09-10.

[5] U.S. Energy Information Administration. Petroleum and other liquids, 2021. URL https://www.eia.gov/dnav/pet/PET_PRI_SPT_S1_D.htm. Accessed: 2021-09-30.

[6] Harish Agarwal, John E. Renaud, Evan L. Preston, and Dhanesh Padmanabhan. Uncertainty quantification using evidence theory in multidisciplinary design optimization. Reliability Engineering and System Safety, 85(1):281–294, 2004. doi: 10.1016/j.ress.2004.03.017.

[7] Buyung Agusdinata. Exploratory Modeling and Analysis: A Promosing Method to Deal with Deep Uncertainty. PhD thesis, 2008.

[8] O. Alozie and M. Diakostefanis. Assessment of degradation equivalent operating time for aircraft gas turbine engines. The Aeronautical Journal, 124(1274):549–580, 2020. doi: 10.1017/aer.2019.15.

[9] Ogechukwu Alozie, Yi-Guang Li, Xin Wu, Xingchao Shong, and Wencheng Ren. An adaptive model-based framework for prognostics of gas pathfaults in aircraft gas turbine engines. International Journal of Health Management, 10(2), 2019. doi: 10.36001/ijphm.2019.v10i2.2728.

[10] Dawn An, Nam H. Kim, and Joo-Ho Choi. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. Reliability Engineering and System Safety, 133:223–236, 2014. doi: 10.1016/j.ress.2014.09.014.

[11] Ha-Rok Bae, Ramana V. Grandhi, and Robert A. Canfield. Epistemic uncertainty quantification techniques including evidence theory for large-scale structures. Computers and Structures, 82 (13-14):1101–1112, 2004. doi: 10.1016/j.compstruc.2004.03.014.

[12] Włodzimierz Balicki, Paweł Glowacki, Stefan Szczecinski, Ryszard Chachurski, and Jerzy Szczecinski. Effect of the atmosphere on the performances of aviation turbine engines. acta mechanica et automatica, 8(2):70–73, 2014. doi: 10.2478/ama-2014-0012.

[13] Malcolm Beynon, Bruce Curry, and Peter Morgan. The dempster–shafer theory of evidence: an alternative approach to multicriteria decision modelling. Omega: The International Journal of Management Science, 28(1):37–50, 2000. doi: 10.1016/S0305-0483(99)00033-X.

[14] Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2007. ISBN 0387310738.

[15] Arnaud Bonnet. Avoiding high speed rejected takeoffs due to egt limit exceedance. https://safetyfirst.airbus.com/avoiding-high-speed-rejected-takeoffs-due-to-egt-limit-exceedance/, 2017. Accessed: 2021-10-05.

[16] J. M. Booker and T. J. Ross. An evolution of uncertainty assessment and quantification. Scientia Iranica, 18(3):669–676, 2011. doi: 10.1016/j.scient.2011.04.017.

[17] E. Borgonovo. A new uncertainty importance measure. Reliability Engineering and System Safety, 92(6):771–784, 2007. doi: 10.1016/j.ress.2006.04.015.

[18] Meherwan P. Boyce and Francisco Gonzalez. A study of on-line and off-line turbine washing to optimize theoperation of a gas turbine. Journal of Engineering for Gas Turbines and Power, 129 (1), 2007. doi: 10.1115/1.2181180.

[19] Vincent Braibant, Adrien Oudshoorn, Chris Boyer, and Franck Delcroix. Nondeterministic "possibilistic" approaches for structural analysis and optimal design. AIAA Journal, 37(10), 2012. doi: 10.2514/2.599.

[20] Andrew J. Brune, Thomas K. West, and Serhat Hosder. Uncertainty quantification of planetary entry technologies. Progress in Aerospace Sciences, 111, 2019. doi: 10.1016/j.paerosci.2019.100574.

[21] Manuel Arias Chao, Chetan S. Kulkarni, Kai Goebel, and Olga Fink. Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics. Prognostic Data Sets, 6(1), 2021. doi: 10.3929/ethz-b-000467626.

[22] Dan Chen and Jianzhong Sun. Fuel and emission reduction assessment for civil aircraft enginefleet on-wing washing. Transportation Research Part D, 65:324–331, 2018. doi: 10.1016/j.trd.2018.05.013.

[23] Sophie Chen, Efstratios Nikolaidis, and Harley H. Cudney. Comparison of probabilistic and fuzzy set methods for designing under uncertainty. In 40th Structures, Structural Dynamics, and Materials Conference and Exhibit, 1999. doi: 10.2514/6.1999-1579.

[24] Wei Chen, Ruichen Jin, and Agus Sudjianto. Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty. Journal of Mechanical Design, pages 953–962, 2005. doi: 10.1115/1.1904642.

