CRANFIELD UNIVERSITY

ALEX GRENYER

ADVANCED UNCERTAINTY QUANTIFICATION WITH DYNAMIC PREDICTION TECHNIQUES UNDER LIMITED DATA FOR INDUSTRIAL MAINTENANCE APPLICATIONS

SCHOOL OF AEROSPACE, TRANSPORT AND MANUFACTURING

PhD
Academic Year: 2017 - 2021

Supervisor: Prof. John Ahmet Erkoyuncu
Associate Supervisor: Dr Yifan Zhao
July 2021

CRANFIELD UNIVERSITY

SCHOOL OF AEROSPACE, TRANSPORT AND MANUFACTURING

PhD

Academic Year 2017 - 2021

ALEX GRENYER

ADVANCED UNCERTAINTY QUANTIFICATION WITH DYNAMIC PREDICTION TECHNIQUES UNDER LIMITED DATA FOR INDUSTRIAL MAINTENANCE APPLICATIONS

Supervisor: Prof. John Ahmet Erkoyuncu
Associate Supervisor: Dr Yifan Zhao
July 2021

This thesis is submitted in partial fulfilment of the requirements for the degree of PhD

© Cranfield University 2021. All rights reserved. No part of this publication may be reproduced without the written permission of the copyright owner.

ABSTRACT

Engineering systems are expected to function effectively whilst maintaining reliability in service. These systems consist of various equipment units, many of which are maintained on a corrective or time-based basis. Challenges to plan maintenance accounting for turnaround times, equipment availability and resulting costs manifest varying degrees of uncertainty stemming from multiple quantitative and qualitative (compound) sources throughout the in-service life.

Under or over-estimating this uncertainty can lead to increased failure rates or, more often, unnecessary maintenance being carried out. As well as the quality availability of data, uncertainty is driven by the influence of expert experience or assumptions and environmental operating conditions. Accommodating for uncertainty requires the determination of key contributors, their influence on interconnected units and how this might change over time.

This research aims to develop a modelling approach to quantify, aggregate and forecast uncertainty given by a combination of historic equipment data and heuristic estimates for in-service engineering systems. Research gaps and challenges are identified through a systematic literature review and supported by a series of surveys and interviews with industrial practitioners. These are addressed by the development of two frameworks: (1) quantify and aggregate compound uncertainty, and (2) predict uncertainty under limited data.

The two frameworks are brought together to produce the Multistep Compound Dynamic Uncertainty Quantification (MCDUQ) app, developed in MATLAB. Results demonstrate effective measurement of compound uncertainties and their impact on system reliability, along with robust predictions under limited data with an immersive visualisation of dynamic uncertainty. The embedded frameworks are each validated through implementation in two case studies. The app is verified with industrial experts through a series of interviews and virtual demonstrations.

Keywords: Engineering systems; Limited data; Uncertainty aggregation; Uncertainty prediction

ACKNOWLEDGEMENTS

I would like to thank everyone who has helped, motivated and encouraged me throughout the course of my research. My foremost thanks go to my primary academic supervisor, Prof. John Erkoyuncu, for his tireless and positive support and motivation from the beginning. My thanks also go to my associate supervisors Prof. Rajkumar Roy (now Dean at City University London), and Dr Yifan Zhao for their encouragement and guidance. I also thank Dr Pavan Addepalli for his help and support in setting up experiments and keeping things grounded. I also thank all members of the former Through-life Engineering Services Centre (TES), now Centre for Digital Engineering and Manufacturing (CDEM) at Cranfield University.

Further gratitude goes to Darren Goodman (BAE Systems) for his support and industrial perspective from the beginning, as well as arranging interviews, securing feedback and organising visits to Portsmouth. I thank Dr Eric Fennessey and Dr Richard (Ed) Nicklin (BAE Systems) for their input and constructive feedback, as well as all industrial practitioners who participated in the interviews and surveys. In addition, I would like to thank Dr Oliver Schwabe (Rolls-Royce Deutschland) for his insights, guidance and collaboration.

I thank my PhD reviewers Dr Glenn Sherwood (Cranfield University), Prof. Harris Makatsoris (formerly Cranfield University, now Kings College London) and Dr Maryam Farsi (Cranfield University) for their constructive feedback, as well as my viva examiners Dr Mey Goh (Loughborough University) and Dr Samir Khan (Cranfield University).

I would also like to recognise the Engineering and Physical Sciences Research Council (EPSRC) for funding this research, as well as my extension funding from UKRI in light of the COVID-19 pandemic.

Finally, I would like to thank all my friends and family who have supported me throughout my research journey and provided the escapism needed to step back and relax.

LIST OF PUBLICATIONS

Journal papers

1. Grenyer, A., Erkoyuncu, J. A., Zhao, Y. and Roy, R. (2021), 'A systematic review of multivariate uncertainty quantification for engineering systems', CIRP Journal of *Manufacturing* Science and Technology, 33, pp. 188–208. DOI: 10.1016/j.cirpj.2021.03.004.

2. Grenyer, A., Schwabe, O., Erkoyuncu, J. A., & Zhao, Y. (2022). Multistep prediction of dynamic uncertainty under limited data. CIRP Journal of Manufacturing Science and Technology, 37, 37–54. https://doi.org/10.1016/j.cirpj.2022.01.002

Conference papers

1. Grenyer, A., Addepalli, S., Zhao, Y., Oakey, L., Erkoyuncu, J. A. and Roy, R. (2018), 'Identifying challenges in quantifying uncertainty: Case study in infrared thermography', *Procedia CIRP*, 73, pp. 108–113. DOI: 10.1016/j.procir.2018.03.301.

2. Farsi, M., Grenyer, A., Sachidananda, M., Sceral, M., Mcvey, S., Erkoyuncu, J. and Roy, R. (2018), 'Conceptualising the impact of information asymmetry on through-life cost: case study of machine tools sector', *Procedia Manufacturing*, 16, pp. 99–106. DOI: 10.1016/j.promfg.2018.10.172.

3. Grenyer, A., Dinmohammadi, F., Erkoyuncu, J. A., Zhao, Y. and Roy, R. (2019), 'Current practice and challenges towards handling uncertainty for effective outcomes in maintenance', *Procedia CIRP*, 86, pp. 282–287. DOI: 10.1016/j.procir.2020.01.024.

4. Grenyer, A., Erkoyuncu, J. A., Addepalli, S. and Zhao, Y. (2020), 'An uncertainty quantification and aggregation framework for system performance assessment in industrial maintenance', *SSRN Electronic Journal*, (November). DOI: 10.2139/ssrn.3718001.

5. Grenyer, A., Schwabe, O., Erkoyuncu, J. A. and Zhao, Y. (2021), 'Dynamic multistep uncertainty prediction in spatial geometry', *Procedia CIRP*, 96, pp. 74–79. DOI: 10.1016/j.procir.2021.01.055.

Submitted papers under peer review

1. Grenyer, A., Erkoyuncu, J. A., Addepalli, S. and Zhao, Y. (2021), 'Compound uncertainty quantification and aggregation (CUQA) for reliability measurement in industrial maintenance', *Reliability Engineering & System Safety.* Submitted 05 March 2021.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AHP	Analytical hierarchy process
AR	Augmented reality
BDL	Bayesian deep learning
BPN	Backpropagation artificial neural network
CES	Complex engineering system
C-MAPSS	Commercial modular aero-propulsion system simulation
CUQA	Compound uncertainty quantification and aggregation
CV	Coefficient of variation
GLUE	General likelihood uncertainty estimation
GSA	Global sensitivity analysis
GSD	Geometric standard deviation
GUM	Guide to the expression of uncertainty measurement
KPI	Key performance indicator
LCA	Life cycle assessment
LHS	Latin Hypercube Sampling
LSA	Local sensitivity analysis
LSTM	Long-short term memory
MAPE	Mean absolute percentage error
MCDUQ	Multistep compound dynamic uncertainty quantification
NN	Neural network
NUSAP	Numeral, unit, spread, assessment, pedigree
PDF	Probability distribution function
PHM	Prognostics and health management
PICOC	Population, intervention, comparison, outcome, context
PSS	Product-service system
RMSE	Root-mean-square error
RNN	Recurrent neural networks
RSS	Root-sum-square
RUL	Remaining useful life
SA	Sensitivity analysis
SALSA	Search, appraisal, synthesis, analysis
SAR	Selected acquisition report
SLR	Systematic literature review
UQ	Uncertainty quantification

CHAPTER 1. INTRODUCTION

1.1 Background

Through-life service contracts deliver levels of availability, affordability and performance for assets operating in various challenging environments consisting of increasingly complex engineering systems (CES). Decisions made planning maintenance for such assets historically requires significant experience and expertise, as well as the use of equipment data that may be inaccurate, sporadic, or outdated. Numerous uncertainties are raised here concerning the validity of expert opinions and accuracy of recorded data, which risk over or under estimation of factors relating to maintenance carried out. Significant costs and delays are risked as a result. Uncertainty is defined in this thesis as the difference between the degree of information required and information held to make a decision concerning a given entity. As well as deviations in quantitative, recorded data, this definition encompasses information sourced from qualitative, subjective opinions, assumptions and environmental factors. The resulting risk is the impact the uncertainty will have on the given entity [1,2,20]. This thesis focuses on uncertainty, not the resulting risk. Modern analytical methods can provide rigorous, self-learning scientific approaches to quantify and predict uncertainty by employing intelligent logical systems to automate and learn from live and historic data to aid decision-making. This has the potential to significantly reduce risk factors while improving performance, efficiency, and safety [5,7].

1.2 Problem statement

Uncertainty quantification (UQ) has been explored in various fields [3,11,15,18]. This is the practice of characterising uncertainties for computational and real-world applications. Many approaches model a particular type of uncertainty from statistical sources under probability theory. Methods to obtain and analyse qualitative attributes often go undefined and unmitigated, which has the potential to increase the occurrence likelihood

of unforeseen events [2]. A holistic aggregation of uncertainty from quantitative and qualitative sources will aid decision-making and reduce under or over estimation of maintenance costs and turnaround time.

Aggregation traditionally considers a summation of best and worst-case scenarios to define boundaries for likely outcomes. This raises the question of whether uncertainties can be aggregated across multiple elements represented through different probability distributions. To do so, a second question is raised of how to standardise and validate qualitative estimates attributed by expert opinion.

To consider how this may change over time, a third question considers how uncertainty can be forecast over the in-service phase of an asset's life cycle. The availability, consistency and accessibility of equipment data can change dynamically over time, necessitating the need for rigorous UQ techniques to optimally incorporate resulting challenges into maintenance planning. This is dependent on the selection of the best applicable UQ method, such as probability theory, Monte Carlo, Bayesian deep learning and neural networks (NN).

This novel research examines the quantification of uncertainty propagated by these challenges, along with those faced in maintenance delivery. This is considered to be unique in literature for this context and offers defined academic contributions around dynamic quantification of technical uncertainties at the equipment-type (ET) level in real-world industrial applications. The ET level considers multiple subsystems that interact with unique availability requirements, prompting a high influence on maintenance expenditure [13]. The increasing complexity of engineering systems makes it progressively difficult to comprehend the impact of uncertainty for alternative maintenance scenarios in Product-Service Systems (PSS). This promotes the need to scientifically quantify uncertainty rather than rely solely on expert opinion that is inherently subjective. There is also a need to implement self-learning systems capable of making predictions and recommendations based on historic data and human input to optimise decision-making for the in-service phase of an asset's life cycle.

1.3 Aim and objectives

The aim of this research is to develop a modelling approach to quantify, aggregate and forecast uncertainty given by a combination of historic equipment data and heuristic estimates for in-service engineering systems.

It is hypothesised that a rigorous and structured approach to quantify, aggregate and forecast technical engineering uncertainties from quantitative and qualitative sources throughout the in-service phase will improve uncertainty management at the ET level for real-world industrial maintenance under limited and sporadic data. To test this hypothesis and deliver on the aim, four key objectives were set:

Objective 1: Map current practice to identify core challenges and resulting uncertainties around equipment cost and availability and how these differ from forecast behaviour within complex engineering systems.

Objective 2: Develop a framework to aggregate uncertainty from quantitative and qualitative sources represented through different probability distributions with an identification of the source of greatest uncertainty.

Objective 3: Develop an approach to predict uncertainty given by limited available data and qualitative factors to relate to equipment cost and availability over the in-service phase.

Objective 4: Validate the final model to assess implementation effectiveness and usability in context.

1.4 Research development

The research presented in this thesis was developed through collaboration between the Through-life Engineering Services Centre (TES) at Cranfield University (UK) and BAE Systems Maritime Services (UK). The research was funded by the Engineering and Physical Sciences Research Council (EPSRC), project reference 1944319, and Doctoral Training Partnership (DTP). Underlying data for each chapter is available on the Cranfield University repository, CORD, under the project title of this thesis.

This thesis is presented as a series of chapters adapted from journal papers. Each chapter begins with a title page consisting of an abstract and details of published or submitted manuscripts, including DOI links. Chapters 3-5 present approaches used to fulfil the objectives. Their discussions and critique are therefore given at the end of the respective chapter, tied together in Chapter 6. References and appendices are provided at the end of the thesis. All experimental work was completed as specified by Alex Grenyer at Cranfield University (UK), with set-up assistance from Dr Pavan Addepalli and suppliers of measurement equipment. Interviews discussed in Chapter 3 were held at BAE Systems Maritime Services, Portsmouth Naval Base (UK), organised by Darren Goodman. Further validation with industrial practitioners was obtained in workshops with the Society for Cost Analysis and Forecasting (SCAF) and at the 36th International Symposium on Military Research (ISMOR). Final validation and verification interviews discussed at the end of Chapter 5 and Chapter 6 were held virtually.

1.5 Thesis structure

A summary of the thesis structure is given in Figure 1.1 with feedforward and feedback loops to illustrate where chapters feed into each other. Chapter 2 presents an in-depth systematic literature review to identify and assess existing methodologies to quantify, aggregate and forecast uncertainty from quantitative and qualitative (compound) sources with a view to better understand the impact on cost and availability to aid decision making throughout the in-service phase. A total of 107 papers were analysed to answer three research questions based on the scope, through which three key research gaps were identified. The review outcome informed the selection of methodologies to develop frameworks that will fulfil objectives and address the research gaps. The literature reviews from the manuscripts on which Chapters 3-5 are based have been incorporated into this chapter. The methodology selection process is detailed in Appendix B.

Chapter 3 examines current practice and challenges in industrial maintenance that exhibit uncertainties around equipment cost and availability, fulfilling Objective 1. Surveys and interviews were held with maintenance managers and validated with wider industrial practitioners. Challenges were examined in practice, considering interlinked systems in the product-service system (PSS) context Core factors that manifest uncertainty in maintenance were identified.

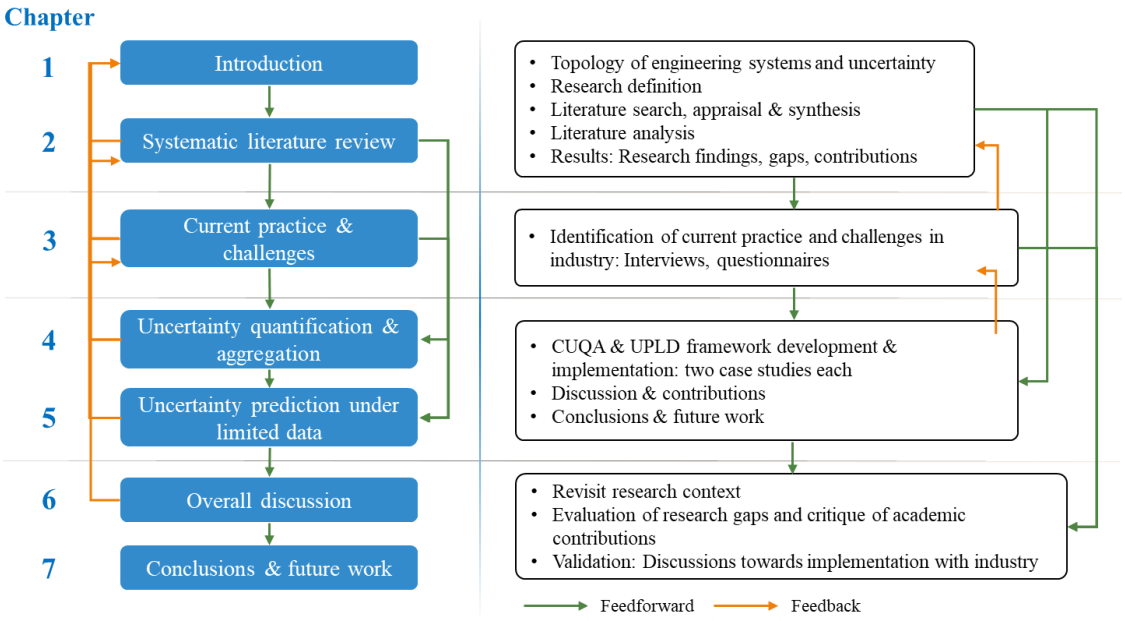


Figure 1.1. Thesis structure

Chapter 4 presents the Compound Uncertainty Quantification and Aggregation (CUQA) framework to fulfil Objective 2. Influenced by the findings of the previous chapters, this framework expanded upon existing techniques to aggregate compound uncertainties through the coefficient of variation (CV) and illustrate which inputs incite the greatest source of uncertainty.

Chapter 5 presents the framework for uncertainty prediction under limited data (UPLD), which embraces the third objective for the in-service life cycle phase. This multistep prediction model combines spatial geometry with long-short term memory (LSTM) neural networks. The framework is designed to be flexible to enable use in a variety of systems and allows the user to tune parameters to enhance the robustness of predictions.

Chapter 6 evaluates the research gaps and critiques the contributions to knowledge. The two frameworks are brought together to form the Multistep Compound Dynamic Uncertainty Quantification (MCDUQ) modelling approach, detailed in Appendix F. Feedback from the industrial sponsor concerning the model's implementation is presented to achieve the fourth objective.

Chapter 7 summarises the key conclusions against the objectives and recommendations of future work in the field of uncertainty quantification, aggregation and prediction. Table 1.1 summarises the thesis plan and the status of paper submissions at the time of writing. A table of experiments performed in the thesis, along with their justification towards the methodology, is given in Table 1.2.

Table 1.1. Thesis plan and status of paper submissions

Chapter	Adapted from paper	Objective	Title	Journal	Status
2	1	1, 2, 3	A systematic review of multivariate uncertainty quantification for engineering systems	CIRP Journal of Manufacturing, Science and Technology	Published 2021 [97]
3	2	1	Identifying challenges in quantifying uncertainty: case study in infrared thermography	Procedia CIRP: IPS2 Conference 2018	Published 2018 [191]
3	3	1	Current practice and challenges towards handling uncertainty for effective outcomes in maintenance	Procedia CIRP: CIRPe Web Conference 2019	Published 2019 [10]
4	4	2	An uncertainty quantification and aggregation framework for system performance assessment in industrial maintenance	SSRN: TES Conference 2020	Published 2020 [30]
4	5	2	Compound uncertainty quantification and aggregation (CUQA) for reliability measurement in industrial maintenance	Reliability Engineering & System Safety	Submitted 2021
5	6	3	Dynamic multistep uncertainty prediction in spatial geometry	Procedia CIRP: CIRPe Web Conference 2020	Published 2020 [196]
5	7	3	Multistep prediction of dynamic uncertainty under limited data	CIRP Journal of Manufacturing, Science and Technology	Published 2022
6	-	1, 2, 3, 4	Discussion of overall work	-	-
7	-	-	Conclusions and future work	-	-

Table 1.2. Table of experiments

Chapter	Case study	Type	Justification
3	Current practice and challenges	Survey / interview	Identification of current practice and challenges that manifest uncertainty in industrial maintenance.
4	Heat exchanger test rig	Lab work, live data recording	Combination of digital and analogue recording devices as well as qualitative factors to manifest compound uncertainty in heat exchanger performance. Demonstration of CUQA framework to assess impact on full system.
4	C-MAPSS turbofan engine degradation	Simulated dataset	Simulated degradation comprising of multiple sensor measurements and qualitative factors. Demonstrated use of CUQA framework to analyse aggregated compound uncertainty over time.
5	US SAR cost variance	Pre-existing dataset	Dataset used to validate initial spatial geometry approach [28]. Applied for UPLD provide comparable consistency in the application and demonstrate the wide applicability of the framework.
5	C-MAPSS turbofan engine degradation (results from Chapter 4)	Calculated uncertainties from simulated dataset	Aggregated and individual uncertainties calculated by the CUQA framework (Chapter 4). Applied to UPLD to further demonstrate the capability to predict uncertainty under limited data.
5	UPLD verification	Survey/interview	Review of the pertinence of the UPLD framework, its benefits and where improvements are required.
5	UPLD approach comparison	Method comparison	Comparison of prediction results using UPLD with other approaches, made by calculating the percentage difference of symmetry from resulting predictions to that observed.
5	UPLD comparison with additional data	Comparison with additial data (simulated time series)	Use of additional, simulated time series data to evaluate effectiveness of the UPLD approach.
6	MCDUQ validation	Survey/interview, demonstration	Demonstration of the final MCDUQ modelling approach with industrial practitioners to provide validation and feedback. Discussions and surveys were held regarding its effectiveness and steps towards industrial implementation.

CHAPTER 2. LITERATURE REVIEW

Abstract

Engineering systems must function effectively whilst maintaining reliability in service. Predicting maintenance costs and asset availability raises varying degrees of uncertainty from multiple sources. Previous reviews in this domain have assessed cost uncertainty and estimation for the entire life cycle. This chapter presents a systematic review to investigate existing methodologies and challenges in uncertainty quantification, aggregation and prediction for modern engineering systems through their in-service life. Approaches to predict uncertainty are hindered chiefly by the quality of available data, experience, and knowledge. A total of 107 papers were analysed to answer three research questions based on the scope, through which three core research gaps are identified. An integrated combination of identified approaches will enhance rigour in uncertainty assessment and prediction. This review contributes a systematic identification and assessment of current practices in uncertainty quantification and scientific methodologies to quantify, aggregate and predict quantitative and qualitative uncertainties to better understand their impact on cost and availability, aiding decision making throughout the in-service phase.

Paper 1 A systematic review of multivariate uncertainty quantification for engineering systems

Published: CIRP Journal of Manufacturing, Science and Technology
DOI: 10.1016/j.cirpj.2021.03.004
Data access: Available upon request

2.1 Introduction

The increasing complexity and dynamic nature of engineering systems drives an inherently high level of uncertainty. Many such complex engineering systems (CES) consist of multiple component parts or subsystems that interact in a collective manner not representative of individual parts [4,16,19]. Examples of complex systems range from biological organisms, global climate and meteorology to bridges, ships and aircraft. Engineering systems are expected to carry out intended functions whilst maintaining reliability in service. It is therefore increasingly challenging to confidently predict availability, cost and performance in various operating conditions [4,6,21]. Decisions made concerning these factors are shrouded in uncertainty, requiring significant experience and expertise, as well as the use of often outdated equipment data. This is typically managed under through-life product-service system (PSS) contracts, where the client makes use of a product in their possession but does not take ownership [23,25,27,29,31].

This review is motivated by the requirement for scientific approaches to quantify, aggregate and forecast technical engineering uncertainties for complex and non-complex engineering systems. These uncertainties impact the ability to effectively carry out maintenance tasks given available techniques and technology to required industry standards. Examples include uncertainties in degradation, no-fault found, obsolescence and failure rates [17,33,35].

It is therefore hypothesised that the utilisation of the above approaches considering a compound aggregation of measured, recorded data (quantitative) and experience-driven opinion or human factors (qualitative) will increase confidence and rigour in determination of the impact of uncertainty over time. There is a requirement to look beyond the probabilistic world and embrace subjective and expert opinions. Such approaches should be applicable for various scenarios where data may be incomplete, inconsistent, inaccessible, and reliant on expert opinion [26,39,52]. In the light of dramatically increasing data volumes and computational capability in engineering systems, rigorous machine learning algorithms should be incorporated to predict how uncertainty may change over the in-service phase [42,44].

Previous reviews in this domain have considered the role of uncertainty estimation in life cycle costing under PSS [26,46,48,50]. The in-service phase covers the largest portion of an asset's lifecycle between contract bidding and disposal. Many approaches to aggregate different types of uncertainty consider a summation of best and worst-case scenarios, represented by probability distributions, to define boundaries for likely outcomes [26]. Inadvertently disregarding the space between these scenarios may result in under or over-estimation and data distortion, adversely impacting decision-making.

This chapter presents a systematic literature review (SLR) to investigate distinct approaches in uncertainty quantification and aggregation that can be applied in a real-world context, in conjunction with how changes in uncertainty can be predicted in-service. Both complex and non-complex engineering systems are considered in this review, with a focus towards CES owing to their increasing relevance within the research scope. The objectives and resulting research questions (RQs) to achieve this are depicted in Section 2.3. The review follows the 4-stage analytical framework composed by Booth, Papaioannou and Sutton [53] to conduct an SLR: search, appraisal, synthesis and analysis (SALSA). This generic approach is well validated and can be applied under varying conditions to provide a clear analysis of literature published in the field of uncertainty and identify research gaps [54,56,58].

The primary contribution of this review is the combined consideration of scientific methodologies to quantify (numerical expression of an entity), aggregate (collation of entities) and predict (likely future outcomes) quantitative and qualitative uncertainties to better understand their impact on cost and availability to aid decision making throughout the in-service phase. A total of 107 papers were analysed to answer three research questions, through which three core research gaps were identified.

The chapter is structured as follows: Section 2.2 discusses a topology of engineering systems and uncertainty, including classification and recognised standards. Section 2.3 defines the research scope and subsequent RQs for the review. The search, appraisal and synthesis stages of the SALSA framework is given in Appendix A. Section 2.4 analyses the findings. Section 2.5 discusses the research findings parallel to the RQs. Section 2.6 concludes the review and identifies research gaps and future work.

2.2 Topology of engineering systems and uncertainty

As stated in Section 2.1, a complex system is comprised of multiple component parts or subsystems interacting linearly or non-linearly, exhibiting a collective behaviour that is distinct from and seldom predictable by that of individual parts or subsystems [4,16,19]. Conversely, a complicated system can be comprised of a myriad of interconnected parts but still exhibit a predictable collective behaviour [16,19]. Complex systems science is a rapidly expanding and evolving field, the theory of which is widely documented [4,6,16,19,21,60,62].

A complex engineering system (CES) is one that is focused on an engineering domain rather than, for example, social, biological or meteorological systems. The inherently complex and dynamic nature of CES manifests high levels of uncertainty. This takes shape in various forms including costing, policymaking, supply chains and technical uncertainties [19,21]. Technical engineering uncertainties within engineering systems set the context for this research, where uncertainty in the performance of one component or subsystem (node) may have knock-on effects with interconnected nodes or the whole system. The level of uncertainty can change throughout the in-service life of each node in an unpredictable and often non-linear manner [16,19]. This calls for adaptive and intelligent approaches to predict uncertainty based on a combination of available data and expert opinion.

There are several definitions and interpretations of uncertainty in literature [1,2,71,72,18,20,64,66–70]. It is defined here as the difference between the amount of information that is required to perform a task and the amount of information already possessed. The relevance of information, or lack of, should be specified concerning the functionality of the organisation or application in question [73]. Uncertainty is caused by variability in the environment, human error and/or human ambiguity (e.g. lack of knowledge) and could cause a negative, positive or neutral impact on overall performance [74]. All of these elements come into play for CES and should be accounted for to avoid unnecessary costs.

The terms error and uncertainty are often used interchangeably. Risk is generally interpreted as purely negative impacts of uncertainty [1,66,67,70,75]. It is important to differentiate these concepts. A statistical error is the (unknown) difference between a

measured value and a true value, following probability distributions. Measurement uncertainty is the lack of information about the magnitude of these errors. Risk is the positive or negative impact specific sources of uncertainty will have on the measurand (the system or components for which uncertainty is being assessed). The degree of uncertainty associated with the measurand can be utilised to aid decision making.

There are two key types of uncertainty described in literature: Type A, which are sourced from quantitative data; and Type B, which make use of qualitative technical and expert knowledge or experience [2,18,52,76–78]. These are further explored in Section 2.4.2. In the context of this research, Type A will hence be referred to as 'quantitative' and Type B as 'qualitative'. Uncertainty can be further characterised as aleatory and epistemic. Epistemic uncertainties are those that could be known in principle but are not known in practice [14,79–82]. This may be due to inaccurate measurements or the measurement model neglecting certain characteristics. Epistemic uncertainties can therefore be reduced by obtaining more data or by refining models. Aleatory uncertainty cannot be reduced as it represents statistical variables that differ each time a given experiment is carried out [7,14,87,88,79–86]. The influence of different types of uncertainty can play a key role in confidence determination for risk and reliability analysis [69].

Further examination can be made by the four '(un)known-(un)known risk quadrants', described in detail by Marshall et al. [89]. These levels of risk identification can be applied to both quantitative and qualitative uncertainty since risk is the impact of uncertainty on the measurand. As their names suggest, 'known knowns' are uncertainty sources that have been taken into account and catered for; 'known unknowns' are understood to exist but their magnitude is not defined; 'unknown knowns' are unidentified sources that may be accounted for through alternate means (possibly by other sources creating information asymmetry [90]); 'unknown unknowns' have not been identified or accounted for and, therefore, pose the greatest risk [21,89,91]. These traits can also represent predictable uncertainties not initially apparent and unpredictable 'black swan' events. Where the amount of information (uncertainty) is known, risk can be reduced. A categorisation of uncertainties centred on the four quadrants based on the nature and source of uncertainty is illustrated in Figure 2.1 [89,91]. An example

uncertainty source for each quadrant is linked to possible types – quantitative, qualitative, epistemic and aleatory.

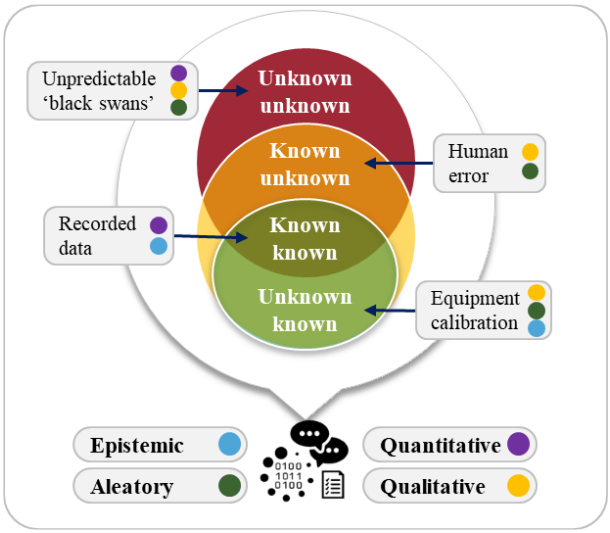


Figure 2.1. Categorisation of uncertainties based on their nature and sources

Frameworks to assess uncertainty in engineering systems, as well as the systems themselves, require a degree of flexibility to accommodate complexity while maintaining a degree of robustness to meet core objectives within specified confidence boundaries [21,31,80]. Flexibility in engineering systems design allows for mitigation in the face of unknown-unknowns, allowing the system to "evolve" when presented with unpredictable challenges to the point of being reconfigurable with high degrees of freedom [21,23,44]. Robust systems are highly reliable within their design scope and predictable range of associated uncertainty [92]. The level of complexity in a robust system is controlled by identifying and mitigating factors that pose the greatest uncertainty [19,23]. The flexibility of machine learning algorithms allows uncertainties to be predicted in a variety of complex domains, examined further in Section 2.4.4 and Chapter 5.

2.3 Research definition

Defining the research scope is necessary to frame clear, answerable questions that formulate the aim and objectives described in Chapter 1; which inform search terms and inclusion/exclusion criteria in the succeeding phases [53]. Various frameworks have been composed to define the research scope and successive research questions (RQs).

The PICOC framework illustrated in Table 2.3 was adopted for this review [53,56,58]. This was selected against others proposed by Booth et al [53,93–95] such as SPICE [94] and CIMO [96] as it provides a transparent and duplicable identification of key concepts to be implemented in the SALSA framework.

Table 2.3. Research scope definition – PICOC framework

Concept	Definition
Population	Uncertainty prediction and assessment; considering the impact attributed by a combination of quantitative and qualitative inputs over the in-service phase of complex or non-complex engineering systems
Intervention	Examination of existing UQ techniques, qualitative assessments, uncertainty prediction, compound uncertainty aggregation for differing probability distributions
Comparison	Current industrial practices – how does the new proposal compare to the existing methods and academic processes?
Outcomes	Determination of relevant probability distributions and guidance on how to quantify uncertainty in context to aid decision making for industrial maintenance
	Identification of methodologies to quantify qualitative uncertainty attributes, combine quantitative and qualitative uncertainties and assess significant correlations
	Identification of methodologies to predict uncertainty through the in-service phase and optimise outputs as new information is acquired
Context	Compound quantification, aggregation and prediction of technical engineering uncertainty for engineering systems in-service, applicable to industrial maintenance

The scope was adapted as more research was uncovered and the author's understanding of the topic grew. The resulting objectives and corresponding RQs are depicted in Table 2.4. These objectives were derived as the basis to achieve the outcomes defined in the PICOC framework to establish key approaches to quantify and predict uncertainty in the maintenance of engineering systems.

Table 2.4. Research objectives and research questions

	Objectives	Research question
1	Identify current practices in the quantification and aggregation of different types of uncertainty	How can compound uncertainties be aggregated and represented through different probability distributions?
2	Identify and assess approaches that could analyse and estimate compound uncertainties for real-world applications	How can qualitative estimates driven by expert opinion and individual experiences be standardised and validated?
3	Explore techniques to predict uncertainty in engineering systems.	How can uncertainty be predicted over the in-service phase of an asset's life cycle and what are the key challenges faced in doing so?

The research search, appraisal and synthesis stages of the SALSA framework are detailed in Appendix A.1 to A.3. A timeline of the review process for the published manuscript [97] is given in Appendix A.4.

2.4 Analysis of synthesised data

This section examines the categorised themes from the synthesis of extracted data (Figure 2.4) to answer the RQs defined in Section 2.3. Thematic analyses are presented to examine the coverage of each theme over the included papers and correlations between them, assessing results from the synthesis. Narrative analysis is presented for each theme to discuss results and case examples. The evidence base from the thematic and narrative analyses are evaluated to answer the RQs in Section 2.5. Any generated hypotheses were grounded to populate an emergent theory. Conclusions are drawn and compared with other studies in the category [53,98].

The identification of uncertainties that influence the measurand will inherently vary in dynamic nature depending on the context of the measurand; be it a simple system under laboratory conditions or a complex engineering system (CES) with a myriad of interconnected subsystems. Section 2.4.1 discusses the allocation of studies in the defined contextual applications. Section 2.4.2 examines RQ1, focusing on the aggregation of uncertainty across multiple elements. Section 2.4.3 looks at the selection and use of relevant probability distributions to conduct the analysis. RQ2 is examined in Section 2.4.2.2, where methods to conduct qualitative uncertainty analysis are discussed. Section 2.4.4 examines RQ3, focusing on uncertainty prediction for the in-service phase of engineering systems with deep learning techniques.

Publication details of year and type for the 107 included papers are illustrated in Figure 2.2 and Figure 2.3. The majority of examined papers were published in 2019-20. A positive linear trend in publications up to the present indicates a growing relevance and interest within the research scope. The term ‘Conference’ includes workshops; ‘Book’ includes book sections and booklets. The majority of examined publications are journal articles, which are identified specifically in Table 2.5. ‘Other’ consists of journal publications featured once.

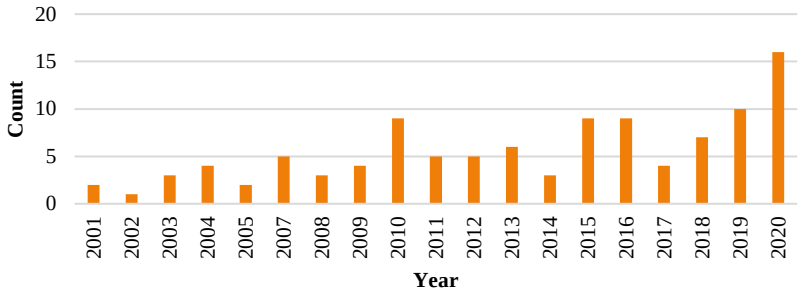


Figure 2.2. Analysis: Included papers – Publication year

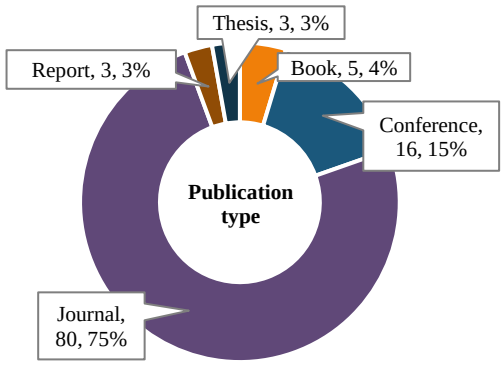


Figure 2.3. Analysis: Included papers – Publication type

Table 2.5. Analysis: Included papers – Featured journal publications

Publication	Papers
CIRP Journal of Manufacturing Science and Technology	11
Reliability Engineering & System Safety	8
Journal of Petroleum Science and Engineering	3
CIRP Annals - Manufacturing Technology	3
Progress in Aerospace Sciences	3
International Journal of Life Cycle Assessment	2
Sustainability	2
International Journal of Production Research	2
Journal of Hydrology	2
Others	44
Total	**80**

2.4.1 Contextual application

The contextual application theme identified in Table A.3 groups publications, as the name suggests, in their applied context. Through the refinement process described in the synthesis, 4 categories were identified: Aerospace & defence (Inc. nuclear weapons and other military applications), Emissions, energy & environment (Inc. oil & gas, meteorology, energy & power, greenhouse gases and coastal models), Manufacturing & maintenance (Inc. optimisation of processes around PSS and in general, structured surfaces, machine tooling and miscellaneous case studies) and Theory (Inc. description and derivation of analytical methods without a specified application). The number and percentage distribution of these applications are illustrated in Figure 2.4.

The majority of included papers examine the theory in uncertainty analysis, aggregation and forecasting (41%). These include statistical analysis, qualitative methods such as the pedigree approach and machine learning and Bayesian reasoning for forecasting. Papers applied in the other three contexts are reasonably distributed.

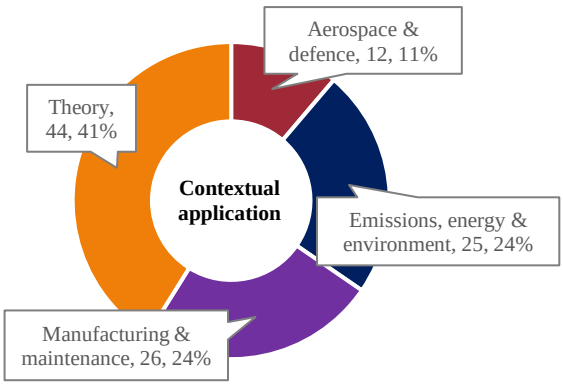


Figure 2.4. Analysis: Contextual application classification of included papers

2.4.2 Uncertainty propagation and simulation techniques

This section examines identified techniques to propagate uncertainty. Findings influenced the composition of the modelling approach in Chapter 4. The percentage of the 107 included papers that make use of or adapt the main techniques identified through the synthesis are illustrated in Figure 2.5, stacked by their contextual application. The ‘other’ category encompasses less used methods used in the research context such as Latin Hypercube sampling and Taylor series expansion.

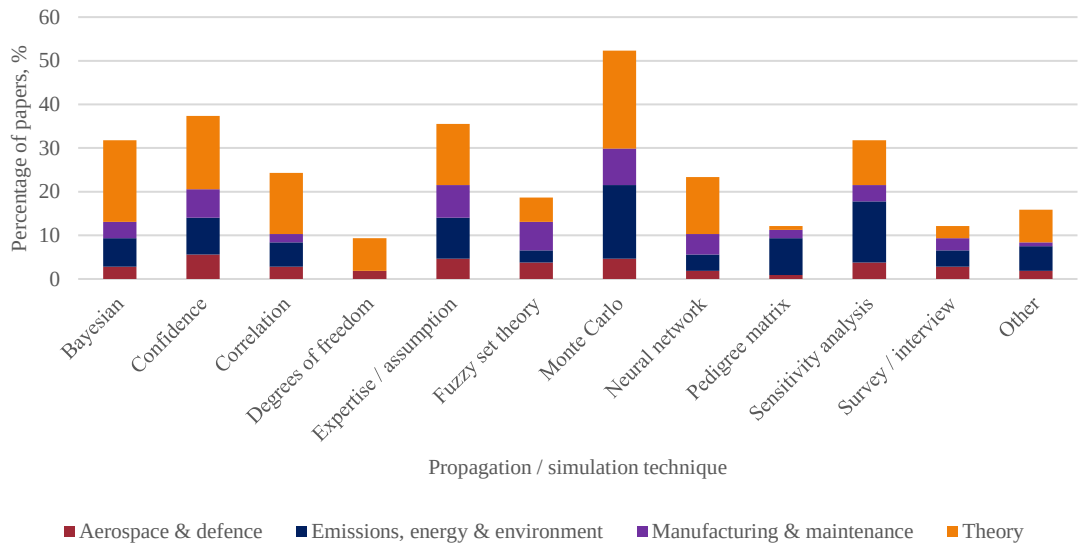


Figure 2.5. Analysis: Percentage of uncertainty propagation and simulation techniques used in included papers by contextual application

The categorised techniques can apply to purely quantitative, (Section 2.4.2.1), qualitative (Section 2.4.2.2) or compound (Section 2.4.2.3) uncertainty quantification and analysis.

The distribution of analysis types by contextual application is shown in Figure 2.6. Purely quantitative analysis is considered by 43 papers (40%), purely qualitative is considered by 23 (21%) and a compound aggregation is considered by 41 (38%). A core objective of this research is to equate qualitative uncertainties in line with quantitative to enable compound aggregation of technical engineering uncertainties in engineering systems.

This consideration is necessary for real-world applications but not essential when considering costing of such systems in theory (further explained in Sections 2.4.2.2 and 2.4.2.3). Terms such as ‘variance’, ‘standard deviation’ and ‘stochastic’ were not included as they were considered too generic. Some commonly used techniques appear to feature less frequently than one might expect (e.g. degrees of freedom in 9% of the 107 papers). The reason for this is that some studies focus on a specific part of the analysis process (e.g. uncertainty source identification through expert opinion or interviews) and so consider other stages to be out of scope.

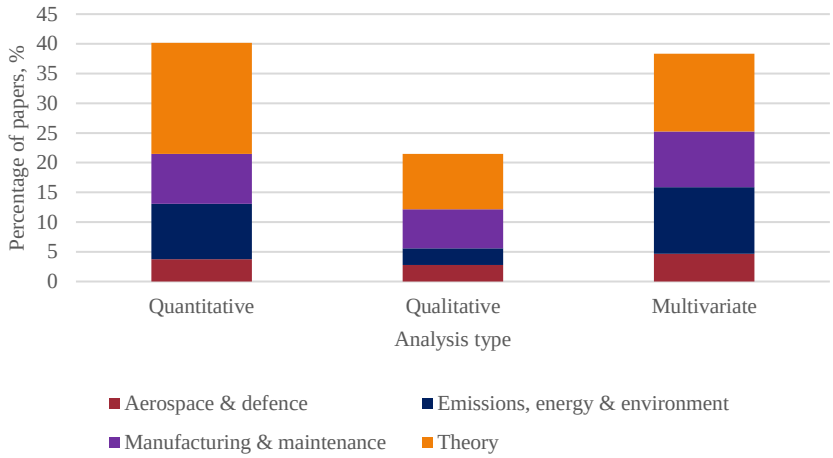


Figure 2.6. Analysis: Percentage of analysis type used in included papers by contextual application

2.4.2.1 Quantitative uncertainty analysis

Purely quantitative uncertainty analysis focuses on epistemic, statistical data. Techniques are discussed in theory below, which are then applied in case examples. Qualitative aspects need to be taken into consideration to be applied to real-world dynamic cases. The most commonly used techniques in the included papers that focus on quantitative analysis are illustrated in Figure 2.7, again stacked by contextual application.

Quantitative uncertainty is statistically equal to the standard deviation of a given dataset, which is equal to the square root of the distribution variance and referred to as the 'standard uncertainty' [2,99]. As seen in Figure 2.7, 40% of the 43 quantitative analysis papers reviewed explicitly use sensitivity analysis and 23% discussed correlation between the inputs.

Correlation accounts for dependencies between input parameters [2,11,37,43,100,101]. The aggregated uncertainty (u_T) due to the uncertainty in quantitative parameters is equal to the root-sum-square (RSS) of those uncertainties (σ) added to significant correlation coefficients (Eq. 2-1) [102]. If parameters are independent ($\rho = 0$), the second half of the equation is zero and cancels out.

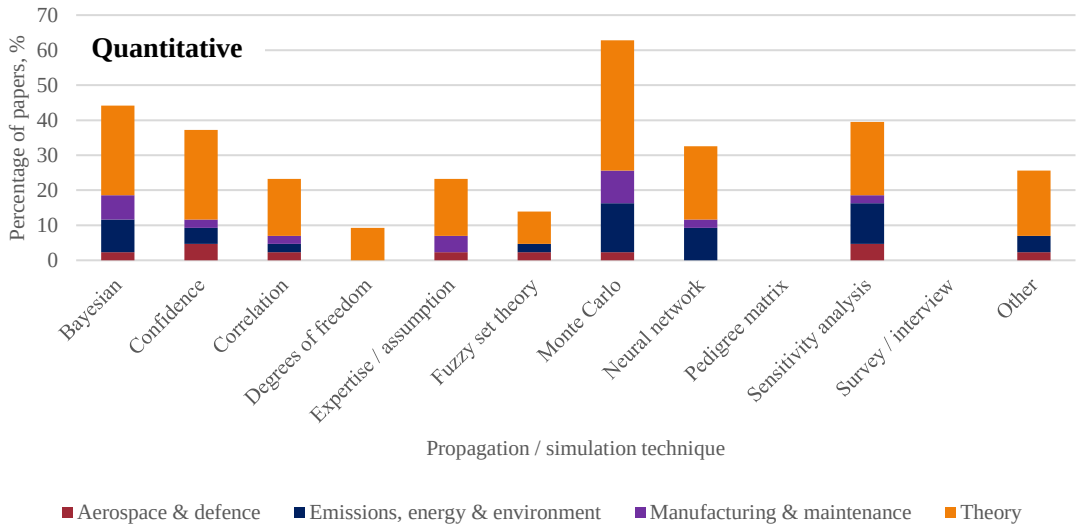


Figure 2.7. Analysis: Percentage of techniques used in included papers for quantitative analysis by contextual application

$$U_T = \sqrt{\sum_{i=1}^{n} \left(\sigma_{x_i}^2 + \sigma_{y_i}^2\right) + 2\left(\rho_{x,y}\, \sigma_x\, \sigma_y\right)} \tag{2-1}$$

The significance of positive and negative correlations on the aggregated uncertainty estimate will vary with system complexity as well as the coefficient value. It is important to remember that correlation is not causation and while two parameters can show a significant correlation, they may not be impacted by one another in practice. This could lead to overestimation of the aggregated uncertainty.

Many potentially identifiable uncertainties will have a negligible impact on the measurand. To maintain focus on uncertainties that have a tangible impact on the system, alongside expert judgement, sensitivity analysis (SA) is conducted across the input parameters [9,80,109,110,81,92,103–108]. SA gives an illustration of relationships between different inputs of various PDFs and parameters. Those with negligible effects can be removed. An important tool in uncertainty assessment, design optimisation and reliability measurement, SA is performed in two ways – local and global. Local sensitivity analysis (LSA) explores the change of the quantity of interest around a certain reference point, such as nominal values via partial derivatives. This is the simplest approach but can prove arduous when applied for a large number of parameters. Global sensitivity

analysis (GSA) studies the effect over the full range of the input space, typically adopting Monte Carlo techniques.

UQ in CES involves the propagation of errors around the sample mean of each parameter via simulation [37]. The three most common and validated propagation techniques are Taylor series expansion, Monte Carlo simulation and Latin Hypercube Sampling (LHS). The widely used propagation of error model uses Taylor series expansion to consider local sensitivity coefficients within the aggregation, given by partial derivatives [2,37,45,102]. While suitable for non-complex models, the use of partial derivatives in complex non-linear models has been shown to give a large degree of error and lead to under or overestimation of propagated uncertainty [45]. LHS migrates simple Monte Carlo to assess convergence of cumulative probability distributions for output variables [9,14,45].

Monte Carlo simulation is by far the most widely used simulation method to evaluate uncertainty; used in 63% of the 43 quantitative papers and 52% of the total 107 included papers. It can be applied to multiple probability distributions for compound analysis, shown to provide effective results in many situations for various combinations and complexities [28,32,34,48,52,72,82,108,111,112]. Extensive sampling of uncertainty ranges for individual variables can be achieved without the use of substitute models [9]. However, it can require significant computational power, with 1,000-10,000 simulation runs generally accepted as appropriate coverage depending on model complexity [111,113].

Clarke et al. [45] applied these propagation techniques in a thermodynamic analysis of heat exchanger designs. This highlighted the need to consider both quantitative and qualitative uncertainty and the identification of parameters that pose the greatest influence on uncertainty through SA [114]. Similarly, Tatara and Lupia [43] examined heat exchanger performance through temperature measurement uncertainty, with a spotlight on the effect data acquisition methods and measurement devices have on the resulting uncertainty. These studies influenced the composition of the study in Section 4.3.1.

Groen [115] compared GSA methods in environmental life cycle assessment, of which Spearman correlation coefficients and Sobol' indices were found to give the best overall

performance. Generally, the best method depends on available data, uncertainty magnitude and the goal of the study. Spearman correlation coefficients assume linearity in the system, which is often not the case in practice. Sobol' indices allow for nonlinearity but assume all parameters to be independent to identify the influence of each input parameter on the output [9,14,122,81,115–121]. Correlation coefficients should ideally be established between input parameters [41,115]. Discounting correlation is acceptable when the sensitivity of parameter x is significantly greater than parameter y, rendering $\rho_{x,y}$ negligible [40]. Where it is not, discounting correlation can lead to under or overestimation of the resulting uncertainty estimate.

Bayesian analysis derives the probability of an event occurring given that a prior event has occurred. This is given as a probabilistic function of the two events occurring independently or together [2,82,123]. Bayesian methods applied in uncertainty prediction are further covered in Section 2.4.4.

In 1995, the International Standards Organisation (ISO) published the Guide to the Expression of Uncertainty Measurement (GUM). Commonly referred to in literature as 'the Guide' or 'GUM', this has seen various updates and expansions since its inception [2,18,32,76,78,124]. The general uncertainty analysis process defined by the GUM involves 5 core stages [2,18,124]: (1) Identify the measurand; (2) Identify uncertainty sources and associated probability distributions; (3) Quantify uncertainties (simulation); (4) Aggregate uncertainties; (5) Report analysis results. While proficient for purely quantitative estimates, the GUM employs coverage factors and confidence limits to accommodate for qualitative or compound estimates. These often lead to underestimation, do not permit flexibility and, therefore, cannot be realistically applied in dynamic, complex engineering systems [92,125].

Since its inception, the GUM has been applied and adapted to assess uncertainty in a range of applications from structured surfaces [11] to micro gear measurement [3], smart grid power systems [105] and risk and reliability assessment in the nuclear weapons sector [85]. Uncertainty typically increases where significant correlation between exists input parameters. Correlation and sensitivity are key considerations for rigorous uncertainty analysis to capitalise on risk with the best possible model representation. Complex system uncertainty analysis involves representations of epistemic and aleatory

uncertainty. For epistemic analysis, uncertainty can be represented through various means including interval analysis, possibility theory, evidence theory and probability theory [16,21,85,126]. Probability theory is the dominant method, but others can be useful in the CES context – examined and compared in Sections 2.4.2.2 and 2.4.4. The main challenges for UQ in these contexts include the aggregation of information from multiple sources and the propagation of complex computational models that incorporate flexibility in design, while holding a degree of robustness to deliver on objectives [16,23,31,85].

2.4.2.2 Qualitative uncertainty analysis

The consideration of qualitative uncertainty factors can have significant effects on the overall estimate. The identification of known qualitative uncertainty sources typically relies on expert opinion. Methods to aid their derivation include surveys, interviews and the pedigree matrix [39,52,74]. Qualitative frameworks are often used in conjunction with quantitative methods such as Monte Carlo and SA in the context of real-world applications. Therefore, the majority of qualitative applied cases are discussed in Section 2.4.2.3, including those considering surveys and interviews. Figure 2.8 shows the distribution of techniques used in purely qualitative analyses.

Expert opinion and Monte Carlo were implemented in 52% and 26% of the 23 qualitative papers respectively. This section will examine commonly used qualitative propagation approaches, namely the pedigree matrix, as well as comparisons between probability theory, evidence theory and fuzzy set theory.

The pedigree approach is a widely renowned and verified approach to equate qualitative estimates in line with quantitative data. First proposed by Funtowicz and Ravetz [127], the approach comprises a matrix to score expert knowledge and opinion according to predefined criteria to permit quantitative reliability assessment.

This has been used in 17% of the 23 papers considering purely qualitative analysis (Figure 2.8), solely applied in the emissions, energy & environment context, and 22% of the 41 papers considering compound analysis (Figure 2.9), applied in all 4 considered contexts, though largely again in emissions, energy & environment. It has also been applied in medical fields and genealogy, largely visualised using decision trees, though these are not examined in the scope of this review [10,32,39,52,104,128].

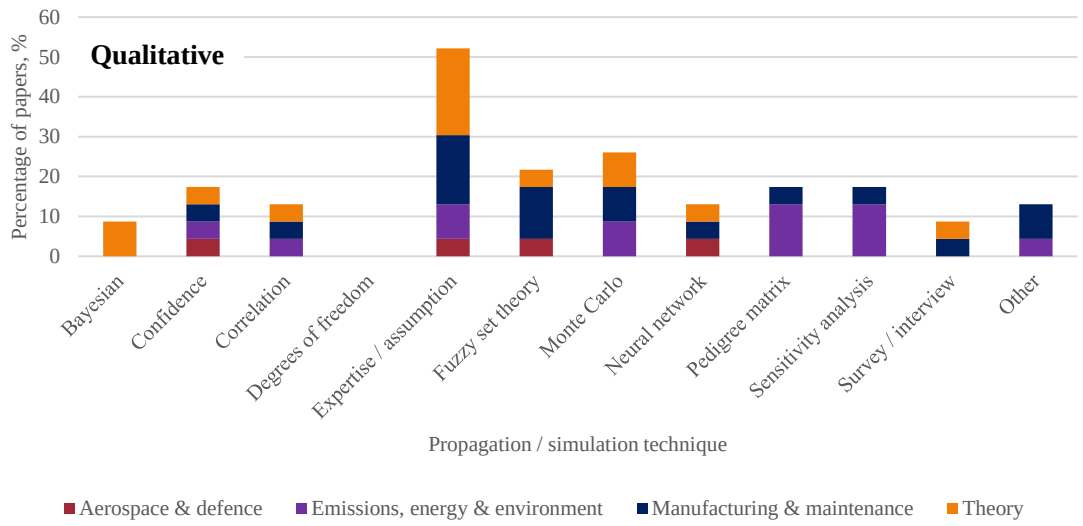


Figure 2.8. Analysis: Percentage of techniques used in included papers for qualitative analysis by contextual application

Pedigree can be applied on its own or through an encompassing approach to standardise combined uncertainty dimensions via 5 qualifiers: Numeral, Unit, Spread, Assessment and Pedigree (NUSAP) [32,51,52,128]. The first 3 terms consider quantitative factors: quantity value, acquisition date and random error of the variance of the dataset (addressed by SA and Monte Carlo simulation), respectively. Implementation of NUSAP is further discussed in Section 2.4.2.3.

Pedigree criteria are defined according to the contextual application of the study [39,52,128]. Qualitative assumptions made in uncertainty analysis can have a significant impact on the resulting estimate, especially in complex systems. The application of the pedigree matrix in complex environmental problems can highlight bias, implausibility, disagreement among stakeholders, limitations and sensitivities (further explored in Chapter 3) [104].

Additional uncertainty propagation approaches include probability theory, evidence theory and fuzzy set theory [72,123]. Probability theory is the ‘classic’ UQ method for input parameters with definable probability distributions, discussed in much of this review. Evidence theory makes use of artificial intelligence (AI) and machine learning to collate evidence from different sources and presents an evaluation to understand if the available evidence is common or contradictory [46,72,129]. Evidence theory can neglect

deterministic decision-making, which considers the outcome alone without associated risk, by keeping an 'open eye' to new information, governed by a belief system to dictate possibility measures [46]. This may be a suitable approach for qualitative reasoning but is less suited to estimating quantitative uncertainty, which is centred on recorded data [72].

Fuzzy set theory is applied in machine learning to assign a grading to input parameters (e.g. a scale of 0 to 1 rather than 0 or 1). This is well suited in cases where recorded data and knowledge is lacking and available data is inherently subjective [14,39,46,72,123,130]. This lack of mediated data is one of the major challenges in UQ for both complex and non-complex engineering systems [4,13,16,26,39,74,130]. Uncertainty analysis where data is scarce benefits greatly from the application of artificial neural networks (NNs). These networks of cooperating input elements are applied to a model and trained to give an optimum output by learning from previous examples [46,131]. NNs are a go-to option for forecasting and prediction tasks to be undertaken – discussed further in Section 2.4.4.

2.4.2.3 Compound uncertainty analysis

The term 'compound' is defined here as the aggregation of uncertainty from quantitative, measured, recorded data and qualitative, experience-driven opinion or human factors. Since qualitative estimates are obtained from technical expert knowledge or experience, they were not initially classed as purely statistical quantities with definable degrees of freedom [2,77]. The GUM proposed coverage factors and confidence limits as methods to accommodate for qualitative or compound estimates. An 'effective' degrees of freedom is applied using the Welch-Satterwhite formula [2,102], though this was later found to lead to underestimation of the combined uncertainty [76,78,125,132,133]. Since then, a range of advanced qualitative, quantitative and compound methods have been proposed to gauge qualitative estimates in a way that can be statistically equal to quantitative estimates.

The percentage of included papers that used compound analysis is shown in Figure 2.9. Expert opinion and assumptions made to carry out the assessment were considered in 39% of the 41 compound analysis papers (discussed in the previous section). The quality of the opinion sways the confidence in the result (considered in 49% of compound

analysis papers), which can be determined through the pedigree matrix (in 22%) and sensitivity analysis (in 32%).

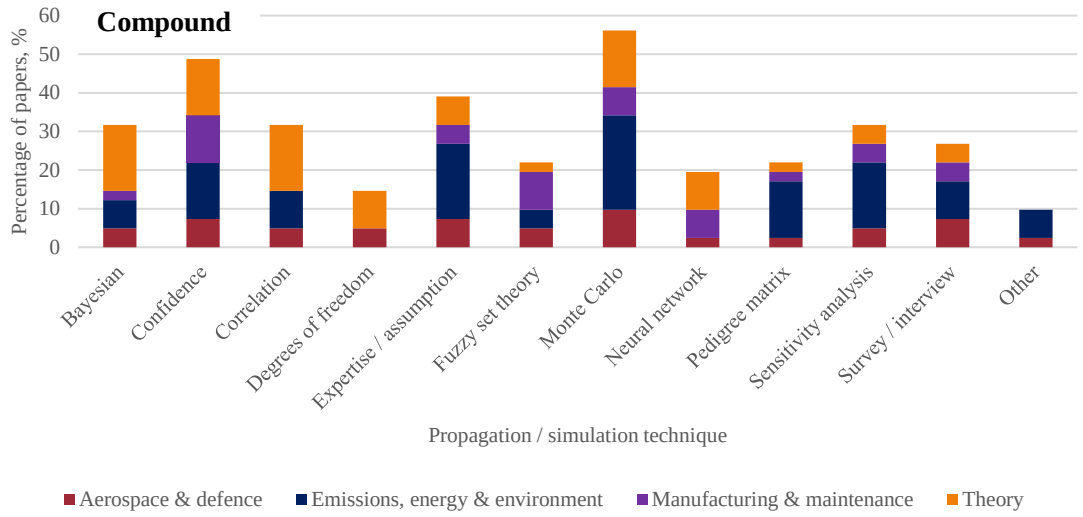


Figure 2.9. Analysis: Percentage of techniques used in included papers for compound analysis by contextual application

The pedigree matrix can be applied to simple calculations and complex models through explicit and systematic reflections on compound uncertainty [39,52,128,134]. Uncertainty estimation in life cycle costing under product-service systems (PSS) is a growing field of interest, where uncertainty changes throughout the life cycle stages [25,26,29,46,48,50,126]. Uncertainty analysis in PSS is examined further in Section 2.4.4. NUSAP has been implemented to estimate uncertainty in cost estimation from different sources at the bidding stage of industrial PSS contracts in the aerospace & defence context [135]. Uncertainties were identified through a predefined classification; commercial, affordability, performance, training, operation, engineering (CAPTOE) [136] and ranked using NUSAP [52].

The incorporation of qualitative estimates with quantitative assessments in the in-service phase of industrial PSS may present challenges due to increasing complexity but can also draw parallels from other phases of the life cycle [25,31]. Additional reviews have analysed value capture for PSS throughout the product life cycle on the transition to servitisation [29], availability support [137] and information flow [50]. Lack of concrete

data and qualitative decisions cause uncertainty that can lead to undesirable results. This also prompts the need for flexibility in PSS under uncertainty [23].

Data quality in life cycle assessments (LCA) is enhanced through a compound consideration of parameters. The use of pedigree and SA allows uncertainty parameters with negligible impact to be eliminated, enabling focus on those that influence the measurand [48,103]. This helps to alleviate the trade-off between accuracy and implementation costs in LCA to identify the most significant input parameters.

Another application domain of compound uncertainty in engineering systems is real-time systems. Largely considered in software engineering, these systems are highly dependent on confident and thorough uncertainty estimates to account for worst-case scenarios [138,139]. Uncertainties considered can range from computational processing times [138,140] to environmental and human factors, such as in virtual reality (VR) applications with remote maintenance [139]. Literature concerning real-time systems in this review is considered under the manufacturing and maintenance context. Real-time systems are inherently complex owing to the range of assumptions taken into account and unpredictable behaviour and interaction of system elements. To obtain confident predictions of worst-case execution times, evolutionary algorithms are employed along with surrogate models, neural networks and regression models – further explored in Section 2.4.4 [140,141].

Further applications of the pedigree matrix and SA, along with Monte Carlo and Taylor Series expansion, are made in the oil & gas sector to estimate uncertainty in greenhouse gas emissions [111]. These highly complex operations consist of compound estimates requiring rigorous estimates. Confidence levels associated with individual sources are dependent on data availability and quality. This process followed the core methods described in the GUM [2,18,111,124]. While applied solely to the oil & gas sector in the examined literature [111], the analysis method should be applicable in broader areas within the research scope.

Ciroth et al. [32] presented a process to improve uncertainty estimation by gauging qualitative uncertainty factors through the pedigree approach for flow data in a multidimensional database. Estimates are attributed by their geometric standard deviation (GSD), where inputs fit the multiplicative lognormal distribution (Eq.2-2) [32,34,36]. The

arithmetic standard deviation used to attribute uncertainty in quantitative data has the disadvantage of relying on the scale (unit) of data in a linear manner [32,34]. Therefore, for the analysis of data from varying sources and measured in different units, uncertainty factors need to be independent of scaling effects – achieved using GSD.

$$\sigma_g = \exp\left(\sqrt{\frac{1}{n} \times \sum_{i=1}^{n} ln\left(\frac{x_i}{\bar{x}_g}\right)^2}\right) \quad (2\text{-}2)$$

Where: σ_g = GSD; n = number of inputs; x_i = dataset; $\bar{x}_g$ = geometric mean of dataset

To enable aggregation where data sources do not follow a lognormal distribution, GSD ratios are obtained via the coefficient of variation (CV) [34,142]. This is a dimensionless measure of variability defined as the ratio between the standard deviation and the mean [142,143]. Muller et al. [34] provided formulas to apply the CV to various distributions to allow the user to select the most appropriate types for analysis.

The CV can be used as a measure of uncertainty for each input and aggregated to give a representative total. The application of CV and pedigree aims to convert quality and lack of knowledge into uncertainty figures [34]. This is a key method to aggregate compound uncertainties through different PDFs, given in Table 2.4, the robustness of which was tested for each parameter PDF using Monte Carlo simulation.

To combine quantitative, recorded parameters with qualitative factors, all parameter uncertainties are converted to their respective CVs according to their PDF type. The arithmetic mean of symmetric PDFs such as Normal and Uniform is equal to the mode and, as such, does not change when uncertainty increases [34]. They can therefore be aggregated additively by RSS (Eq.2-1). Lognormal distributions are asymmetric; the arithmetic mean will change with increasing or decreasing uncertainty. CVs represented by the lognormal distribution, CV_{Ln}, are aggregated multiplicatively by Eq. 2-3 [34]. To combine these with symmetric distributions, a new arithmetic mean needs to be calculated to account for the shifting uncertainty, given by Eq. 2-4 [34].

Table 2.4. Probability distribution function (PDF) and relative coefficient of variation (CV) calculations [32,34]

Distribution	Parameters	Deterministic value	PDF	CV calculation
Lognormal	x: Input dataset μ_g: Geometric mean σ_g: Geometric standard deviation (GSD)	Median: μ_g	$f(x,\mu_g,\sigma_g) = \frac{exp\left(-\frac{(\ln x - \ln \mu_g)^2}{2\,ln^2\,\sigma_g}\right)}{\sqrt{2\pi} ln \sigma_g}$	$CV = \sqrt{exp(ln^2\,\sigma_g) - 1}$
Normal	x: Input dataset μ: Arithmetic mean σ: Arithmetic standard deviation	Mean: μ	$f(x,\mu,\sigma) = \frac{exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}{\sigma\sqrt{2\pi}}$	$CV = \frac{\sigma}{\mu}$
Uniform	x: Input dataset a: Minimum value b: Maximum value	Mean: $\frac{a+b}{2}$	$\begin{cases} f(x,a,b) = \frac{1}{b-a}\ for\ a < x < b \\ otherwise, f(x,a,b) = 0 \end{cases}$	$CV = \frac{b-a}{\sqrt{3}(b+a)}$
Triangular	x: Input dataset a: Minimum value b: Maximum value c: Most likely value	Most likely value: c	$\begin{cases} f(x,a,b,c) = \frac{2(x-a)}{(b-a)(c-a)}\ for\ a < x < c \\ f(x,a,b,c) = \frac{2(b-x)}{(b-a)(b-c)}\ for\ c < x < b \\ otherwise, f(x,a,b,c) = 0 \end{cases}$	$CV = \frac{\sqrt{a^2+b^2+c^2-ab-ac-cb}}{\sqrt{2}(a+b+c)}$

$$CV_{Ln} = \sqrt{\prod_{i=1}^{n}(CV_i^2 + 1) - 1} \tag{2-3}$$

$$\mu_T CV_T = \mu\sqrt{CV_{Sym}^2 + CV_{Logn}^2} \tag{2-4}$$

Qualitative uncertainties given by subjective opinion are intuitively correlated in terms of rank rather than linear relationships [9]. Spearman's rank correlation (ρ) is, therefore, best suited to consider the correlation between compound uncertainties (x, y) – given by Eq. **2-5**. The proposed approach to aggregate compound uncertainty is discussed in Chapter 4.

$$\rho_{x,y} = \frac{\sum_{i=1}^{n}[\rho(x_i) - \bar{\rho}(x)][\rho(y_i) - \bar{\rho}(y)]}{\sqrt{\sum_{i=1}^{n}[\rho(x_i) - \bar{\rho}(x)]^2 \cdot \sum_{i=1}^{n}[\rho(y_i) - \bar{\rho}(y)]^2}} \tag{2-5}$$

2.4.3 Probability distributions for uncertainty analysis

The selection of the most appropriate PDF depends on the nature of each input parameter (quantitative or qualitative sources) and how it is recorded [39,144]. The most common types of PDF used in the included papers are stacked by their contextual application in Figure 2.10.

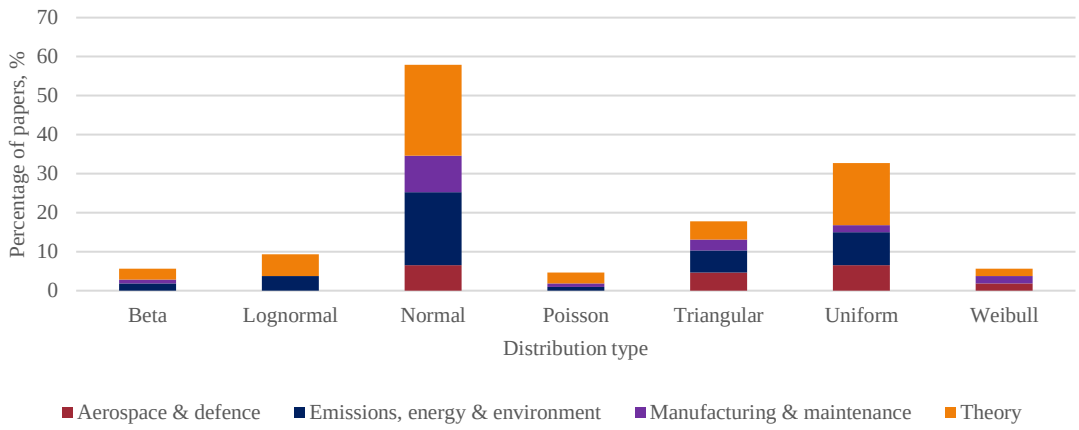


Figure 2.10. Analysis: Percentage of PDFs used in included papers by contextual application

Statistical measured data is typically represented by the normal (Gaussian) distribution, used in 58% of the 107 examined papers, or lognormal in 9%. Uniform distributions are considered in 33% of papers. When recording data, an individual digital readout has a uniformly distributed uncertainty, since it is on or off. The values of the readout are represented by a different distribution, depending on how it was recorded.

Several publications therefore considered more than one type of distribution. Table 2.7 describes the main distributions identified in the papers, adapted from Stockton and Wang [144], Everitt and Skrondal [68], and Erkoyuncu [39].

The Weibull distribution is used in reliability modelling and analysis for life cycle forecasting [26,33,145]. This could be an important distribution choice when considering forecasting uncertainty, however, it was only considered in 6% of papers included in this review.

2.4.4 Uncertainty assessment, prediction and forecasting

This section of the analysis focuses on how uncertainty can be modelled and predicted over the in-service phase of an asset's life cycle and where these are or can be applied to complex and non-complex engineering systems. The term 'assessment' is a judgement of value or quality based on available information. A forecast is the determination of future outcomes based on historic and new data (Bayesian), while a prediction is an indication of a future event with or without prior information [68,142].

The majority of reviewed manuscripts in the PSS context centre around cost estimation [13,23,25,26,28,29,46,48,74,123]. The in-service phase of PSS covers the largest portion of the life cycle situated between contract bidding and disposal. This phase calls for numerous equipment considerations including reliability, flexibility, availability and maintainability to ensure the asset is fit for purpose [46,135]. Each of these considerations raise challenges which promote numerous uncertainties, covered in Section 2.4.2.3.

Table 2.7. Analysis: Comparison of commonly used PDFs [39,68,144]

Distribution	Parameters	Application	Advantage	Disadvantage
Beta	Lower and upper range plus 2 shaping parameters	Variability over a fixed range	Highly flexible distribution	Requires additional estimation points to shape appropriately
Lognormal	Mean and Log. of standard deviation	Nonlinear, skewed ranges	Works well for factors that interact in a multiplicative manner	Can be difficult to express standard deviation Criticised for giving over estimated probability
Normal	Mean and standard deviation	Standard distribution is considered as standard uncertainty of the estimate	Works well for symmetrical data	Not as applicable for defining risk, which is usually asymmetrical
Triangular	Minimum, maximum and mean	Used when most likely value is distinguished	Simple and intuitive, can be skewed or symmetric	Points are highly absolute Can lead to under or over estimation as confidence levels cannot be stated
Uniform	Minimum and maximum	Constant data flow or where shape is unknown	Very simple to use	High risk of over or under estimation
Weibull	Scale and shape parameters $W(L, \alpha, \beta)$ is an open-ended distribution with location L, scale parameter α, and shape parameter β	Reliability modelling and analysis, life cycle forecasting	Can take the form of multiple distributions, depending on the value of β	Parameter selection can be inaccurate – leads to underestimation

An overwhelming issue in the forecasting of equipment states and related maintenance is the quality and availability of data [10,74,84,139,146]. To make accurate and robust predictions, a degree of historic data is required. Where this does not exist, the solution is generally to model the physical system and obtain data through simulation. As systems grow in complexity, robust and dependable models are harder and more expensive to produce [59,146]. Schwabe et al. [28] stated that the ability to quantify and forecast cost uncertainty is often limited by minimal measurement points, lack of experience, unknown history and low data quality. This precipitates innovation hesitancy in the face of an ever-increasing rise in technological innovation [147].

In the CES context, sampling rates are rarely consistent and feature a highly variable number of signals from different system units [10]. It then falls to subjective opinions and assumptions of experts – designers and manufacturers – to determine when maintenance will be required, equipment needed, time scales and resulting costs. This naturally places a large degree of uncertainty on the accuracy and robustness of such predictions, which must be quantified and considered.

Three prominent areas that have seen advancements in forecasting and prediction capability in recent years considering uncertainty are remaining useful life (RUL) prediction [59,146,148], cost estimation [13,25,26,28,48] and meteorology [149]. RUL prediction is a central task for maintenance practices of CES [65]. There are a myriad of RUL prediction approaches, notably reviewed by Lei et al. [146]. While the theory, general implementation and evaluation metrics of many approaches are open source, several industries have developed their own protected approaches for their specific requirements. Uncertainty is a significant point of vulnerability in long-term RUL predictions. Bayesian filtering algorithms are typically applied, the most prominent of which are Kalman filters; only suitable for linear systems and variations thereof for non-linear systems [146]. A more flexible algorithm is the particle filter, designed for use with nonlinear systems, which has become a widely used method for performing real-time uncertainty assessment in RUL predictions [146,148,150,151].

Another key tool in forecasting is deep learning, which makes use of neural networks (NNs) to learn from existing data. The quality of data ultimately determines the quality of the result. Applications are covered in detail for RUL prediction by Lei et al. [146]. NNs are composed of multiple layers, allowing them to learn complex non-linear relationships. Bayesian deep learning (BDL) and variations thereof have been widely applied to forecast future events given existing data and update when presented with new data [42,88,152–158]. Deep learning models are only as accurate as the data they are trained on and, as such, typically require large datasets with defined trends over time [42]. They must therefore be flexible to consider all data properties necessary to achieve robust predictions. Flexible models can make better predictions, but all predictions involve assumptions that manifest uncertainty [42,97,152].

The end of Section 2.4.2.2 identified the endorsement of NNs to aid uncertainty analysis for complex engineering systems. The terms and qualities identified in the synthesis to represent uncertainty assessment, prediction and forecasting are illustrated in Figure 2.11 and stacked by contextual application.

Uses of NNs and Bayesian techniques from Figure 2.5 and selected distributions from Figure 2.10 are included for comparison. Life cycles of products or services were considered in 39% of the 107 included papers, with 21% considering NNs and 32% considering Bayesian techniques.

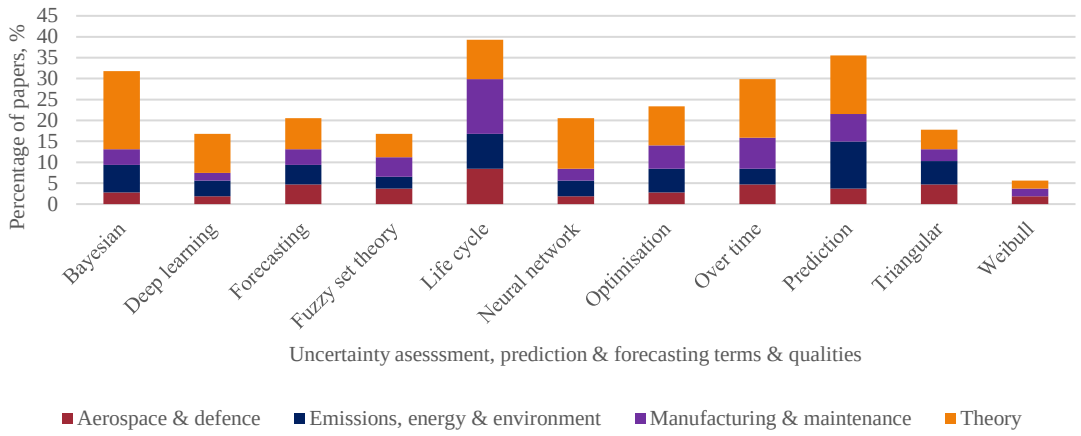


Figure 2.11. Analysis: Percentage of terms and attributes in included papers for uncertainty assessment and forecasting

To give greater confidence in estimates such as maintenance costing, backpropagation algorithms can be applied to further improve the quality of NN training [33,44,141,149,157,159,160]. Applications were reviewed in terms of their learning capability and reliability in uncertainty prediction. Stochastic models calculated from steady-state probabilities do not necessarily reflect reality since maintenance policies can take several years to stabilise [33].

Naturally, the structure of NNs and training options applied have a significant impact on prediction accuracy for specific applications [146]. Determination of parameters that result in minimal prediction error can be achieved through hyperparameter tuning, often performed via a grid search technique [65,161]. Bayesian deep learning (BDL) is one of the most popular techniques to learn from and forecast data trends

[42,88,152,153,158,162]. However, this approach requires significant modification models, adopting variation inference instead of backpropagation. Modification of deep learning models as more data becomes available can make implementation more complex and require extensive computation time [149,158,161].

This issue can be mitigated in part by dropout training, applied by Gal [158] in a method to approximate Bayesian inference in Gaussian processes (GPs) in deep neural networks and more generally by Cicuttin et al. [163] and Srivastava et al. [161]. GPs are highly flexible non-parametric models widely used for regression and classification, growing in complexity in line with the density of training data [42,152,164,165]. Defined as a layer within the network structure, dropout randomly sets input sequences below a defined probability to zero. This alters the underlying network structure for each iteration to prevent overfitting [158,163]. The uncertainty assessed by Gal [158] was in the deep learning process itself, not the resulting uncertainty interval. These methods still require enough prior data of sufficient quantity and quality to fulfil the Central Limit Theorem, where the normalised sum of variables will tend towards a normal distribution [2,26,28,145,146].

The General Likelihood Uncertainty Estimation (GLUE) method uses Bayesian inference to assess uncertainty in model predictions. Largely applied in hydrology and meteorology, the method uses ensemble forecasting of weighted parameter sets to identify the contribution level of each set for a forecasted point in time [119,155,166,167].

Wang et al. [149] proposed a deep uncertainty quantification (DUQ) prediction model to learn from historic data through a negative log-likelihood error (NLE) calculation to forecast weather patterns. The combination of deep learning and UQ was shown to improve generalisation of point estimation compared to RMSE calculation to forecast multi-step meteorological time series but is best suited to scenario modelling in meteorology.

Recurrent neural networks (RNNs) are a form of NN with a feedback loop to better capture non-linear relationships. Long short-term memory (LSTM) networks are a type of RNN increasingly used in sequential time-series forecasting and RUL prediction, the theory of which is widely covered in literature [59,63,65,160,168]. A key advantage of LSTMs over other types of NN is their ability to use gates to avoid vanishing or exploding

gradients, increasing prediction accuracy [59,160,168]. Wu et al. [65] applied a vanilla LSTM model to predict RUL and identify physical degradation mechanisms, the parameters of which were defined through hyperparameter tuning. Shi and Chehade [59] proposed a Dual-LSTM framework to predict uncertain change points from which degradation accelerates and heath indexes that can be used to determine RUL in real-time. Both studies were compared with, and found to outperform, benchmark methods. A common trait among the examined publications is the use of the C-MAPSS turbofan engine degradation dataset to test and demonstrate RUL prediction with proposed methodologies [59,61,63,65,169,170]. Further applications of this dataset are examined and ranked by Ramasso [57]. Different approaches to account for uncertainty in the datasets are also covered – the most popular being probability theory. The turbofan engine degradation dataset is also used in this research – detailed in Chapter 5.

Uncertainty in cost estimation is largely examined in the context of product-service systems (PSS) [13,25,26,28,48,74,97]. Multiple equipment uncertainties arise in this context including reliability, availability, and maintainability. Smart [142] applied Bayes' Theorem to estimate costs from trends with minimal data points. Existing data and assumptions were combined with limited real-time data to produce accurate forecasts with a degree of confidence.

Schwabe et al. [28,131] devised an approach to estimate cost uncertainty under limited data. The topology and symmetry of variance data is given by its geometric shape at the time of estimation. This was initially driven by the idea that most statistical conclusions obtained arithmetically can also be achieved by geometry, which can simplify otherwise complex conclusions [28,171]. Rather than interdependencies between individual data points, spatial geometry describes the behaviour of a space created by connecting outlying data points, represented in vector space in a point cloud around an origin, forming a regular cyclical polygon. A shape with greater symmetry requires less information to be described. A positive correlation was therefore hypothesised between symmetry and information entropy. The symmetry of the space created by each input dimension (cost variances) was able to predict future development without requiring significant volumes of data [26,28]. This is a promising approach to predict uncertainty under limited data.

There is limited literature on holistic, compound cost uncertainty estimation for the in-service phase of PSS [13,46]. Guidance is scarce to aid the selection of suitable uncertainty modelling methods such as NN, BDL and fuzzy set theory, which in themselves generally only consider epistemic forms of uncertainty [42,46,72,123,131].

2.5 Research results and discussion

The final phase of the review methodology discusses the research methodology and results conducted through the SALSA framework [53]. An evaluation of the validity of research methods adopted and findings culminated throughout the review is given in Appendix A.5. Research questions 1 and 2 share many similarities and are discussed in Section 2.5.1, summarised in Table 2.6. Research question 3 is discussed in Section2.5.2, summarised in Table 2.7. Section 2.5.3 summarises the core contributions to knowledge from findings of the research questions.

2.5.1 Discussion of findings for research questions 1 and 2

How can compound uncertainties be aggregated and represented through different probability distributions?

The analysis of papers to answer this question is presented in Sections 2.4.2 and 2.4.3. Quantitative uncertainty analysis considers an aggregation of input parameter uncertainty whose value is derived from statistical data. Sensitivity analysis and Monte Carlo simulation are used to propagate uncertainty ranges over multiple PDFs along with correlation between inputs and respective degrees of freedom. The majority of solely quantitative approaches follow the standard GUM method, or an adaption thereof.

The main qualitative analysis techniques combined the pedigree matrix, largely integrated in NUSAP, with quantitative assessment methods such as quantitative risk assessments and LCA. The former appreciated the need for compound considerations but there were no examples found of a combined approach. The latter applied SA to eliminate negligible inputs to alleviate the trade-off between measurement accuracy and implementation costs. However, uncertainty over the life cycle was considered constant, when in reality it is likely to fluctuate. The compound aggregation of quantitative and qualitative uncertainty

is essential in real-world contextual applications to provide estimates of cost, availability and reliability with high levels of confidence.

The selection of the most appropriate PDF to represent a given uncertainty source is crucial in the analysis process [39,144]. This can be achieved visually by comparing fits against a plotted histogram of the data. Attributing qualitative factors as geometric standard deviation (GSD) enables the quantification and aggregation of compound uncertainties through an amalgamation of the pedigree matrix, Monte Carlo simulation and coefficient of variation (CV) [32,34]. This method can be applied to a range of symmetric and asymmetric PDFs. While formulae to denote inputs of varying PDFs by their respective CVs are defined, a method to aggregate CVs from a mix of symmetric and asymmetric PDFs in a compound manner is unclear. This is necessary to establish compound uncertainty estimates represented by different PDFs with a high degree of confidence.

The compound aggregation approach can be used in GSA to calculate sensitivity indices. Correlations should be considered where suitable to avoid under or overestimation in the estimate. However, the majority of applied studies assume input variables to be independent. It is logical to assume there will be significant correlations between quantitative, measured variables and the qualitative influence on how those variables are recorded. Incorporation with qualitative uncertainties requires further research at this stage [9,118,120,121]. The risks in ignoring correlation in uncertainty propagation and SA are explored extensively by Groen [40,115]. The consideration of correlation through the sampling GSA approach allows for increased accuracy in the determination of which variables have the most significant impact on the overall uncertainty, and is therefore incorporated in Chapter 4 [40,41]. The ability to consider PDFs other than normal will further enhance this capability in the aggregation modelling approach.

Findings were qualified by referred sources and standardised methods for quantitative and qualitative uncertainty analysis. The probity of the amalgamation of these methods is considered unbiased since it can, in theory, be applied to multiple PDFs in multiple contexts. It also fulfils the outcome of the PICOC framework to determine relevant probability distributions and methods to quantify uncertainty that can be applied in industrial maintenance. Other approaches examined were only applied in theory,

prompting the need for further research in applied fields. Alternative techniques may exist that were not covered in this review. This can be down to the probity of the initial search string and robustness of the elimination process.

How can qualitative estimates driven by expert opinion and individual experiences be standardised and validated?

The analysis of papers to answer this question is presented in Section 2.4.2.2. Qualitative approaches applied in real-world cases are used in conjunction with quantitative methods such as Monte Carlo and SA. The pedigree matrix is one of the most widely used methods to validate qualitative attributes such as expert opinion and experience [32,34,39,52,127,128]. This requires the definition of pedigree criteria upon which the experience or qualifications of an 'expert' are scored and aggregated to attribute a quantitative measure of uncertainty. These criteria can be defined through surveys and interviews with industrial practitioners and academics. This approach has been adapted and implemented in a range of fields for various purposes [32,52,72,103,104,111,128]. Expert opinion and individual experiences can be validated against defined pedigree criteria to provide a standardised representation of uncertainty.

The definition of criteria alleviates bias in the approach, though this should be made by a diverse selection of suitably qualified individuals. The pedigree approach was the only qualitative technique explored in detail as it was deemed best suited and widely accepted to fulfil the desired application. Other approaches or adaptations of pedigree may warrant further investigation, but the application through GSD and CV proposed by Ciroth [32] and Muller [34] appear best suited to fulfil RQs 1 and 2. These factors also achieve the outcome portion of the PICOC framework in Table 2.3 to identify methodologies to quantify qualitative uncertainty attributes and combine them with quantitative uncertainties.