[25] Roger Cooke. The anatomy of the squizzel: The role of operational definitions in representing uncertainty. Reliability Engineering System Safety, 85(1-3):313–319, 2004. doi: 10.1016/j.ress.2004.03.019.

[26] Henrique Luiz Correa. Linking Flexibility, Uncertainty and Variability in Manufacturing Systems: Managing Un-Planned Change in the Automative Industry. Avebury, 1994. ISBN 1856286207.

[27] Gregory G. Dess and Donald W. Beard. Dimensions of organizational task environments. Administrative Science Quarterly, 29(1):52–73, 1984. doi: 10.2307/2393080.

[28] I. S. Diakunchak. Performance deterioration in industrial gas turbines. Journal of Engineering for Gas Turbines and Power, 114(2):161–168, 1992. doi: 10.1115/1.2906565.

[29] N. C. Do and S. Razavi. Correlation effects? a major but often neglected component in sensitivity and uncertainty analysis. Water Resources Research, 56(3), 2015. doi: 10.1029/2019WR025436.

[30] D. Douglas-Smith, T. Iwanaga, B. F. W. Croke, and A. J. Jakeman. Certain trends in uncertainty and sensitivity analysis: An overview of software tools and techniques. Environmental Modelling and Software, 124(19), 2020. doi: 10.1016/j.envsoft.2019.104588.

[31] David Draper. Assessment and propagation of model uncertainty. Journal of the Royal Statistical Society, 57(1):45–70, 1995. doi: 10.2307/2346087.

[32] Liu Du and Kyung K. Choi. An inverse analysis method for design optimization with both statistical and fuzzy uncertainties. In ASME 2006 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, pages 1225–1234, 2008. doi: 10.1115/DETC2006-99731.

[33] Xiaoping Du and Wei Chen. Methodology for managing the effect of uncertainty in simulation-based design. AIAA Journal, 38(8), 2000. doi: 10.2514/2.1125.

[34] M. G. Dunn, C. Padova, J. E. Moller, and R. M. Adams. Performance deterioration of a turbofan and a turbojet engine upon exposure to a dust environment. Journal of Engineering for Gas Turbines and Power, 109(3):336–343, 1987. doi: 10.1115/1.3240045.

[35] Felix Döring, Stephan Staudacher, Christian Koch, and Matthias Weißschuh. Modeling particle deposition effects in aircraft engine compressors. Journal of Turbomachinery, 139(5), 2016. doi: 10.1115/1.4035072.

[36] Bradley Efron and Trevor Hastie. Computer Age Statistical Inference: Algorithms, Evidence, and Data Science. Cambridge University Press, USA, 2016. ISBN 1107149894.

[37] O. F. Eker, F. Camci, and I. K. Jennions. A similarity-based prognostics approach for remaining useful life prediction. In European Conference of the Prognostics and Health Management Society, volume 2, 2014. doi: 10.36001/phme.2014.v2i1.1479.

[38] Michael S. Eldred and Laura P. Swiler. Efficient algorithms for mixed aleatory-epistemic uncertainty quantification with application to radiation-hardened electronics, 2009. Technical Report.

[39] Michael Scott Eldred and John Burkardt. Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification. In 47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition, 2009. doi: 10.2514/6.2009-976.

[40] Amare D. Fentaye, Aklilu T. Baheta, Syed I. Gilani, and Konstantinos G. Kyprianidis. A review on gas turbine gas-path diagnostics:state-of-the-art methods, challenges and opportunities. Aerospace, 6(7), 2019. doi: 10.3390/aerospace6070083.

[41] GasTurb. Gasturb 14 manual. https://www.gasturb.de/Downloads/Manuals/GasTurb14.pdf, 2021. Accessed: 2021-07-18.

[42] Daniel Giesecke and Uyioghosa Igie. Performance and techno-economic investigation of on-wing compressor wash for a short-range aero engine. In ASME Turbo Expo 2012: Turbine Technical Conference and Exposition, volume 1, pages 235–244, 2012. doi: 10.1115/GT2012-68995.

[43] Graham Glen and Kristin Isaacs. Estimating sobol sensitivity indices using correlations. Environmental Modelling Software, 37:157–166, 2012. doi: https://doi.org/10.1016/j.envsoft.2012.03.014.

[44] Stéphane Hallegatte, Ankur Shah, Robert Lempert, Casey Brown, and Stuart Gill. Investment decision making under deep uncertainty – application to climate change. Technical report, The World Bank, 2012. URL http://hdl.handle.net/10986/12028.