Table 2.6. Identified approaches to resolve RQ1 and RQ2

Problem / Approach	RQ1: Aggregation of compound uncertainties represented through different probability distributions	RQ2: Standardise and validate qualitative estimates
GUM method [2,18,32,76,78,124]	Standardised methods for quantitative aggregation (standard deviation) Gives standard 5-step process to identify, quantify and combine uncertainties Uses effective degrees of freedom for qualitative aggregation, leads to underestimation Widely used with small variations in multiple applications	Use of effective degrees of freedom via Welch-Satterwhite formula can lead to underestimation of combined uncertainty – improved method presented by Willink [125]
NUSAP [52,128]	Can be applied to simple calculations and complex models Found to improve the depiction of uncertainty through visualisation and background knowledge compared to quantitative risk assessments Not clear how quantitative and qualitative estimates were combined explicitly	Uses pedigree to attribute qualitative estimates in a quantitative manner, suited to a broad range of applications
Geometric standard deviation (GSD) and Coefficient of variation (CV) [32,34,142,143]	Estimates are represented under the lognormal distribution as GSD to eliminate scaling effects from different types of data CV enables aggregation of quantitative and qualitative uncertainties represented by different PDFs	Uses pedigree to attribute qualitative estimates via GSD
Willink method [78,125]	Fits quantitative estimates to qualitative by attributing a known parent distribution to quantitative "Proposed method improves performance when some error components are drawn from non-normal distributions whose variances are obtained by non-statistical means"	Qualitative estimates represented by known variance and 'coefficient of excess' Removes bias of overall variance estimate
Top-down approach AKA: Nordtest approach, Single-lab validation [133]	Broad level – does not go far into measurement procedure and does not attempt to quantify all uncertainty sources individually, contrary to GUM, but follows the same 5-step process Instead, uncertainty sources are quantified in large "batches" via components that take several uncertainty sources into account Uncertainty obtained characterises analysis procedure rather than an explicit result	Considers uncertainty component by possible bias – determined against an uncertain reference value

2.5.2 Discussion of findings for research question 3

How can uncertainty be predicted over the in-service phase of an asset's life cycle and what are the key challenges faced in doing so?

The analysis of papers to answer this question is presented in Section 2.4.4. The quality and availability of data is the greatest driver of uncertainty in the forecasting and prediction of equipment states, RUL, and determination of when and how maintenance should be carried out [85,172]. Growing complexity in engineering systems makes precise modelling of physical systems harder and more expensive to produce in order to obtain reliable simulated data [59,146]. These challenges limit the ability to optimally train networks through probabilistic Bayesian learning, which reduces confidence and robustness in associated uncertainty estimates.

Intelligent learning techniques are increasingly used to flexibly forecast uncertainty in a range of fields, though applied methods for in-service maintenance are limited. RUL prediction is a key determinate for maintenance scheduling and costing in CES [65,146]. Variants of RNN are widely used but require sufficient data with which to train networks to make accurate and robust predictions. This also limits the robustness of probabilistic methods such as BDL [42].

Hyperparameter tuning is not suitable for regular updates to network architectures owing to significant computation time when comparing multiple training options and network structures [168]. It can, however, can provide an effective starting point to make initial predictions. Dropout training can help improve prediction robustness by preventing overfitting, as well as updating the LSTM state at each prediction step as more data becomes available. Alternative approaches to predict uncertainty under limited data have been proposed such as deep uncertainty quantification (DUQ) [149], drop out learning [158] and spatial geometry [28].

Uncertainty manifested under limited data and assumptions as discussed above should be predicted to allow decision-makers to plan with greater confidence. Doing so will reduce under or over estimation. Uncertainties that may pose an undesirable risk at a given point in time can be mitigated to reduce the likelihood of unforeseen costs and delays [97].

Predictions need to be robust and as accurate as possible despite being produced under limited data, where traditional probabilistic methods are not applicable.

Findings for RQ3 may be considered biased towards the context of cost estimation in PSS [23,25,29,50,137]. Additional research is needed to examine how the assessed deep learning approaches can be applied for uncertainty assessment in industrial maintenance under limited data [130,139]. This requires a compound aggregation at present and a prediction of how the uncertainty may change through the in-service phase, considered for individual system components and as a whole. This achieves the final outcome of the PICOC framework to identify methods to predict uncertainty and the core challenge of limited data. Predictions can be utilised by decision-makers to mitigate uncertainty, reducing the likelihood of unforeseen costs and delays.

Table 2.7. Identified approaches to resolve RQ3

Problem / Method	RQ3: Forecasting uncertainty over the in-service phase of an asset's life cycle
Fuzzy set theory [14,39,46,72,123,130]	Function assigns a grade between 0 and 1 to each input parameter of a set, as opposed to Boolean that are 0 or 1 Suitable for qualitative reasoning, not for estimating quantitative uncertainty. Often recommended in cases where recorded data and knowledge is lacking, and available data is inherently subjective. Used alongside NNs to aid uncertainty analysis
Neural network (NN) with Backpropagation (BPN) [39,42,85,140–142,144,158,173]	A flexible network of cooperating processing elements to give an output. Applied to a model and 'trained' to give an optimum output Backpropagation computes the gradient of the loss function and uses it to change input parameters to reduce mistakes and optimise the output Other applications reviewed regarding learning capability and reliability in uncertainty prediction, giving greater confidence in maintenance cost estimates BPN addresses stabilisation of maintenance policies based on steady-state probabilities from stochastic models at inception that may not reflect reality for forecasts.
GLUE method [119,155,166,167]	Uses Bayesian inference and ensemble forecasting to assess uncertainty and contribution (sensitivity) of factors for a forecasted point in time Monte Carlo simulation provides information to the decision-maker on expected uncertainty with a degree of confidence. Allows for identification of differences in model performance and quantification of parameter-induced uncertainty
Deep uncertainty quantification (DUQ) [149]	Combines deep learning and UQ to forecast multi-step meteorological time series Uncertainty is incorporated straight into a loss function and is directly optimised through backpropagation

	Improves generalisation compared to mean squared error (MSE) and mean absolute error (MAE)
	BPN incorporates uncertainty directly into loss function for direct optimisation
	Regression is solved as a mapping problem rather than curve fitting and so cannot be naturally applied to multi-step timer-series forecasting
Dropout as Bayesian approximation	Theoretical framework casting dropout training in deep NNs as approximate Bayesian inference in deep Gaussian processes
[158]	Bayesian models require significant modification to train deep models, making them harder to implement and computationally slower
	Dropout training used to approximate Bayesian inference in Gaussian processes
	Approximate Bayesian inference updates probability as more evidence becomes available
	Considerable improvement in predictive log-likelihood and RMSE compared to existing state-of-the-art methods such as BDL
LSTM networks	Feedback loop to better capture non-linear relationships
[59,65]	Use gates to avoid vanishing or exploding gradients, increasing prediction accuracy
Spatial geometry	Forecasts cost uncertainty for a given point in time where available data is scarce, determined by the geometric symmetry of cost variance data at the time of estimation
[28,131]	Represents uncertainty in a vector space, aggregated to give probable cost variance in state space.
	Propagation described through the symmetrical relationship between cost variance data at a given point in time set apart from 0.
	Alternative to traditional parametric techniques where available data is not sufficient to fulfil the Central Limit Theorem

2.5.3 Research questions contribution to knowledge

The analysis of synthesised literature to answer RQ1 in Section 2.4.2 summarised the key UQ approaches used to undertake purely quantitative, purely qualitative and compound analysis. The importance of considering correlation and sensitivity was highlighted. Section 2.4.3 identified PDFs best suited for uncertainty analysis applicable to industrial maintenance. Standardisation of qualitative factors to answer RQ2 in Sections 2.4.2.2 and 2.4.2.3 highlighted the use of the pedigree matrix to assign scores corresponding to uncertainty intervals [39,52]. These are attributed by their geometric standard deviation (GSD) to combine with quantitative estimates. To gauge these on an equivalent scale for aggregation, the respective coefficient of variation (CV) of each input is used as the uncertainty measure [32,34]. Systems in the reviewed context of emissions, energy & environment are inherently complex. Methods used must be flexible and therefore likely to be transferable to industrial maintenance.

The analysis to answer RQ3 in Section 2.4.4 highlighted the use of deep learning to predict uncertainty. Methods to predict individual and aggregated uncertainty manifested by data availability, quality, experience and knowledge over time should be applicable under limited data where traditional probabilistic Bayesian learning cannot be applied. Approaches summarised in Table 2.7 should be explored to make confident predictions of which uncertainties will pose undesirable risk throughout the in-service phase.

2.6 Conclusions and future work

The purpose of this review was to investigate distinct methodologies used to quantify, aggregate and predict uncertainty for real-world applications. Knowledge gaps within the research scope were highlighted, prompting the future research direction for dynamic uncertainties manifested in engineering systems to optimise performance and availability for the in-service phase.

Section 2.1 hypothesised that current approaches considering a compound aggregation of factors will increase confidence and rigour in determining the impact of uncertainty over time under limited available data. The methodologies identified above for compound aggregation in theoretical and real-world applications, along with deep learning techniques to predict uncertainty have been shown to achieve this and consequently prove the hypothesis to be true.

Conclusions drawn from the discussion of approaches prove that the aggregation and prediction of uncertainty are hindered by the quality of available data, experience and knowledge. Modern engineering systems feature a myriad of subsystems interacting simultaneously and nonlinearly with each other with levels of importance dependent on operational conditions and system environment. Limited data concerning the optimisation of such systems and interactions between them increases uncertainty throughout their in-service life. The in-service life typically spans several years, prompting a need for robust predictions of technical engineering uncertainties relating to cost and equipment availability. These systems typically operate under product-service system (PSS) contracts with multiple stakeholders, which presents challenges to confidently and accurately determine the level of uncertainty at present or in the future. Uncertainties

related to stakeholder relationships are largely tied into supply chains and therefore out of scope for this research [23,50].

From the findings of this review in answering the three research questions, three core research gaps were identified:

1. Approaches to quantify and aggregate compound uncertainties represented by different distributions, considering dependencies between them, applicable to increasingly complex engineering systems.
2. Application of GSA to determine the impact of individual uncertainties on the aggregated total, accounting for compound parameters and significant correlation.
3. Limited approaches to predict uncertainty in engineering systems with complex and non-complex entities under limited data, and to do this without the need to produce complicated and expensive models of physical systems.

Future work to close the first and second gaps is recommended to develop robust frameworks that consider dependencies between compound inputs within increasingly complex system boundaries and identify which inputs have the greatest influence on the aggregated uncertainty. Flexibility in engineering systems design allows unpredictable unknown-unknowns to be mitigated (Figure 2.1), which should be reflected in UQ frameworks. While many UQ approaches exist for purely quantitative scenarios, standardised methodologies to quantify compound uncertainty are limited in the manufacturing and maintenance context, especially for the in-service phase. The suitability of the pedigree matrix to determine qualitative uncertainty in the context of the research questions proves promising for the research direction.

Future work to fulfil the third gap can be achieved through a combination of deep learning LSTM networks and spatial geometry. This will allow uncertainty to be forecast for real-world applications, incorporating complex and non-complex entities. LSTMs can make flexible forecasts based on prior data and update when new data becomes available. Spatial geometry offers a novel approach to predict uncertainty under limited data, though not yet applied outside of cost estimation. The push to develop deep learning methods to predict uncertainty is gathering importance as data volumes, computational capability and complexity in engineering systems increases.

Maintenance processes can be simulated through surrogate models, incorporating the identified challenges to execute frameworks to quantify, aggregate and predict resulting uncertainties. Simulated data can then be incorporated to train developed frameworks to confidently aggregate and predict compound uncertainty.

CHAPTER 3. CURRENT PRACTICE AND CHALLENGES

Abstract

Complex engineering systems present a wealth of uncertainties in factors from performance measurements to maintainability and through-life characteristics. A quantifiable understanding of these uncertainties is vital to system optimisation and plays a key role in decision-making processes for manufacturing organisations worldwide; impacting profit, product availability and manufacturing efficiency. The influence of 32 categorised uncertainty factors is assessed through a questionnaire completed by nine experienced maintenance managers from a leading defence company. The pedigree approach is applied to score the validity of respondents' answers according to their experience and job role to normalise scores. Results are discussed in interviews with respondents along with current practice in, and ways to improve, uncertainty assessment. Six core challenges are verified with 40 practitioners from various industrial backgrounds. From the interviews, it is deemed that a holistic view of heuristic and statistical attributes ultimately allows for more accomplished decision-making but requires trade-offs between quality and cost over the asset's life cycle.

Paper 2 Identifying challenges in quantifying uncertainty: Case study in infrared thermography

Published: Procedia CIRP, IPS2 Conference 2018
DOI: 10.1016/j.procir.2018.03.301
Data access: 10.17862/cranfield.rd.11961435.v2

Paper 3 Current practice and challenges towards handling uncertainty for successful outcomes in maintenance

Published: Procedia CIRP, CIRPe Web Conference 2019
DOI: 10.1016/j.procir.2020.01.024
Data access: 10.17862/cranfield.rd.11949021.v3

3.1 Introduction

Decision-making in industrial maintenance is typically based on two broad factors: recorded data and subjective expert opinions. The prior presents hard facts, subject to a degree of uncertainty that can be quantified statistically by the standard deviation of the dataset. The latter attributes qualitative uncertainty by what traits qualify someone as an expert and the basis of their view to establish its validity. Data recording methods, accuracy of equipment used, or maintainer performance are rarely considered as an attribute to overall uncertainty. Once statistical uncertainty estimates are obtained from recorded data, it is necessary to also question how these recordings were made, their accuracy and how such approaches may differ in various operating conditions. In complex engineering systems (CES), decisions made for one component or subsystem can have unforeseen effects on others. An example of one such system is the maintainer. The degree of uncertainty associated with the maintainer's discretion in the quality of maintenance carried out is significant due to the number of variables that may influence their decisions; such as training level, measurement quality and environmental conditions.

A combination of the hard facts and subjective opinion needs to be considered to make informed and effective decisions leading to prosperous outcomes in maintenance. Some cases require more expertise; some require more data. The question here is whether a holistic view of these uncertainties can improve decision-making capabilities and reduce through-life costs as well as unforeseen challenges.

In the context of industrial product-service systems (PSS), maintenance responsibilities are shifted back from the client to the product provider (contractor) [174]. These service contracts are increasing in scale and complexity, now accommodating highly complex and dynamic systems. Operational life cycles of such systems promote extensive relationships between the contractor and client. The availability, reliability and maintainability of these systems and equipment is therefore essential in logistical contracts and through-life support services. Some significant maintenance technologies that support these services are non-destructive testing (NDT) and degradation assessment, repair and remote maintenance that sustain maintenance activities [174,175]. These

should therefore be profitable to the contractor, but also ensure supply chain sustainability and customer affordability [176].

The approach to a specific maintenance task by a contracted maintainer may differ from that of the client's maintainer on the same task. Decisions made here raise several qualitative uncertainties from both sides that are naturally problematic to quantify.

This chapter presents a survey questionnaire to examine and rank prominent factors that influence uncertainty in maintenance based on literature and input from industry experts. Uncertainties are considered in the context of maintenance for CES, considering the dynamic nature of system requirements over time and the effect these may have on the through-life maintainability of CES from the perspective of industrial PSS.

Respondent qualities are attributed in a pedigree assessment. Results are reviewed and discussed in a series of semi-structured interviews and validated with wider industrial practitioners before producing a refined survey and pedigree criteria. Results are ranked using the well-established Analytical Hierarchy Process (AHP) to determine areas facing the most significant challenges and uncertainties.

3.2 Research background

3.2.1 How does uncertainty affect industrial PSS?

Cost assessments for the service and support of long-term projects is a challenge shrouded in uncertainty owing to the variable nature of such services and unpredictable changes in customer requirements [46,137]. Further uncertainties are found in highly variable equipment usage rates, lack of information to make accurate forecasts, importance of creating the right incentives around long-term maintenance and accurately predicting schedules [177]. These uncertainties present an inherent degree of risk to industrial PSS, which can be utilised as a measure of future uncertainties in achieving performance within defined cost, schedule and performance constraints [178].

3.2.2 Decision-making techniques

Saaty's [179] AHP has been extensively implemented and validated to prioritise alternative options via a set of evaluation criteria. Pairwise comparisons are applied to

each criterion in a set of matrices to generate weighted scores, which are then aggregated to give a global indication of the best or most popular option [103,135,180,181].

Other multi-criteria decision making (MCDM) methods such as TOPSIS and PROMETHEE can be applied in tandem with AHP to compare complex parameters such as algorithms through fuzzy theory [182–185]. Other qualitative approaches such as SWOT (strengths, weaknesses, opportunities, threats) analysis can be used to quickly identify risks and factors influencing uncertainties in a group setting, but may result in a plethora of factors that can't be accurately summarised in a quantitative manner with resources available [46,136,178]. AHP is therefore adopted in this study to identify the most significant challenges with a high level of accuracy.

3.3 Survey questionnaire – core challenges influencing uncertainty

A survey questionnaire was composed to rank prominent factors that influence uncertainty in maintenance based on literature and input from industry experts to gather heuristic data on challenges in industrial maintenance and the underlying uncertainty propagation. This consisted of scored and open questions. Scored questions were used to gain quantitative values denoting the degree to which, for example, the respondent agreed with a given statement. Open questions only restricted by the topic allowed respondents to give a clearer view of their true opinions and lead to further discussions and a clearer understanding of the subject.

The questionnaire comprised of four core sections:

- Current practice; considering the nature of long-term projects and resulting relationships with clients and external contractors, maintenance procedures in complex systems and the impact of maintainer wellbeing on maintenance quality
- Influencing factors; ranking of uncertainties arising in long-term projects according to the influence they have on the quality and effectiveness of maintenance carried out
- Data handling; systems and methods used to manage maintenance data and techniques used to influence decisions for future projects
- Risk and mitigation; strategies used to reduce uncertainty in maintenance

Nine responses were obtained from a leading defence company. Respondents scored 32 factors according to their influence on uncertainty on eight-point Likert scales (0-7) from "no influence" to "high influence" to avoid the neutral middle point, with a '0' option for 'no effect' [68,186]. These were refined and adapted by respondents and the author from a list defined by Erkoyuncu et al. [74], divided into 5 categories: commercial, affordability, maintainer performance, operational and engineering – illustrated in Figure 3.1 [135]. Respondents were each assigned a random ID to protect their anonymity. Respondent years of experience in current and relevant previous roles are illustrated in Figure 3.2.

3.3.1 Pedigree assessment

The pedigree matrix scores qualitative, expert opinion against predefined criteria to permit quantitative reliability assessment [52,127]. These criteria are defined according to the contextual application of the study [39,52,128]. The criteria were scored according to (1) years of experience in current role, (2) years of relevant experience prior to current role and (3.1-5) years of experience working on 5 select ship classes. Each criterion adhered to the same 1 to 5 scale: 1 = <5 years, 2 = 5-9 years, 3 = 10-14 years, 4 = 15-20 years, 5 = >20 years. Explicit roles were not included here to uphold anonymity.

An example of pedigree scores for two respondents is shown in Table 3.1. The weighted mean of these scores was used as a scaling factor to attribute proportionate scoring to their survey answers. These are compared with the mean scores in Figure 3.3. The weights of each criterion were defined by the author and are in themselves inherently subjective.

Commercial	1	Labour efficiency
	2	Customer equipment usage
	3	KPI specifications
	4	Stability of requirements
	5	Primary contractor relationship with customer
	6	Primary contractor relationship with OEM
	7	Accuracy & availability of technical data (concerning IPR, etc.)
	8	Communication between shareholders
Affordability	9	Customer ability to spend
	10	Customer willingness to spend
	11	Availability of equipment
Maintainer performance	12	Ability to screen candidates in training
	13	Availability of suitably qualified maintainers
	14	Knowledge and experience to perform a given task
	15	Material readiness state awareness
	16	Commitment to record data in relevant data banks
	17	Response to working environment (temp., confined spaces, etc.)
Operational	18	Quality of documentation / information from OEM
	19	Availability of resources to support maintenance
	20	OEM logistics (i.e. supply of parts)
	21	Complexity of equipment
	22	Quality of components and manufacturing
	23	Mean time between failure (MTBF) data
	24	Supply chain logistics
	25	Sufficiency of spare parts storage (on the shelf)
Engineering	26	Accuracy of cost estimation
	27	Confidence that reference books are reviewed and up-to-date
	28	Technology integration (availability of system interrogation software)
	29	Data reliability and quality
	30	Efficiency of engineering effort
	31	System capability upgrades
	32	Level of obsolescence (component, system or process)

Figure 3.1. Survey: Influential factors for uncertainty in industrial maintenance [135]

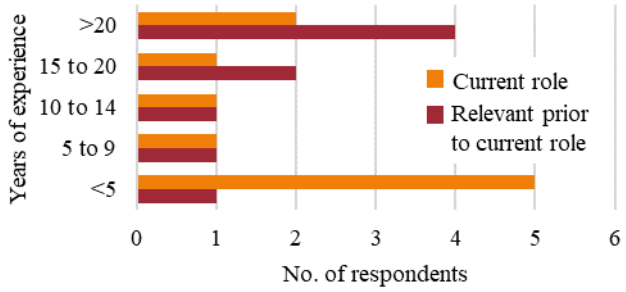


Figure 3.2. Survey: Respondent years of experience

Table 3.1. Survey: Example pedigree scores for two respondents

ID	(1)	(2)	(3.1)	(3.2)	(3.3)	(3.4)	(3.5)	(3.6)	Mean	W. mean
R1	5	5	-	2	2	4	2	5	3.57	3.85
R2	5	2	-	-	5	-	-	1	4.00	2.54

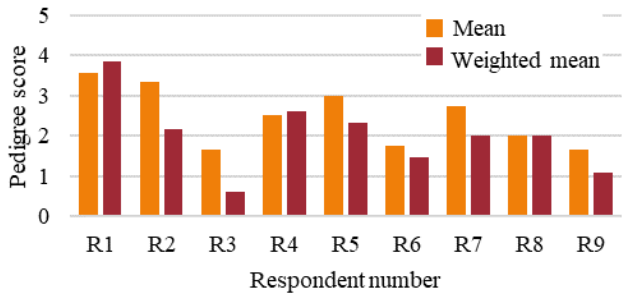


Figure 3.3. Survey: Mean and weighted mean comparison of pedigree scores for all respondent attributes

The mean and range for each influencing factor and category were evaluated in MS Excel. This is represented for all factors in Figure 3.4 numbered in the x-axis corresponding to Figure 3.1. Agreement between respondents is represented by the range, where a high range reflects high disagreement. These can be influenced by a specific project and not necessarily reflect their overall view. Factors that showed contrasting levels of agreement between the respondents are summarised below.

- **High influence on uncertainty, high levels of agreement:** Ability to screen candidates in training (12); Quality of information from OEM (18); Data reliability & quality (29)
- **High influence, high disagreement:** Customer ability to spend (9); Availability of resources to support maintenance (19); Supply chain logistics (24)
- **Low influence, high disagreement:** Labour efficiency (1); KPI specs (3); MTBF data (23)

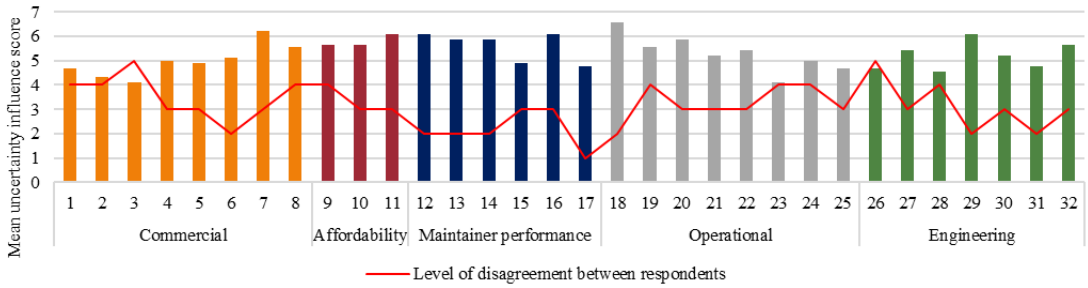


Figure 3.4. Survey results: Mean score for influence on uncertainty for all factors with level of disagreement

3.3.2 AHP implementation

AHP estimates relative magnitudes of inputs through pairwise comparisons [179]. These were represented in a positive reciprocal matrix adopting an algorithm defined by Erkoyuncu [39] for each of the 5 categories. Detailed results of the AHP are available in the supplementary data. The resulting weights highlighted the most prominent factors in each category, which were elaborated on in the interviews.

3.3.3 Interviews with industry

Survey results were analysed and discussed in a series of semi-structured interviews with respondents to obtain subjective views across maintenance departments. This structure allowed discussion of relevant topics while permitting respondents to provide further detail on their viewpoint from the survey [74,135,187]. Strategies and examples from literature [186–189] were used to structure and phrase the questions to obtain relevant information that can then be put forward to compose a framework capable of predicting the level of subsequent uncertainty influenced by challenges raised. Respondents were assured that responses would be handled confidentially and would not be linked to individuals. Where necessary, probes and prompts were used to encourage further responses and greater clarity [187]. The interview questions are depicted in Appendix C.

3.3.4 Core challenges summary

Core challenges that influence uncertainty prediction in maintenance, as highlighted from the questionnaire and interviews, can be summarised in six factors as follows:

Intellectual property rights (IPR), where modern systems are comprised of a vast number of components, many of which can only be maintained by the OEM due to IPR. This yields a degree of information asymmetry leading to uncertainty around the accuracy and availability of technical data; validated by the 'OEM logistics' factor having the single greatest influence on uncertainty in the survey. If a specialist maintainer cannot be sent out to fix the component, significant delays could ensue.

Maintainer performance, where levels of knowledge and experience can have a significant impact on maintenance quality and material state awareness. Additional time pressures and individual attitudes impact effort put into completing a task. Naval ships are deployed for several months at a time, whereas platforms such as aircraft are flown for a matter of hours and undergo rigorous maintenance checks between sorties. Over time, each ship on deployment naturally develops its own 'crew culture'. This has a core influence on maintainer attitude and affects the quality to which they conduct and record maintenance activities. Dockside maintainers would then not hold accurate data on the material state of a given part. If a problem was found the part would have to be replaced, accumulating unplanned costs and delays.

Quality of information, where documentation on maintenance procedures from OEMs is not well maintained. Books of Reference (BoR) are reviewed every 5 years, yet some date back to 1995. This can influence KPI specifications for a given platform, further raising uncertainty in maintenance procedures. In ship support, Job Instruction Cards (JIC), customer instructions and OEM documentation often lack detail. This exaggerates issues in data application for industrial and managerial support. Maintenance scheduling can then be affected, causing components to be maintained on a reactive basis rather than preventive. Materials and parts are not always available on the shelf when they should be and a robust system to purchase these materials is not in place. A range of data management systems are used for different ship platforms. For some, data is not necessarily recorded by the required party. Managers only get half the picture.

Resistance to change, where what is expected by the customer goes against what is or can be provided by the primary contractor. Many maintenance tasks need to be sub-contracted to a third-party OEM, which the primary contractor has no control over. That OEM could be operating under a one-off contract to maintain a specific part or system.

Significant uncertainties are raised here for the primary contractor as the time schedule and cost incurred from the third-party OEM cannot be finalised until the contract is completed, which may have knock-on effects for interconnected systems.

Stakeholder communication, where subcontractors may be fully qualified to sign off work done but cannot due to conflicts of interest, so the same task is repeated, resulting in unnecessary time and cost losses. An example was given in the interviews where two maintainers who have not conversed did not know the current material state or planned maintenance schedule of systems that connect at a platform level. The asset, maintained by the OEM, was rendered obsolete by ship staff while on deployment. It therefore missed a planned maintenance period when in dock, meaning the ship could not carry out its tasked duties.

Technology integration, where the exponential progression of technology means that training may not have kept up and software required to interrogate a system for diagnostic checks is not held by maintainers. New builds often have maintenance procedures locked in the design phase. Older platforms experience multiple upgrades over their lifetime which can result in examples such as seven different ship types under one platform grouped into a maintenance procedure, even though procedures for each type are different. Customer requirements may also change through design and upgrade programmes, which induce substantial costs and schedule delays.

A summary of the six core factors that influence uncertainty in industrial maintenance for industrial PSS and current approaches to maintenance is represented by Figure 3.5 in a broad sense between the OEM, contractor and client. The outer blue factors elaborate on areas where uncertainty is manifested, as discussed in the interviews.

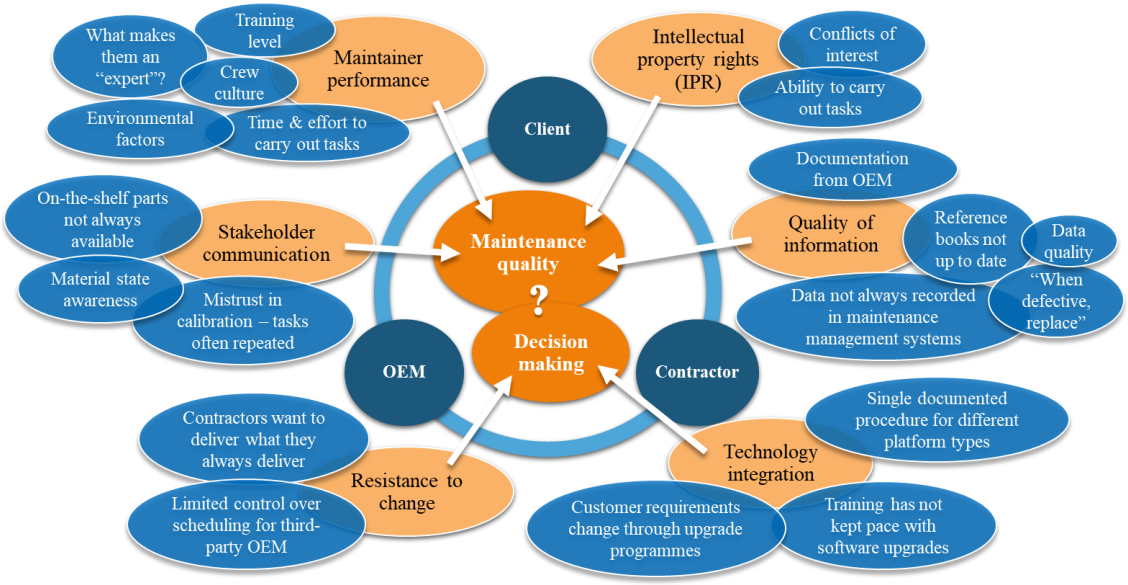


Figure 3.5. Survey results: Core factors influencing uncertainty in industrial maintenance

3.3.5 Wider industrial input

A live survey was carried out with industry practitioners and cost estimators at a workshop on modelling risk and uncertainty. The six core challenges identified were presented using Mentimeter live voting software. Respondents were asked if they considered a combination of quantitative, statistical (Type A) and qualitative, heuristic (Type B) uncertainty in their work and to identify their background, achieving 58 responses.

Segmentation of respondents according to their answer to the first two questions is illustrated in Figure 3.6 (unknown means the first question was unanswered). 41% of respondents were from the defence sector, 16% from aerospace and 24% cost analysists. A near 50:50 division of backgrounds was found and was relatively equal across each sector. Finally, respondents ranked the six challenges according to their influence on uncertainty, which gained 40 responses. The weighted mean score of each factor is shown in Figure 3.7, with an area plot for response distribution on the Likert scales. Higher weights are applied to data points with more responses, illustrated by peaks in the distributions.

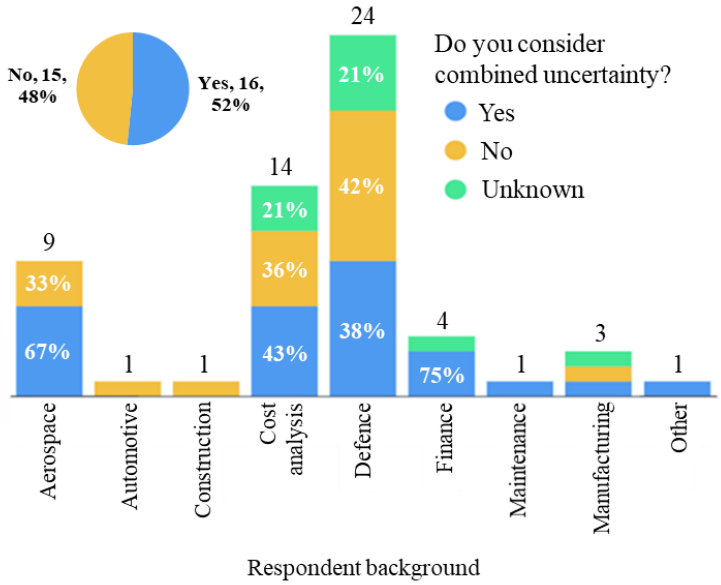


Figure 3.6. Live survey results: Respondent background according to whether they consider

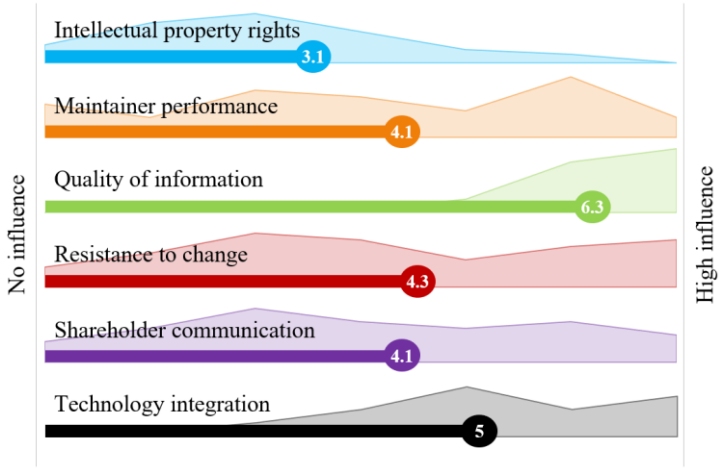


Figure 3.7. Live survey results: Subjective opinions on the core factors influencing uncertainty

Quality of information showed the greatest influence on uncertainty, with a weighted mean score of 6.3. This is followed by technology integration, with a weighted mean score of 5. Maintainer performance, resistance to change and stakeholder communication were found to have a relatively wide distribution spread, indicating disagreement between respondents. However, maintainer performance shows a higher distribution towards 'high influence'. As before, disagreement can be due to respondents' own comparative experiences in their industry in general or on a specific project they are working on. IPR showed the lowest influence, with a weighted mean of 3.1.

3.4 Discussion and conclusions

This chapter aimed to identify and rank core factors that influence uncertainties originating from challenges in the maintenance of complex assets under industrial PSS. Maintenance managers from a leading defence company completed a survey questionnaire identifying these factors. An assessment of the validity of their responses was made through defined pedigree criteria, the results of which were applied to each respondent to normalise their answers. Results were discussed and developed in a series of semi-structured interviews. Mean scores for each factor were weighted using AHP to identify the most influential factors. Core challenges were discussed in Section 3.3.4.

The derivation of pedigree criteria is inherently subjective. The criteria selected for this study (Section 3.3.1) were deemed, through the interviews and academic input, most applicable to score a level of expertise to respondents. Ranking more detailed qualifications against each other adds levels of complexity deemed out of scope for this study.

The AHP allowed factors to be weighed against each other within the survey categories. From this, the six core challenges were determined. These were validated through wider industrial input in a live survey, where the quality of information was deemed the most influential factor on uncertainty.

A shared understanding of material state across all departments is required to fill gaps in the supply chain, improve communication between stakeholders, overall decision-making and cost-effectiveness of ship support. Maintenance regimes used by the contractor or client may also differ, therefore holding a greater degree of uncertainty. In CES, where a change in uncertainty in one system has an unknown impact on another, this issue is amplified as different components may be maintained by different parties in the same system. There are approximately 300 different data repositories in use across the studied company, the majority of which are not linked and consist of numerous duplicate entries [190]. This includes DRACAS (Data Reporting, Analysis and Corrective Action System) and UMMS (Unit Maintenance Management System), where data may not be recorded in a useable fashion. This restricts the availability of data concerning specific components and the impact they will have on interlinked systems [191].

A common support model (CSM) is under development to tackle this challenge, featuring five management disciplines for through-life ship support: enterprise, class, design, maintenance and equipment [190]. These are endorsed by a complex web of information and knowledge management that is historically subject to a degree of asymmetry. This was made apparent in the interviews and previous studies across industrial sectors [48,192,193].

Stress levels and working conditions further influence uncertainty, as a heightened degree of each will negatively impact the quality of maintenance. It is incredibly difficult to obtain data on maintainer wellbeing as unions do not like to give or authorise the collection of such information. In many cases, more attention is given to critical and complex components. Non-critical components therefore receive less attention. For example, bypass valves could be considered non-essential until they fail.

A combined understanding of the impact of qualitative and quantitative uncertainty on system performance will provide a holistic picture allowing for more informed and effective decisions leading to prosperous outcomes in maintenance, but this comes at a cost. Budgets can be set for this with the 'spend to save' approach or set aside lump sums for unforeseen circumstances. Ultimately, a trade-off is required.

This study can be extended in several ways for further research. First, a broader framework can be developed to identify contributing factors in a given system, define them as quantitative (statistical) or qualitative (heuristic), identify acceptable uncertainty parameters for each element and combine the total subsystem uncertainties to gain a more holistic, quantitative picture. Second, the interrelationship between criteria can be incorporated and modelled through other quantitative and qualitative techniques such as the Analytic Network Process (ANP) [181] and PROMETHEE. Third is to develop analytical frameworks in order to better understand potential impacts of uncertainty and the ability to manage them should they arise.

CHAPTER 4. COMPOUND UNCERTAINTY QUANTIFICATION AND AGGREGATION

Abstract

The mounting increase in technological complexity of modern engineering systems demands rigorous determination of equipment availability and turnaround time whilst allowing for overruns and unforeseen costs. Quality and availability of quantitative data, as well as qualitative expert opinion and experience, expose uncertainties that can result in under or overestimation of the above factors. Quantifying such uncertainty should consider inter-connected components and associated processes from a combination of quantitative and qualitative (compound) perspectives. This chapter presents a Compound Uncertainty Quantification and Aggregation (CUQA) framework to determine the compound output along with an assessment of which parameters contribute the greatest uncertainty through global sensitivity analysis. This will provide maintenance planners with a confident, comprehensive view of parameters surrounding the above factors to improve decision-making capabilities. The framework was validated by assessing compound uncertainties in two case studies: a bespoke heat exchanger test rig and a simulated turbofan engine. The results demonstrate an effective measurement of compound uncertainty through the CUQA framework and the impact on system reliability. Further work will derive methods to predict uncertainty through the in-service phase of an asset's life cycle and its incorporation with more complex case studies.

Paper 4 An uncertainty quantification and aggregation framework for system performance assessment in industrial maintenance

Published: SSRN, TES Conference 2020
DOI: 10.2139/ssrn.3718001
Data access: 10.17862/cranfield.rd.12906443.v1

Paper 5 Compound uncertainty quantification and aggregation (CUQA) for reliability measurement in industrial maintenance

Submitted: Reliability Engineering and System Safety
Data access: 10.17862/cranfield.rd.13550561

4.1 Introduction

Uncertainty quantification (UQ) concerning the maintenance of engineering systems is growing in recognition and rigour as the complexity of such systems surges in the modern world. Complex engineering systems (CES) are comprised of multiple sub-elements including equipment and operators that interact simultaneously and nonlinearly with each other and the environment on multiple levels [2,3]. Consideration of the relationships between elements is vital to understand emergent behaviour to aid decision making [4]. Complex systems science is a field in itself, the theory of which is widely discussed in literature [4,6,8,9] but is out of scope for this research.

The maintenance of complex and non-complex engineering systems exhibit a range of uncertainties from interconnected factors such as quality and availability of quantitative equipment data and the qualitative influence of operators, expert opinion, experience and environmental conditions [10]. These uncertainties are represented by varying probability distribution functions (PDFs) and can lead to under or overestimation of maintenance costs, reliability measurement, equipment availability and delays in maintenance scheduling. Recent research in CES has explored UQ in micro gear measurements [3], structured surfaces using metrological characteristics [11], correlation uncertainty in gear conformity [12], grey-box energy models for office buildings [14], uncertainty in disassembly line design [15] and others reviewed in various related studies. Many of these approaches only consider quantitative uncertainty given by variability in measured data, rather than the compound aggregation of quantitative and qualitative uncertainties [3,11,14,15]. Methodologies to do this are growing in many areas, but are limited from an industrial maintenance perspective. This is necessary to obtain a comprehensive understanding of system reliability, as well as the inherent risks and knock-on effects imposed by altering elements within the system. Limited research guiding the aggregation of compound uncertainty sets the focus for this chapter.

A 6-step framework is presented to quantify and aggregate compound uncertainties to enhance system performance assessment. This will provide maintenance planners with a comprehensive view of parameters surrounding the above factors to improve decision-making capabilities.

The proposed framework is detailed in Section 4.2 along with key mathematical formulae, functions and assumptions made. Section 4.3 applies the framework to two case studies: a bespoke heat exchanger test rig comprised of multiple sub-systems, developed at Cranfield University [17], and a simulated dataset for turbofan engine degradation. Individual uncertainties from quantitative and qualitative sources and correlations between them are assessed and aggregated to give a confident indication of system performance. Section 4.4 discusses the results, strengths and limitations of the framework along with conclusions and future work in this area.

4.2 Compound uncertainty quantification & aggregation (CUQA) framework

Every measurement or estimate is subject to a degree of error, which in turn contributes a level of uncertainty. Quantifying this uncertainty enables a thorough assessment of the scale of risk inflicted on the system by each component [2,18]. This chapter contributes a holistic assessment of compound uncertainties in dynamic data represented by different distributions with an integrated assessment of correlations and sensitivity. This addresses research gaps 1 and 2 identified in Chapter 2, achieved through a 6-step modelling approach developed in MATLAB, described below and illustrated in Figure 4.1.

The framework was designed as an extension and amalgamation of existing methodologies from literature [2,18,20,22,24,26,28]. An initial version was presented in Grenyer et al. [30]. Here it is further developed and validated in two case studies, considering key parameter variables identified within the system. This chapter has been submitted as a manuscript and is under peer review at the time of writing. The framework steps were developed from the traditional approach in the GUM, extended to consider compound uncertainty and GSA, detailed as follows:

Step 1: Outline system setup & uncertainty sources. Inputs are grouped according to their uncertainty type – quantitative or qualitative. This includes all measured data, assumptions made and environmental predictions. Distribution types are established by ‘goodness-of-fit’ tests. Selected types are indexed for later calculation.