[45] Li-Ping He, Hong-Zhong Huang, Li Du, Xu-Dong Zhang, and Qiang Miao. A review of possibilistic approaches to reliability analysis and optimization in engineering design. In International Conference on Human-Computer Interaction, 2007. doi: 10.1007/978-3-540-73111-5_118.

[46] Yanyan He, Mahsa Mirzargar, and Robert M. Kirby. Mixed aleatory and epistemic uncertainty quantification using fuzzy set theory. International Journal of Approximate Reasoning, 66(3), 2015. doi: 10.1016/j.ijar.2015.07.002.

[47] Felix O. Heimes. Recurrent neural networks for remaining useful life estimation. International Conference on Prognostics and Health Management, 2008. doi: 10.1109/PHM.2008.4711422.

[48] J. C. Helton and F. J. Davis. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. 2003. doi: 10.2172/806696. Technical Report.

[49] J. C. Helton, J. D. Johnson, and W. L. Oberkampf. An exploration of alternative approaches to the representation of uncertainty in model predictions. Reliability Engineering and System Safety, 85(1-3):39–71, 2004. doi: 10.1016/j.ress.2004.03.025.

[50] J. C. Helton, J. D. Johnson, C. J. Sallaberry, and C. B. Storlie. Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliability Engineering System Safety, 91(10-11):1175–1209, 2006. doi: 10.1016/j.ress.2005.11.017.

[51] Jon Herman and Will Usher. Salib: An open-source python library for sensitivity analysis. Journal of Open Source Software, 2(9):97, 2017. doi: 10.21105/joss.00097.

[52] F. Owen Hoffman and Jana S. Hammonds. Propagation of uncertainty in risk assessments: The need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. Risk Analysis, 14(5):707–712, 1994. doi: 10.1111/j.1539-6924.1994.tb00281.x.

[53] Chao Hu, Byeng D. Youn, Pingfeng Wang, and Joung Taek Yoon. Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. Reliability Engineering and System Safety, 103:120–135, 2012. doi: 10.1016/j.ress.2012.03.008.

[54] Hong-Zhong Huang, Xudong Zhang, De-Biao Meng, Yu Liu, and Yan-Feng Li. A new multidisciplinary design optimization method accounting for discrete and continuous variables under aleatory and epistemic uncertainties. International Journal of Computational Intelligence Systems, 5(1):93–110, 2012. doi: 10.1080/18756891.2012.670524.

[55] I. Hutter. Engine deterioration and maintenance actions. https://www.icao.int/Meetings/EnvironmentalWorkshops/Documents/ICAO-TransportCanada-2006/Hutter.pdf, 2006. Presentation. Accessed: 2021-08-15.

[56] Uyioghosa Igie. Gas turbine compressor fouling and washing in power and aerospace propulsion. Journal of Engineering for Gas Turbines and Power, 139(12), 2017. doi: 10.1115/1.4037453.

[57] Bertrand Iooss and Paul Lemaître. A review on global sensitivity analysis methods. Operations Research and Computer Science Interfaces Series, Springer, 59, 2015. doi: 10.1007/978-1-4899-7547-8_5.

[58] Rogerio Ishibashi and Cairo L. Nascimenta Jr. Gfrbs-phm: A genetic fuzzy rule-based system for phm with improved interpretability. In IEEE Conference on Prognostics and Health Management, 2013. doi: 10.1109/ICPHM.2013.6621419.

[59] Dominik Janzing, Lenon Minorics, and Patrick Blobaum. Feature relevance quantification in explainable ai: A causal problem. 2019. URL https://arxiv.org/abs/1910.13413.

[60] Lawrence R. Jauch and Kenneth L. Kraft. Strategic management of uncertainty. Academy of Management Review, 11(4), 1986. doi: 10.5465/amr.1986.4283934.

[61] J.C.Helton, J. D. Johnson W. L. Oberkampf, and C. B. Storlie. A sampling-based computational strategy for the representation of epistemic uncertainty in model predictions with evidence theory. Computer Methods in Applied Mechanics and Engineering, (37-40):3980–3998, 2007. doi: 10.1016/j.cma.2006.10.049.

[62] Cedric Y. Justin, Elena Garcia, and Dimitri N. Mavris. Aircraft valuation: A network approach to the evaluation of aircraft for fleet planning and strategic decision making. 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, 2010. doi: 10.2514/6.2010-9061.

[63] Cedric Yves Justin and Dimitri N. Mavris. Aircraft and engine economic evaluation for fleet renewal decision-making and maintenance contract valuation. Proceedings of the Institution of Mechanical Engineers Part G Journal of Aerospace Engineering, 229(11), 2015. doi: 10.1177/0954410014564403.