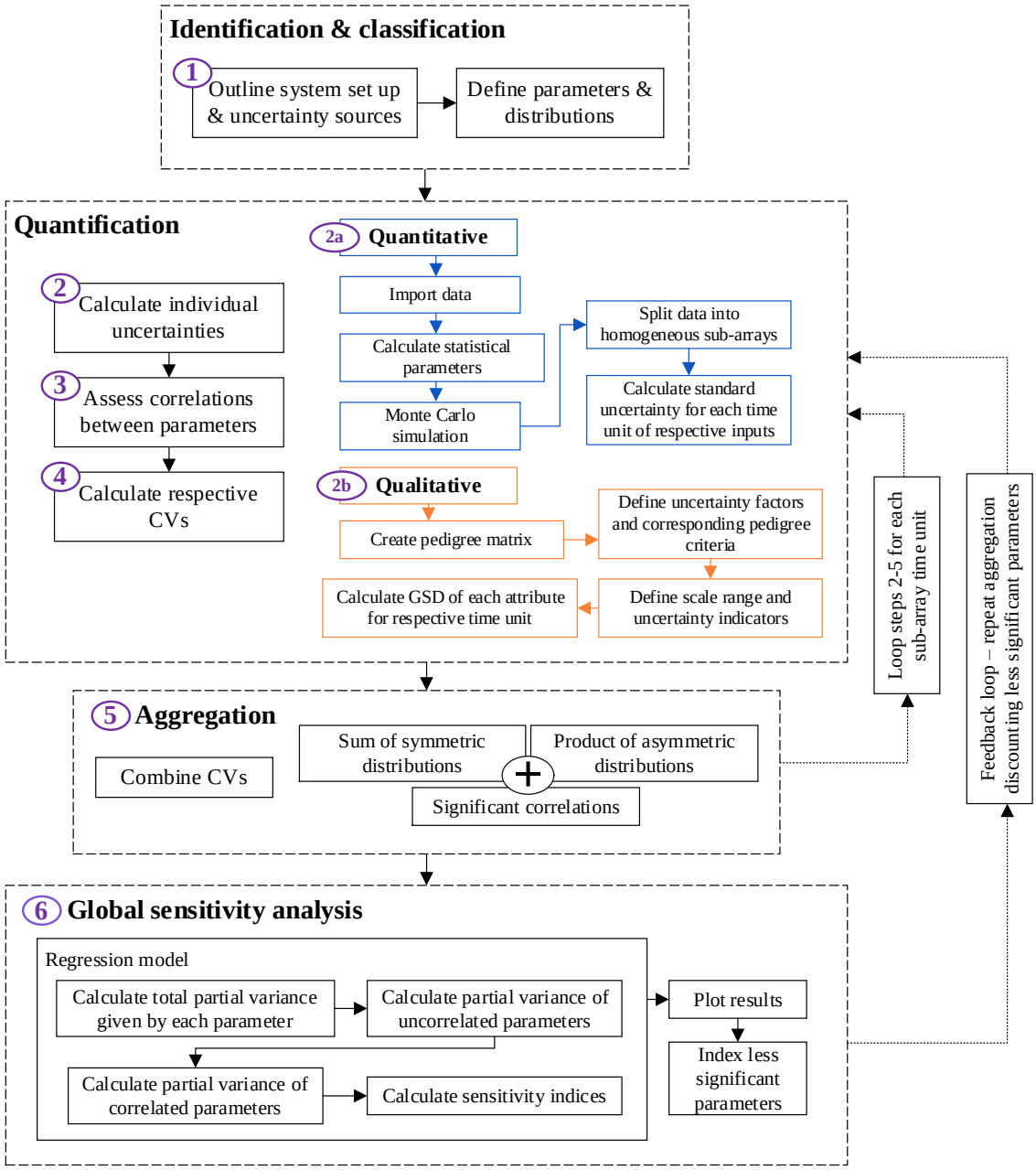


Figure 4.1. CUQA framework overview

Step 2: Calculate individual uncertainties. Statistical parameters are calculated for each input according to their relative distribution via Monte Carlo simulation and the pedigree matrix. These are grouped for each subsystem; for which the standard uncertainties and correlations are determined separately before combining with the whole system, elaborated as follows:

Step 2a: Quantitative, recorded data is concatenated in a cell array to allow inputs with a varying number of data points to be considered. Any non-numeric values are removed. Monte Carlo simulations are run for the relative indexed PDF over a user-defined number of points (default 10,000) or to the size of the largest input parameter. This propagates input data to a homogeneous array size. In order to consider the uncertainty in the measured values, each dataset (X_i) is split into sub-arrays over the recorded time period. The number of rows for each sub-array (S_i) can be selected by the user or defined automatically. Possible values for S_i are defined by the number of factors (N_f) in the value of the length of the dataset (dim (X_i)). The automatic selection is given by Eq.4-1. This aims to select the middle factor, providing enough values to determine the uncertainty at each point while allocating enough sub-arrays to determine the change in uncertainty for the recorded period. Each dataset is then reshaped according to Eq.4-2, where $S_{i,j}$ is the reshaped sub-array dimension.

$$S_i = \begin{cases} \left[\left(\frac{N_f}{2}\right) + 1\right], N_f < 10 \\ \left[\left(\frac{N_f}{2}\right)\right], \quad N_f \geq 10 \end{cases}, \mathrm{S_i} \geq 1 \tag{4-1}$$

$$X_i \in \mathbb{R}^{\dim(X_i)} \rightarrow X_i \in \mathbb{R}^{S_{i,j}} \tag{4-2}$$

The arithmetic and geometric mean and deviation are calculated for each sub-array and the full dataset, along with maximum and minimum values of each input variable. The standard deviation of each time unit is then calculated using the simulated data for each distribution type. For lognormal variables, the mean and standard deviation is given as geometric. Normal and uniform distribution variables are arithmetic [32]. To visualise the data, boxplots for each sub-array are overlaid on the initial dataset. These plots give more detailed information than standard error bars on the change in uncertainty over time with dynamic datasets.

Step 2b: Qualitative factors are defined through pedigree criteria. Based on the example implemented by Ciroth [32], the matrix defines uncertainty indicators based on expert judgement. Criteria are defined for each score for each factor, which relates to predefined case-dependent uncertainty measures. The ideal case has a pedigree score of 1, corresponding to minimal uncertainty. Scores of 2-n have progressively higher uncertainties owing to their representative criteria. While there is no limit to the number

of scores, typically a maximum of 5-7 is used. The scores for each factor correspond to an uncertainty indicator, the GSD of which is obtained from one or multiple sources (interviews, surveys etc.). These scores will not be fixed over time, and so are pseudo-randomly applied ±1 of the defined score for each sub-array. If the uncertainty indicators are obtained from a single source, the GSD is given as its square root. If they are obtained from multiple sources, the GSD is given by Eq. 2-2, modelled by the lognormal distribution [32,34,36]. The GSD of less ideal indicators is given as a ratio of the calculated GSD and that of the ideal score for each input, meaning that it is always equal to or greater than 1 [32].

Step 3: Determine significant correlations between input parameters. To best determine correlation, input parameters must be of equal length. For quantitative data, initial recordings prior to Monte Carlo are sampled to the size of the largest parameter length around their respective PDF type. Qualitative parameters are sampled using their uncertainty score as the respective mean and GSD as standard deviation under a lognormal distribution to achieve a homogeneous sample size. Spearman's correlation coefficient ρ (Eq. 2-5) is calculated between each pairwise input parameter, along with their corresponding p-values. These are the result of the null hypothesis significance test that determines whether what is observed in the data sample is likely to be true for a wider population. A default significance level (α) of 0.05 determines that for p-values $< \alpha$, there is 5% chance that a significant correlation does not exist between those parameters [9,37]. In addition, an ideal limit to define significant coefficient magnitude is defined by the user as, ρ_{lim} and cut-off, ρ_{cutoff}. If there is not at least one pairwise coefficient for which the absolute value $|\rho| > \rho_{lim}$, the ideal ρ_{lim} is reduced in increments of 0.01 via a 'while' loop until the condition is true or the defined ρ_{cutoff} is reached. This enables the user to define the degree of correlation to be included in the aggregation with the assurance that the resulting coefficients are statistically significant. Corresponding input parameters for which the final condition is true are plotted in a correlation matrix and stored for use in Step 5. This matrix provides a visualisation of correlation magnitude for each parameter with a significantly correlated pair [38].

Step 4: **Calculate the CV for each input.** Uncertainties from different data types represented by different PDFs must be considered on an equal scale in order to be aggregated. This is achieved through the CV, explained in Section 2.4.2.3, the formulae for which are given in Table 2.4 [34]. These are calculated within the framework by a sequential algorithm according to the specified input and distribution type. Summary tables are then generated for the compound inputs and correlation, as calculated in Steps 2-3.

Step 5: Aggregate respective CVs and correlated parameters. As discussed in Section 2.4.2.1, symmetric distributions are aggregated additively by RSS (Eq. 2-1). Asymmetric distributions, given by lognormal distributions, CV_{Ln}, are aggregated multiplicatively by Eq. 2-3 [34]. The framework splits the calculated CVs of quantitative inputs according to the distribution type. The sum of symmetric attributes is added to the product of lognormal attributes. Comparing this with Eq. 2-1, the aggregated uncertainty is given by CV_T in Eq. 4-3:

$$CV_T = \sqrt{\sum_{i=1}^{n}(CV_{sym}^2) + \left(\prod_{i=1}^{n}(CV_{Ln}^2 + 1) - 1\right) + 2\sum_{i=1}^{n}\left(\rho_{x,y}\, CV_x\, CV_y\right)} \tag{4-3}$$

Where $(\rho_{x,y}\, CV_x\, CV_y)$ is the Spearman correlation coefficient of 2 parameters x and y multiplied by their respective CV.

Individual CVs are plotted as bars against the aggregated total, along with a colour bar to visualise the acceptability of relative factors according to predefined scales. The correlation coefficient standardizes the variables and is therefore unaffected by changes in scale or units. The formulae allow the aggregated CV of quantitative and qualitative data to be determined as a measure of total uncertainty. Given that CV is the ratio between the standard deviation and the mean, the output follows a normal distribution. The uncertainty can therefore be expressed back as the standard deviation via Eq. 4-4.

$$\sigma_T = \sqrt{\sum_{i=1}^{n}(\sigma_i)^2} = \sqrt{\sum_{i=1}^{n}(\mu_i CV_i)^2} \tag{4-4}$$

Steps 2-5 are repeated for each sub-array unit. Summary variables including the individual and aggregated CV are stored and used to calculate the sensitivity indices in Step 6.

Step 6: Conduct GSA and visualise results. The relative influence of individual uncertainties on the aggregated total is calculated as the response vector over each sub-array time unit. The sampling approach proposed by Groen [40], influenced by Xu and Gertner [41], is applied to consider the effect of correlated parameters using an adjusted regression model. Results are visualised by a 3D bar plot to show dependant and independent effects against the total, with the same colour scale applied as for Step 5 to illustrate the severity. A feedback loop is then taken back to Step 2 where parameters with total effects below a defined threshold (default 5%) are discounted. The aggregated uncertainty and sensitivity indices are updated to determine the parameters contributing the greatest impact to the aggregated uncertainty, visualised in the same manner.

4.3 Stepped implementation and results of CUQA framework

4.3.1 Case study 1: Heat exchanger test rig

The framework was first applied to a bespoke heat exchanger test rig, developed from an initial design by Addepalli et al. [17] with the installation of a motorised pump and digital sensors. The combination of digital and analogue recording, along with qualitative factors discussed below, manifests compound uncertainty in heat exchanger performance. These uncertainties need to be quantified and aggregated to assess their impact on the system, assessed via the heat transfer coefficient [43,45,47]. This is calculated with the resulting uncertainty, derived alongside the CUQA framework as follows:

Step 1: Outline system setup & uncertainty sources. The system comprised of a hot closed-loop system and a cold open-loop system, illustrated in Figure 4.2 (notation defined in Table 4.2). Component specifications are described in Table 4.1.

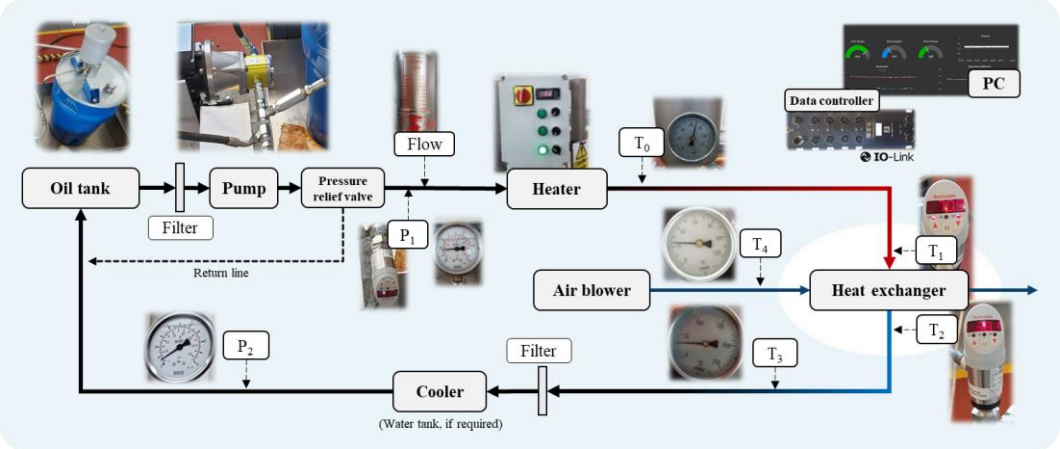


Figure 4.2. Heat exchanger test rig: System design [17]

Table 4.1. Heat exchanger test rig: Component specifications of the initial design

Component	Specification
Oil	Aero shell turbine 500
Pump	Vivoil X2P4702EBBA motorised pump
Heater	3 connected units controlled by 3 switches, temp. indicated by probe
Heat exchanger	Jaguar oil cooler, plate-fin type
Temperature sensors	Barksdale BTS38GVM0050M1
Pressure sensor	Barksdale BPS38GVM0010B
IO-Link master	Pepperl+Fuchs ICE2-8IOL-G65L-V1D

The experimental setup comprised of seven quantitative parameters, summarised in Table 4.2 along with their corresponding reading interval and error, and five qualitative factors: (1) Reliability of data, (2) Basis of estimate, (3) Reading accuracy, (4) Environmental conditions and (5) Sample size – each modelled by the lognormal distribution. Oil temperature at the inlet (T_1) and outlet (T_2) was measured by dual temperature sensors. A constant flow rate was maintained by a motorised pump. Oil pressure (P_1) was regulated by a pressure relief valve, recorded by a dual pressure sensor at the pump outlet. The sensors fed real-time data to the PC controller via IO-Link, logged to a CSV file in 1-second intervals along with a timestamp.

Table 4.2. Heat exchanger test rig: Uncertainty sources – measured parameters

Parameter	Reading type	PDF	Reading interval	Reading error
T_1, Sensor, hot fluid temp. into HEx (°C)	Digital	Lognormal	0.1°C	± 0.1°C
T_2, Sensor, hot fluid temp. out of HEx (°C)	Digital	Lognormal	0.1°C	± 0.1°C
T_3, Dial, hot fluid temp. out of HEx (°C)	Analogue	Normal	5°C	± 2°C
T_4, Dial, cold fluid temp. (air blower) (°C)	Analogue	Uniform	2°C	± 0.5°C
P_1, Sensor, hot fluid pressure pre-HEx (bar)	Digital	Lognormal	0.01 bar	± 0.01 bar
P_2, Dial, hot fluid pressure post-HEx (bar)	Analogue	Uniform	0.5 bar	± 0.3 bar
$\dot{V}$, Volumetric flow rate of hot fluid (L/min)	Analogue	Uniform	5 L/min	± 2 L/min

The heat transfer coefficient is given by the heat load Q of the hot (h) and cold (c) fluid (Eq. 4-5):

$$\begin{aligned} Q_h &= \dot{m}_h \cdot cp_h \cdot (T_{h\,In} - T_{h\,Out}) \\ Q_c &= \dot{m}_c \cdot cp_c \cdot (T_{c\,Out} - T_{c\,In}) \end{aligned} \tag{4-5}$$

Where $\dot{m}$ = mass flow rate, given by the product of the volumetric flow rate $\dot{V}$ and density ρ; cp = specific heat capacity; and $T_{In} - T_{Out}$ is the fluid temperature differential in and out of the heat exchanger.

The heat balance error and composite heat load considering associated uncertainty are given by Eq. 4-6 and Eq. 4-7 respectively, as derived by Tatara and Lupia [43]. Contributing measurement uncertainties and additional qualitative bias in the system were calculated separately using the propagation of error method [37].

While $|HBE| < |U_{HBE}|$ (Eq. 4-8), the overall heat transfer coefficient can be found and associated measurement uncertainties are considered valid [43]. The focus of this study was on the uncertainty in the measured values over time, not the uncertainty of the overall recording period.

$$HBE = \frac{Q_h - Q_c}{Q_h} \cdot 100\% \tag{4-6}$$

$$Q = \frac{Q_c U_{Q_h}^2 + Q_h U_{Q_c}^2}{U_{Q_h}^2 + U_{Q_c}^2} \tag{4-7}$$

$$U_{HBE} = 100\% \cdot \frac{Q_h}{Q_c}\sqrt{\left(\frac{U_{m_h}}{m_h}\right)^2 + \left(\frac{U_{T_{hIn}}}{T_{hIn}-T_{hOut}}\right)^2 + \left(\frac{-U_{T_{hOut}}}{T_{hIn}-T_{hOut}}\right)^2 + \left(\frac{U_{m_c}}{m_c}\right)^2 + \left(\frac{U_{T_{cIn}}}{T_{cIn}-T_{cOut}}\right)^2 + \left(\frac{-U_{T_{cOut}}}{T_{cIn}-T_{cOut}}\right)^2} \quad (4\text{-}8)$$

The heating system was set to switch off at 80°C to prevent overheating. However, due to its design, the heater was not able to sustain the temperature at 0.02°C/min for 10mins, as recommended by Tatara and Lupia [43] to determine steady-state. While this is unsuitable for thorough thermodynamic assessment of heat transfer efficiency from the heat exchanger, it contributes further qualitative uncertainty to the system, which is reflected in the application of the CUQA framework.

The steady-state region was therefore defined by the time of the first and last peak temperature readings at T_1. Two cycles were completed, with a total of 85 minutes recorded: a total of 5590 data points for the three digital parameters. The temperature recorded at T_1 had an overall range of 6.8°C and 1.2°C at T_2 over the recorded period. The pressure, P_1 was set at 1.8 bar, following a lognormal distribution with a range of 0.32 bar.

Aside from these readings, all variable measurements were recorded via in-line analogue dials. Many of these dials gave readings on different interval scales, varying measurement accuracy, and therefore resulted in an increased uncertainty. Additional attributes such as parallax error and ambient temperature further increase uncertainty in the measurement.

The volumetric flowrate $\dot{V}_h$ of the oil (hot fluid) was held at 5 L/min (0.83×10^{-3} m^3/s) with a uniform distribution. A reading error of ±2L/min was assigned owing to the scale of the flowmeter. At a maximum temperature of 80°C, $\rho \approx 0.95$ kg/L (950 kg/m^3). Therefore, $\dot{m}_h$ for the hot fluid = 0.08kg/s. cp_h is given as 1800 J/kg.C. For the air (cold fluid), $\dot{m}_c$ was given as 1.12 kg/s and cp_c as 1005 J/kg.C. Further thermodynamic analysis involving parameters such as oil viscosity and temperature loss through connecting pipes were out of scope for the framework application. The uncertainty contributed by these factors was factored into the pedigree matrix.

Step 2a: Calculate quantitative uncertainties. A summary of the seven quantitative parameters is given in Table 4.4. Summary statistics from the logged data for T_1, T_2 and P_1 are given by the boxplots in Figure 4.3. Outliers are values greater than $q_3 + w(q_3 - q_1)$ or less than $q_1 - w(q_3 - q_1)$, where w is the maximum whisker length, 1.5 times the interquartile range, and q_1 & q_3 are the 25th and 75th quartiles of the respective dataset [49].

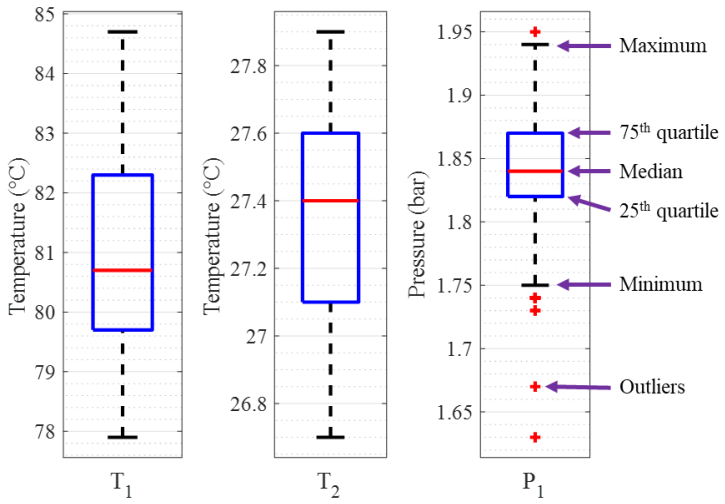


Figure 4.3. Heat exchanger test rig: Boxplots for T_1, T_2 and P_1

The three digitally recorded parameters were split into 65 homogeneous sub-arrays over the 5590 data points. The overlaid boxplots are shown in Figure 4.4, plotted over the time series of the logged data. Owing to the multimodal shape of the data, the sub-array standard deviation for T_1 is low to negligible at the peaks and troughs and high for temperature increases or decreases. The temperature at T_2 is more constant respective to T_1 showing a step change over time owing to the heat transfer coefficient of the heat exchanger.

The greater the sub-array size (S_i) the greater the uncertainty in the measurement. This is illustrated in Appendix D (Figure D.1) for all possible factors (left), with a focus on S_i values of 0-130 and the automatically selected value, 86, highlighted (right). This procedure enables a mean uncertainty estimate to be obtained where the recorded data is not able to meet the criterion for steady-state readings.

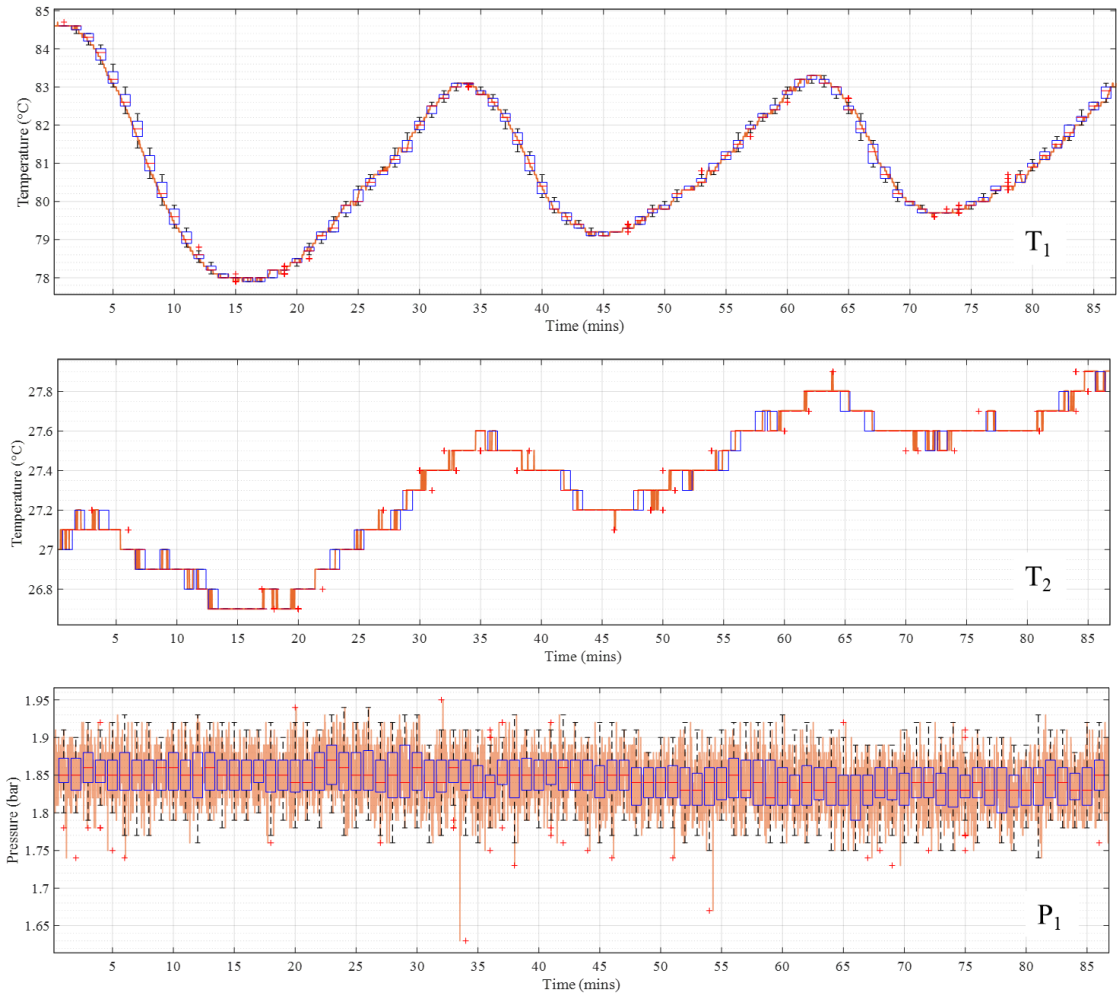


Figure 4.4. Heat exchanger test rig: Sub-array boxplots over time-series data

As S_i increases, the number of sub-arrays decreases, resulting in greater uncertainty. This is considered by the 'basis of estimate' factor in the pedigree matrix.

The four remaining quantitative parameters were acquired by analogue dials with varying reading intervals (Table 4.2). These were taken every 30mins over the recording period, resulting in limited data in comparison to the automated recording. Using Monte Carlo simulation, the readings were propagated to match the array size of the three digital parameters according to their statistical range and rounded to their corresponding reading intervals.

Step 2b: Calculate qualitative uncertainties. The 5 qualitative factors were scored by defined pedigree criteria detailed in Table 4.3. These were based on adjusted examples from literature to apply to the case study [10,51,52]. Uncertainty indicators for each factor for increasing pedigree scores corresponding to the criteria are illustrated in Figure 4.5. For this case study, the uncertainty indicators were obtained from a single source (the authors opinion) and applied to the full dataset.

Table 4.3. Heat exchanger test rig: Pedigree criteria

Score	1	2	3	4	5
Reliability of data	Data is < 2 months old and/or recorded by fully calibrated sensor or fully qualified person	Data is < 6 months old and/or recorded by fully qualified person but sensor requires recalibration	Data is < 12 months old and/or recorded by experienced person but sensor requires recalibration	Data is > 12 months old and/or recorded by experienced person, sensor accuracy unknown	Age or source of data unknown or > 12 months old
Basis of estimate	Best possible data, use of historical field data, validated tools and independently verified data, given by fully qualified person	Smaller sample of historic data, parametric estimates, internally verified data, some experience in the area	Limited available data, unverified, inexperienced opinions	Incomplete data, small sample, educated guesses, indirect approximate rule of thumb estimate	No experience in the data
Reading accuracy	Measurements taken using fully calibrated and accurate equipment: ±0.01°C, ±0.1 bar	Measurements taken using recently calibrated but less accurate equipment: ±0.1°C, ±0.5 bar	Measurements taken using recently calibrated but less accurate equipment: >±1°C, >±2 bar	Measurements taken using accurate equipment that may need recalibrating	Measurements taken using un-calibrated and inaccurate equipment
Environmental conditions	Data recorded under specific consistent conditions or a specified range of conditions from area under study	Data recorded in generally consistent conditions with fluctuations specified	Data recorded in generally consistent conditions, changes not specified	Data recorded in a range of unspecified conditions	Data from unknown or distinctly different areas
Mean sample size	> 1000	> 100	> 50	< 50	Unknown

Their GSD is therefore given as the square root of the uncertainty indicator. These scores will not remain fixed over time and are therefore pseudo-randomly applied ±1 of the defined score circled in Figure 4.5 for each sub-array.

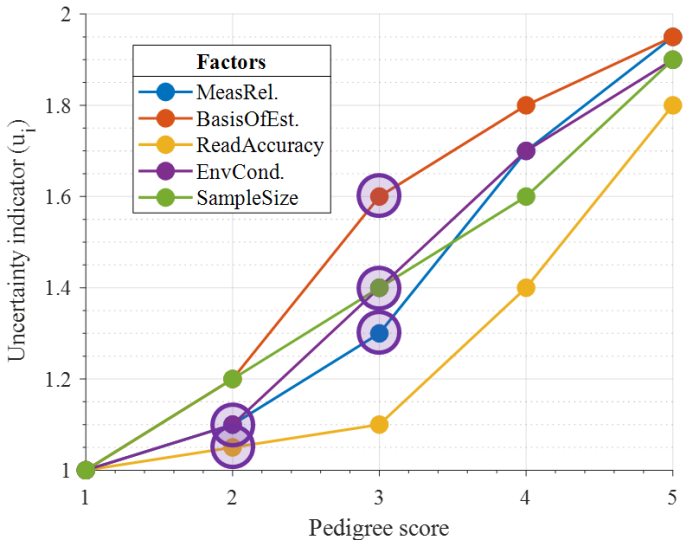


Figure 4.5. Heat exchanger test rig: Uncertainty indicators for increasing pedigree scores

The resulting CV calculated in Step 4 was significantly greater than that of the lognormal recorded data. This is most likely due to the small number of data points in the sub-arrays. To give a closer comparison of the uncertainty, the pedigree factors were rescaled by Eq. 4-9. The following results up to Step 6 illustrate an example for the first sub-array time unit.

$$U_{i\_scaled} = \frac{(U_i - 1)}{10} + 1 \tag{4-9}$$

Where U_i = uncertainty indicator

Step 3: Assess correlations between parameters. The ideal limit of ρ was set to 0.5, with a cut-off at 0.2. Naturally, significant positive correlation was identified between T_1 and T_2, highlighted in red (Figure 4.6). The negative correlation to P_1 reflects the pressure drop due to oil viscosity with increasing temperature. This shows the effectiveness of selecting the desired ρ limit to remove minor correlations from the analysis.

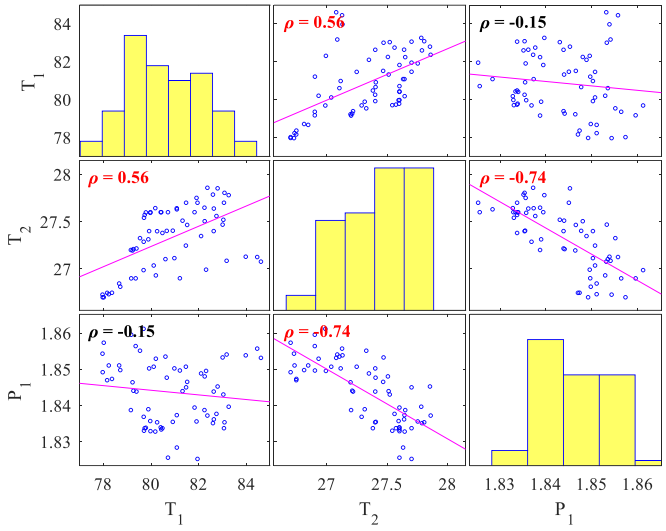


Figure 4.6. Heat exchanger test rig: Significant correlations for which $|\rho| \geq 0.5$

Step 4: Calculate respective CVs. The summary tables with calculated CV for each input are given in Table 4.4 and Table 4.5 for the quantitative and qualitative factors respectively.

Table 4.4. Heat exchanger test rig: Recorded data and calculated parameters

Parameter	Reading interval	Reading error	Dist.	Mean	Standard deviation	Min	Max	CV
T_1 (°C)	0.1°C	± 0.1°C	Ln	80.8654	1.0209	77.9628	84.6012	0.0207
T_2 (°C)	0.1°C	± 0.1°C	N	27.3305	0.3351	26.7000	27.8581	0.0123
T_3 (°C)	5.0°C	±2.0°C	Ln	24.6000	2.8810	20.0000	28.0000	0.1171
T_4 (°C)	2.0°C	±0.5°C	U	21.6000	0.8944	20.0000	22.0000	0.0000
P_1 (bar)	0.5 bar	±1.0 bar	N	1.8436	0.0088	1.8252	1.8612	0.0048
P_2 (bar)	0.5 bar	±0.3 bar	U	0.9000	0.2236	0.5000	1.0000	0.0000
Flow (L/min)	5 L/min	± 2 L/min	U	4.9575	0.0253	4.9343	4.9865	0.0000

Table 4.5. Heat exchanger test rig: Pedigree factors with relating GSD and CV

Factor	Distribution	Pedigree score	Uncertainty indicator	GSD	CV
Meas. Relbl.	Lognormal	2	1.1	1.0488	0.0477
Basis of Est.	Lognormal	2	1.2	1.0954	0.0914
Read Accuracy	Lognormal	1	1.0	1.0000	0.0000
Envir. Cond.	Lognormal	2	1.1	1.0488	0.0477
Sample Size	Lognormal	3	1.4	1.1832	0.1694

Step 5: **Combine CVs.** The combined CV of each PDF is calculated by Eq. 4-3 and summarised in Table 4.6, aggregated for symmetric and asymmetric distributions and total CV with correlation between T_1 and T_2 – given in the table as $2\left(\rho_{T_1,T_2} \cdot CV_{T_1} \cdot CV_{T_2}\right)$.

Table 4.6. Heat exchanger test rig: CV aggregation results

PDF	CV comb.	CV agg.	Corr.	CV_T
Ln recorded	0.0207	0.2256	0.0001	0.2593
Ln pedigree	0.2050		0.0011	
Norm. recorded	0.1179	0.1179		
Uni. recorded	0.0000			

The visualisation in Figure 4.7 illustrates the relative CV of each quantitative (blue), qualitative (orange) and correlated (purple) input against the aggregated total (cream) for one of the 86 sub-array time units. When calculated for only the quantitative parameters, the aggregated CV fell to 0.1293; a percentage decrease of 50.1% for the example time unit. This illustrates the significance of accounting for qualitative factors alongside quantitative parameters – providing a holistic view of factors that manifest uncertainty in the system. While the depiction of these factors is subjective, the compound consideration reduces the risk of underestimating the aggregated uncertainty, which can occur if only accounting for quantitative parameters [32]. Individual uncertainties are then expressed as variances by the square of Eq. 4-4 to feed into Step 6. The change in individual and aggregated CV over all time units for $S_i = 65$ (86 sub-arrays) is given in Figure 4.8a and compared with $S_i = 215$ (26 sub-arrays) in Figure 4.8b. This demonstrates the effect of sub-array size on the resulting uncertainty estimate.

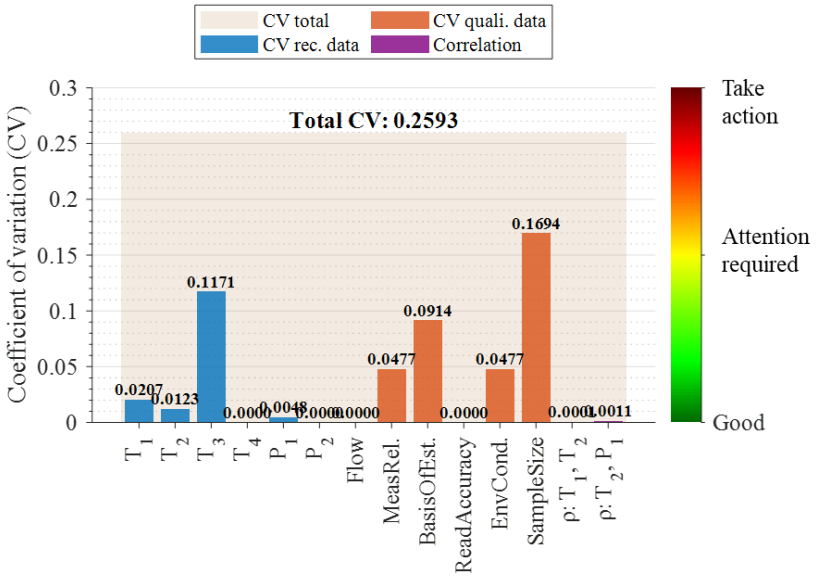


Figure 4.7. Heat exchanger test rig: Aggregated total CV against individual factors for one time unit

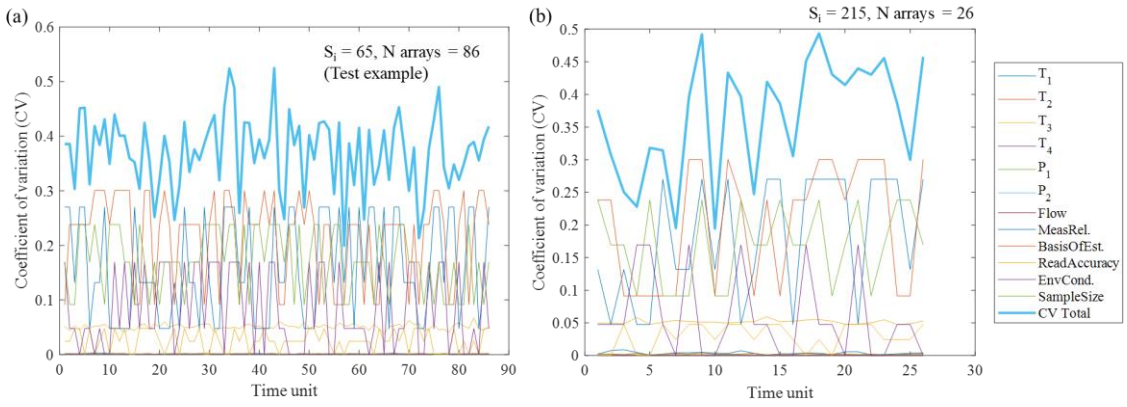


Figure 4.8. Heat exchanger test rig: Aggregated total CV against individual factors over all time units for $\boldsymbol{S_i = 65}$ (a) and $\boldsymbol{S_i = 215}$ (b)

Calculating the heat load parameters from Eqs. 4-5 to 4-8 gives [43]: Q_h = 366.2MW, U_{Qh} = 16.31MW, Q_c = 4.52kW, U_{Qc} = 97.56W and resulting Q = 4.52kW. The heat balance error (HBE) = 99.98% and composite load uncertainty U_{HBE} = 311%. This passes the validity test given by as $|HBE| < |U_{HBE}|$, indicating the measurements are valid.

Step 6: GSA and visualisation. The relative influence of individual uncertainties on the aggregated total is plotted in Figure 4.9a. The uncertainty in T_3, the oil temperature after being cooled by the heat exchanger, has an overwhelmingly greater effect (76%) on the aggregated uncertainty than any other parameter. This is due to the large error margin of ± 2°C given by the reading interval on the dial. If T_3 is discounted, along with parameters with an impact below 5% (uniformly distributed), the basis of the estimate was deemed to have the greatest effect at 56% (Figure 4.9b). The influence of T_1 and T_2 is minimal due to the comparatively equal deviation for each sub-array time unit.

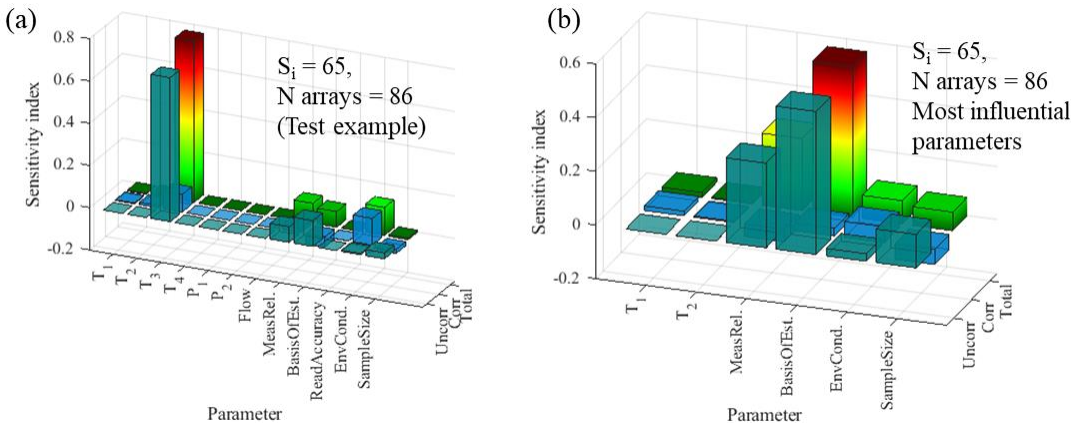


Figure 4.9. Heat exchanger test rig: Sensitivity effects of individual to aggregated uncertainty over all time units for all factors (a) and most influential parameters (b)

Altering the pedigree score allocation of the qualitative factors impacts the degree of uncertainty each factor will contribute to the aggregated total, according to the defined uncertainty indicators in Figure 4.5. Applying higher pedigree scores will apply a higher representative level of uncertainty. The difference between one uncertainty indicator to another will influence the respective factor's sensitivity index owing to the pseudorandom score allocation. Increasing the degree of allocation (e.g., from ±1 to ±2) will also influence the respective sensitivity indices, though this was not deemed necessary in this study for the score range of 1-5. While the uncertainty indicator scores are subjective, they are expected to increase linearly or exponentially. Therefore, lower scores will have less influence on the aggregated total.

4.3.2 Case study 2: Turbofan engine degradation

The framework was applied to a turbofan engine degradation dataset simulated from the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) tool, developed by NASA [55,57]. This publicly available dataset has been widely applied in prognostics and health management (PHM) [57,59,61,63,65]. The C-MAPSS data consists of four datasets simulated under different operating conditions. The FD001 training dataset, simulating degradation of the high-pressure compressor (HPC), was applied to the CUQA framework to analyse the aggregated uncertainty in the measurements over time:

Step 1: Outline system setup & uncertainty sources. The FD001 dataset consisted of 21 sensors measuring temperature, pressure and speed for 100 engine units, each with a random start time and normal operating level, running to failure. For this study, one engine unit was selected with 192 cycles to failure. The system design is illustrated in Figure 4.10.

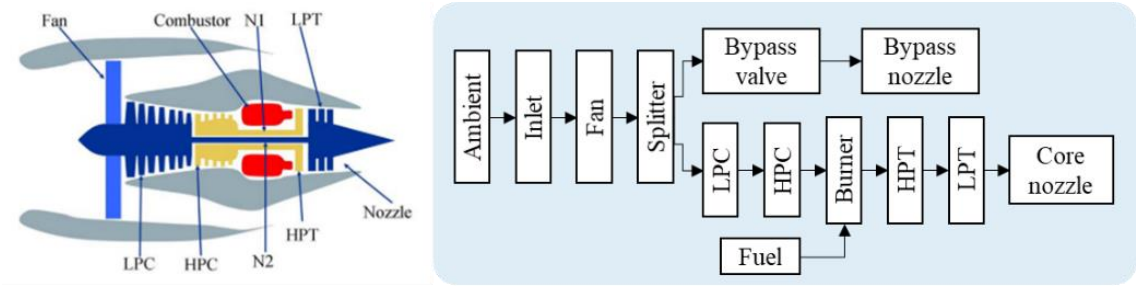


Figure 4.10. C-MAPSS turbofan engine: System design as simulated in C-MAPSS [55]

Previous work using this dataset focused on remaining useful life (RUL) prediction [59,65]. In these studies, sensor data was divided into three categories according to the data trend; ascending, descending and irregular/constant. Data that does not exhibit an ascending or descending trend over time (uniform) is not viable for RUL prediction and was therefore discounted from the dataset. The previous case study showed that constant, uniform parameters do not contribute to the uncertainty. Therefore, the same approach is applied here. A description of the 14 included sensors is given in Table 4.7.

Step 2a: Calculate quantitative uncertainties. The sensor data was indexed and divided into 16 sub-arrays consisting of 12 rows by Eq. 4-1. The mean and deviation of each array

were calculated up to the point of failure. This is illustrated for 4 of the 14 inputs in Figure D.2. A comparison of sub-array size to the mean deviation is given in Figure D.3. Other than for the derivation of pedigree factors in Step 2b, the illustrated results up to Step 6 give an example for the first sub-array unit. A summary of the quantitative sensor data for this example is given in Table 4.9.

Table 4.7. C-MAPSS turbofan engine: Detailed description of sensors [55]

Sensor number	Notation	Description	Unit
2	T24	Total temperature at LPC inlet	°R (Rankine scale)
3	T30	Total temperature at HPC inlet	°R
4	T50	Total temperature at LPT inlet	°R
7	P30	Total pressure at HPC outlet	psi abs. (pounds per square inch, abs.)
8	Nf	Physical fan speed	rpm (revolutions per minute)
9	Nc	Physical core speed	rpm
11	Ps30	Static pressure at HPC outlet	psi abs.
12	Phi	Ratio of fuel flow to Ps30	psi
13	NRf	Corrected fan speed	rpm
14	NRc	Corrected core speed	rpm
15	BPR	Bypass ratio	–
17	htBleed	Bleed enthalpy	–
20	W31	HPT coolant bleed	lbm/s (pound mass per second)
21	W32	LPT coolant bleed	lbm/s

Step 2b: Calculate qualitative uncertainties. Random noise models were used to propagate qualitative factors associated with the simulated data with a mix of distributions to give realistic results [55,59]. This was given as a combination of three core factors applied to all sensors: manufacturing and assembly variations (resulting in varying degrees of initial wear), process noise (factors not taken into account in modelling) and measurement noise. More in-depth factors concerning maintenance between flights and environmental operating conditions could be considered in practice. For this study, they are incorporated in the three core factors for the simulated data, scored against the pedigree criteria detailed in Table 4.8 [55]. Uncertainty indicators for each factor are illustrated in Figure 4.11, with GSD given as the square root of the uncertainty indicator.