[64] Armen Der Kiureghian and Ove Ditlevsen. Aleatory or epistemic? does it matter? Structural Safety, 31(2):105–112, 2008. doi: 10.1016/j.strusafe.2008.06.020.

[65] V. Kreinovich and S. A. Ferson. A new cauchy-based black-box technique for uncertainty in risk analysis. Reliability Engineering System Safety, 85(1-3):267–279, 2004. doi: 10.1016/j.ress.2004.03.016.

[66] Patrick Kreiser and Louis Marino. Analyzing the historical development of the environmental uncertainty construct. Management Decision, 2002. doi: 10.1108/00251740210441090.

[67] S. Kucherenko, S. Tarantola, and P. Annoni. Estimation of global sensitivity indices for models with dependent variables. Computer Physics Communications, 183(4):937–946, 2012. doi: 10.1016/j.cpc.2011.12.020.

[68] Joey Van Kuiken, Bea Yu, and Daniel Telford. Managing uncertainty in reliability analysis with bayesian inference and uncertainty propagation. AIAA Journal, 2012. doi: 10.2514/6.2010-1762.

[69] R. Kurz and K. Brun. Degradation in gas turbine. ASME Journal of Engineering for Gas Turbines and Power, 123(1):70–77, 2001. doi: 10.1115/1.1340629.

[70] Sang Gyu Kwak and Jong Hae Kim. Central limit theorem: the cornerstone of modern statistics. Korean Journal of Anesthesiology, 70(2):144–156, 2017. doi: 10.4097/kjae.2017.70.2.144.

[71] Jan H. Kwakkel and Erik Pruyt. Exploratory modeling and analysis, an approach for model-based foresightunder deep uncertainty. Technological Forecasting & Social Change, 80(3):419–431, 2013. doi: 10.1016/j.techfore.2012.10.005.

[72] Jan H. Kwakkel, Warren E. Walker, and Vincent A. W. J. Marchau. Classifying and communicating uncertainties in model-based policy analysis. International Journal of Technology Policy and Management, 10(4):299–315, 2010. doi: 10.1504/IJTPM.2010.036918.

[73] A. N. Lakshminarasimha, M. P. Boyce, and C. B. Meher-Homji. Modeling and analysis of gas-turbine performance deterioration. Journal of Engineering for Gas Turbines and Power, 116(1), 1994. doi: 10.1115/1.2906808.

[74] Yaguo Lei, Naipeng Li, Liang Guo, Ningbo Li, Tao Yan, and Jing Lin. Machinery health prognostics: A systematic review from dataacquisition to rul prediction. Mechanical Systems and Signal Processing, 104:799–834, 2018. doi: 10.1016/j.ymssp.2017.11.016.

[75] G. Li, H. Rabitz, P. E. Yelvington, O. O. Oluwole, F. Bacon, C. E. Kolb, and J. Schoendorf. Global sensitivity analysis for systems with independent and/or correlated inputs. The Journal of Physical Chemistry A, 114(19):6022–6032, 2010. doi: 10.1021/jp9096919.

[76] Linxia Liao and F. Kottig. Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. IEEE Transactions on Reliability, 2014. doi: 10.1109/TR.2014.2299152.

[77] Kaibo Liu, Nagi Z. Gebraeel, and Jianjun Shi. A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis. IEEE Transactions on Automation Science and Engineering, 10(3):652–664, 2013. doi: 10.1109/TASE.2013.2250282.

[78] Kaibo Liu, Abdallah Chehade, and Changyue Song. Optimize the signal quality of the composite health index via data fusion for degradation modeling and prognostic analysis. IEEE Transactions on Automation Science and Engineering, 14(3):1504 – 1514, 2017. doi: 10.1109/TASE.2015.2446752.

[79] Scott Lundberg. Shap. https://github.com/slundberg/shap, 2021. URL https://github.com/slundberg/shap. Github Repository: Issue 1382.

[80] Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions. 2017. URL https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

[81] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. From local explanations to global understanding with explainable ai for trees. Nature Machine Intelligence, 2(1):56–67, 2020. doi: 10.1038/s42256-019-0138-9.

[82] Luna Majumder and S. S. Rao. Interval-based multi-objective optimization of aircraft wings under gust loads. AIAA Journal, 47(3), 2002. doi: 10.2514/1.37224.

[83] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 21(2): 239–245, 1979. doi: 10.2307/1268522.

[84] Ineke Susanne Maria Meijer. Uncertainty and Entrepreneurial Action. PhD thesis, 2008.

[85] N. Metropolis. The beginning of the monte carlo method. Los Alamos Science Special Issue, 15: 125–130, 1987.