As for the previous study, the scores are pseudo-randomly applied ±1 of the defined score circled in Figure 4.11 for each sub-array, scaled by Eq. 4-9.

Step 3: Assess correlations between parameters. Each sub-array consists of 12 data points. The ideal limit of ρ therefore needed to be set to a high level of 0.8, with a cut-off at 0.6. No significant correlations were present above 0.8, so the value was reduced incrementally to 0.78, for which significant correlation was detected between the pressure at the HPC outlet and turbine core speed (Figure 4.12a). While it is logical to expect a positive relationship between these parameters, notable in the plot, it was not maintained through the other 15 sub-arrays. This does not mean the relationship was not present, but that other dependencies were more prevalent below the limit of 0.8. When run for all data points, a positive trend was identified between the physical and corrected core speed of the engine (Figure 4.12b).

Table 4.8. C-MAPSS turbofan engine: Pedigree criteria

Score	1	2	3	4	5
Manufacturing and assembly variations	Negligible range of initial wear on components, not contributing to engine efficiency	Minimal range in initial wear on engine components	Notable range in initial wear on engine components, occasional reduction in engine efficiency	Notable range in initial wear on engine components, regular reduction in engine efficiency	High range in initial wear on engine components, high variance in engine efficiency
Process noise	Negligible trend in degradation trajectory, no noise	Minor trend in degradation trajectory, minimal noise	Minor trend in degradation trajectory, manageable noise	Significant trend in degradation trajectory, variable noise	Highly contaminated degradation trajectory
Measurement noise	Negligible sensor noise, no impact	Minimal sensor noise, minor impact, predictable trend	Notable random complex sensor noise, measurable impact	Significant random complex sensor noise, inaccurate impact measurement	High random complex sensor noise, tangible point estimate unobtainable

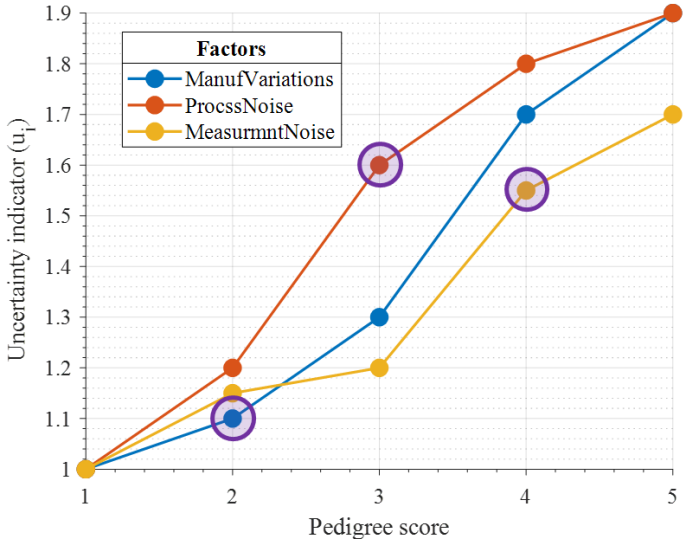


Figure 4.11. C-MAPSS turbofan engine: Uncertainty indicators for increasing pedigree scores

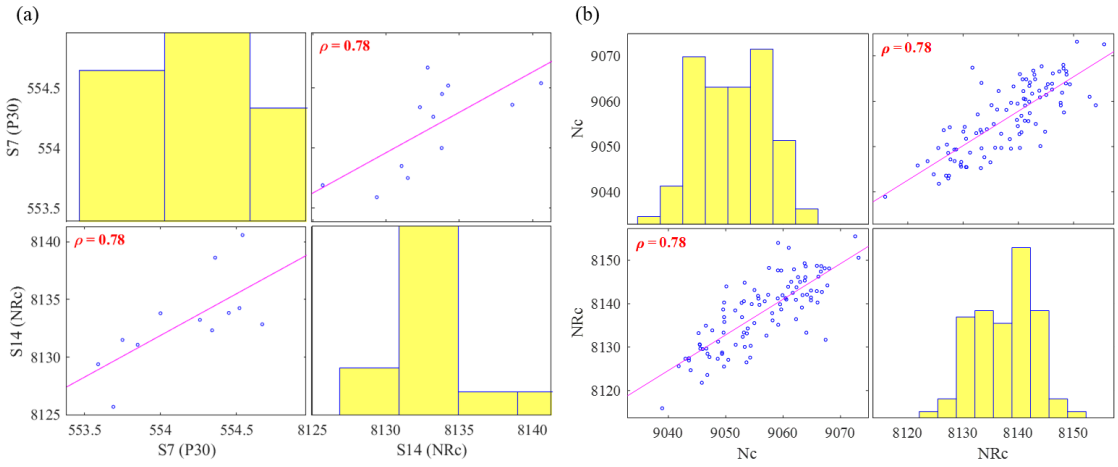


Figure 4.12. C-MAPSS turbofan engine: Significant correlations for which $|\boldsymbol{\rho}| \geq \mathbf{0.6}$

Step 4: Calculate respective CVs. Summary tables with calculated CV for each input are given in Table 4.9 and Table 4.10 for the quantitative and qualitative factors respectively. The majority of factors here are lognormally distributed by the goodness of fit tests.

Table 4.9. C-MAPSS turbofan engine: Recorded data and calculated parameters

Parameter	Distribution	Mean	Deviation	Min	Max	CV
S2 (T24)	Ln	642.20	1.0004	641.71	642.56	0.0004
S3 (T30)	Ln	1586.85	1.0026	1581.75	1592.32	0.0026
S4 (T50)	Ln	1400.76	1.0021	1394.80	1406.22	0.0021
S7 (P30)	Ln	554.17	1.0007	553.59	554.67	0.0007
S8 (Nf)	Ln	2388.05	1.0000	2388.00	2388.11	0.0000
S9 (Nc)	N	9049.55	4.9243	9040.80	9059.13	0.0005
S11 (Ps30)	Ln	47.25	1.0029	47.03	47.49	0.0029
S12 (Phi)	Ln	522.05	1.0008	521.40	522.86	0.0008
S13 (NRf)	Ln	2388.04	1.0000	2388.01	2388.08	0.0000
S14 (NRc)	Ln	8133.09	1.0005	8125.69	8140.58	0.0005
S15 (BPR)	Ln	8.41	1.0027	8.37	8.43	0.0027
S17 (htBleed)	Ln	391.75	1.0022	390.00	393.00	0.0022
S20 (W31)	Ln	38.99	1.0018	38.88	39.10	0.0018
S21 (W32)	Ln	23.40	1.0021	23.31	23.48	0.0021

Table 4.10. C-MAPSS turbofan engine: Pedigree factors with related GSD and CV

Factor	Distribution	Pedigree score	Uncertainty indicator	GSD	CV
ManufVariations	Lognormal	2	1.01	1.0488	0.0477
ProcssNoise	Lognormal	3	1.06	1.0954	0.0914
MeasurmntNoise	Lognormal	4	1.05	1.3038	0.2701

Step 5: **Combine CVs.** The combined CV is summarised in Table 4.11, aggregated for symmetric and asymmetric distributions and total CV with correlation. The visualisation in Figure 4.13 illustrates the relative CV of each input against the aggregated total for the example time unit.

Table 4.11. C-MAPSS turbofan engine: CV aggregation results

PDF	CV comb.	CV agg.	Corr.	CV_T
Ln recorded	0.00639	0.29049	2.7168e-07	0.2905
Ln pedigree	0.29042			
Norm. recorded	0.000544	0.000544		
Uni. recorded	0			

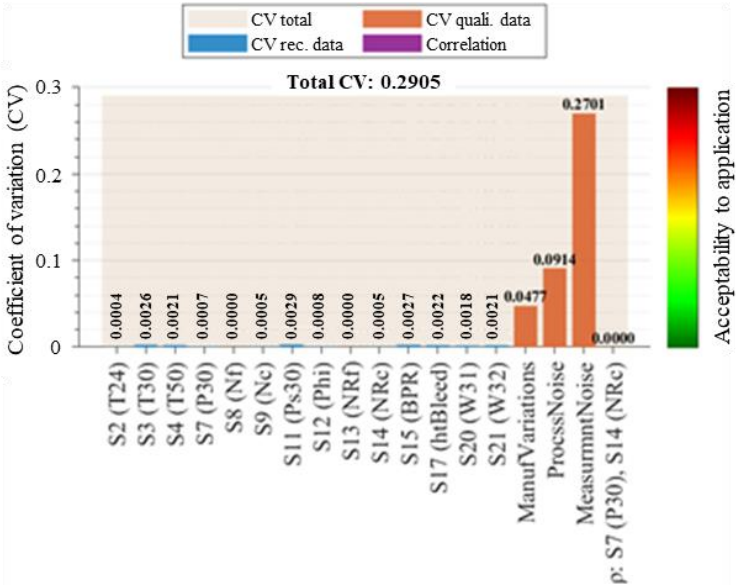


Figure 4.13. C-MAPSS turbofan engine: Aggregated total CV against individual factors for one time unit

For the example time unit, the measured data has minimal uncertainty compared to the qualitative factors. Discounting the qualitative factors here resulted in a 97.8% decrease in the aggregated CV from 0.2905 to 0.0065. The minimal quantitative uncertainty is due to the spread of the 12 data points in the sub-array. Increasing the number of data points increases the mean uncertainty depending on the variability in the dataset, but reduces the number of sub-arrays (Figure D.3). The change in individual and aggregated CV over all time units for $S_i = 12$ (16 sub-arrays) is given in Figure 4.14a and compared with $S_i = 3$ (64 sub-arrays) in Figure 4.14b.

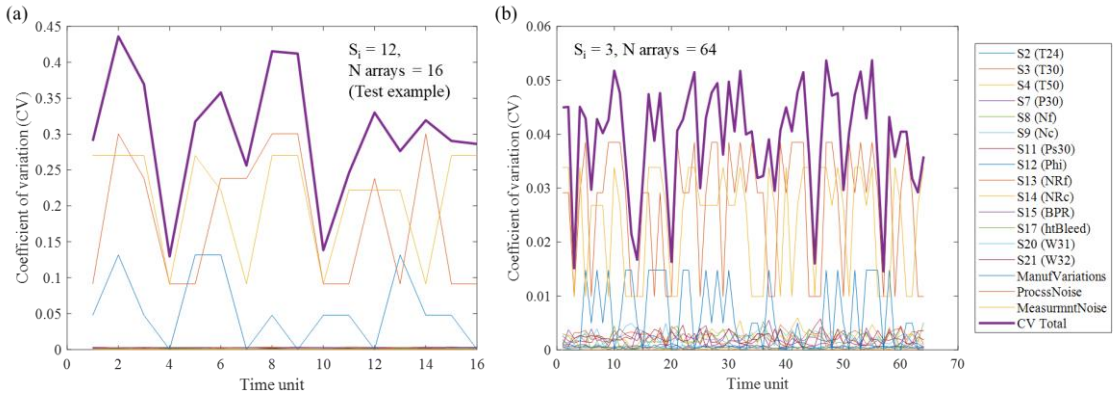


Figure 4.14. C-MAPSS turbofan engine: Aggregated total CV against individual factors over all time units for $\boldsymbol{S_i = 12}$ (a) and $\boldsymbol{S_i = 3}$ (b)

Step 6: GSA and visualisation. The relative influence of individual uncertainties on the aggregated total is plotted in Figure 4.15a and results after discounting the less influential parameters factors (Figure 4.15b). The quantitative parameters have a greater influence than the qualitative factors, despite them having a lower CV for all sub-array units. The most influential parameter uncertainty was T50 (temperature at LPT inlet) at 37%. Discounting parameters with an impact < 5% results in Nc (turbine core speed) having a dominating influence, while T50 dropped to 9%. This is again due to the variation in the data points of each sub-array.

As for case study 1, the difference between one uncertainty indicator to another, defined in Figure 4.11, will influence the respective factor's sensitivity index owing to the pseudorandom score allocation.

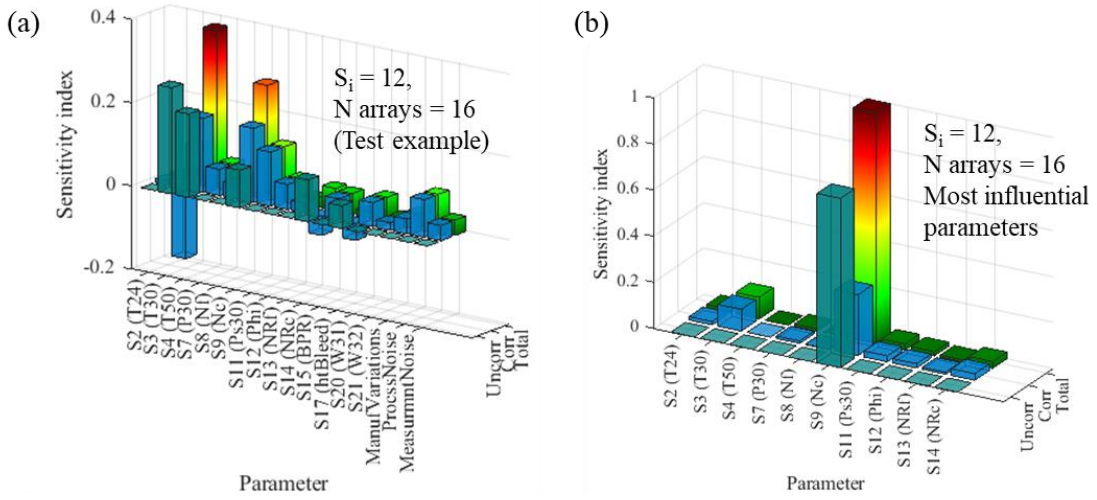


Figure 4.15. C-MAPSS turbofan engine: Sensitivity effects of individual to aggregated uncertainty over all time units for all factors (a) and most influential parameters (b)

4.4 Discussion and conclusions

The CUQA framework presented in this chapter was designed to enhance system reliability measurement in a manner applicable to complex and non-complex engineering systems through quantification and aggregation of compound uncertainties. These develop as a result of recording methods and assumptions made about the system and are modelled by different distribution types. The framework builds on existing literature to aggregate compound uncertainty considering dependant variables in the analysis, as well

as identification of the greatest contributing factors through GSA. Benefits of this framework include enhancements to performance assessment and corresponding maintenance planning for complex and non-complex engineering systems and respective subsystems.

The framework was first applied to a bespoke heat exchanger test rig, which contributed a range of uncertainties that impacted measurement quality and accuracy. Three distributions were considered: lognormal, normal and uniform. All qualitative factors were lognormal [32]. The measured parameters were deemed valid, though true steady-state was not obtainable owing to the heating system [43]. The second case study implemented a simulated engine degradation dataset [55]. The majority of the selected sensors exhibited a lognormal distribution up to failure. The following paragraphs critique the effectiveness of the framework through the results of the two case studies, concluding with a summary of the contributions and recommendations for future work.

The CUQA framework is capable of assessing uncertainty for nonhomogeneous input data. The user can view and select the best-suited distribution for each input via 'goodness-of-fit' tests. While effective for a small number of inputs, an automated method would prove more efficient for more complex systems. Monte Carlo simulation was used in Step 2a to give a homogeneous array size, enabling level consideration of each input. Monte Carlo was selected due to its flexibility with multiple distributions [9]. The inherently random nature of the simulation, though within respective distribution parameters, causes different results each time the experiment is run, which may impact the accuracy of parameter values. Other techniques such as Latin Hypercube Sampling (LHS) and Taylor series expansion may provide samples tighter to the respective mean, but do not show the same flexibility as Monte Carlo for multiple distribution types.

Splitting the input data into sub-arrays enabled uncertainty in the measured values to be determined over time. The greater the number of rows in each sub-array, the fewer arrays are allocated over the time series. The more arrays allocated, the more loops are performed between Steps 2 to 5, increasing execution time. It is therefore necessary to find a balance with optimum values in each sub-array, which was the purpose of the automatic selection by Eq. 4-1 (comparisons of mean deviation with increasing sub-array size are illustrated in Appendix D for the two case studies). Input parameters that do not

maintain a positive or negative trend require more sub-arrays to account for their variation. The framework allocates the same number of sub-arrays to each input to maintain equal consideration throughout the analysis. Flexible size allocation by individual input trend or average variance rather than sample size warrants further investigation.

Step 2b defined uncertainty indicators associated with qualitative inputs. These are ideally defined by multiple sources such as surveys, interviews and historical trends. The mean indicator is taken to calculate the geometric standard deviation (GSD). Naturally, high uncertainty reflects low confidence in the measured parameter. While the use of GSD overcomes scale dependency in measured data, the resulting coefficient of variation (CV) was found to be considerably lower than that of normally distributed data and the qualitative factors attributed by the pedigree matrix. This is due to the number of data points in the sub-array unit. Uncertainty indicators for the qualitative factors were initially assigned on a scale between 1 and 2, and the square root calculated to give the GSD [32]. These were rescaled by Eq. 4-9 to give a more equal comparison to the quantitative data. This would however artificially reduce the aggregated total, and saw normally distributed parameters such as T_3 in case study 1 attributing the greatest influence over the aggregated total.

Significant correlations between input variables are defined via Spearman's rank coefficient in Step 3. The ability to define the ideal coefficient limit allows the user to define the desired level of detail of dependant variables. This can have a significant impact on the resulting estimate. The dependencies identified between parameter values did not impact the aggregated total of the two case studies in Step 5. However, the influence attributed by individual CVs to the aggregated total in Step 6 was shown to exhibit dependencies that warrant further investigation. Stronger dependencies between parameter values will have a greater influence on emergent behaviour in more complex systems.

The CV was adopted as the uncertainty measure in Step 4 to allow inputs of varying distribution types to be represented on an equal scale, enabling effective uncertainty quantification. Representing uncertainty by the CV proved effective to aggregate uncertainties represented by different distributions in Step 5, but further research is

required into the scaling of geometric against arithmetic standard deviations. Acceptable levels of uncertainty are user-defined according to the application and visualised by the colour scale. Conversion of further distribution types such as Weibull and non-parametric derivations will allow for the consideration of more complex datasets. Aggregating the individual CVs by a combination of the propagation of error method for symmetric CVs and the product of asymmetric CVs allowed an aggregated total estimate to be obtained. This can be used to determine how the aggregated uncertainty changes over time, which is converted back to the standard deviation and used as the response vector in Step 6.

Global sensitivity analysis (GSA) was employed to identify which individual uncertainties contribute the greatest influence to the aggregated total. The sampling method was applied by Groen [40] using matrix-based LCA. It was applied in this study using the individual uncertainties of each sub-array as the inputs and the aggregated uncertainty at each point as the response. It was deemed the best suited GSA method for the CUQA framework because it can be implemented with relatively small datasets and illustrates the influence of correlated and uncorrelated uncertainties against the total effects. While the sub-array derivation in Step 2a is more accurate with a greater number of rows in each sub-array, the number of sub-arrays affects the quality of the GSA over each unit. The removal of factors that do not contribute to the aggregated total (uniformly distributed or negligible for each iteration) allowed for a focused analysis on influential parameters in a second pass through the feedback loop. The risks formed as a result of these uncertainties can then be mitigated. More in-depth GSA at each time unit using methods such as Sobol indices would require derivation of model process equations for the system application, which is out of scope for this study.

Compared to complex engineering systems used in operational environments, case study 1 represented a relatively simple laboratory system set-up but served to prove the functionality of the CUQA framework as it exhibited uncertainties akin to those faced in such environments and presents comparable challenges to UQ. While the coefficients of correlated parameters fell between negligible error margins for both case studies with minimal risk, they may have a significant impact in real-world environments where operating conditions such as atmospheric temperatures or wind speeds will impact the accuracy of recorded data or subjective opinion.

The core contributions of the CUQA framework are:

1. Use of CV to enable effective quantification and aggregation of compound uncertainties represented by different distribution types
2. Assessment of correlation between compound parameters
3. GSA for dependant compound parameters
4. Intuitive visualisation of results – most significant parameters, greatest effects

The authors propose future work to derive uncertainty from non-parametric and stochastic distributions through clustering techniques. Further assessment of aggregated compound uncertainty is necessary, incorporating additional distribution types and improving the rigour of the GSA approach in variance decomposition for each sub-array time unit. The emergent behaviour of uncertainties should be forecast through the in-service life to determine when and where further mitigation may be required.

CHAPTER 5. UNCERTAINTY PREDICTION UNDER LIMITED DATA

Abstract

Engineering systems are growing in complexity, requiring increasingly intelligent and flexible methods to account for and predict uncertainties in service. Increasing complexity manifests varying degrees of quantitative and qualitative uncertainty over time, driven by the quality and availability of data, experience and knowledge of system performance. Under limited available data, existing approaches seldom consider how the resulting uncertainty may change over time, leading to under or over estimation of factors including maintenance costs, equipment availability and failure rates. To that end, this chapter presents a framework for dynamic uncertainty prediction under limited data (UPLD). The theory of spatial geometry is incorporated with LSTM networks to enable multistep prediction. This provides an element of self-validation with uncertainty given in real-time by the symmetry of the geometric shape area, given in vector space. The framework was tested and validated through two case studies: US SAR cost uncertainty data and simulated degradation of a turbofan engine. Results demonstrate robust prediction of trends in limited and dynamic uncertainty data with parallel determination of geometric symmetry at each point in time. Immersive visualisation of dynamic uncertainty is presented. Future work is recommended to explore alternative network architectures suited to limited data and development for applications of visualisation in augmented reality.

Paper 6 Initial conference paper: Dynamic multistep uncertainty prediction in spatial geometry

Published: Procedia CIRP, CIRPe Web Conference 2020
DOI: 10.1016/j.procir.2021.01.055
Data access: 10.17862/cranfield.rd.12906716.v1

Paper 7 Multistep prediction of dynamic uncertainty under limited data

Published: CIRP Journal of Manufacturing, Science and Technology
DOI: 10.1016/j.cirpj.2022.01.002
Data access: 10.17862/cranfield.rd.14381987

5.1 Introduction

The growing complexity in engineering systems manifests a range of uncertainty surrounding in-service maintenance. Such systems are comprised of various equipment units, many of which are maintained on a corrective or time-based basis. Unexpected failures outside planned maintenance periods require reactive maintenance to repair or replace units. Sampling rates of maintenance data in this context are often sporadic due to manual recording methods and disjointed signals from equipment units. The resulting quality and availability of data, as well as the influence of expert experience, assumptions, and environmental operating conditions, drive uncertainty that increases the likelihood of under or overestimating factors such as turnaround times, equipment availability and resulting costs [13,194,195]. This can lead to increased failure rates or, more often, unnecessary maintenance carried out. Accommodating for uncertainty requires the determination of key contributors, their influence on interconnected units how this might change over time.

Limited available or poor-quality data directly hinders forecast accuracy and robustness. Once quantified, predictions of the uncertainty in such data and assumptions made surrounding it can enhance decision-making capabilities for the maintenance of increasingly complex systems and equipment units. This sets the motivation of the chapter, which presents a framework to predict dynamic uncertainty under limited available time-series data. The framework is designed to be embedded in a range of systems such as engines, radar, and heating systems as well as uncertainty in associated maintenance costs. The aim is not to mitigate or reduce the uncertainty, but to provide a holistic view as to which factors require mitigation or may become an issue in the future.

The framework structure is detailed in Section 5.2, along with key mathematical formulae and assumptions made. Section 5.3 applies the framework to two use cases: US SAR cost uncertainty data and simulated turbofan engine degradation. Results of each step are given to illustrate the multistep prediction for each long-short term memory (LSTM) network allocation. Section 5.4 discusses the strengths and limitations of each step of the framework, while Section 5.5 summarises the study along with future work in this area.

5.2 Framework overview: Uncertainty prediction under limited data (UPLD)

This chapter contributes a framework for dynamic uncertainty prediction under limited data (UPLD). Spatial geometry is combined with LSTMs to enable covariant analysis of dynamic variables within state space, whereby a change in one variable will affect another. For each time step of the input sequence, the network learns to predict the value of the next time step. This work builds on a conceptual model presented in Grenyer et al. [196]. Here it is further developed and validated through two case studies. This chapter has been submitted as a manuscript and is under peer review at the time of writing. The steps were developed from emerging studies in literature utilising LSTM networks to forecast time-series data, extended to consider the geometric symmetry between input variances to improve prediction robustness under limited data. This addresses the third research gap from Chapter 2 and is achieved through a 5-step framework developed in MATLAB, described below and illustrated in Figure 5.1.

Step 1. Evaluate input topology. To examine interactions, uncertainties and knock-on effects within the system, its topology must first be defined. Input uncertainty data is given as a time series of changing variance, formatted as row vectors where each column represents one time unit. The number of rows gives the number of input dimensions. The variance data is scaled according to the range of each input dimension i over each time slice j by Eq. 5-1, where n is the number of input dimensions. Under limited data a robust standard deviation cannot be applied, making traditional standardisation methods with mean and standard deviation undesirable [28].

$$dScaled_{i,j} = \frac{data_{i.j} - dMin_i}{dRange_i} \cdot \left(1 - \frac{1}{n}\right) + \frac{1}{n} \tag{5-1}$$

Where: $dScaled$ = scaled dataset; $data$ = initial dataset; $dMin$ = minimum value of each input over the time series; $dRange$ = range of each input over the time series

The scaled variance data is split into training and test data according to a defined partition. The default partition is set to 60% to allow for a comparable proportion of observed and predicted values to determine prediction accuracy and robustness with varying input dimensions. The number of training steps is given by Eq. 5-2.

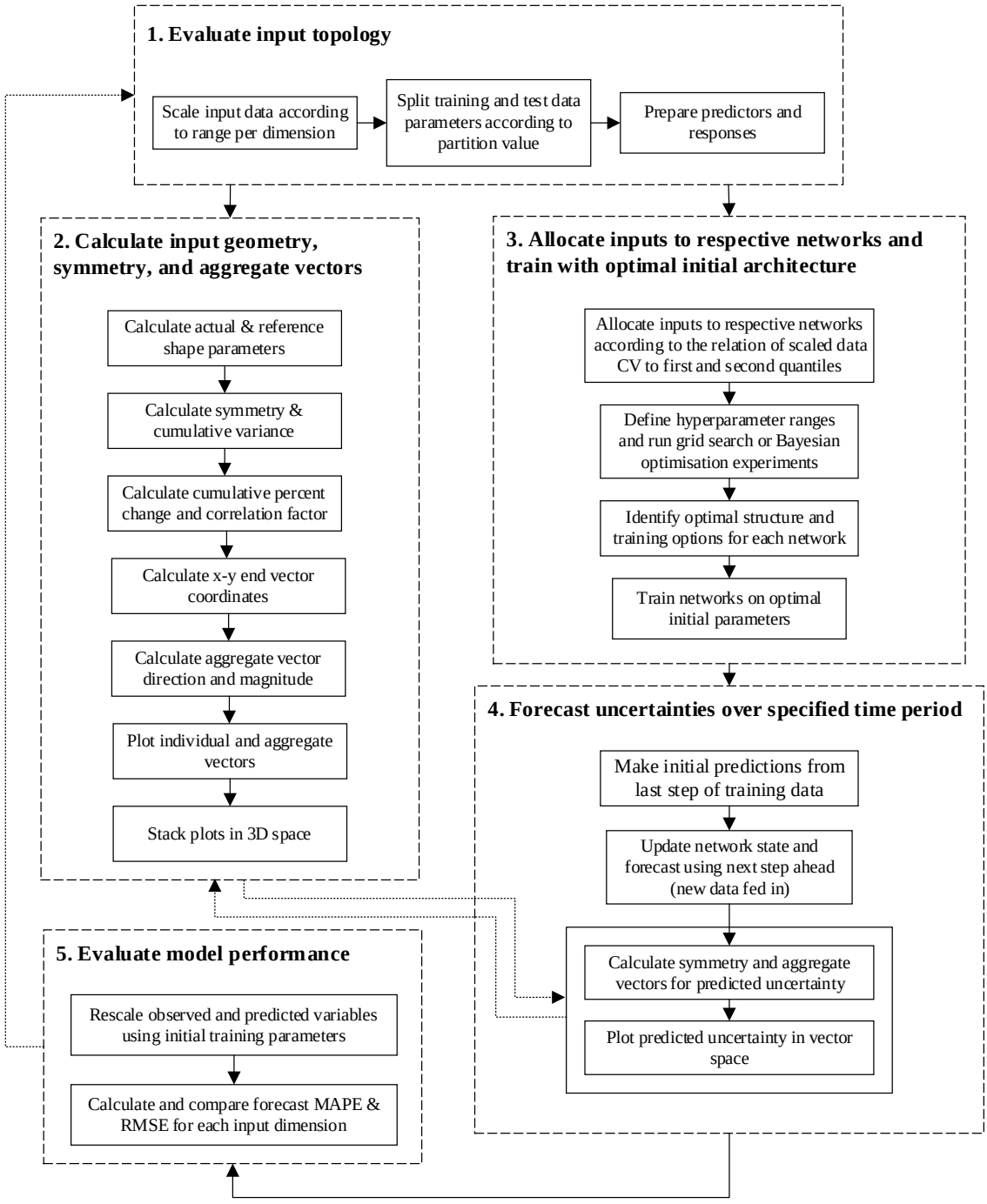


Figure 5.1. UPLD Framework overview

$$nStepsTrain = \lfloor P \cdot nStepsTotal \rfloor \tag{5-2}$$

Where: P = Partition; $nStepsTrain$ = Number of training steps; $nStepsTotal$ = Total number of time steps

The input to be forecast is the scaled training data. The forecast data is then compared against the test data to determine prediction error. In parallel, the symmetry and aggregate vectors are calculated for the training, test and predicted data, detailed in the next step, and compared in the same manner. The LSTM networks then take the next time step, update the network state and corresponding prediction.

Step 2. Calculate input geometry, symmetry, and aggregate vectors. Spatial geometry determines an uncertainty range based on the geometric symmetry between input variances for each available time unit via polar force-field analysis in vector space. The procedures in this step to calculate symmetry and vector coordinates are based on previous work by Schwabe et al. [28].

Symmetry is defined as the relationship between the actual shape area of the evaluated time slice and the maximum possible area from the created geometry, illustrated in Figure 5.2 by an example time slice with six input dimensions [28]. For each calculation in this step, the radial degree between each input vector and their input order (dimensional sequence) is kept constant [131].

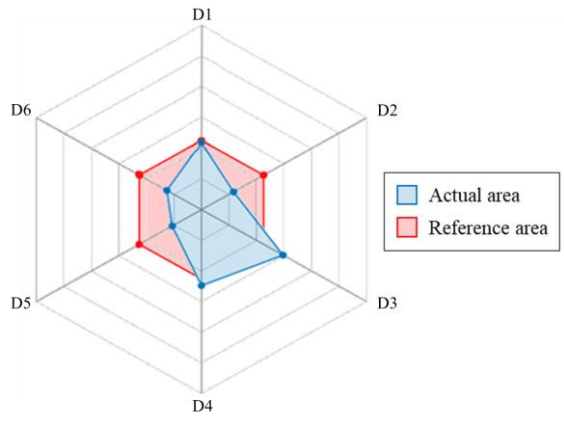


Figure 5.2. Spatial geometry actual vs. reference shape area example

Coordinate data points for vertices of the actual area shape are given by the scaled input variances for each time slice. The space between each vector dimension (D) out from the origin is a triangle (six in Figure 5.2). The sum of each triangle's area gives the full actual shape area. This is calculated by Eq. 5-3, where a_i and b_i are the respective magnitudes of each vector that make up the triangle sides and rad is the radial degree. The sum of the outer face lengths then gives the shape perimeter, calculated by Eq. 5-4.

$$Area_{Act} = \sum_{i=1}^{n} \left(\frac{1}{2} \cdot a_i \cdot b_i \cdot \sin(rad) \right) \tag{5-3}$$

$$Perimeter_{Act} = \sum_{i=1}^{n} \sqrt{a_i^2 + b_i^2 - 2(a_i \cdot b_i) \cdot \cos(rad)} \tag{5-4}$$

The reference shape perimeter is calculated by the mean of the outer face lengths multiplied by the number of input dimensions. This creates a regular polygon, for which the apothem (line from centre to midpoint of each side) is given by Eq. 5-5. This is then used to calculate the reference shape vertex magnitude (Eq. 5-6), which in turn is used to calculate the reference shape area (Eq. 5-7). The symmetry between the actual and reference shape areas is then calculated by Eq. 5-8 [28]. Spatial geometry uses a ring topology, analysing the linear progression of symmetry. There is a positive correlation between the percent change in the cumulative increase of actual area and symmetry. The correlation factor for each time slice can be used to determine an uncertainty metric against the baseline estimate. This is the most likely or best guess value of the data point from which the input variance is obtained, explored further in Section 5.3.

$$Apothem_{Ref} = \sqrt{FaceLength_{Ref}^2 - \left(\frac{FaceLength_{Ref}^2}{4} \right)} \tag{5-5}$$

$$Vertex_{Ref} = \sqrt{Apothem_{Ref}^2 + \left(\frac{FaceLength_{Ref}}{2} \right)^2} \tag{5-6}$$

$$Area_{Ref} = \frac{1}{2} \cdot n \cdot Vertex_{Ref}^2 \cdot sin\,(rad) \tag{5-7}$$

$$Symmetry = \frac{Area_{Act}}{Area_{Ref}} \tag{5-8}$$

To plot the change in shape geometry over time, X and Y end vector coordinates for each dimension, *i* over the time period, *j* are obtained by Eq. 5-9, iterated through each radial degree around the unit circle [196]. The sum of these points identifies the aggregate vector (Eq. 5-10), whose magnitude is given by Eq. 5-11.

$$absEndX_{i,j} = cos(rad) \cdot dScaled_{i,j}$$
$$absEndY_{i,j} = sin(rad) \cdot dScaled_{i,j} \quad (5\text{-}9)$$

$$aggVectX_j = \sum_{i=1}^{n} absEndX_i$$
$$aggVectY_j = \sum_{i=1}^{n} absEndY_i \quad (5\text{-}10)$$

$$aggVectMag_j = \sqrt{aggVectX_j^2 + aggVectY_j^2} \quad (5\text{-}11)$$

The aggregate vector magnitude and degree are assumed to represent the source of greatest uncertainty for each time slice. The resulting plots for each time unit are stacked to illustrate a dynamic change in the uncertainty of each input and aggregated vectors over time. An example illustration is given in Figure 5.3 for six input dimensions, with aggregate vectors removed for illustrative purposes.

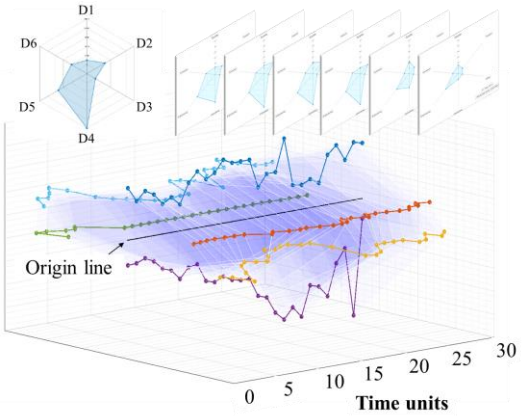


Figure 5.3. Stacked plot example [196]

Step 3. Allocate inputs to respective networks and train with optimal initial architecture. While some inputs are relatively constant, others can vary significantly over the time series. The mix of dynamic and comparatively constant trends on a single network limits that network's ability to accurately and robustly forecast future time steps. To reduce under or over estimation, training is split across three networks with different architectures and initial training options.

Different parameters are applied for different ranges of data according to the relation of the coefficient of variation (CV) of the scaled data to the first and second quantiles of

each input dimension. CV is a dimensionless measure of relative variability, given by the ratio of the standard deviation to the mean [30,34]. This is illustrated in Figure 5.4, with the networks hence referred to as "LSTM networks".

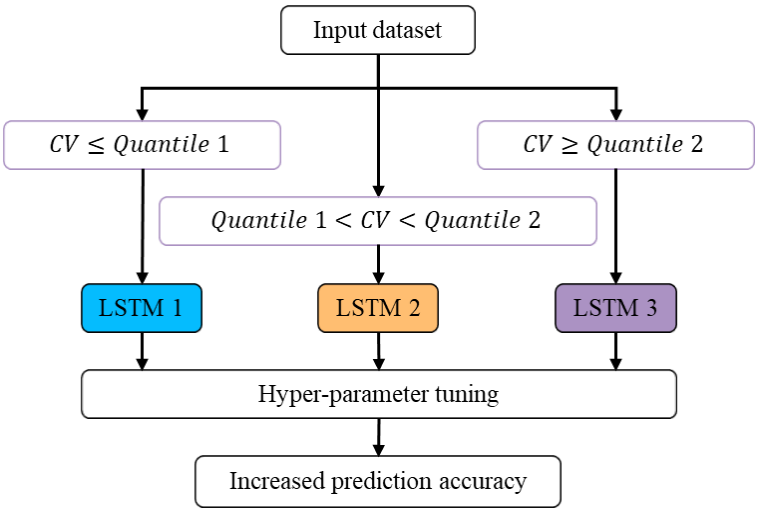


Figure 5.4. LSTM network allocation according to input parameters

Each network has a variable structure and range of training options to best suit the variability in the data applied to it. The best of these, i.e., the combination that gives the lowest prediction error, is determined through hyperparameter tuning according to the mean absolute percentage error (MAPE), discussed further in Step 5.

As shown in Figure 5.5, the variable structure for each network consists of 1-3 LSTM layers, each with 100-250 hidden units, a rectified linear unit (ReLU) layer, dropout layer (50-90%) and a regression output layer. Increasing the number of LSTM layers can make predictions more robust but also increases computation time [168]. The ReLU layer simply sets any value less than zero to zero, avoiding the vanishing gradient problem found in tanh and sigmoid functions [163,164]. The dropout layer then sets input sequences below a defined probability to zero to prevent overfitting [158,163]. The fully connected layers compile all neurons in the previous layer to a defined output size. The final regression output layer computes the half-mean-squared-error loss of the output.

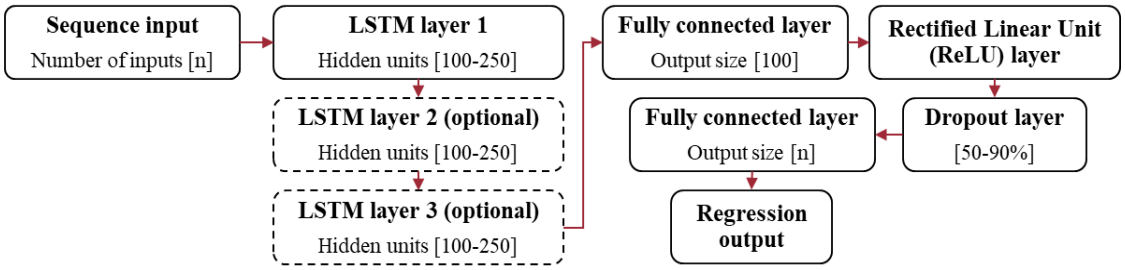


Figure 5.5. Hyperparameter setup metrics: Network structure

The range of training options compared by hyperparameter tuning is denoted in Table 5.1. Three solvers are compared: Adaptive moment estimation (Adam), Stochastic gradient descent with momentum (SGDM) and root mean square propagation (RMS prop). Each of these are variations of gradient descent algorithms that update network parameters (weights and biases) to minimize prediction error by taking steps towards the negative gradient of the loss function [59,160,163]. The number of epochs is the number of full passes over the training data. The learning rate controls the changes made to the model for every epoch.

Table 5.1. Hyperparameter setup metrics: Training options

Training options	Value range
Solver	Adam, SGDM, RMS prop
Max. Epochs	150-250
Initial learn rate	0.005-0.01
Learn rate drop factor	0.1-0.5

The optimal network structure and training options can be found by two methods: an exhaustive grid search, comparing every possible combination with set interval ranges for the parameters, or by Bayesian optimisation, where the software selectively alters a specified range of hyperparameters to minimise or maximise a selected evaluation metric. The three LSTM networks are then trained sequentially using the optimal hyperparameters.

Step 4. Forecast uncertainties over specified time period. The scaled variance data is forecast using the trained networks from the partition to the end of the initial dataset. Initial predictions are made using the last time step of the training response. When making

predictions using standardised data (according to mean and deviation) the same training data parameters are used for the test (observed) data [59,196]. Therefore, this approach uses the same range and minimum parameters from the training data to compare observed data against predicted data. The corresponding symmetry and aggregate vectors are calculated in parallel by Step 2 and compared in the same manner. The network state is updated to use observed values at each step in place of the predicted values to increase robustness [196]. The observed and predicted data is then plotted in vector space, stacked for each time slice.

Step 5. Evaluate model performance. Prediction error between the observed O_i and predicted P_i uncertainty can be attributed to the model parameters, unexpected changes in the inputs (causing no clear trend) and the amount of data available. Prior to performance evaluation, all variables are rescaled up to their original values by rearranging Eq. 5-1, using the initial range and min. parameters, given by Eq. 5-12. Common evaluation metrics are root-mean-square error (RMSE) (Eq. 5-13), mean absolute percentage error (MAPE) (Eq. 5-14) and custom score functions. RMSE is widely used in RUL prediction and regression problems. MAPE is a widely applied evaluation metric to determine forecast accuracy and robustness, providing a distinct percentage evaluation. As new data is recorded, the framework loops back to Step 1 to reassess the input topology and feeds through to performance evaluation.

$$data_{i,j} = \frac{dScaled_{i,j} - \frac{1}{n}}{\left(1 - \frac{1}{n}\right)} \cdot dRange_i + dMin_i \tag{5-12}$$

$$RSME = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \tag{5-13}$$

$$MAPE = \frac{1}{\mathrm{n}} \cdot \sum_{i=1}^{n} \left| \frac{O_i - P_i}{O_i} \right| \tag{5-14}$$

5.3 Framework implementation and results

5.3.1 Case study 1: US SAR data

The initial spatial geometry approach utilised US Department of Defense Air Force Selected Acquisition Report (SAR) summary tables to test and validate the method [28,131]. This is made up of a mixture of summary cost data over various phases of product life cycles in aerospace, land and sea sectors. Cost variances used were considered significant enough to require monitoring by stakeholders [28]. The same dataset is applied here to provide comparable consistency in the application and demonstrate the wide applicability of the framework.

Step 1. Evaluate input topology. Annual cost variances in US $ Mil are given over the life cycle of a range of US Air Force military platforms for a 28-year period from 1986-2013. Further detail is given by Schwabe et al. [28,131]. The data is categorized into 6 cost variance factors and formatted as absolute integers as [196]:

- Quantity: Change in the number of units of an end item of equipment.
- Schedule: Change in procurement or delivery schedule, completion date, development, or production milestone.
- Engineering: Alterations to physical or functional characteristics of a system.
- Estimating: Correction of previous estimating errors or refinements of current estimates.
- Other: Unforeseeable events not covered in any other category (e.g., natural disaster or strike).
- Support: Cost changes for support equipment of major hardware items not included in other costs.

Step 2. Calculate input geometry, symmetry, and vector coordinates. Following Eqs. 5-2 to 5-8, the resulting change in the actual area, reference area and symmetry over time is illustrated in Figure 5.6. Initial observations can be made here to highlight the reduction in symmetry through to 2005, indicating an increase in the amount of information required to describe the shape. From here, the symmetry fluctuates up to the end of the observed period.