[86] Frances J. Milliken. Three types of perceived uncertainty about the environment: State, effect, and response uncertainty. Academy of Management, 12(1), 1987. doi: 10.2307/257999.

[87] Ramin Moghaddass and Ming J. Zuo. An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process. Reliability Engineering System Safety, 124:92–104, 2014. doi: 10.1016/j.ress.2013.11.006.

[88] Christoph Molnar. Interpretable Machine Learning. 2019. ISBN 9780244768522. URL https://christophm.github.io/interpretable-ml-book/.

[89] Claudio Moraga. Introduction to fuzzy logic. Facta universitatis - series: Electronics and Energetics, 18:319–328, 2005. doi: 10.2298/FUEE0502319M.

[90] Max D. Morris. Factorial sampling plans for preliminary computational experiments. Technometrics, 33(2):161–174, 1991. doi: 10.2307/1269043.

[91] Zissimos P. Mourelatos and Jun Zhou. Reliability estimation and design with insufficient data based on possibility theory. AIAA Journal, 43(8), 2005. doi: 10.2514/1.12044.

[92] Zissimos P. Mourelatos and Jun Zhou. A design optimization method using evidence theory. Journal of Mechanical Design, 128(4):901–908, 2006. doi: 10.1115/1.2204970.

[93] Muhammad Naeem. Impacts of low-pressure (lp) compressors' fouling of aturbofan upon operational-effectiveness of a military aircraft. Applied Energy, 85(4):243–270, 2008. doi: 10.1016/j.apenergy.2007.02.011.

[94] R. B. Nelsen. An Introduction to Copulas (second ed.). Springer Series in Statistics, 2006.

[95] Eddie K. H. Ng. Kernel-based Copula Processes. PhD thesis, 2010.

[96] William Oberkampf and Jon Helton. Investigation of evidence theory for engineering applications. AIAA Journal, 2012. doi: 10.2514/6.2002-1569.

[97] William Oberkampf, Jon Helton, and Kari Sentz. Mathematical representation of uncertainty. AIAA Journal, 2012. doi: 10.2514/6.2001-1645.

[98] William L. Oberkampf and Christopher J. Roy. Verification and Validation in Scientific Computing. 2010. doi: 10.1017/CBO9780511760396.

[99] A. B. Owen. On dropping the first sobol' point. 2020. URL https://arxiv.org/abs/2008.08051.

[100] Art B. Owen. Sobol' indices and shapley value. Journal of Uncertainty Quantification, 2:245–251, 2014. doi: 10.1137/130936233.

[101] Inseok Park and Ramana V. Grandhi. Quantification of model-form and parametric uncertainty using evidence theory. Structural Safety, 39:44–51, 2012. doi: 10.1016/j.strusafe.2012.08.003.

[102] Yu Peng, Hong Wang, Jianmin Wang, Datong Liu, and Xiyuan Peng. A modified echo state network based remaining useful life estimation approach. In IEEE Conference on Prognostics and Health Management, 2012. doi: 10.1109/ICPHM.2012.6299524.

[103] J. Peter and Eric Savin. Generalized polynomial chaos and stochastic collocation method for uncertainty quantification in aerodynamics, 2018. URL https://core.ac.uk/display/101751997. (Presentation), Accessed on 05/12/2021.

[104] E. Plischke, E. Borgonovo, and C. L. Smith. Global sensitivity measures from given data. European Journal of Operational Research, 226(3):536–550, 2013. doi: 10.1016/j.ejor.2012.11.047.

[105] A. A. Pohya, K. Wicke, and T. Kilian. Introducing variance-based global sensitivity analysis for uncertainty enabled operational and economic aircraft technology assessment. Aerospace Science and Technology, 2021. Submitted.

[106] Ahmad Ali Pohya, Jennifer Wehrspohn, Robert Meissner, and Kai Wicke. A modular and lifecycle based discrete event simulation framework for quantifying the true value of aircraft technologies, maintenance strategies, and operational procedures. Advances in Engineering Software, 2021.

[107] Steven W. Popper, Robert J. Lempert, and Steven C. Bankes. Shaping the futrue. Scientific American, 2005.

[108] Chris Rackauckas. Global sensitivity analysis: Lecture notes. https://mitmath.github.io/18337/lecture17/global_sensitivity, 2020. Last accessed on: 05/08/2021.

[109] Emmanuel Ramasso, Michèle Rombaut, and Noureddine Zerhouni. Joint prediction of continuous and discrete states in time-series based on belief functions. IEEE Transactions on Cybernetics, 43(1):37–50, 2012. doi: 10.1109/TSMCB.2012.2198882.

[110] S. S. Rao and L. Berke. Analysis of uncertain structural systems using interval analysis. AIAA Journal, 35(4), 1997. doi: 10.2514/2.164.