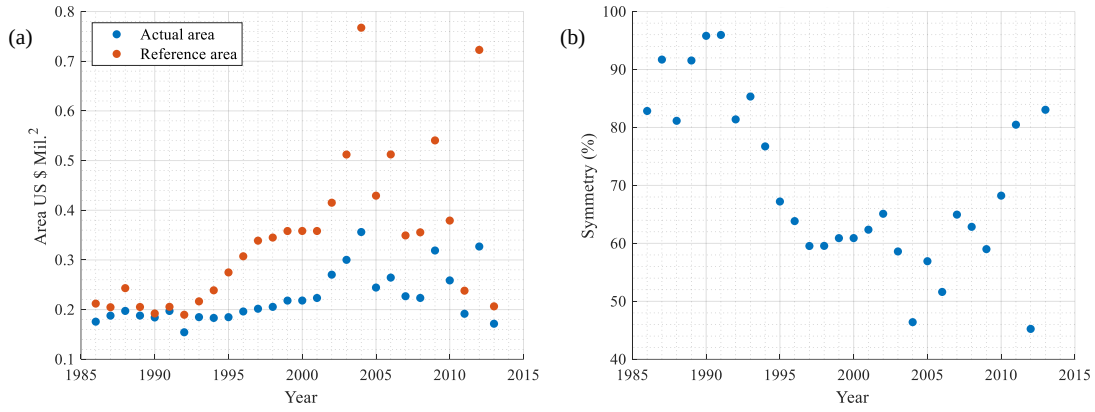


Figure 5.6. SAR data: (a) Change in actual and reference shape area over time and (b) change in symmetry

The cumulative increase of actual area and symmetry gave a linear trend for the observed period. The percent change for this increase between each unit is plotted in Figure 5.7a. This displays a negative exponential trend with a correlation coefficient of 0.95 (Figure 5.7b).

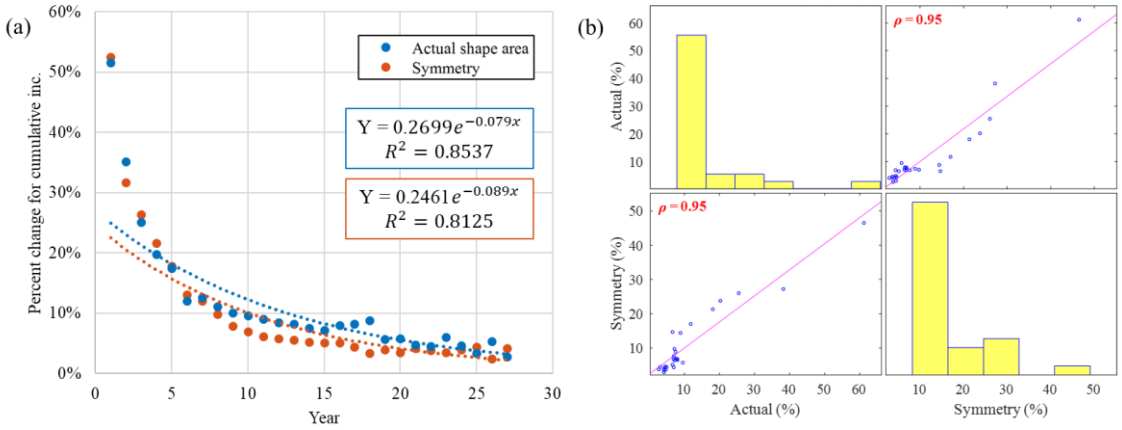


Figure 5.7. SAR data: (a) Percent change for cumulative increase and (b) correlation matrix

The gradient and intercept values from the actual area and symmetry trend line equations were plugged in for 100 time units. Their correlation factor, given by the $actualArea/symmetry$ is illustrated in Figure 5.8 with an R^2 value of 1. The shaded area shows the region of available data given by the 28 time units.

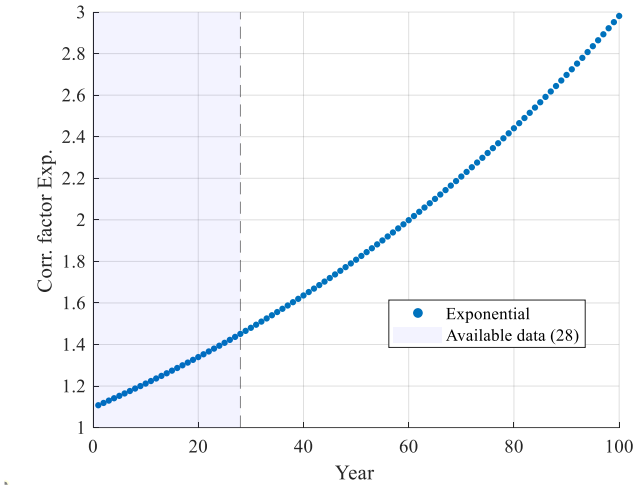


Figure 5.8. SAR data: Actual vs. symmetry correlation factor for exponential trend

An interesting phenomenon occurs when taking a linear trend line from Figure 5.7a and calculating the correlation factor in the same manner (Figure 5.9) where a lognormal relationship is displayed. The asymptote where Y=1 appears to meet the x-axis just prior to where the available data ends. The reasons for this warrant further investigation but are out of scope for this study.

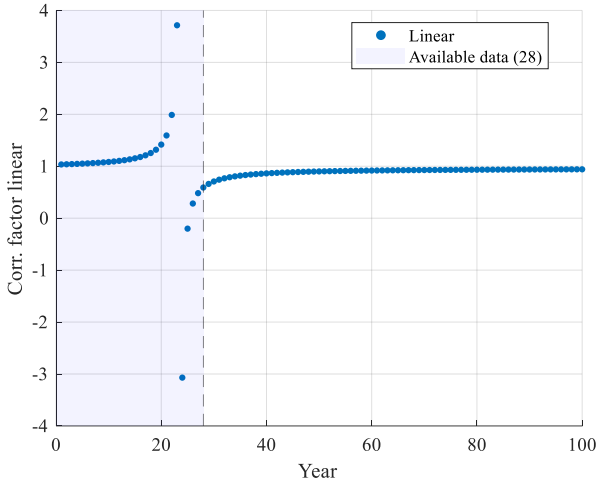


Figure 5.9. SAR data: Actual vs. symmetry correlation factor if assuming linear trend in cumulative % increase

Next, the X and Y end vector coordinates are calculated for each input dimension over the 28-year time period. The resulting endpoints and aggregated vectors for each input dimension are stacked and plotted in Figure 5.10. The dynamic shape area is shown in

Figure 5.10a by the white lines and blue fill between each vector coordinate. The aggregate vector magnitude in Figure 5.10b is visualised by the end marker, scaled up 40x for illustration. It can be seen here that the estimating factor prompts the greatest variance over the analysed period.

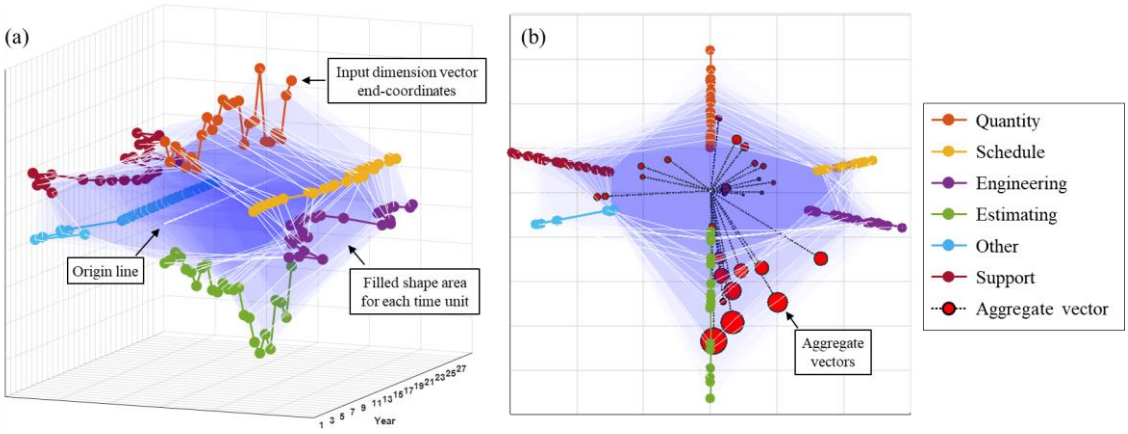


Figure 5.10. SAR data: (a) Stacked vector 3D plot and (b) face-on with aggregated vectors over 28-year period

It should also be noted that the radial degree between each input dimension is kept constant – in this case 1.0472 radians (60°). The apparent difference between e.g., Quantity-Schedule and Schedule-Engineering is due to the scaling of the figure produced in MATLAB.

Step 3. Allocate inputs to respective networks and train with optimal initial architecture. The range and deviation of each input used to train the network varies significantly. Using the CV as the deterministic parameter allows inputs with higher variation to be trained separately from those with lower deviation. Summary statistics are illustrated in Figure 5.11 and categorised into the relevant LSTM networks according to Figure 5.4. For the training data, this fits the Engineering factor into the first network, Quantity, Schedule, Estimating and Support into the second and Other into the third.

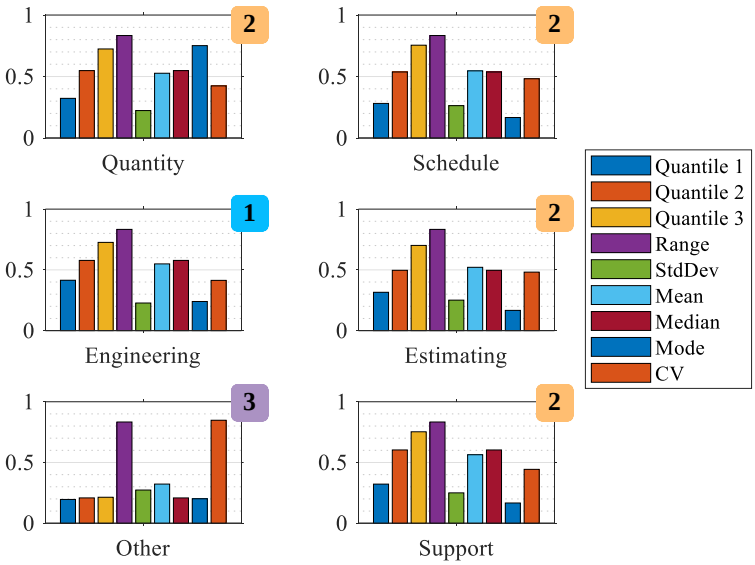


Figure 5.11. SAR data: Summary statistics for each input and corresponding LSTM network allocation

Hyperparameters were filtered through Bayesian optimisation to minimize the resulting average MAPE between inputs for the respective networks, using defined ranges (e.g., 100-250 hidden units). A maximum sweep time of 2hrs was set, which gave approximately 200 runs. The resulting initial network structure is illustrated in Figure 5.12. The optimal initial network training options are given in Table 5.2. For all three networks, the learning rate schedule was set to 'Piecewise' and gradient threshold to 1.

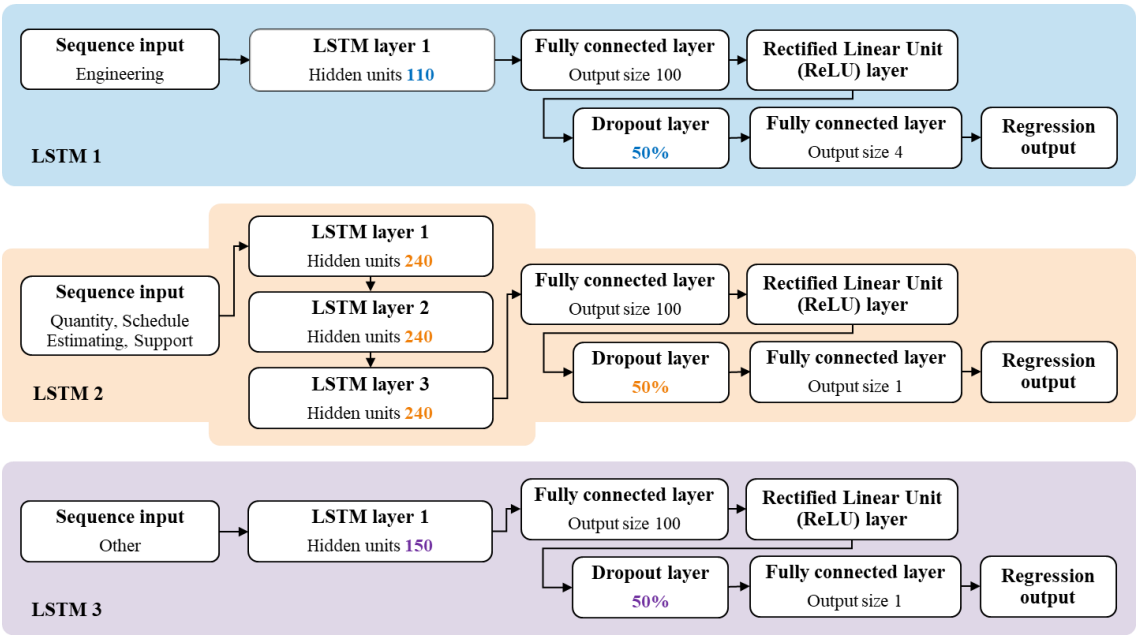


Figure 5.12. SAR data: LSTM network input allocation and structure following hyperparameter tuning

Table 5.2. SAR data: Defined training options following hyperparameter tuning

Training options	LSTM 1	LSTM 2	LSTM 3
Solver	SGDM	Adam	Adam
Max. Epochs	235	200	130
Initial Learn Rate	0.003	0.002	0.002
Learn rate drop factor	0.14	0.11	0.10

Step 4. Forecast uncertainties over specified time period. Initial predictions made using the trained networks are shown in Figure 5.13a, where the solid lines are the training data for each input dimension (partitioned at 60% of the full dataset), the dashed lines are the predictions, and the thin dotted lines are the actual (observed) data for the test period.

After resetting the network state, predictions were updated for each time step accounting for the previous step in Figure 5.13b. Following the updated prediction, it was observed that values for the Estimating factor now lie within the observed range. More detail in the prediction error can be seen in Figure 5.15.

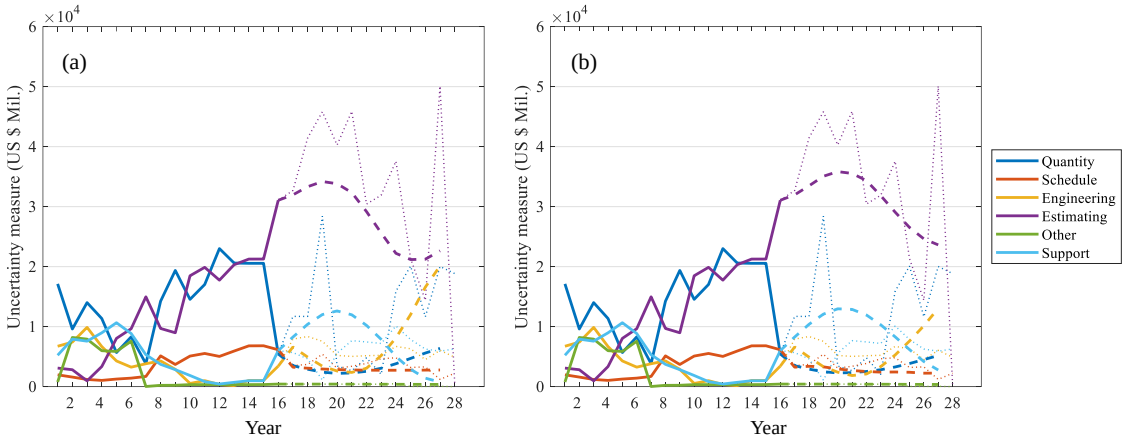


Figure 5.13. SAR data: Observed vs. predicted uncertainty for (a) initial forecast and (b) updated forecast

The corresponding symmetry, aggregate vectors and stacked vector plot built in Step 2 were calculated and updated in Table 5.3 and Figure 5.14. The percentage difference between the observed and predicted symmetry and aggregate vector magnitudes over the test period is directly influenced by the prediction error for each input dimension.

Table 5.3. SAR data: Observed vs. predicted symmetry and aggregate vectors

	Symmetry				Aggregate vector magnitude			
Time	Observed	Predicted	Diff.	% Diff.	Observed	Predicted	Diff.	% Diff.
18	81.44	87.71	6.27	7%	0.51	0.65	0.14	24%
19	76.35	71.30	5.05	7%	0.34	0.65	0.31	62%
20	64.47	60.25	4.23	7%	0.59	0.66	0.07	11%
21	53.23	60.66	7.43	13%	0.67	0.65	0.03	4%
22	62.47	65.30	2.82	4%	0.38	0.61	0.23	47%
23	55.55	74.49	18.94	29%	0.45	0.55	0.10	21%
24	71.55	88.60	17.05	21%	0.30	0.49	0.20	50%
25	76.11	97.51	21.40	25%	0.03	0.48	0.45	177%
26	80.37	80.68	0.31	0%	0.02	0.55	0.53	185%
27	54.18	42.64	11.54	24%	0.49	0.66	0.17	29%
28	72.53	27.33	45.20	91%	0.34	0.77	0.43	78%

Significant percentage difference in the aggregate vector magnitude for years 25 and 26 is due to the change in shape area, illustrated in the stacked 3D plot. Initial training data was removed for the illustration. Due to the significant difference in variance magnitude of the Estimating factor to all other factors, the aggregate vector direction is relatively unchanged other than in years 25 and 26. Further evaluation is made in the final step.

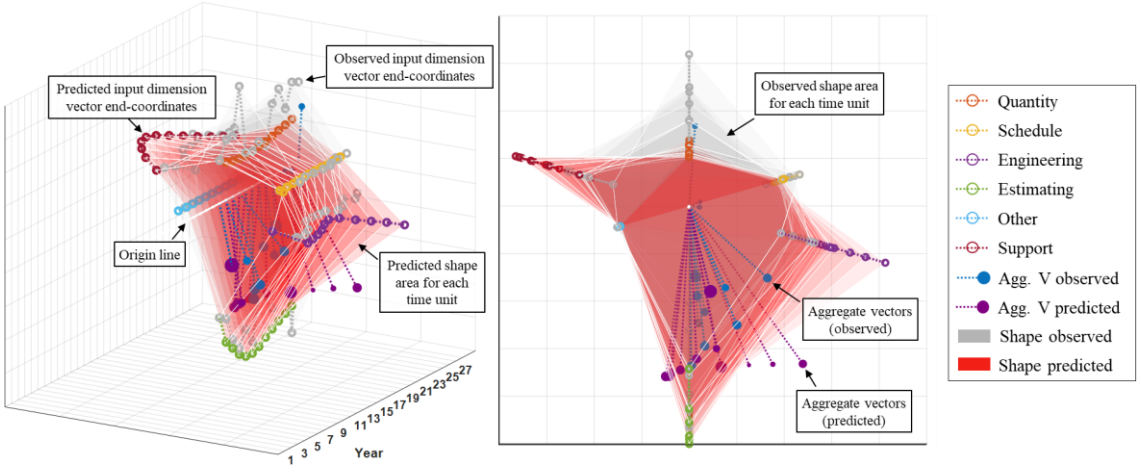


Figure 5.14. SAR data: Stacked 3D vector plot including observed and predicted data

Step 5. Evaluate model performance. The difference in the observed and predicted data is illustrated in Figure 5.15, scaled back up to the original variances, with corresponding line and stem plots for each input dimension. The stem plots show the difference in the observed to predicted data. Prediction error is noticeably variable over the time period for each of the six input dimensions. This is due to the quantity of data on which the networks were trained and the unpredictable peaks and troughs in the observed data.

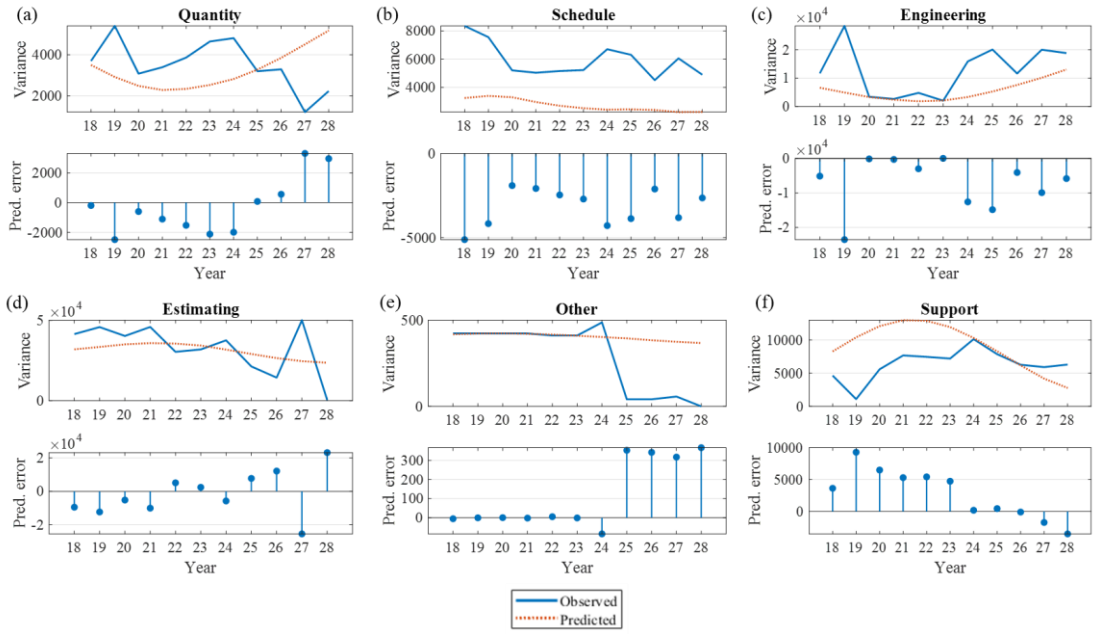


Figure 5.15. SAR data: Observed vs. predicted variance over the test period for each input dimension (a-f)

For example, the Schedule factor (Figure 5.15b) is underestimated but the overall downward trend is picked up in the prediction. The observed data of the Engineering and Estimating factors (Figure 5.15c-d) is scaled to 1.0×10^4 US $Mil. The relatively constant variance from year 20 to year 23 is accurately predicted, but the sudden increase was not predictable in the training data. As the model updated the multistep prediction, the increasing trend was identified up to year 28. Similarly, the Estimating factor (Figure 5.15d) was able to predict the overall downward trend in the test data period but not the sudden changes in variance.

The MAPE and RMSE are calculated in Table 5.4. The lowest MAPE was observed in the Engineering factor due to low prediction error in year 20 to year 23 and the following positive trend. While the estimating factor appears to hold the trend of the observed data, the scale of the variance means it has the highest MAPE and RMSE. The Other factor holds constant up to year 23 before an unexpected dive, which the network was not able to account for in the prediction.

Table 5.4. SAR data: MAPE and RMSE of observed vs. predicted values over the test period

Input	MAPE	RMSE
Quantity	140%	2003
Schedule	92%	3834
Engineering	64%	8363
Estimating	867%	17110
Other	381%	208
Support	248%	4200

These sudden changes and the scale in the observed variance data directly impact the mean prediction error, causing the high variation in MAPE and RMSE over the test period. While the predictions cannot be considered accurate, the ability to reflect the observed trends despite outliers in the observed data allows predictions to be deemed robust [197].

5.3.2 Case study 2: Turbofan engine degradation

As discussed in Section 2.4.4, a number of studies have applied a turbofan engine degradation dataset to forecast RUL using LSTMs, as well as other areas of prognostics and health management (PHM) [55,57,59,61,63,65]. Simulated using the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) tool, this publicly available dataset consists of four degradation scenarios. The FD001 training set was selected for this study because it consists of a range of quantitative data measured by sensors and qualitative factors given as noise. Uncertainties in the data and assumptions made were calculated by splitting the data into sub-arrays in Section 4.3.2. The resulting uncertainty data over 16 time cycles was applied to the framework to further demonstrate the capability to predict uncertainty under limited data.

Step 1. Evaluate input topology. The initial dataset consisted of 21 sensors measuring temperature, pressure and speed for 100 engine units, each with a random start time and normal operating level, running to failure. Previous studies using this dataset discounted any uniform sensor data as these will not change in any forecasts made or contribute to the uncertainty. The same approach is applied here, as well as discounting parameters whose individual uncertainty has a minimal impact on the aggregated uncertainty or

overall forecast. A description of the resulting 10 input dimensions forecast is given in Table 4.7. As for the SAR data case study, a partition of 60% was applied to split the training and test data, which was then scaled according to the range of the training data. The C-MAPSS dataset does consist of defined training and test sensor data for RUL prediction. The focus of this study is to forecast the uncertainty of that data, where there is limited previous data on which to base predictions. Using the training set from the database was therefore deemed sufficient.

Table 5.5. C-MAPSS data: Description of input dimensions to be forecast [55]

Sensor number	Notation	Description
3	T30	Total temperature at HPC inlet
4	T50	Total temperature at LPT inlet
7	P30	Total pressure at HPC outlet
9	Nc	Physical core speed
11	Ps30	Static pressure at HPC outlet
12	Phi	Ratio of fuel flow to Ps30
14	NRc	Corrected core speed
15	BPR	Bypass ratio
20	W31	HPT coolant bleed
-	-	Process noise

Step 2. Calculate input geometry, symmetry, and vector coordinates. The actual and reference shape areas and resulting symmetry over time are illustrated in Figure 5.16, given by Eqs. 5-2 to 5-8. For the 16 cycles observed, a trend cannot be identified for the shape areas or symmetry.

As for the SAR data, the percent change for the cumulative increase in actual area and symmetry shows a negative exponential trend (Figure 5.17a) with a highly significant correlation coefficient of 0.97 (Figure 5.17b).

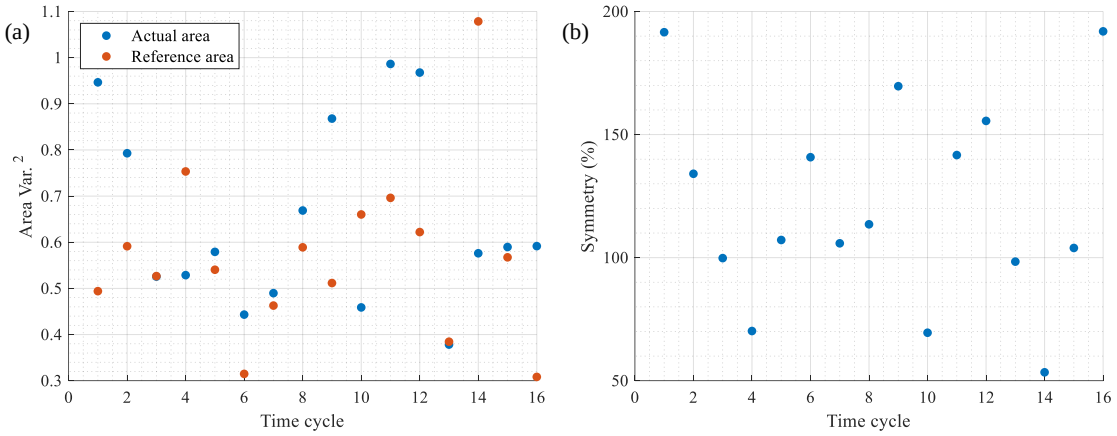


Figure 5.16. C-MAPSS data: (a) Change in actual and reference shape area over time and (b) change in symmetry

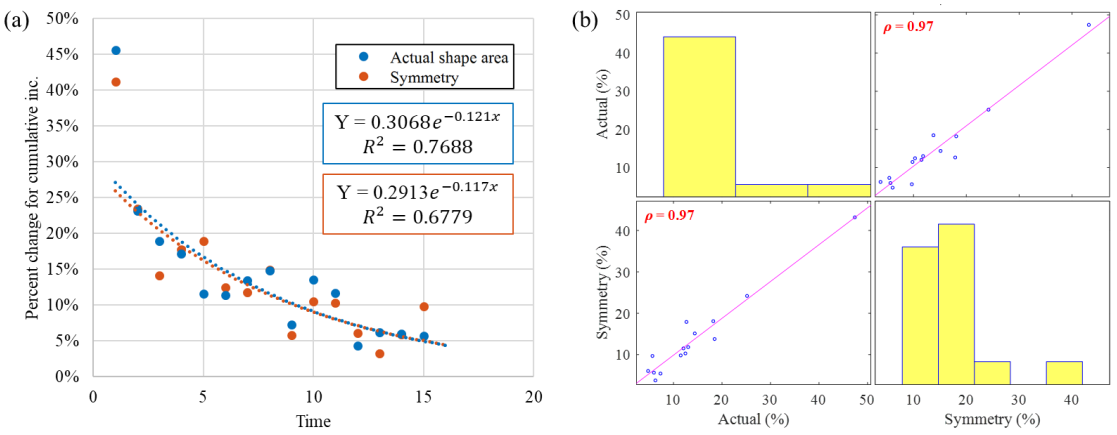


Figure 5.17. C-MAPSS data: (a) Percent change for cumulative increase and (b) correlation matrix

Following the line equations given by the actual area and symmetry, the correlation factor over 100 units given by the $actualArea/symmetry$ is illustrated in Figure 5.18 with an R^2 value of 1. The same asymptote trait occurred as for the SAR data when calculating the correlation factor assuming a linear trend line, occurring where the available data ends. The 16 time units signify uncertainty in the data up to engine failure. This plot is purely illustrative to expand the decreasing correlation factor. A key difference here to the SAR data is the opposite (negative) trend. As the variation and corresponding

uncertainty in the sensor data increases up to failure, the relation of the geometric shape area to its symmetry reduces.

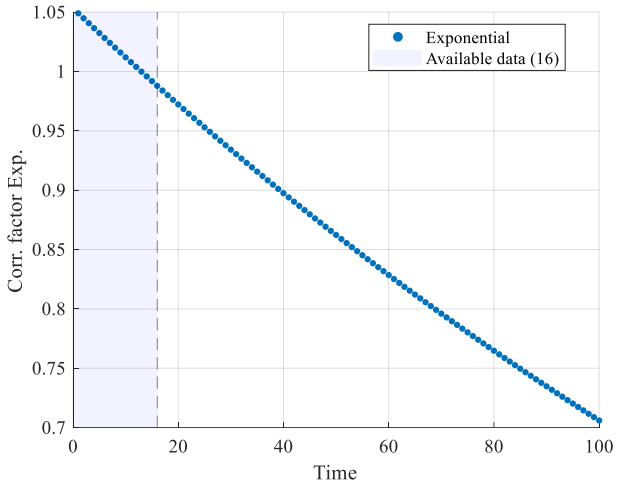


Figure 5.18. C-MAPSS data: Actual vs. symmetry correlation factor for exponential trend

The coordinate endpoints and aggregated vectors over the 16 time cycles for each input dimension are stacked and plotted in Figure 5.19. The shape area is starkly different here compared to the SAR data. This is due not only to the four additional input dimensions but also the contrast in variability between the dimensions about the radial degree (0.63 radians, 36°). The aggregate vector magnitude in Figure 5.10b is scaled up 10x for illustration and tend towards the low-pressure turbine inlet temperature (S4).

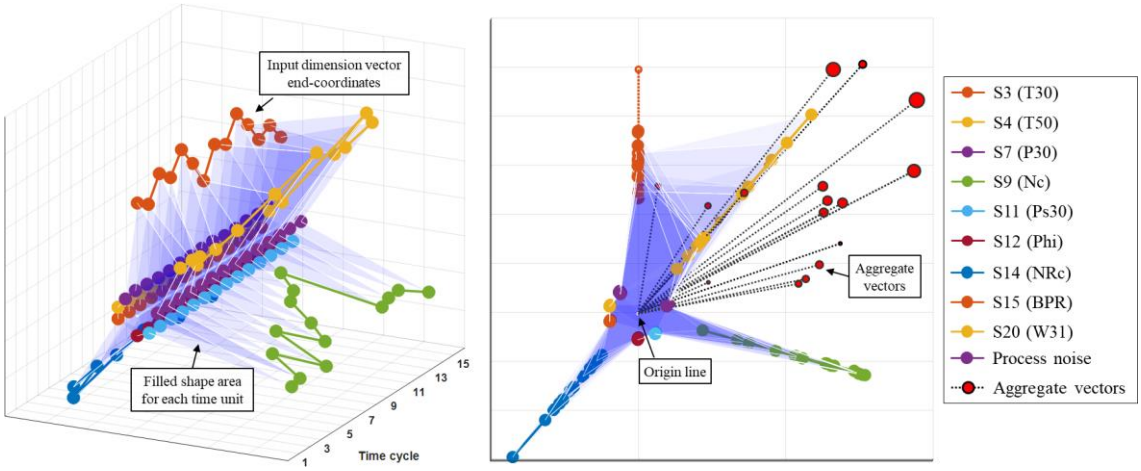


Figure 5.19. C-MAPSS data: Stacked vector 3D plot and face-on with aggregated vectors over 16 time units

Step 3. Allocate inputs to respective networks and train with optimal initial architecture. The variation in the uncertainty data is illustrated in the summary statistics in Figure 5.20. Categorised by Figure 5.4 according to the respective CV, the majority of dimensions fell into LSTM network 3. The turbine core speed (S9) was placed in LSTM 1 and the high-pressure turbine coolant bleed (S20) in LSTM 2.

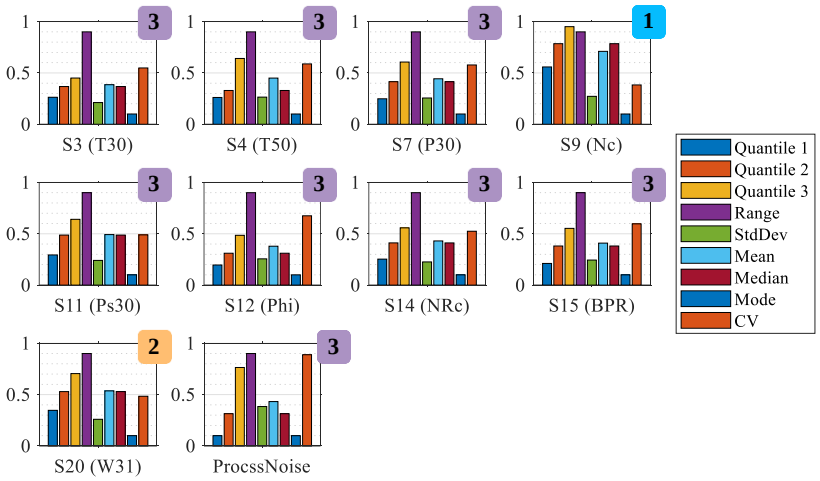


Figure 5.20. C-MAPSS data: Summary statistics for each input and corresponding LSTM network allocation

The hyperparameter ranges applied to identify the optimal initial network structure are the same limits as for the SAR data, trained using Bayesian optimisation for the same maximum of 2hrs. The resulting structure is summarised by Figure 5.21 and training options in Table 5.6. For all three networks, the learning rate schedule was set to Piecewise and gradient threshold to 1.

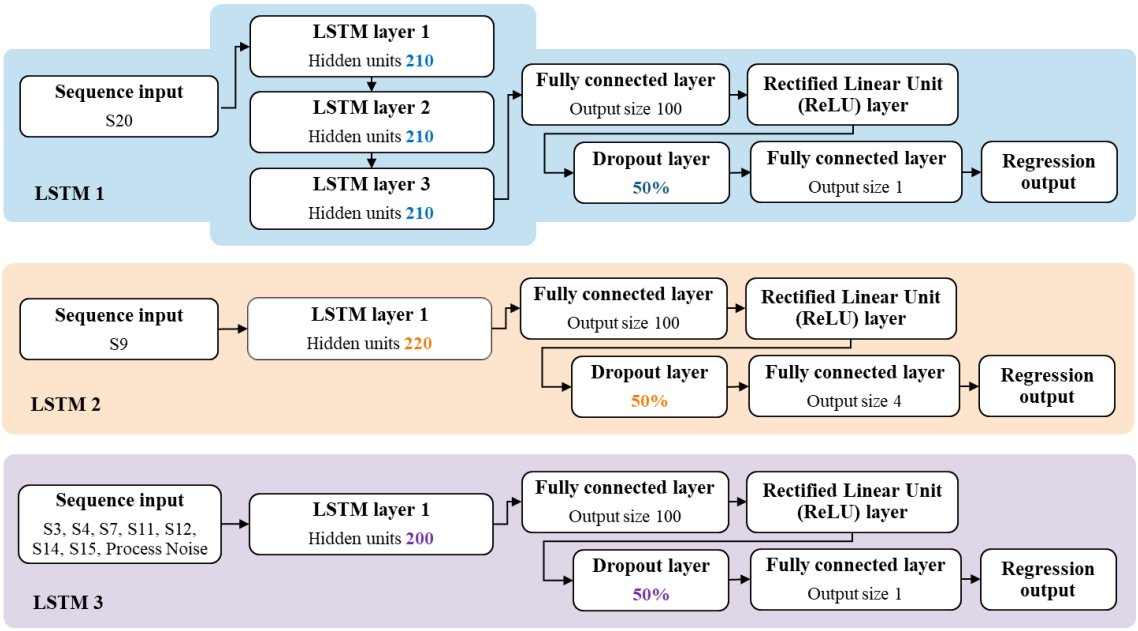


Figure 5.21. C-MAPSS data: LSTM network input allocation and structure following hyperparameter tuning

Table 5.6. C-MAPSS data: Defined training options following hyperparameter tuning

Training options	LSTM 1	LSTM 2	LSTM 3
Solver	Adam	Adam	Adam
Max. Epochs	220	180	120
Initial Learn Rate	0.009	0.009	0.021
Learn rate drop factor	0.23	0.061	0.138

Step 4. Forecast uncertainties over specified time period. Initial predictions made using the trained networks are shown in Figure 5.22a. As for the SAR data, the solid lines are the training data, dashed lines are the predictions and thin dotted lines are the observed data for the test period. Figure 5.22b shows the predictions after updating for each time step to account for the previous step. Changes in the predicted values are examined in Step 5. While not immediately noticeable in the plots, a reduction is noted in S3 (blue) from time cycle 12, a reduction in the negative gradient in S20 (orange) and a constant period in S7 between cycles 13 and 14.

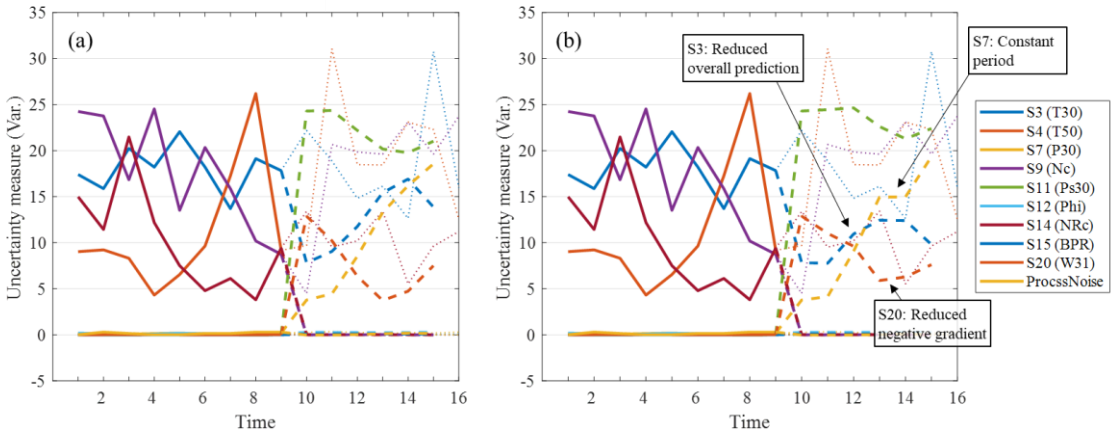


Figure 5.22. C-MAPSS data: Observed vs. predicted uncertainty for (a) initial forecast and (b) updated forecast

The corresponding symmetry and aggregate vectors from Step 2 are compared in Table 5.7 for each time step and through the stacked vector plot in Figure 5.23. The large difference in symmetry for time cycles 13 and 14 is due to the exploding gradients within the networks, which a higher dropout percentage could avoid. However, when tested at 80% rather than 50% the networks were found to give a constant line as the prediction. Further testing of combinations and degrees of dropout layers may alleviate the errors within the LSTM.

Table 5.7. C-MAPSS data: Observed vs. predicted symmetry and aggregate vector magnitude

	Symmetry				Aggregate vector magnitude			
Time	Observed	Predicted	Diff.	% Diff.	Observed	Predicted	Diff.	% Diff.
11	141.67	149.23	7.56	5%	1.28	0.33	0.95	118%
12	155.55	196.87	41.32	23%	0.83	0.40	0.42	69%
13	98.38	165.01	66.63	51%	0.75	0.56	0.19	29%
14	53.42	124.97	71.56	80%	1.10	0.78	0.32	34%
15	103.93	101.93	2.01	2%	1.27	0.87	0.40	37%
16	191.93	149.66	42.27	25%	0.74	0.86	0.12	15%

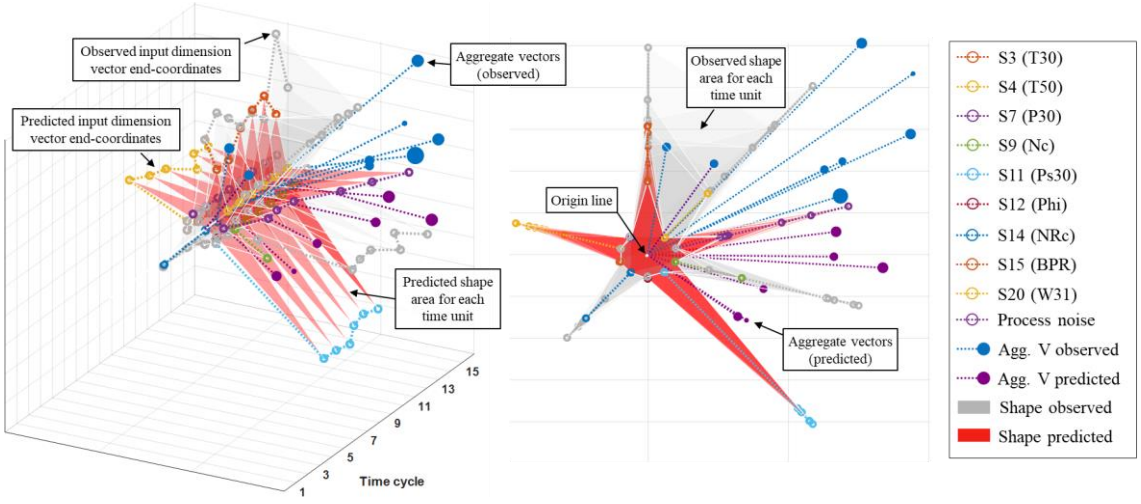


Figure 5.23. C-MAPSS data: Stacked 3D vector plot for observed and predicted data

The stacked vector plot demonstrates this further (initial training data was removed for the illustration). Notable differences in the observed and predicted data can be seen where the grey shape area (observed) is not covered by the red area (predicted). Caused mainly by S4, S9 and S14, these errors alter the resulting symmetry and corresponding aggregate vector magnitude and direction towards different factors.

Step 5. Evaluate model performance. The difference in the rescaled observed and predicted data is illustrated in Figure 5.24 in the line and stem plots for each input dimension. The significant prediction error for a number of factors is most likely due to the very limited number of steps on which it was trained and the lack of defined trends. Development of the network allocation methodology or inclusion of additional networks to train further variabilities in the data may improve robustness in the prediction. As for the previous case study, the multistep prediction was not able to pick up sudden changes in the variance data. Variances considered here are magnitudes smaller than those used in the previous case study and propagated over a smaller time period.

The tracing of overall positive or negative trends in the test data where they are apparent, such as for S7 (Figure 5.24c), S14 (g) and S15 (h), and predicting within the range boundaries of the observed data is therefore considered a satisfactory result. This case study illustrates the pitfalls of making predictions based on very limited data.

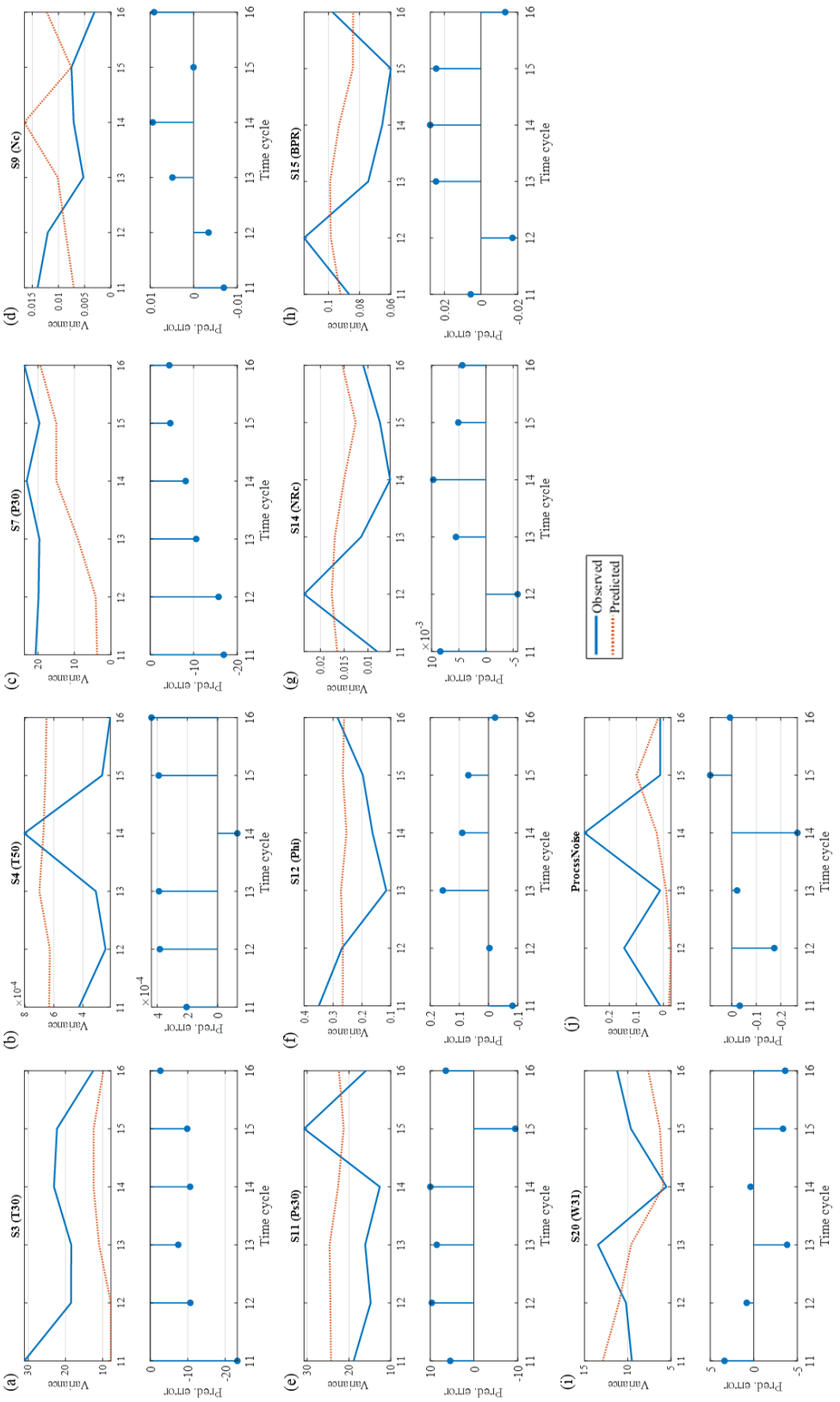


Figure 5.24. C-MAPSS data: Observed vs. predicted variance over test period for each input dimension (a-j)

The MAPE and RMSE over the observed period for each input dimension are given in Table 5.8. The range and scale of variances are comparatively small against the SAR data in Section 5.3.1, which is where the MAPE demonstrates the ability to better compare the prediction errors. Large variation in prediction error is due to unpredictable changes in the data. This naturally drives up the prediction error, as seen in time cycle 14 in S4 (Figure 5.24b). Even if the remainder of the test period has a very low prediction error, that increased error will increase the overall MAPE. As for Section 5.3.1, predictions are robust as they reflect observed trends despite outliers and limited data on which to train [197].

Table 5.8. C-MAPSS data: MAPE and RMSE of observed vs. predicted values over test period

Input	MAPE	RMSE
S3 (T30)	23%	12.52
S4 (T50)	70%	0.00
S7 (P30)	30%	2.43
S9 (Nc)	46%	0.01
S11 (Ps30)	26%	7.25
S12 (Phi)	20%	0.09
S14 (NRc)	40%	0.01
S15 (BPR)	16%	0.02
S20 (W31)	20%	3.30
ProcssNoise	314%	0.12

5.4 Discussion

The following paragraphs critique the framework steps through the results of the two case studies, concluding with an examination of industrial applications. Input uncertainty data for both case studies were given as a time series of variances, formatted as row vectors. The use of case studies in distinct domains demonstrated the framework's flexibility to be embedded in different systems.

The data was scaled according to the range of each input dimension. As stated in Section 5.2, the number of time steps available under limited data is unlikely to provide a robust deviation measure required for traditional standardisation methods. The scaling equation

(Eq. 5-1) can theoretically be applied to any format of data such as standard deviation or raw sensor data. The useability and results of using such formats in the framework have not been explored and may warrant further investigation. The scaled data was split into training and test data according to a defined partition, set to 60% to provide a comparable proportion of observed and predicted values to determine robustness of predictions. A lower partition would reduce the amount of data on which to train the networks, leading to reduced robustness, while an increased partition would reduce the data on which to test and update the networks and make predictions beyond the available time period. Comparisons with a varying partition would be beneficial for cases with a larger forecastable period, though up to a point the available data may no longer be considered "limited" and more traditional statistical approaches can come into play. Schwabe et al. [131] highlighted that at least 42 discrete time units are required for each parameter to make forecasts with statistical certainty under Kolmogorov complexity theory. Even when working with "big data", parametrics and statistics must be treated with caution prior to validation when significant correlations are unknown.

Spatial geometry was used as the uncertainty descriptor because of its ability to propagate interdependent cost uncertainties under limited data. Connecting outlying data points in vector space formed geometric shapes for each time slice, the area of which was used to determine the symmetrical relationship between inputs. This enabled a simplification of what may otherwise be complex conclusions [28,171]. The greater the symmetry, the greater the information entropy and therefore representative uncertainty for a given time slice. Symmetry and respective vector coordinates were calculated in Step 2 of the framework and run in parallel with Step 4 as uncertainties were predicted through the LSTM. The aggregated vectors for each time slice illustrated the greatest source of uncertainty and gave an indication of shape change for the next time interval, stacked in a point cloud in 3D space.

The visualisation provided an immersive view of shape change over time as well as the source of greatest uncertainty via the aggregated vector. These uncertainties require the most attention; be it mitigation, exploitation or simply increased awareness [196]. Employment of the 'shape' of data through spatial geometry for forecasting against the correlation of individual data points is a significant novelty in the developing field of big

data analysis. Live and continuous forecasts are beneficial to industry in several areas including maintenance planning and digital twins in the face of mounting increases in technological complexity.

To calculate the symmetry and aggregate vectors and build the 3D visualisation, defined parameters had to be fixed while others were allowed to change over time. Summarised in Table 5.9, key fixed parameters were the radial degree and dimensional sequence of inputs, while changeable parameters included the vector coordinates of each dimension over time. Altering the sequence of input dimensions would change the magnitude and direction of the aggregate vector but should maintain the shape area. Adding a new input dimension part-way into the space will alter the radial degree and require rescaling of the full dataset. Repercussions and allowances for altering fixed parameters warrant further research for spatial geometry.

Table 5.9. Spatial geometry taxonomy for fixed and changeable parameters

Fixed parameters	Translation space	Changeable parameters (over time)
• Scaling equation for all input dimensions • Shape area calculation • Symmetry calculation • Radial degree between inputs • Dimensional sequence of inputs • Origin location	• Layout/plotting functions for visualisation • Computational complexity	• Input dimension vector coordinates • Shape area • Symmetry • Aggregate vector direction and magnitude • Forecast most likely variance

The third step allocated the scaled inputs to one of three networks according to their coefficient of variation (CV) over the time period, then used hyperparameter tuning to define the optimal initial network structure and training options to yield robust predictions. The CV was used as the deterministic measure for network allocation because it provides a dimensionless measure of relative variability. Alternative measures such as the mean are affected by outliers, while the mode and standard deviation are not suitable for small sets of data. Other methods to define bins in which to allocate input dimensions should be explored, such as interquartile range, and variable allocation methods based on the amount of available data and respective variability. The allocation of input dimensions

to the three networks has a significant effect on the robustness of resulting predictions, making this one of the most important steps of the framework.

Initial parameters for the adaptable network architecture were defined through hyperparameter tuning. For both case studies, Bayesian optimisation was used to minimise the MAPE by comparing a defined range of parameters. Experiments were run for a maximum of 2hrs for each network. This computation time is not viable for regular updates when new data becomes available so was only used to gain an optimal initial setup. This does not necessarily give the best possible initial setup as not every combination can be tested with the time frame. An exhaustive grid search comparing each parameter iteration would not be viable without extensive computing power.

The allocation approach is similar to the semi-double-loop learning concept proposed by Putnik et al. [198]. This was used to select the best learning models for predictive maintenance scenarios. This method could prove effective in further development of the UPLD framework, where the application of double-loop learning principles in reinforcement learning would be used to help allocate input parameters and define initial network architecture.

The defined ranges for training options and network structure are detailed in Section 5.2. Different combinations will generate different results. Additional LSTM layers will typically improve prediction accuracy but take longer to train. The hidden units of each LSTM layer are equal. Comparisons of different sizes in each layer and additional training options such as mini-batch size may improve results with reduced computation time.

The final output layer of each network was a regression layer. Regression typically relies on statistical data sufficient to fulfil the Central Limit Theorem. Under limited data scenarios, this is not the case without artificial propagation through Monte Carlo simulation [83,145,196]. This may therefore lead to reduced robustness in predictions based on the available training data. Alternative, custom output layers should be explored to provide more robust predictions.

Uncertainty was predicted in Step 4 using the trained networks. The symmetry and individual and aggregate vectors were calculated for the predicted data via Step 2 running in parallel with the LSTMs. It is important to note the distinction of the forecast direction

where the LSTMs forecast through the time axis, while the symmetry and aggregate vectors are calculated for each predicted time unit. Predictions were rescaled to the original input variances and plotted to illustrate the difference in the observed and forecast uncertainty. The dropout layer prevented overfitting for each input parameter, but when set too high resulted in near-constant predicted values and reduced accuracy. The multistep model updated predictions as new data was fed in.

The initial, observed and predicted data were plotted in the stacked 3D vector space. These plots provided an immersive view of the shape area through time as well the aggregate vector magnitude and direction. However, the visualisation can become chaotic when too many parameters are visualised at once. Further developments detailed in Appendix F therefore allow selected parameters to be visualised and removed, as well as value labels and altering the shape area fill transparency.

The fifth and final step of the framework evaluated model performance via the MAPE and RMSE. Other evaluation metrics such as custom scoring functions should be developed to gauge the quality of uncertainty prediction and develop a methodology to identify areas where more data is required to allow comprehensive decisions to be made concerning equipment availability, turnaround time and unforeseen costs through the system life cycle.

The pertinence of the framework was discussed with key personnel from a leading defence company in four hours of semi-structured interviews. These included some participants from the initial studies detailed in Chapter 3. The questions posed are given in Appendix E. A large degree of uncertainty is portioned to numerous data repositories, maintenance formats and failure modes for different platforms. Sampling rates of maintenance data from different systems can have unpredictable gaps and varying sampling rates. The quality of signal reconstruction and determination of operational defects is used to determine when maintenance will be required. Rates of degradation or identification of other failure modes it is not always achievable. It was agreed that continuous forecasting of uncertainties resulting from these traits is vital to facilitate dependable maintenance costing and ensure equipment availability. Further work towards implementation is discussed at the end of the next section.

A direct comparison of the UPLD framework with traditional, probabilistic forecasting methods such as regression is not suitable because they are designed for large volumes of data that fulfil the Central Limit Theorem (as discussed above and in Chapter 2). Such models aim to forecast statistical data, not the uncertainty in that data and surrounding qualitative factors. They are therefore not appropriate under limited data scenarios.

A comparison of the percentage difference in symmetry given by predicted variables over the test period of case study 2 against that observed is therefore plotted in Figure 5.25. The results from Table 5.7 and Table 5.8 using UPLD are plotted against predictions made by linear regression and exponential smoothing, using the training data from case study 2. Predictions were also updated (upd) as per the UPLD framework by including the data of the previous time step for each iteration. The symmetry for the resulting predictions was calculated according to Step 2 of the UPLD framework to provide comparative data. The UPLD LSTM gives the lowest percentage difference to the observed symmetry, thus outperforming the other methods.

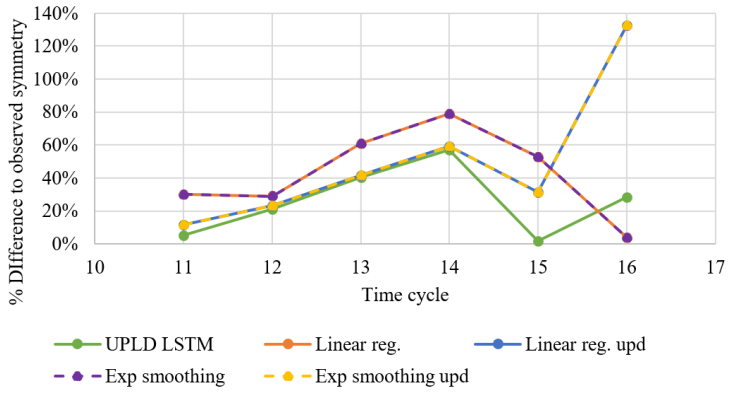


Figure 5.25. Forecast method comparison – percentage difference of observed and predicted symmetry

Further validation of the UPLD framework was made by testing its effectiveness with non-limited data. A simulated time series dataset was applied, consisting of 12 parameters over 50 time steps to provide enough points to consider statistical analysis under Kolmogorov complexity theory [6]. The percentage difference of the observed symmetry to the symmetry of data predicted by the UPLD LSTM is plotted in Figure 5.26 and compared against the same methods as in Figure 5.25. Plots of symmetry and correlation factor are available in the online supplementary data. The UPLD LSTM demonstrates an

overall similar performance to the other methods up to time unit 47, where the other methods see an increased percentage difference to the observed symmetry. This further demonstrates the capabilities of the LSTM to make robust predictions of time series data. Larger time series would not be considered 'limited data' and are therefore out of scope for the application of the UPLD framework.

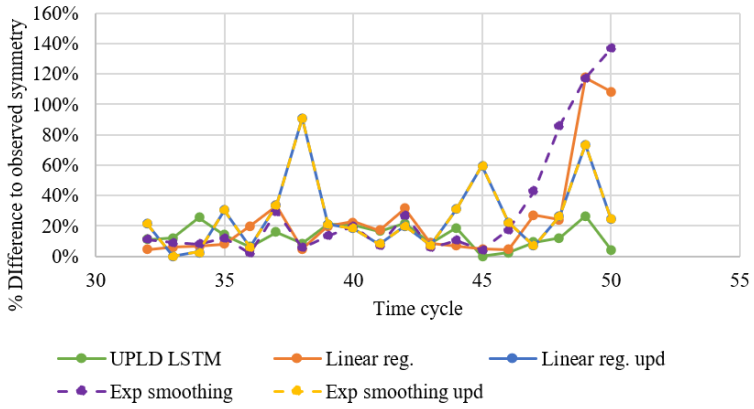


Figure 5.26. Forecast method comparison for extended time series data – percentage difference of observed and predicted symmetry

5.5 Conclusions and future work

This chapter presented a framework to predict dynamic uncertainty exhibited under limited data (UPLD) for the maintenance of increasingly complex engineering systems. These uncertainties arise as a result of data quality and availability, operating conditions and assumptions made surrounding maintenance. Coded in MATLAB, the framework was designed to be embedded in a variety of systems, building on supporting literature to develop a flexible forecasting model capable of making predictions under limited data from complex and non-complex factors without the need to develop precise models of physical systems. LSTMs were applied in parallel with spatial geometry to predict uncertainty in time-series data through the geometric symmetry between input dimensions. Additional benefits include the ability to update uncertainty predictions as new data becomes available by comparing initial predictions against the observed data and projecting forecasts through the visualisation of polar force fields in 3D vector space. This allows factors that may require future mitigation to be identified, which can, in turn,

reduce under or over estimation of turnaround times, equipment availability and resulting costs.

The framework was applied to two case studies in different contexts: Annual cost variances for a range of US Air Force military platforms (SAR data) [28,131] and precalculated uncertainties from a turbofan engine degradation simulation (C-MAPSS data) [55,57]. The SAR data consisted of six input dimensions with a widespread of variances over a 28-year period. The C-MAPSS data consisted of 10 dimensions made up of sensor data and process noise, with variances precalculated for 16 time cycles, determined initially from raw sensor data for RUL prediction [55]. Section 2.4.4 highlighted the wide use of LSTMs for RUL prediction, for which many studies use the C-MAPSS dataset. While the study does not use the dataset directly, findings on uncertainty in sensor and noise data will impact the determination of the acceptable range on which decisions are made when planning maintenance for related systems.

Key findings of this research are:

- Employment of the 'shape' of data to describe uncertainty by the geometric symmetry between inputs for each point in time provided discernible information to determine and predict equipment health under limited data.
- Allocation of inputs to one of three networks according to their variation enabled improved definition of initial network architecture and more robust predictions.
- As technological complexity grows, live and continuous forecasts of uncertainty manifested by data quality and availability are of great benefit to industry.

The core contributions of the UPLD framework are:

1. Robust prediction of uncertainty under limited data
2. Adaptable allocation of inputs to networks with variable structure and training options
3. Initial technique for immersive visualisation of dynamic uncertainties and shape change with an indication of magnitude and direction of the greatest contributing factor

The authors propose future work to simulate and interpolate input data to fill gaps in signal data. Implications of altering fixed parameters within spatial geometry merit further research. The impact of changing the dimensional sequence of inputs (input order

around the origin) for each time unit to a variable rather than a constant parameter is being investigated. Prediction robustness of the LSTM networks may be improved by exploring alternatives to the regression output layer. Alternative evaluation metrics such as custom scoring functions should be developed to gauge prediction quality and identify where additional data is required.

In terms of implementation, the framework enables forecasting under limited data, though prediction robustness is highly dependent on the LSTM network architecture each respective input dimension is assigned to. Further development of the approach should explore implementation in real-time applications to receive live equipment data to update predictions to ultimately provide uncertainty predictions on factors such as RUL. The interoperability of such implementation will depend on data sampling rates, computational processing times and varying environmental and human factors [97,138,140]. Development of the visualisation for applications in AR will enhance useability and allow the user to access additional state information for a given point in time. In addition, suitable approaches to mitigate, tolerate or exploit uncertainty through deep learning according to the magnitude should be explored.

CHAPTER 6. OVERALL DISCUSSION

6.1 Introduction

This chapter discusses the core research findings by revisiting the research context and initial research gaps identified in Chapter 2 as well as the elaborated gaps addressed in subsequent chapters. The methodology selection approach and research findings are evaluated along with a critique of the contributions to knowledge. Finally, the benefits to industry are discussed with the composition of a MATLAB-based application tool applying the two frameworks to quantify, aggregate and forecast uncertainty.

6.2 Research context revisited

The overall aim of the research was to develop a modelling approach capable of learning from a combination of historic equipment data and qualitative estimates to allow the user to quantify, aggregate and forecast uncertainty through the in-service life of engineering systems.

Engineering systems are expected to function effectively whilst maintaining reliability in service. These systems consist of various equipment units, many of which are maintained on a corrective or time-based basis. Challenges to confidently and accurately plan maintenance, accounting for turnaround times, equipment availability and resulting costs, manifest varying degrees of uncertainty stemming from multiple quantitative and qualitative sources throughout the in-service life.

Under or overestimating this uncertainty can ultimately lead to increased failure rates or, more often, unnecessary maintenance carried out. As well as the quality and availability of data, uncertainty is driven by the influence of expert experience or assumptions and environmental operating conditions. Accommodating for uncertainty requires the determination of key contributors, their influence on interconnected units and how this

might change over time. In addition, sporadic sampling rates of maintenance data owing to manual recordings manifest increased uncertainty in equipment and system performance.

6.3 Evaluation of research gaps and critique of academic contributions

Chapter 2 presented a systematic review of multivariate uncertainty quantification for engineering systems. This contributed an identification and assessment of scientific methodologies to (1) quantify uncertainty manifested by purely quantitative, purely qualitative and compound factors, and (2) forecast that uncertainty for the in-service phase with the application of emerging deep learning techniques.

The review identified three core research gaps:

1. Approaches to quantify and aggregate compound uncertainties represented by different distributions, considering dependencies between them, applicable to increasingly complex engineering systems.
2. Application of GSA to determine the impact of individual uncertainties on the aggregated total, accounting for compound parameters and significant correlation.
3. Limited approaches to predict uncertainty in engineering systems with complex and non-complex entities under limited data, and to do this without the need to produce complicated and expensive models of physical systems.

An integrated combination of identified approaches was seen to enhance rigour in uncertainty assessment and forecasting to better understand the impact on cost and availability, which will aid decision-making throughout the in-service phase.

Chapter 3 sought to establish the current practice and challenges in industrial maintenance concerning uncertainty. Six core challenges were identified and verified with practitioners from various industrial backgrounds. These supported the findings of the systematic review concerning the second research question. Three challenges were deemed in scope of the research: maintainer performance (or skill), quality of information and stakeholder communication. A holistic view of quantitative and qualitative attributes ultimately allowed for more accomplished decision-making. However, trade-offs

between quality and cost of implementation over the asset's life cycle play a significant role in the applicability of such considerations [10].

The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) method was applied to identify the best-suited methodologies to address the research gaps, detailed in Appendix B [182–184]. The research gaps identified in Chapter 2 and findings from Chapter 3 were addressed through the development of two frameworks. These used an amalgamation of the methods highlighted in the TOPSIS approach, explored by further review of emerging literature in the above chapters.

6.3.1 Compound uncertainty quantification and aggregation

The Compound Uncertainty Quantification and Aggregation (CUQA) framework (Chapter 4) sought to aggregate compound uncertainties given by quantitative data and qualitative estimates, each represented by different PDFs [30]. This addressed the first and second research questions from Chapter 2. Benefits of the framework included enhancements to performance assessment and corresponding maintenance planning for complex and non-complex engineering systems and respective subsystems.

Further review of emerging literature and development of the framework unveiled two elaborated research gaps:

1. Approaches to quantify and aggregate compound uncertainties represented by different distributions, considering dependencies between them, applicable to increasingly complex engineering systems.
2. Application of global sensitivity analysis (GSA) to determine the impact of individual uncertainties on the aggregated total, accounting for compound parameters and significant correlation.

To resolve these gaps, the CUQA framework made four key academic contributions:

1. Use of coefficient of variation (CV) to enable effective quantification and aggregation of compound uncertainties represented by different distribution types.
2. Assessment of correlation between compound parameters.
3. GSA for dependant compound parameters.
4. Intuitive visualisation of results – most significant parameters, greatest effects.

Deriving the uncertainty measure as the CV proved effective for aggregation of uncertainties represented by different PDFs, but further research into the scaling of geometric against arithmetic standard deviations is required. Aggregating the individual CVs by a combination of the propagation of error method for symmetric CVs and the product of asymmetric CVs allowed an aggregated total estimate to be obtained. This can be used to determine how the aggregated uncertainty changes over time.

Dependencies between compound parameters were not found to impact the aggregated total for the two case studies. However, the influence attributed by individual CVs to the aggregated total was shown to exhibit dependencies that warrant further investigation. Such dependencies may have a significant impact in real-world environments where operating conditions such as atmospheric temperatures or wind speeds impact the accuracy of recorded data or subjective opinion. As discussed in Section 4.4, the case studies served to prove the functionality of the CUQA framework, exhibiting uncertainties akin to those faced in operational environments and comparable challenges to UQ. User-defined ideal limits to identify significant correlations between compound parameters enabled the definition of desired levels of detail for dependant variables. Stronger dependencies between parameter values will have a greater influence on emergent behaviour in more complex systems.

The GSA method applied by Groen [40] was deemed the best-suited approach for the CUQA framework because it can be implemented with relatively small datasets and illustrated the influence of dependant and independent uncertainties against the aggregated total. Intuitive visualisation of the results at each stage further boosted framework useability and enabled rapid identification of uncertainties outside of acceptable levels and where mitigation is required.

6.3.2 Uncertainty prediction under limited data

The second framework (Chapter 5) – uncertainty prediction under limited data (UPLD) – addressed the third initial research question and the third gap from Chapter 2 [97]. The critical research gaps were a lack of approaches to predict uncertainty in engineering systems with complex and non-complex entities under limited data, and to do this without the need to produce complicated and expensive models of physical systems.

The utilisation of spatial geometry in parallel with LSTMs enabled multistep prediction of dynamic uncertainty under limited data. Further benefits of the UPLD framework include forecast projection through the visualisation of polar force fields in 3D vector space. Factors that may require future mitigation can then be identified, ultimately leading to reduction of under or over estimation of turnaround times, equipment availability and resulting costs.

The UPLD framework made three contributions to knowledge:

1. Robust prediction of uncertainty under limited data
2. Adaptable allocation of inputs to networks with variable structure and training options
3. Initial technique for immersive visualisation of dynamic uncertainties and shape change, with an indication of magnitude and direction of the greatest contributing factor

Employment of the 'shape' of data to describe uncertainty by the geometric symmetry between inputs for each point in time provided discernible information to determine and predict equipment health under limited data. This enabled a simplification of what may otherwise be complex conclusions [28,171].

The framework was designed to be embedded in a variety of existing applications, demonstrated by the use of two case studies from distinct domains. Allocation of inputs to one of three networks according to their variation enabled a more robust definition of initial network architecture and more robust predictions.

The immersive visualisation in vector space enabled interactive depiction of dynamic shape area and the source of greatest uncertainty. As technological complexity grows, live and continuous forecasts of uncertainty manifested by data quality and availability are of great benefit to industry.

6.4 Discussions towards implementation with industry

The two frameworks have been brought together to produce the Multistep Compound Dynamic Uncertainty Quantification (MCDUQ) application tool, detailed in Appendix F. This has been developed in MATLAB using the app designer platform to boost useability. Methodologies were adapted and expanded from those identified in literature

[2,24,26,28,34,52,149,158] and interviews within the defence sector. This makes a key contribution to optimise uncertainty management capabilities for real-world industrial applications.

A summary presentation was held with the industrial sponsor to present the MCDUQ tool via the developed app, discuss application benefits and areas where further development was required. The eight attendees included maintenance managers present in the interviews detailed in Chapter 3, as well as data scientists and modellers from the company. The respondents first gauged two statements on a 7-point scale according to whether they agreed or disagreed:

1. The approach and visualisation of results is intuitive and points to where mitigation may be needed in the future
2. The approach will improve decision-making for maintenance practices

The results in Figure 6.1 show both statements scored a mean of 4.4 out of 7 with a wide distribution, indicating a mix of opinions between respondents. This is due to different respondents having different priorities concerning the contextual application and their previous experience with the approaches, such as participation in the interviews in Chapter 3 and Chapter 5.

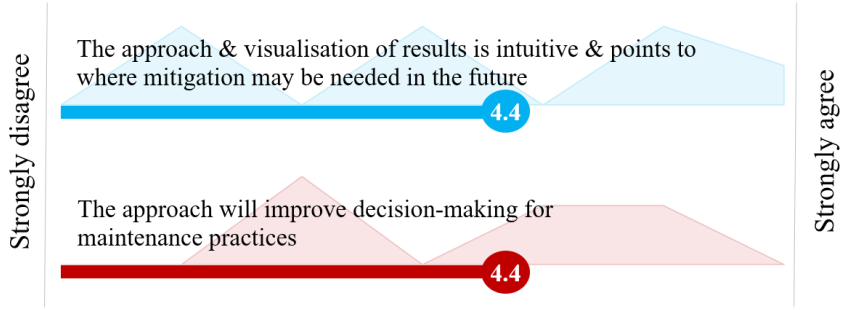


Figure 6.1. Validation: Mentimeter survey

These results were further explored through two open-ended questions. The first was "What do you think the benefits of the approach would be?". The MCDUQ tool was labelled as part of the journey to true condition-based maintenance, allowing greater assurance of platform availability. Higher confidence in estimates concerning specific equipment will allow action priorities to be assigned to tasks. The application tool could also help identify the likelihood of equipment failures between maintenance periods. For

the industrial sponsor, maintenance is usually focused within planned ‘docked’ periods, with only low-level tasks completed otherwise.

The second question was “Are there any additions or amendments that could be made to improve the approach before applying it?”. The main concerns were in the qualification of the novel approaches used in the tool to real-world instances and how to build trust in their application. ‘Real’ data from the company should be fed into the model to better demonstrate its capabilities. As discussed in previous chapters, the confidentiality of such data renders this a ‘catch 22’ situation.

An aggregated and individual picture of uncertainty at a given point in time, determination of greatest or most undesirable sources of uncertainty and forecasts of how these may change through the in-service phase are of great benefit to assess system performance. However, a key element raised was how uncertainty in measured parameters and assumptions can be linked to uncertainty prediction of factors such as equipment availability. As discussed in Chapter 5, data sampling rates and gaps in data play a key role in determining uncertainty. Reliable remaining useful life (RUL) estimation is a major goal for maintenance planning. A holistic picture of the uncertainty surrounding it will be of great benefit.

While the model implements a novel combination of methodologies to consider increasingly complex data, outputs should be displayed in as simple a manner as possible so they can be easily actionable by maintainers. The colour bar given in the CUQA model in both the CV aggregation and sensitivity plot does this well, though parameter ranges that determine colour bounds of ‘good’ and ‘bad’ should have a rigorous determination procedure for each application. The 3D visualisation for the UPLD model provides useful information for data analysts and planners that would see greater usability in AR.

Further discussions held with the company’s data science team explored data integration and comparison with current practice to assess and forecast uncertainty for maintenance. One of the greatest challenges here is the quality and availability of data [10,74,84]. Even if big data is available, parametrics and statistical assessments must be treated with caution without validation and established correlations. While it was agreed to be an area of interest that requires attention for increasingly complex systems, the cost of implementation and ways to interpolate gaps in data raise barriers to the development of

a fully integrated model. As systems grow in complexity and variability, methodologies to quantify, aggregate and forecast uncertainty must be flexible to accommodate a range of input dimensions, scaling and sampling rates.

The verification and validation approach through these sessions was considered suitable under the circumstances as they had to be held virtually [37]. In addition, the implementation capabilities of the CQUA and UPLD frameworks are demonstrated through case studies in their respective chapters. Further assessment to consider usability in practice will require further development of the application tool to simplify the process of adding new datasets that currently require manual coding and formatting. The approaches used were deemed robust and beneficial to fulfil the academic requirements. However, further development is required ahead of implementation.

Overall, a combined understanding of the impact of compound uncertainty on system performance will provide a holistic picture allowing for more informed and effective decisions. The ability to forecast such uncertainty given by limited and sporadic data will improve decision-making capabilities, though the level of confidence and the range of uncertainty must be taken into consideration to increase awareness of under or over estimation. Alternative approaches involve simulating maintenance data, whether by Monte Carlo techniques for individual parameters or agent-based modelling of maintenance scenarios. Developing models that are representative of increasingly complex systems requires significant computing power and cost. Acting upon forecast uncertainty over time creates a trade-off between the cost of corresponding risk mitigation and acceptance of the possibility of delays and increased maintenance costs through unexpected failures or unnecessary work carried out.

CHAPTER 7. CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

The overall findings of this thesis deliver a novel combination of approaches to (1) quantify and aggregate uncertainty contributed by quantitative and qualitative sources, then (2) predict how that uncertainty may change over time under limited data scenarios. The presented frameworks are designed to be applicable for increasingly complex engineering systems. The initial hypothesis that this approach can improve uncertainty management at the equipment-type level for real-world industrial maintenance is proven by the fulfilment of the research objectives as follows:

Objective 1: Map current practice to identify core challenges and resulting uncertainties around equipment cost and availability and how these differ from forecast behaviour within complex engineering systems.

- Methodologies to quantify and aggregate purely quantitative uncertainty is well versed in literature and applied in practice. Uncertainty attributed by qualitative, subjective opinions is seldom taken into consideration alongside quantitative data. This can lead to under or over estimation of uncertainty and determinate factors for a given system (Chapter 2 and Chapter 3)
- Key challenges in the maintenance of increasingly complex systems that manifest uncertainty include intellectual property rights (IPR), maintainer performance, quality of information, resistance to change, stakeholder communication and technology integration – with quality of information being the greatest driver (Chapter 3). Since IPR, resistance to change and technology integration are largely tied into other themes such as supply chain management, they were considered out of scope for this thesis.

- A variety of support systems and data repositories are used to report failures and track corrective actions. Many are not linked, resulting in duplicate entries, gaps in historic data and further assumptions having to be made (Chapter 3)
- The resulting limited data scenario prohibits traditional forecasting methods as associated systems grow in complexity, calling for greater flexibility in methods used to assess and forecast uncertainty for maintenance planning (Chapter 2)
- A holistic picture of the impact of qualitative and quantitative uncertainty on system performance can enable more informed and effective decision-making, but a trade-off is required with implementation costs under a 'spend to save' approach or set aside lump sums for unforeseen circumstances (Chapter 3)

Objective 2: Develop a framework to aggregate uncertainty from quantitative and qualitative sources represented through different probability distributions with an identification of the source of greatest uncertainty.

- Existing methodologies to quantify and aggregate uncertainty from purely quantitative, purely qualitative and compound perspectives were identified and ranked using TOPSIS (Chapter 2 and Appendix B)
- The compound uncertainty quantification and aggregation (CUQA) framework was developed, incorporating the top-ranked approaches (Chapter 4):
  - Uncertainty identification from a component to system level
  - Improved rigour in uncertainty assessment for industrial maintenance
  - Consideration of different probability distributions through use of coefficient of variation (CV)
  - Derivation of dependencies between parameters and identification of greatest contributing factors through global sensitivity analysis (GSA)
- Limitations of the CUQA framework that prompt future work in Section 7.2 are:
  - Input data distributions are defined manually
  - Number of data points in sub-array units can result in disproportionately low CV for lognormally distributed parameters compared to qualitative factors
  - Only tested for normal, lognormal and uniform distributions

Objective 3: Develop an approach to predict uncertainty given by limited available data and qualitative factors to relate to equipment cost and availability over the in-service phase.

- Existing methodologies were identified to forecast uncertainties in maintenance that influence equipment cost and availability (largely in the product-service systems (PSS) context), ranked using TOPSIS (Chapter 2 and Appendix B)
- Methodologies to predict uncertainty and remaining useful life (RUL) were assessed for use with limited maintenance data in increasingly complex engineering systems (Chapter 2)
- The framework for uncertainty prediction under limited data (UPLD) was developed, incorporating the top-ranked approaches (Chapter 5):
  - Robust prediction of trends in limited and dynamic uncertainty data given by compound attributes
  - Parallel determination of geometric symmetry at each point in time
  - Immersive visualisation of dynamic uncertainties and shape change with an indication of magnitude and direction of the greatest contributing factor
- The model can be embedded in a range of applications, including uncertainty in equipment availability and costs, but a direct connection is not made between these and compound uncertainty in the presented work
- Limitations of the UPLD framework that prompt future work in Section 7.2 are:
  - Partition of training/observed to test/predicted data is fixed to 60%. This is variable within the developed application, but not examined in the presented work
  - Number and order of input parameters is fixed for all predictions – new inputs cannot be added over time
  - Methodology to allocate input dimensions to LSTM networks has a significant effect on the robustness of resulting predictions, making this one of the most important steps of the framework
  - Use of MAPE and RMSE as evaluation metrics resulted in significant prediction errors. This was due to sudden changes in the observed data for which predictions could not account

Objective 4. Validate the final model to assess implementation effectiveness and usability in context

- Development of the MCDUQ app boosts framework usability, allowing results and forecasts to be viewed in an intuitive manner.
- Surveys held with industrial practitioners deemed the methodologies applied in the model of great benefit to assess system performance. However, to be truly implemented the app would likely require recoding in alternative programming languages such as Python and in C# for AR applications.
- A key subject was the application and qualification of the novel approach to real-world scenarios and establish trust in the methodology. The use of 'real' data would provide a better demonstration of the MCDUQ app's capabilities. This was not possible for this thesis owing to confidentiality restrictions.
- A holistic picture of uncertainty surrounding RUL prediction is of great benefit, taking qualitative factors, data sampling rates and gaps in data into consideration.
- The next major step highlighted in the surveys was the linking of uncertainties calculated and predicted via the MCDUQ app with factors such as equipment availability, turnaround times and costing.

7.2 Future work

Recommendations for future research as a result of the studies undertaken in the above chapters are listed below:

Uncertainty quantification and aggregation

- Further assessment of aggregated compound uncertainty, incorporating additional distribution types such as Weibull, Gamma and Beta. These may allow for greater flexibility in highly variable time series data but require additional points to shape and scale accurately.
- Use of clustering techniques to derive and classify uncertainty from non-parametric and stochastic distributions. While elements of the CUQA framework are automated in the MCDUQ app, intelligent learning techniques to identify the most appropriate sub-array size allocation or representative distributions are not featured. These can enhance usability and robustness of the aggregated uncertainty for more complex inputs.

- Improved rigour of the GSA approach in variance decomposition for each sub-array time unit. The GSA approach applied in the CUQA framework can be implemented with relatively small datasets and illustrate dependencies of individual uncertainties against the total effects. A more rigorous depiction of the greatest contributing factors may be achieved through derivation of Sobol indices. However, this would require additional derivation of model/system process equations for each application (such as heat transfer and energy loss). While out of scope for this research, the model can be further developed to incorporate such equations in the first step of the CUQA framework.

Uncertainty prediction under limited data

- Simulate and interpolate input data to fill gaps in signal data. While the UPLD framework is able to predict uncertainty for small datasets, it has not been applied where data may be missing for certain time units. Gaps in data can be filled by interpolation.
- Explore implications of altering fixed parameters within spatial geometry. Determination of shape area and symmetry relies on fixed parameters such as the radial sequence between input dimensions. The impact of changing the input order around the origin for each time unit to a variable rather than a constant parameter is under investigation. Optimum input orders are not yet defined. Derivation may be possible through testing the results of every possible order and taking the mean result, though this would become cumbersome with an increasing number of inputs.
- Explore alternatives to regression output layers for the LSTM networks to improve robustness in predictions. Regression output layers were used to provide the prediction results. Typically, these would be based on sufficient data to fulfil the Central Limit Theorem. Under limited data, the use of regression contributes to prediction errors. Alternative, custom output layers should be explored to provide greater robustness.
- Develop the 3D visualisation for applications in AR. This will allow further detail to be provided on each input node through time in an intuitive manner. The uncertainty range can be illustrated via fan plots, as well as proposals for mitigation and correlated factors.

Overall

- Test the frameworks with more complex inputs. The case studies applied in this thesis to test and validate the CUQA and UPLD frameworks may be considered comparatively simple in terms of the nature of inputs and qualitative factors such as operational environments and assumptions grouped under 5-point scoring in the pedigree approach. While the uncertainties exhibited are akin to those faced in various environments and present comparable challenges to quantification and forecasting, further testing in real-world scenarios will allow for further development.
- Integration of strategies to mitigate, tolerate or exploit uncertainty with suggestions of how to manage different levels for each factor. Where the quantified, aggregated or forecast uncertainties lie outside of an acceptable level, approaches to manage them should be made available to the user. These approaches could be predefined for different boundaries in the initial identification of inputs. Visualisation of such approaches could be integrated as part of the AR interface.
- Simulate maintenance processes through surrogate models to replicate challenges identified in Chapter 3. Data collected from simulations can then be incorporated to train developed frameworks to aggregate and forecast compound uncertainty with greater confidence where real-world data is not obtainable.
- Uncertainties calculated and predicted using the amalgamation of methods presented in this thesis need to be connected to their impact on factors such as equipment availability, turnaround times and costing to enhance decision-making capabilities in maintenance planning.

REFERENCES

[1] Andretta, M. (2014), 'Some Considerations on the Definition of Risk Based on Concepts of Systems Theory and Probability', *Risk Analysis*, 34 (7), pp. 1184–1195. DOI: 10.1111/risa.12092.

[2] NASA (2010), *Measurement Uncertainty Analysis Principles and Methods, NASA-HDBK-8739.19-3: NASA Measurement Quality Assurance Handbook - Annex 3*. NASA. Available at: https://standards.nasa.gov/standard/osma/nasa-hdbk-873919-3.

[3] Lanza, G. and Viering, B. (2011), 'A novel standard for the experimental estimation of the uncertainty of measurement for micro gear measurements', *CIRP Annals - Manufacturing Technology*, 60 (1), pp. 543–546. DOI: 10.1016/j.cirp.2011.03.062.

[4] Newman, M. E. J. (2011), 'Complex Systems: A Survey', *American Journal of Physics*, 79 (8), pp. 800–810. DOI: 10.1119/1.3590372.

[5] Takata, S., Kirnura, F., van Houten, F. J. A. M., Westkamper, E., Shpitalni, M., Ceglarek, D. and Lee, J. (2004), 'Maintenance: Changing Role in Life Cycle Management', *CIRP Annals - Manufacturing Technology*, 53 (2), pp. 643–655. DOI: 10.1016/S0007-8506(07)60033-X.

[6] Stevens, R. (2008), 'Profiling Complex Systems', in *2nd Annual IEEE Systems Conference*. Montreal, Que., pp. 1–6. DOI: 10.1109/SYSTEMS.2008.4519017.

[7] Thunnissen, D. P. (2003), 'Uncertainty classification for the design and development of complex systems', in *Proceedings of the 3rd Annual Predictive Methods Conference*. California, USA. Available at: https://www.semanticscholar.org/paper/Uncertainty-Classification-for-the-Design-and-of-Thunnissen/11f0bbaf7c69351e6cee61f02176d9400b706d65.

[8] Efthymiou, K., Mourtzis, D., Pagoropoulos, A., Papakostas, N. and Chryssolouris, G. (2016), 'Manufacturing systems complexity analysis methods review', *International Journal of Computer Integrated Manufacturing*, 29 (9), pp. 1025–1044. DOI: 10.1080/0951192X.2015.1130245.

[9] Helton, J. C. and Davis, F. J. (2003), 'Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems', *Reliability Engineering & System Safety*, 81 (1), pp. 23–69. DOI: 10.1016/S0951-8320(03)00058-9.

[10] Grenyer, A., Dinmohammadi, F., Erkoyuncu, J. A., Zhao, Y. and Roy, R. (2019), 'Current practice and challenges towards handling uncertainty for effective outcomes in maintenance', *Procedia CIRP*, 86, pp. 282–287. DOI: 10.1016/j.procir.2020.01.024.

[11] MacAulay, G. D. and Giusca, C. L. (2016), 'Assessment of uncertainty in structured surfaces using metrological characteristics', *CIRP Annals - Manufacturing Technology*, 65 (1), pp. 533–536. DOI: 10.1016/j.cirp.2016.04.068.