[111] S. S. Rao and Lingtao Cao. Optimum design of mechanical systems involving interval parameters. Journal of Mechanical Design, 124(3):465–472, 2002. doi: 10.1115/1.1479691.

[112] S. Razavi. What do we mean by sensitivity analysis? the need for comprehensive characterization of global sensitivity in earth and environmental systems models. Water Resources Research, 51 (5), 2015. doi: 10.1002/2014WR016527.

[113] S. Razavi, A. Jakeman, A. Saltelli, C. Prieur, B. Iooss, E. Borgonovo, E. Plischke, S. Lo Piano, T. Iwanaga, W. Becker, S. Tarantola, J. H. A. Guillaume, J. Jakeman, H. Gupta, N. Melillo, G. Rabitti, V. Chabridon, Q. Duan, X. Sun, S. Smith, R. Sheikholeslami, N. Hosseini, M. Asadzadeh, A. Puy, S. Kucherenko, and H. R. Maier. The future of sensitivity analysis: An essential discipline for systems modeling and policy support. Environmental Modelling Software, 137, 2021. doi: https://doi.org/10.1016/j.envsoft.2020.104954.

[114] Jens Christian Refsgaard, Jeroen P. van der Sluijs, Anker Lajer Højberg, and Peter A. Vanrolleghem. Uncertainty in the environmental modelling process – a framework and guidance. Environmental Modelling and Software, 22(11):1543–1556, 2007. doi: 10.1016/j.envsoft.2007.02.004.

[115] Helen M. Regan, Mark Colyvan, and Mark A. Burgman. A taxonomy and treatment of uncertainty for ecology and conservation biology. Ecological Applications, 12(2):618–628, 2002. doi: 10.1890/1051-0761(2002)012[0618:ATATOU]2.0.CO;2.

[116] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. 2016. URL https://arxiv.org/abs/1602.04938.

[117] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Anchors: High-precision model-agnostic explanations. In In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, 2018. URL https://homes.cs.washington.edu/~marcotcr/aaai18.pdf.

[118] Martijn Roelofs and Roelof Vos. Technology evaluation and uncertainty-based design optimization: A review. American Institute of Aeronautics and Astronautics Inc., 2018. doi: 10.2514/6.2018-2029.

[119] Timothy J. Ross, Jane M. Booker, and W. Jerry Parkinson. Fuzzy Logic and Probability Applications: Bridging the Gap. 2002. ISBN 978-0-89871-525-5. doi: 10.1137/1.9780898718447.

[120] William D. Rowe. Understanding uncertainty. Risk Analysis, 14(5):743–750, 1994. doi: 10.1111/j.1539-6924.1994.tb00284.x.

[121] Christopher J. Roy and William L. Oberkampf. A comprehensive framework for verification, validation, and uncertaintyquantification in scientific computing. Computer Methods in Applied Mechanics and Engineering, 200:2131–2144, 2010. doi: 10.1016/j.cma.2011.03.016.

[122] A. Saltelli, S. Tarantola, and K. P. S. Chan. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics, 41(1):39–56, 1999. doi: 10.1080/00401706.1999.10485594.

[123] Andrea Saltelli and Paola Annoni. How to avoid a perfunctory sensitivity analysis. Environmental Modelling and Software, 25(12):1508–1517, 2010. doi: 10.1016/j.envsoft.2010.04.012.

[124] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Caribonia, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. Global Sensitivity Analysis: The Primer. 2008. ISBN 978-0-470-05997-5.

[125] Andrea Saltelli, Ksenia Aleksankina, William Becker, Pamela Fennell, Federico Ferretti, Niels Holst, Sushan Li, and Qiongli Wu. Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. Environmental Modelling and Software, 114: 29–39, 2019. doi: 10.1016/j.envsoft.2019.01.012.

[126] H. I. H. Saravanamuttoo, H. Cohen, and G. F. C. Rogers. Gas Turbine Theory: 5th Edition. Pearson, 2001.

[127] Abhinav Saxena, Kai Goebel, Don Simon, and Neil Eklund. Damage propagation modeling for aircraftengine run-to-failure simulation. International Conference on Prognostics and Health Management, 2008. doi: 10.1109/PHM.2008.4711414.

[128] O. Schwabe, E. Shehab, and J. Erkoyuncu. Uncertainty quantification metrics for whole product life cycle cost estimates in aerospace innovation. Progress in Aerospace Sciences, 2015. doi: 10.1016/j.paerosci.2015.06.002.

[129] Ralf Seemann. Modeling the life cycle cost of jet engine maintenance, 2010. M.Sc. Thesis.

[130] Ralf Seemann, Thomas Schilling, Stephan Langhans, and Volker Gollnick. Modeling the life cycle cost of jet engine maintenance. In 60. Deutscher Luft- und Raumfahrtkongress (DLRK), 2011. URL https://elib.dlr.de/75837/.

[131] Glenn Shafer. A Mathematical Theory of Evidence. 1976. ISBN 9780691100425.

[132] L. S. Shapley. A value for n-person games. Contributions to the Theory of Games (AM-28), 1953. doi: 10.1515/9781400881970-018.

[133] Youngwon Shin and Justin Wu. A hybrid possibilistic-probabilistic analysis framework for uncertainty management. In 44th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2003. doi: 10.2514/6.2003-1573.

[134] Nozer D. Singpurwalla and Jane M. Booker. Membership functions and probability measures of fuzzy sets. Journal of the American Statistical Association, 99(467):867–877, 2004. doi: 10.1198/016214504000001196.

[135] Daniel J. C. Skinner, Sophie A. Rocks, Simon J. T. Pollard, and Gillian H. Drew. Identifying uncertainty in environmental risk assessments: The development of a novel typology and its implications for risk characterization. Human and Ecological Risk Assessment, 20(3):607–640, 2014. doi: 10.1080/10807039.2013.779899.

[136] Natasha Smith and Sankaran Mahadevan. Probabilistic methods for aerospace system conceptual design. Journal of Spacecraft and Rockets, 40(3), 2012. doi: 10.2514/2.3961.

[137] M. J. Sobel, J. G. Szmerekovsky, and V. Tilson. Scheduling projects with stochastic activity duration to maximize expected net present value. European Journal of Operational Research, 198 (3):697–705, 2009. doi: 10.1016/j.ejor.2008.10.004.

[138] I. M. Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Mathematics and Computers in Simulation, 55(1-3):271–280, 2001. doi: 10.1016/S0378-4754(00)00270-6.

[139] I. M. Sobol and S. Kucherenko. Derivative based global sensitivity measures and their link with global sensitivity indices. Mathematics and Computers in Simulations, 79(19):3009–3017, 2009. doi: 10.1016/j.matcom.2009.01.023.

[140] Prabhu Soundappan, Efstratios Nikolaidis, Raphael T. Haftka, Ramana Grandhi, and Robert Canfiel. Comparison of evidence theory and bayesian theory for uncertainty modeling. Reliability Engineering System Safety, 85(1-3):295–311, 2004. doi: 10.1016/j.ress.2004.03.018.

[141] T. J. Sullivan. Introduction to Uncertainty Quantification. 2015. URL https://www.zib.de/sullivan/pdf/bms_uq_notes_00-05.pdf.

[142] Jianzhong Sun, Hongfu Zuo, Wenbin Wang, and Michael G. Pecht. Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance. Mechanical Systems and Signal Processing, 28, 2012. doi: 10.1016/j.ymssp.2011.09.029.

[143] S. Tarantola, D. Gatelli, and T. A. Mara. Random balance designs for the estimation of first order global sensitivity indices. Reliability Engineering and System Safety, 91(6):717–727, 2006. doi: 10.1016/j.ress.2005.06.003.

[144] Lufthansa Technik. Price list line maintenance services, 2012.

[145] J. Teng, A. J. Jakeman, J. Vaze, B. F. W. Croke, D. Dutta, and S. Kim. Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. Environmental Modelling and Software, 90:201–216, 2017. doi: 10.1016/j.envsoft.2017.01.006.

[146] Jean-Yves Tissot and Clémentine Prieur. Bias correction for the estimation of sensitivity indices based on random balance designs. Reliability Engineering System Safety, 107:205–213, 2012. doi: 10.1016/j.ress.2012.06.010.

[147] Laura Uusitalo, Annuka Lehikoinen, Inari Helle, and Kai Myrberg. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environmental Modelling and Software, 63:24–31, 2014. doi: 10.1016/j.envsoft.2014.09.017.

[148] Rens van de Schoot, Jaap Denissen, Franz J. Neyer, David Kaplan, Jens B. Asendorf, and Marcel A. G. van Aken. A gentle introduction to bayesian analysis: Applications to developmental research. Child Development, 85(3):842–860, 2013.

[149] Sebastien Da Veiga, Francois Wahl, and Fabrice Gamboa. Local polynomial estimation for sensitivity analysis on models with correlated inputs. Technometrics, 51(4):452–463, 2008. doi: 10.1198/TECH.2009.08124.

[150] Giulia Vilone and Luca Longo. Explainable artificial intelligence: a systematic review. 2020. URL https://arxiv.org/abs/2006.00093.

[151] W. E. Walker, P. Harremoes, J. Rotmans, J. P. Van Der Sluijs, M. B. A. Van Asselt, P. Janssen, and M. P. Krayer Von Krauss. Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. Integrated Assessment, 4(1):5–17, 2003. doi: 10.1076/iaij.4.1.5.16466.

[152] Hai Wang and David A. Sheen. Combustion kinetic model uncertainty quantification, propagationand minimization. Progress in Energy and Combustion Science, 47, 2015. doi: 10.1016/j.pecs.2014.10.002.

[153] Pingfeng Wang and Byeng Dong Youn. Efficient bayesian reliability analysis and design with a user-defined confidence level. In Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, 2010. doi: 10.2514/6.2010-2590.

[154] Pingfeng Wang, Byeng D. Youn, Zhimin Xi, and Artemis Kloess. Bayesian reliability analysis with evolving, insufficient, and subjective data sets. Journal of Mechanical Design, 131(11), 2009. doi: 10.1115/1.4000251.

[155] Tobias Wensky, Lutz Winkler, and Jens Friedrichs. Environmental influences on engine performance degradation. In ASME Turbo Expo 2010: Power for Land, Sea, and Air, pages 249–254, 2010. doi: 10.1115/GT2010-22748.

[156] P. Wiederkehr. Global sensitivity anlaysis with dependent inputs. Master's thesis, 2018. URL https://sudret.ibk.ethz.ch/education/student-projects/masterTheses/p--wiederkehr.html.

[157] Zhimin Xi, Rong Jing, Pingfeng Wang, and Chao Hu. A copula-based sampling method for data-driven prognostics and health management. In IEEE Conference on Prognostics and Health Management (PHM), 2013. doi: 10.1109/ICPHM.2013.6621450.

[158] Yinjiao Xing, Qiang Miao, K. L. Tsui, and Michael Pecht. Prognostics and health monitoring for lithium-ion battery. In Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics, 2011. doi: 10.1109/ISI.2011.5984090.

[159] Jiuping Xu and Lei Xu. Health management based on fusion prognostics for avionics systems. Journal of Systems Engineering and Electronics, 22(3):428–436, 2011. doi: 10.3969/j.issn.1004-4132.2011.03.010.

[160] Jiuping Xu, Yusheng Wang, and Lei Xu. Phm-oriented integrated fusion prognostics foraircraft engines based on sensor data. IEEE Sensors Journal, 14(4):1124–1132, 2014. doi: 10.1109/JSEN.2013.2293517.

[161] Wen Yao, Xiaoqian Chen, Wencai Luo, Michel van Tooren, and Jian Guo. Review of uncertainty-based multidisciplinary design optimization methodsfor aerospace vehicles. Progress in Aerospace Sciences, 47(6):450–479, 2011. doi: 10.1016/j.paerosci.2011.05.001.

[162] Mustagime Tülin Yildirim and Bülent Kurt. Aircraft gas turbine engine health monitoring system by real flight data. International Journal of Aerospace Engineering, 2018. doi: 10.1155/2018/9570873.

[163] Ilker Yilmaz. Evaluation of the relationship between exhaustgas temperature and operational parameters incfm56-7b engines. In Proceedings of the Institution of Mechanical Engineers Part G Journal of Aerospace Engineeering, 2009. doi: 10.1243/09544100JAERO474.

[164] Byeng D. Youn, Kyung K. Choi, Liu Du, and David Gorsich. Integration of possibility-based optimization and robust design for epistemic uncertainty. Journal of Mechanical Design, 129(8):876–882, 2006. doi: 10.1115/1.2717232.

[165] Jianbo Yu. Aircraft engine health prognostics based on logistic regression with penalization regularization and state-space-based degradation framework. Aerospace Science and Technology, 68, 2017. doi: 10.1016/j.ast.2017.05.030.

[166] L. A. Zadeh. Fuzzy sets. Information and Control, 8(3):338–353, 1965. doi: 10.1016/S0019-9958(65)90241-X.

[167] L. A. Zadeh. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1(1):3–28, 1978. doi: 10.5555/310817.310820.

[168] A. V. Zaita, G. Buley, and G. Karlsons. Performance deterioration modeling in aircraft gas turbine engines. Journal of Engineering for Gas Turbines and Power, 120(2):344–349, 1998. doi: 10.1115/1.2818128.

[169] Thomas A. Zang, Michael Hemsch, Mark W. Hilburger, Sean P. Kenny, Peiman Maghami, W. Hefferson Stroud, Sharon L. Padula, and James M. Luckring. Needs and opportunities for uncertainty-based multidisciplinary design methods for aerospace vehicles, 2002. URL https://ntrs.nasa.gov/citations/20020063596.