[12] Dantan, J. Y., Vincent, J. P., Goch, G. and Mathieu, L. (2010), 'Correlation uncertainty—Application to gear conformity', *CIRP Annals - Manufacturing Technology*, 59 (1), pp. 509–512. DOI: 10.1016/j.cirp.2010.03.040.

[13] Erkoyuncu, J. A., Khan, S., Eiroa, A. L., Butler, N., Rushton, K. and Brocklebank, S. (2017), 'Perspectives on trading cost and availability for corrective maintenance at the equipment type level', *Reliability Engineering & System Safety*, 168, pp. 53–69. DOI: 10.1016/j.ress.2017.05.041.

[14] Shamsi, M. H., Ali, U., Mangina, E. and O'Donnell, J. (2020), 'A framework for uncertainty quantification in building heat demand simulations using reduced-order grey-box energy models', *Applied Energy*, 275 (05), p. 115141. DOI: 10.1016/j.apenergy.2020.115141.

[15] Bentaha, M. L., Battaïa, O., Dolgui, A. and Hu, S. J. (2014), 'Dealing with uncertainty in disassembly line design', *CIRP Annals - Manufacturing Technology*, 63 (1), pp. 21–24. DOI: 10.1016/j.cirp.2014.03.004.

[16] Mayfield, M., Punzo, G., Beasley, R., Clarke, G., Holt, N. and Jobbins, S. (2018), *Challenges of complexity and resilience in complex engineering systems*, *ENCORE Network+*. ENCORE Network+ White Paper. Available at: https://civil-struct.dept.shef.ac.uk/encore/public/storage/rnJInf6iHsqXZcXyAw0IsXb9tdAQnnJS2BTts4QA.pdf.

[17] Addepalli, S., Eiroa, D., Lieotrakool, S., François, A.-L., Guisset, J., Sanjaime, D., Kazarian, M., Duda, J., Roy, R. and Phillips, P. (2015), 'Degradation Study of Heat Exchangers', *Procedia CIRP*, 38, pp. 137–142. DOI: 10.1016/j.procir.2015.07.057.

[18] Lequin, R. M. (2004), 'Guide to the Expression of Uncertainty of Measurement: Point/Counterpoint', *Clinical Chemistry*, 50 (5), pp. 977–978. DOI: 10.1373/clinchem.2003.030528.

[19] ElMaraghy, W., ElMaraghy, H., Tomiyama, T. and Monostori, L. (2012), 'Complexity in engineering design and manufacturing', *CIRP Annals - Manufacturing Technology*, 61 (2), pp. 793–814. DOI: 10.1016/j.cirp.2012.05.001.

[20] Grote, G. (2009), *Management of Uncertainty: Theory and Application in the Design of Systems and Organisations*. Edited by R. Roy. Zurich: Decision Engineering, London: Springer.

[21] McManus, H. and Hastings, D. (2005), 'A Framework for Understanding Uncertainty and its Mitigation and Exploitation in Complex Systems', *INCOSE International Symposium*, 15 (1), pp. 484–503. DOI: 10.1002/j.2334-5837.2005.tb00685.x.

[22] Limbourg, P. and de Rocquigny, E. (2010), 'Uncertainty analysis using evidence theory – confronting level-1 and level-2 approaches with data availability and computational constraints', *Reliability Engineering & System Safety*, 95 (5), pp. 550–564. DOI: 10.1016/j.ress.2010.01.005.

[23] Richter, A., Sadek, T. and Steven, M. (2010), 'Flexibility in industrial product-service systems and use-oriented business models', *CIRP Journal of Manufacturing Science and Technology*, 3 (2), pp. 128–134. DOI: 10.1016/j.cirpj.2010.06.003.

[24] Helton, J. C. and Johnson, J. D. (2011), 'Quantification of margins and uncertainties: Alternative representations of epistemic uncertainty', *Reliability Engineering & System Safety*, 96 (9), pp. 1034–1052. DOI: 10.1016/j.ress.2011.02.013.

[25] Bertoni, A. and Bertoni, M. (2020), 'PSS cost engineering: A model-based approach for concept design', *CIRP Journal of Manufacturing Science and Technology*, 29, pp. 176–190. DOI: 10.1016/j.cirpj.2018.08.001.

[26] Schwabe, O., Shehab, E. and Erkoyuncu, J. (2015), 'Uncertainty quantification metrics for whole product life cycle cost estimates in aerospace innovation', *Progress in Aerospace Sciences*, 77, pp. 1–24. DOI: 10.1016/j.paerosci.2015.06.002.

[27] Baxter, S., Killoran, A., Kelly, M. P. and Goyder, E. (2010), 'Synthesizing diverse evidence: the use of primary qualitative data analysis methods and logic models in public health reviews', *Public Health*, 124 (2), pp. 99–106. DOI: 10.1016/j.puhe.2010.01.002.

[28] Schwabe, O., Shehab, E. and Erkoyuncu, J. A. (2016), 'A framework for geometric quantification and forecasting of cost uncertainty for aerospace innovations', *Progress in Aerospace Sciences*, 84, pp. 29–47. DOI: 10.1016/j.paerosci.2016.05.001.

[29] Matschewsky, J., Lindahl, M. and Sakao, T. (2020), 'Capturing and enhancing provider value in product-service systems throughout the lifecycle: A systematic approach', *CIRP Journal of Manufacturing Science and Technology*, 29, pp. 191–204. DOI: 10.1016/j.cirpj.2018.08.006.

[30] Grenyer, A., Erkoyuncu, J. A., Addepalli, S. and Zhao, Y. (2020), 'An uncertainty quantification and aggregation framework for system performance assessment in industrial maintenance', *SSRN Electronic Journal*, (November). DOI: 10.2139/ssrn.3718001.

[31] Hochdörffer, J., Buergin, J., Vlachou, E., Zogopoulos, V., Lanza, G. and Mourtzis, D. (2018), 'Holistic approach for integrating customers in the design, planning, and control of global production networks', *CIRP Journal of Manufacturing Science and Technology*, 23, pp. 98–107. DOI: 10.1016/j.cirpj.2018.07.004.

[32] Ciroth, A., Muller, S., Weidema, B. and Lesage, P. (2016), 'Empirically based uncertainty factors for the pedigree matrix in ecoinvent', *The International Journal of Life Cycle Assessment*, 21 (9), pp. 1338–1348. DOI: 10.1007/s11367-013-0670-5.

[33] Adasooriya, N. D. (2016), 'Fatigue reliability assessment of ageing railway truss bridges: Rationality of probabilistic stress-life approach', *Case Studies in Structural Engineering*, 6 (1), pp. 1–10. DOI: 10.1016/j.csse.2016.04.002.

[34] Muller, S., Lesage, P., Ciroth, A., Mutel, C., Weidema, B. P. and Samson, R. (2016), 'The application of the pedigree approach to the distributions foreseen in ecoinvent v3', *International Journal of Life Cycle Assessment*, 21 (9), pp. 1327–1337. DOI: 10.1007/s11367-014-0759-5.

[35] Muchiri, P. N., Pintelon, L., Martin, H. and Chemweno, P. (2014), 'Modelling maintenance effects on manufacturing equipment performance: Results from simulation analysis', *International Journal of Production Research*, 52 (11), pp. 3287–3302. DOI: 10.1080/00207543.2013.870673.

[36] Limpert, E., Stahel, W. A. and Abbt, M. (2001), 'Log-normal Distributions across the Sciences: Keys and Clues', *BioScience*, 51 (5), p. 341. DOI: https://doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2.

[37] Coleman, H. W. and Steele, W. G. (2009), *Experimentation, Validation, and Uncertainty Analysis for Engineers*, *AIChE Symposium Series*. Hoboken, NJ, USA: John Wiley & Sons, Inc. DOI: 10.1002/9780470485682.

[38] Mathworks documentation (2012), *Corrplot*, *Econometrics Toolbox*. Available at: https://uk.mathworks.com/help/econ/corrplot.html.

[39] Erkoyuncu, J. A. (2011), *Cost uncertainty management and modelling for industrial product-service systems*. PhD thesis. Cranfield University, Cranfield. Available at: http://dspace.lib.cranfield.ac.uk/handle/1826/5624.

[40] Groen, E. A. and Heijungs, R. (2017), 'Ignoring correlation in uncertainty and sensitivity analysis in life cycle assessment: what is the risk?', *Environmental Impact Assessment Review*, 62, pp. 98–109. DOI: 10.1016/j.eiar.2016.10.006.

[41] Xu, C. and Gertner, G. Z. (2008), 'Uncertainty and sensitivity analysis for models with correlated parameters', *Reliability Engineering & System Safety*, 93 (10), pp. 1563–1573. DOI: 10.1016/j.ress.2007.06.003.

[42] Ghahramani, Z. (2015), 'Probabilistic machine learning and artificial intelligence', *Nature*, 521 (7553), pp. 452–459. DOI: 10.1038/nature14541.

[43] Tatara, R. and Lupia, G. (2012), 'Assessing heat exchanger performance data using temperature measurement uncertainty', *International Journal of Engineering, Science and Technology*, 3 (8), pp. 1–12. DOI: 10.4314/ijest.v3i8.1.

[44] Ahmed, R., Sreeram, V., Mishra, Y. and Arif, M. D. (2020), 'A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization', *Renewable and Sustainable Energy Reviews*, 124, p. 109792. DOI: 10.1016/j.rser.2020.109792.

[45] Clarke, D. D., Vasquez, V. R., Whiting, W. B. and Greiner, M. (2001), 'Sensitivity and uncertainty analysis of heat-exchanger designs to physical properties estimation', *Applied Thermal*

Engineering, 21 (10), pp. 993–1017. DOI: 10.1016/S1359-4311(00)00101-0.

[46] Erkoyuncu, J. A., Roy, R., Shehab, E. and Cheruvu, K. (2010), 'Understanding service uncertainties in industrial product-service system cost estimation', *International Journal of Advanced Manufacturing Technology*, 52 (9–12), pp. 1223–1238. DOI: 10.1007/s00170-010-2767-3.

[47] Thulukkanam, K. (2013), *Heat Exchanger Design Handbook*. 2nd edn. CRC Press. DOI: 10.1201/b14877.

[48] Goh, Y. M., Newnes, L. B., Mileham, A. R., McMahon, C. A. and Saravi, M. E. (2010), 'Uncertainty in through-life costing-review and perspectives', *IEEE Transactions on Engineering Management*, 57 (4), pp. 689–701. DOI: 10.1109/TEM.2010.2040745.

[49] Langford, E. (2006), 'Quartiles in Elementary Statistics', *Journal of Statistics Education*, 14 (3). DOI: 10.1080/10691898.2006.11910589.

[50] Madenas, N., Tiwari, A., Turner, C. J. and Woodward, J. (2014), 'Information flow in supply chain management: A review across the product lifecycle', *CIRP Journal of Manufacturing Science and Technology*, 7 (4), pp. 335–346. DOI: 10.1016/j.cirpj.2014.07.002.

[51] Durugbo, C., Erkoyuncu, J. A., Tiwari, A., Alcock, J. R., Roy, R. and Shehab, E. (2010), 'Data uncertainty assessment and information flow analysis for product-service systems in a library case study', *International Journal of Services Operations and Informatics*, 5 (4), p. 330. DOI: 10.1504/IJSOI.2010.037002.

[52] Van Der Sluijs, J. P., Craye, M., Funtowicz, S., Kloprogge, P., Ravetz, J. and Risbey, J. (2005), 'Combining Quantitative and Qualitative Measures of Uncertainty in Model-Based Environmental Assessment: The NUSAP System', *Risk Analysis*, 25 (2), pp. 481–492. DOI: 10.1111/j.1539-6924.2005.00604.x.

[53] Booth, A., Papaioannou, D. and Sutton, A. (2012), *Systematic approaches to a successful literature review*, *Educational Psychology in Practice*. London: Sage.

[54] Grant, M. J. and Booth, A. (2009), 'A typology of reviews: An analysis of 14 review types and associated methodologies', *Health Information and Libraries Journal*, 26 (2), pp. 91–108. DOI: 10.1111/j.1471-1842.2009.00848.x.

[55] Saxena, A., Goebel, K., Simon, D. and Eklund, N. (2008), 'Damage propagation modeling for aircraft engine run-to-failure simulation', in *2008 International Conference on Prognostics and Health Management*, pp. 1–9. DOI: 10.1109/PHM.2008.4711414.

[56] Fernández del Amo, I., Erkoyuncu, J. A., Roy, R., Palmarini, R. and Onoufriou, D. (2018), 'A systematic review of Augmented Reality content-related techniques for knowledge transfer in maintenance applications', *Computers in Industry*, 103, pp. 47–71. DOI:

10.1016/j.compind.2018.08.007.

[57] Ramasso, E. and Saxena, A. (2020), 'Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets', *International Journal of Prognostics and Health Management*, 5 (2), pp. 1–15. DOI: 10.36001/ijphm.2014.v5i2.2236.

[58] Palmarini, R., Erkoyuncu, J. A., Roy, R. and Torabmostaedi, H. (2018), 'A systematic review of augmented reality applications in maintenance', *Robotics and Computer-Integrated Manufacturing*, 49, pp. 215–228. DOI: 10.1016/j.rcim.2017.06.002.

[59] Shi, Z. and Chehade, A. (2021), 'A dual-LSTM framework combining change point detection and remaining useful life prediction', *Reliability Engineering & System Safety*, 205 (October 2020), p. 107257. DOI: 10.1016/j.ress.2020.107257.

[60] Bar-Yam, Y. (2003), 'When systems engineering fails-toward complex systems engineering', in *IEEE International Conference on Systems, Man and Cybernetics*. Washington, DC, USA, pp. 2021–2028. DOI: 10.1109/ICSMC.2003.1244709.

[61] Behera, S., Misra, R. and Sillitti, A. (2021), 'Multiscale deep bidirectional gated recurrent neural networks based prognostic method for complex non-linear degradation systems', *Information Sciences*, 554, pp. 120–144. DOI: 10.1016/j.ins.2020.12.032.

[62] Kwon, K., Ryu, N., Seo, M., Kim, S., Lee, T. H. and Min, S. (2020), 'Efficient uncertainty quantification for integrated performance of complex vehicle system', *Mechanical Systems and Signal Processing*, 139, p. 106601. DOI: 10.1016/j.ymssp.2019.106601.

[63] Chen, Z., Wu, M., Zhao, R., Guretno, F., Yan, R. and Li, X. (2021), 'Machine Remaining Useful Life Prediction via an Attention-Based Deep Learning Approach', *IEEE Transactions on Industrial Electronics*, 68 (3), pp. 2521–2531. DOI: 10.1109/TIE.2020.2972443.

[64] Walker, W. E., Harremoës, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B. A., Janssen, P. and Krayer von Krauss, M. P. (2003), 'Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support', *Integrated Assessment*, 4 (1), pp. 5–17. DOI: 10.1076/iaij.4.1.5.16466.

[65] Wu, Y., Yuan, M., Dong, S., Lin, L. and Liu, Y. (2018), 'Remaining useful life estimation of engineered systems using vanilla LSTM neural networks', *Neurocomputing*, 275, pp. 167–179. DOI: 10.1016/j.neucom.2017.05.063.

[66] Perminova, O., Gustafsson, M. and Wikström, K. (2008), 'Defining uncertainty in projects – a new perspective', *International Journal of Project Management*, 26 (1), pp. 73–79. DOI: 10.1016/j.ijproman.2007.08.005.

[67] Ward, S. and Chapman, C. (2003), 'Transforming project risk management into project uncertainty

management', *International Journal of Project Management*, 21 (2), pp. 97–105. DOI: 10.1016/S0263-7863(01)00080-1.

[68] Everitt, B. S. and Skrondal, A. (2010), *The Cambridge Dictionary of Statistics*. 4th edn. Cambridge: Cambridge University Press.

[69] Kiureghian, A. and Ditlevsen, O. (2009), 'Aleatoric or Epistemic? Does it matter?', in *Special Workshop on Risk Acceptance and Risk Communication*, pp. 105–112. DOI: 10.1016/j.strusafe.2008.06.020.

[70] Krane, H. P., Johansen, A. and Alstad, R. (2014), 'Exploiting Opportunities in the Uncertainty Management', in *Procedia - Social and Behavioral Sciences*, pp. 615–624. DOI: 10.1016/j.sbspro.2014.03.069.

[71] Rowe, W. D. (1994), 'Understanding Uncertainty', *Risk Analysis*, 14 (5), pp. 743–750. DOI: 10.1111/j.1539-6924.1994.tb00284.x.

[72] Datong Liu, Yue Luo and Yu Peng (2012), 'Uncertainty processing in prognostics and health management: An overview', in *Proceedings of the IEEE 2012 Prognostics and System Health Management Conference*, pp. 1–6. DOI: 10.1109/PHM.2012.6228860.

[73] Kreiser, P. and Marino, L. (2002), 'Analyzing the historical development of the environmental uncertainty construct', *Management Decision*, 40 (9), pp. 895–905. DOI: 10.1108/00251740210441090.

[74] Erkoyuncu, J. A., Durugbo, C. and Roy, R. (2013), 'Identifying uncertainties for industrial service delivery: A systems approach', *International Journal of Production Research*, 51 (21), pp. 6295–6315. DOI: 10.1080/00207543.2013.794316.

[75] Savage, S. (2002), 'The flaw of averages', *Harvard Business Review*, 80 (11), p. 20. DOI: 10.1111/j.1539-6924.2009.01326.x.

[76] Willink, R. (2005), 'A procedure for the evaluation of measurement uncertainty based on moments', *Metrologia*, 42 (5), pp. 329–343. DOI: 10.1088/0026-1394/42/5/001.

[77] Ratcliffe, C. and Ratcliffe, B. (2015), *Doubt-Free Uncertainty In Measurement*. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-12063-8.

[78] Willink, R. (2013), *Measurement Uncertainty and Probability*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139135085.

[79] Soundappan, P., Nikolaidis, E., Haftka, R. T., Grandhi, R. and Canfield, R. (2004), 'Comparison of evidence theory and Bayesian theory for uncertainty modeling', *Reliability Engineering & System Safety*, 85 (1–3), pp. 295–311. DOI: 10.1016/j.ress.2004.03.018.

[80] Chalupnik, M. J., Wynn, D. C. and Clarkson, P. J. (2009), 'Approaches to mitigate the Impact of Uncertainty in Development Processes', in *16th International Conference on Engineering Design*. 24-27 August, Stanford, USA, pp. 459–470.

[81] Brune, A. J., West, T. K. and Hosder, S. (2019), 'Uncertainty quantification of planetary entry technologies', *Progress in Aerospace Sciences*, 111, p. 100574. DOI: 10.1016/j.paerosci.2019.100574.

[82] McFarland, J. and DeCarlo, E. (2020), 'A Monte Carlo framework for probabilistic analysis and variance decomposition with distribution parameter uncertainty', *Reliability Engineering & System Safety*, 197 (November 2019), p. 106807. DOI: 10.1016/j.ress.2020.106807.

[83] Schwabe, O., Shehab, E. and Erkoyuncu, J. A. (2015), 'Long Tail Uncertainty Distributions in Novel Risk Probability Classification', *Procedia CIRP*, 28, pp. 191–196. DOI: 10.1016/j.procir.2015.04.033.

[84] Xu, Y., Elgh, F., Erkoyuncu, J. A., Bankole, O., Goh, Y., Cheung, W. M., Baguley, P., Wang, Q., Arundachawat, P., Shehab, E., Newnes, L. and Roy, R. (2012), 'Cost Engineering for manufacturing: Current and future research', *International Journal of Computer Integrated Manufacturing*, 25 (4–5), pp. 300–314. DOI: 10.1080/0951192X.2010.542183.

[85] Helton, J. C. and Johnson, J. D. (2011), 'Quantification of margins and uncertainties: Alternative representations of epistemic uncertainty', *Reliability Engineering & System Safety*, 96 (9), pp. 1034–1052. DOI: 10.1016/j.ress.2011.02.013.

[86] Helton, J. C. and Oberkampf, W. L. (2004), 'Alternative representations of epistemic uncertainty', *Reliability Engineering & System Safety*, 85 (1–3), pp. 1–10. DOI: 10.1016/j.ress.2004.03.001.

[87] Igusa, T., Buonopane, S. G. and Ellingwood, B. R. (2002), 'Bayesian analysis of uncertainty for structural engineering applications', *Structural Safety*, 24, pp. 165–186. DOI: 10.1016/S0167-4730(02)00023-1.

[88] Kwon, Y., Won, J. H., Kim, B. J. and Paik, M. C. (2020), 'Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation', *Computational Statistics and Data Analysis*, 142, p. 106816. DOI: 10.1016/j.csda.2019.106816.

[89] Marshall, A., Ojiako, U., Wang, V., Lin, F. and Chipulu, M. (2019), 'Forecasting unknown-unknowns by boosting the risk radar within the risk intelligent organisation', *International Journal of Forecasting*, 35 (2), pp. 644–658. DOI: 10.1016/j.ijforecast.2018.07.015.

[90] Farsi, M., Grenyer, A., Sachidananda, M., Sceral, M., Mcvey, S., Erkoyuncu, J. and Roy, R. (2018), 'Conceptualising the impact of information asymmetry on through-life cost: case study of machine tools sector', *Procedia Manufacturing*, 16, pp. 99–106. DOI: 10.1016/j.promfg.2018.10.172.

[91] Povey, A. C. and Grainger, R. G. (2015), 'Known and unknown unknowns: Uncertainty estimation in satellite remote sensing', *Atmospheric Measurement Techniques*, 8 (11), pp. 4699–4718. DOI: 10.5194/amt-8-4699-2015.

[92] Azene, Y. T., Roy, R., Farrugia, D., Onisa, C., Mehnen, J. and Trautmann, H. (2010), 'Work roll cooling system design optimisation in presence of uncertainty and constrains', *CIRP Journal of Manufacturing Science and Technology*, 2 (4), pp. 290–298. DOI: 10.1016/j.cirpj.2010.06.001.

[93] Booth, A. (2006), 'Clear and present questions: formulating questions for evidence based practice', *Library Hi Tech*, 24 (3), pp. 355–368. DOI: 10.1108/07378830610692127.

[94] Booth, A. (2004), 'Formulating answerable questions', in Booth, A. and Brice, A. (eds) *Evidence Based Practice for Information Professionals: A handbook*. London: London: Facet Publishing, pp. 61–70.

[95] Petticrew, M. and Roberts, H. (2006), *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford: Blackwell.

[96] Denyer, D. and Tranfield, D. (1987), 'Producing a systematic literature review', *Communication Education*, 36 (1), pp. 1–1. DOI: 10.1080/03634528709378635.

[97] Grenyer, A., Erkoyuncu, J. A., Zhao, Y. and Roy, R. (2021), 'A systematic review of multivariate uncertainty quantification for engineering systems', *CIRP Journal of Manufacturing Science and Technology*, 33, pp. 188–208. DOI: 10.1016/j.cirpj.2021.03.004.

[98] Suri, H. and Clarke, D. (2009), 'Advancements in Research Synthesis Methods: From a Methodologically Inclusive Perspective', *Review of Educational Research*, 79 (1), pp. 395–430. DOI: 10.3102/0034654308326349.

[99] Bell, S. (2001), *A Beginners Guide to Uncertainty of Measurement, Measurement Good Practice Guide*. Teddington: National Physical Laboratory.

[100] Minkina, W. and Dudzik, S. (2009), *Infrared Thermography: Errors and Uncertainties*. John Wiley and Sons.

[101] Groen, E. A. (2016), *An uncertain climate: The value of uncertainty and sensitivity analysis in environmental impact assessment of food*. PhD thesis. Wageningen University. DOI: 10.18174/375497.

[102] Castrup, H. (2004), 'Estimating and combining uncertainties', *8th Annual ITEA Instrumentation Workshop. Lancaster, CA*. Available at: www.isgmax.com.

[103] Baek, C.-Y., Park, K.-H., Tahara, K. and Chun, Y.-Y. (2017), 'Data Quality Assessment of the Uncertainty Analysis Applied to the Greenhouse Gas Emissions of a Dairy Cow System', *Sustainability*, 9 (10), p. 1676. DOI: 10.3390/su9101676.

[104] Kloprogge, P., van der Sluijs, J. P. and Petersen, A. C. (2011), 'A method for the analysis of assumptions in model-based environmental assessments', *Environmental Modelling & Software*, 26 (3), pp. 289–301. DOI: 10.1016/j.envsoft.2009.06.009.

[105] Pourmousavi, S. A., Behrangrad, M., Ardakani, A. J. and Nehrir, M. H. (2017), 'Ownership Cost Calculations for Distributed Energy Resources Using Uncertainty and Risk Analyses', *Papers 1709.08023, arXiv.org*, pp. 1–8. Available at: http://arxiv.org/abs/1709.08023.

[106] Borgonovo, E. (2007), 'A new uncertainty importance measure', *Reliability Engineering & System Safety*, 92 (6), pp. 771–784. DOI: 10.1016/j.ress.2006.04.015.

[107] Tarantola, S. and Koda, M. (2010), 'Improving Random Balance Designs For The Estimation Of First Order Sensitivity Indices', *Procedia - Social and Behavioral Sciences*, 2 (6), pp. 7753–7754. DOI: 10.1016/j.sbspro.2010.05.212.

[108] Mohammadi, A. and Raisee, M. (2019), 'Efficient uncertainty quantification of stochastic heat transfer problems by combination of proper orthogonal decomposition and sparse polynomial chaos expansion', *International Journal of Heat and Mass Transfer*, 128, pp. 581–600. DOI: 10.1016/j.ijheatmasstransfer.2018.09.031.

[109] Valdez, A. R., Rocha, B. M., Chapiro, G. and Weber dos Santos, R. (2020), 'Uncertainty quantification and sensitivity analysis for relative permeability models of two-phase flow in porous media', *Journal of Petroleum Science and Engineering*, 192, p. 107297. DOI: 10.1016/j.petrol.2020.107297.

[110] Venturin, M., Turchetti, L. and Liberatore, R. (2020), 'Uncertainty quantification in a hydrogen production system based on the solar hybrid sulfur process', *International Journal of Hydrogen Energy*, 45 (29), pp. 14679–14695. DOI: 10.1016/j.ijhydene.2020.03.200.

[111] LEVON Group (2015), *Addressing uncertainty in oil and natural gas industry greenhouse gas inventories*. London: IPIECA/API. Available at: https://www.ipieca.org/resources/good-practice/addressing-uncertainty-in-oil-and-natural-gas-industry-greenhouse-gas-inventories-technical-considerations-and-calculation-methods/.

[112] Cremon, M. A., Christie, M. A. and Gerritsen, M. G. (2020), 'Monte Carlo simulation for uncertainty quantification in reservoir simulation: A convergence study', *Journal of Petroleum Science and Engineering*, 190, p. 107094. DOI: 10.1016/j.petrol.2020.107094.

[113] Fleeter, C. M., Geraci, G., Schiavazzi, D. E., Kahn, A. M. and Marsden, A. L. (2020), 'Multilevel and multifidelity uncertainty quantification for cardiovascular hemodynamics', *Computer Methods in Applied Mechanics and Engineering*, 365, p. 113030. DOI: 10.1016/j.cma.2020.113030.

[114] Vasquez, V. R. and Whiting, W. B. (2005), 'Accounting for Both Random Errors and Systematic Errors in Uncertainty Propagation Analysis of Computer Models Involving Experimental

Measurements with Monte Carlo Methods', *Risk Analysis*, 25 (6), pp. 1669–1681. DOI: 10.1111/j.1539-6924.2005.00704.x.

[115] Groen, E. A., Bokkers, E. A. M., Heijungs, R. and de Boer, I. J. M. (2017), 'Methods for global sensitivity analysis in life cycle assessment', *The International Journal of Life Cycle Assessment*, 22 (7), pp. 1125–1137. DOI: 10.1007/s11367-016-1217-3.

[116] Saltelli, A. and Bolado, R. (1998), 'An alternative way to compute Fourier amplitude sensitivity test (FAST)', *Computational Statistics and Data Analysis*, 26 (4), pp. 445–460. DOI: 10.1016/S0167-9473(97)00043-1.

[117] Sobol, I. M. (1993), 'Sensitivity analysis for nonlinear mathematical models', *M. V. Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, Moscow*, 1 (4), pp. 407–414. DOI: 10.18287/0134-2452-2015-39-4-459-461.

[118] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S. (2007), *Global Sensitivity Analysis. The Primer*, *Global Sensitivity Analysis. The Primer*. Chichester, UK: John Wiley & Sons, Ltd. DOI: 10.1002/9780470725184.

[119] Saltelli, A. (2002), 'Sensitivity analysis for importance assessment', *Risk Analysis*, 22 (3), pp. 579–590. DOI: 10.1111/0272-4332.00040.

[120] DeCarlo, E. C., Mahadevan, S. and Smarslok, B. P. (2018), 'Efficient global sensitivity analysis with correlated variables', *Structural and Multidisciplinary Optimization*, 58 (6), pp. 2325–2340. DOI: 10.1007/s00158-018-2077-1.

[121] Iooss, B. and Lemaître, P. (2015), 'A Review on Global Sensitivity Analysis Methods', in Dellino, G. and Meloni, C. (eds) *Operations Research/Computer Science Interfaces Series*. Springer, Boston, MA (Uncertainty Management in Simulation - Optimization of Complex Systems), pp. 101–122. DOI: 10.1007/978-1-4899-7547-8_5.

[122] Patelli, E., Pradlwarter, H. J. and Schuëller, G. I. (2010), 'Global sensitivity of structural variability by random sampling', *Computer Physics Communications*, 181 (12), pp. 2072–2081. DOI: 10.1016/j.cpc.2010.08.007.

[123] Li, Y., Chen, J. and Feng, L. (2013), 'Dealing with Uncertainty: A Survey of Theories and Practices', *IEEE Transactions on Knowledge and Data Engineering*, 25 (11), pp. 2463–2482. DOI: 10.1109/TKDE.2012.179.

[124] Ellison, S. and Williams, A. (eds) (2012), *Quantifying Uncertainty in Analytical Measurement*. 3rd edn, *EURACHEM/CITAC working group*. 3rd edn. EURACHEM/CITAC working group. Available at: https://www.eurachem.org/index.php/publications/guides/quam.

[125] Willink, R. (2013), 'An improved procedure for combining Type A and Type B components of

measurement uncertainty', *International Journal of Metrology and Quality Engineering*, 4 (1), pp. 55–62. DOI: 10.1051/ijmqe/2012038.

[126] Cardin, M.-A., Nuttall, W. J., de Neufville, R. and Dahlgren, J. (2007), 'Extracting Value from Uncertainty: A Methodology for Engineering Systems Design', *INCOSE International Symposium*, 17 (1), pp. 668–682. DOI: 10.1002/j.2334-5837.2007.tb02903.x.

[127] Funtowicz, S. O. and Ravetz, J. R. (1990), *Uncertainty and Quality in Science for Policy*. Edited by W. Leinfellner and G. Eberlein. Dordrecht: Springer Netherlands. DOI: 10.1007/978-94-009-0621-1.

[128] Berner, C. L. and Flage, R. (2016), 'Comparing and integrating the NUSAP notational scheme with an uncertainty based risk perspective', *Reliability Engineering & System Safety*, 156, pp. 185–194. DOI: 10.1016/j.ress.2016.08.001.

[129] Farsi, M., Erkoyuncu, J. A., Steenstra, D. and Roy, R. (2019), 'A modular hybrid simulation framework for complex manufacturing system design', *Simulation Modelling Practice and Theory*, 94, pp. 14–30. DOI: 10.1016/j.simpat.2019.02.002.

[130] Schönmann, A., Dengler, C., Reinhart, G. and Lohmann, B. (2018), 'Anticipative strategic production technology planning considering cyclic interactions', *CIRP Journal of Manufacturing Science and Technology*, 23, pp. 118–127. DOI: 10.1016/j.cirpj.2018.07.002.

[131] Schwabe, O. (2018), *Geometrical framework for forecasting cost uncertainty in innovative high value manufacturing*. PhD thesis. Cranfield University, Cranfield. Available at: https://dspace.lib.cranfield.ac.uk/handle/1826/13616.

[132] Castrup, H. (2001), 'An Investigation into Estimating Type B Degrees of Freedom', *Bakersfield: Integrated Sciences Group*, pp. 1–5. Available at: www.isgmax.com.

[133] Magnusson, B., Näykki, B., Hovind, H. and Krysell, M. (2012), *Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories*. 4th edn. TR 537. Nordtest, Oslo. Available at: www.nordtest.info/wp/2017/11/29/handbook-for-calculation-of-measurement-uncertainty-in-environmental-laboratories-nt-tr-537-edition-4/.

[134] Van der Sluijs, J. P., Potting, J., Risbey, J., Van Vuuren, D., De Vries, B., Beusen, A., Heuberger, P., Corral Quintana, S., Funtowicz, S. 0., Kloprogge, P., Nuijten, D., Petersen, A. and Ravetz, J. (2002), *Uncertainty assessment of the IMAGE/TIMER B1 CO 2 emissions scenario, using the NUSAP method*. 410 200 104. Dutch National Research Programme on Global Air Pollution and Climate Change. Available at: www.nusap.net/workshop/report/finalrep.pdf.

[135] Erkoyuncu, J. A., Durugbo, C., Shehab, E., Roy, R., Parker, R., Gath, A. and Howell, D. (2013), 'Uncertainty driven service cost estimation for decision support at the bidding stage', *International Journal of Production Research*, 51 (19), pp. 5771–5788. DOI: 10.1080/00207543.2013.794318.

[136] Randhawa, J. S. and Ahuja, I. S. (2017), 'Examining the role of 5S practices as a facilitator of business excellence in manufacturing organizations', *Measuring Business Excellence*, 21 (2), pp. 191–206. DOI: 10.1108/MBE-09-2016-0047.

[137] Datta, P. P. and Roy, R. (2010), 'Cost modelling techniques for availability type service support contracts: A literature review and empirical study', *CIRP Journal of Manufacturing Science and Technology*, 3 (2), pp. 142–157. DOI: 10.1016/j.cirpj.2010.07.003.

[138] Bate, I., Griffin, D. and Lesage, B. (2020), 'Establishing Confidence and Understanding Uncertainty in Real-Time Systems', in *Proceedings of the 28th International Conference on Real-Time Networks and Systems*. New York, NY, USA, pp. 67–77. DOI: 10.1145/3394810.3394816.

[139] Laplante, P. A. (2004), 'The certainty of uncertainty in real-time systems', *IEEE Instrumentation and Measurement Magazine*, 7 (4), pp. 44–50. DOI: 10.1109/MIM.2004.1383464.

[140] Shah, S. A. B., Rashid, M. and Arif, M. (2020), 'Estimating WCET using prediction models to compute fitness function of a genetic algorithm', *Real-Time Systems*, 56 (1), pp. 28–63. DOI: 10.1007/s11241-020-09343-2.

[141] Rashid, M., Shah, S. A. B., Arif, M. and Kashif, M. (2020), 'Determination of worst-case data using an adaptive surrogate model for real-time system', *Journal of Circuits, Systems and Computers*, 29 (1), p. 2050005. DOI: 10.1142/S021812662050005X.

[142] Smart, C. (2014), 'Bayesian parametrics: how to develop a CER with limited data and even without data', *International Cost Estimating and Analysis Association*, pp. 1–23. Available at: http://www.iceaaonline.com/ready/wp-content/uploads/2014/07/PA-3-Paper-Bayesian-Parametrics-Developing-a-CER-with-Limited-Data-and-Even-Without-Data.pdf.

[143] Hochbaum, D. S. and Wagner, M. R. (2015), 'Production cost functions and demand uncertainty effects in price-only contracts', *IIE Transactions (Inst Ind Engrs)*, 47 (2), pp. 190–202. DOI: 10.1080/0740817X.2014.938843.

[144] Stockton, D. and Wang, Q. (2004), 'Developing cost models by advanced modelling technology', *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 218 (2), pp. 213–224. DOI: 10.1243/095440504322886532.

[145] Schwabe, O., Shehab, E. and Erkoyuncu, J. (2015), 'Geometric Quantification of Cost Uncertainty Propagation: A Case Study', *Procedia CIRP*, 37, pp. 158–163. DOI: 10.1016/j.procir.2015.08.078.

[146] Lei, Y., Li, N., Guo, L., Li, Ningbo, Yan, T. and Lin, J. (2018), 'Machinery health prognostics: A systematic review from data acquisition to RUL prediction', *Mechanical Systems and Signal Processing*, 104, pp. 799–834. DOI: 10.1016/j.ymssp.2017.11.016.

[147] Lanza, G. and Rühl, J. (2009), 'Simulation of service costs throughout the life cycle of production

facilities', *CIRP Journal of Manufacturing Science and Technology*, 1 (4), pp. 247–253. DOI: 10.1016/j.cirpj.2009.06.004.

[148] Jouin, M., Gouriveau, R., Hissel, D., Péra, M.-C. and Zerhouni, N. (2016), 'Particle filter-based prognostics: Review, discussion and perspectives', *Mechanical Systems and Signal Processing*, 72–73, pp. 2–31. DOI: 10.1016/j.ymssp.2015.11.008.

[149] Wang, B., Lu, J., Yan, Z., Luo, H., Li, T., Zheng, Y. and Zhang, G. (2019), 'Deep uncertainty quantification: a machine learning approach for weather forecasting', in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, USA, pp. 2087–2095. DOI: 10.1145/3292500.3330704.

[150] Chen, X., Yu, J., Tang, D. and Wang, Y. (2012), 'A novel PF-LSSVR-based framework for failure prognosis of nonlinear systems with time-varying parameters', *Chinese Journal of Aeronautics*, 25 (5), pp. 715–724. DOI: 10.1016/S1000-9361(11)60438-X.

[151] Cerdeira, P. B. O., Galvão, R. K. H. and Malère, J. P. P. (2013), 'Particle filter prognostic applied in landing gear retraction', *PHM 2013 - Proceedings of the Annual Conference of the Prognostics and Health Management Society 2013*, (2011), pp. 616–623. DOI: https://doi.org/10.36001/phmconf.2013.v5i1.2180.

[152] Maier, M., Zwicker, R., Akbari, M., Rupenyan, A. and Wegener, K. (2019), 'Bayesian optimization for autonomous process set-up in turning', *CIRP Journal of Manufacturing Science and Technology*, 26, pp. 81–87. DOI: 10.1016/j.cirpj.2019.04.005.

[153] Fujishima, M., Narimatsu, K., Irino, N., Mori, M. and Ibaraki, S. (2019), 'Adaptive thermal displacement compensation method based on deep learning', *CIRP Journal of Manufacturing Science and Technology*, 25, pp. 22–25. DOI: 10.1016/j.cirpj.2019.04.002.

[154] Ahmed, T. and Srivastava, A. (2016), 'Predicting human interest: an application of artificial intelligence and uncertainty quantification', *Journal of Uncertainty Analysis and Applications*, 4 (1), p. 9. DOI: 10.1186/s40467-016-0051-2.

[155] Beven, K. and Freer, J. (2001), 'Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology', *Journal of Hydrology*, 249 (1–4), pp. 11–29. DOI: 10.1016/S0022-1694(01)00421-8.

[156] Yang, L., Guo, Y. and Kong, Z. (2019), 'On the performance evaluation of a hierarchical-structure prototype product using inconsistent prior information and limited test data', *Information Sciences*, 485, pp. 362–375. DOI: 10.1016/j.ins.2019.02.018.

[157] Hariri-Ardebili, M. A. and Barak, S. (2020), 'A series of forecasting models for seismic evaluation of dams based on ground motion meta-features', *Engineering Structures*, 203, p. 109657. DOI: 10.1016/j.engstruct.2019.109657.

[158] Gal, Y. and Ghahramani, Z. (2016), 'Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning', in *Proceedings of the 33rd International Conference on Machine Learning*. New York, USA, pp. 1050–1059. Available at: http://proceedings.mlr.press/v48/gal16.pdf.

[159] Song, T., Ding, W., Liu, H., Wu, J., Zhou, H. and Chu, J. (2020), 'Uncertainty quantification in machine learning modeling for multi-step time series forecasting: Example of recurrent neural networks in discharge simulations', *Water (Switzerland)*, 12 (3). DOI: 10.3390/w12030912.

[160] Mujeeb, S., Javaid, N., Ilahi, M., Wadud, Z., Ishmanov, F. and Afzal, M. (2019), 'Long short-term memory: a new price and load forecasting scheme for big data in smart cities', *Sustainability*, 11 (4), p. 987. DOI: 10.3390/su11040987.

[161] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014), 'Dropout: A simple way to prevent neural networks from overfitting', *Journal of Machine Learning Research*, 15, pp. 1929–1958.

[162] Zhu, Y., Zabaras, N., Koutsourelakis, P. S. and Perdikaris, P. (2019), 'Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data', *Journal of Computational Physics*, 394, pp. 56–81. DOI: 10.1016/j.jcp.2019.05.024.

[163] Cicuttin, A., Crespo, M. L., Mannatunga, K. S., Garcia, V. V., Baldazzi, G., Rignanese, L. P., Ahangarianabhari, M., Bertuccio, G., Fabiani, S., Rachevski, A., Rashevskaya, I., Vacchi, A., Zampa, G., Zampa, N., Bellutti, P., Picciotto, A., Piemonte, C. and Zorzi, N. (2016), 'A programmable System-on-Chip based digital pulse processing for high resolution X-ray spectroscopy', in *2016 International Conference on Advances in Electrical, Electronic and Systems Engineering (ICAEES)*, pp. 520–525. DOI: 10.1109/ICAEES.2016.7888100.

[164] Gal, Y. (2016), *Uncertainty in Deep Learning*. PhD thesis. Cambridge: University of Cambridge. Available at: http://mlg.eng.cam.ac.uk/yarin/thesis/thesis.pdf.

[165] Radaideh, M. I. and Kozlowski, T. (2020), 'Surrogate modeling of advanced computer simulations using deep Gaussian processes', *Reliability Engineering & System Safety*, 195, p. 106731. DOI: 10.1016/j.ress.2019.106731.

[166] Li, L., Xia, J., Xu, C.-Y. and Singh, V. P. (2010), 'Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models', *Journal of Hydrology*, 390 (3–4), pp. 210–221. DOI: 10.1016/j.jhydrol.2010.06.044.

[167] Simmons, J. A., Harley, M. D., Marshall, L. A., Turner, I. L., Splinter, K. D. and Cox, R. J. (2017), 'Calibrating and assessing uncertainty in coastal numerical models', *Coastal Engineering*, 125, pp. 28–41. DOI: 10.1016/j.coastaleng.2017.04.005.

[168] Jozefowicz, R., Zaremba, W. and Sutskever, I. (2015), 'An empirical exploration of Recurrent

Network architectures', *32nd International Conference on Machine Learning, ICML 2015*, 3, pp. 2332–2340.

[169] Wang, S., Zhang, X., Gao, D., Chen, B., Cheng, Y., Yang, Y., Yu, W., Huang, Z. and Peng, J. (2018), 'A Remaining Useful Life Prediction Model Based on Hybrid Long-Short Sequences for Engines', in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1757–1762. DOI: 10.1109/ITSC.2018.8569668.

[170] Sun, J., Zuo, H., Wang, W. and Pecht, M. G. (2012), '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, pp. 585–596. DOI: 10.1016/j.ymssp.2011.09.029.

[171] Porter, T. (2004), *Karl Pearson: The Scientific Life in a Statistical Age*. Oxford: Princeton University Press. Available at: http://www.jstor.org/stable/j.ctt4cgc9j.

[172] Roy, R., Shaw, A., Erkoyuncu, J. A. and Redding, L. (2013), 'Through-life engineering services', *Measurement and Control (United Kingdom)*, 46 (6), pp. 172–175. DOI: 10.1177/0020294013492283.

[173] Amin, M. S. R. (2015), 'The pavement performance modeling: deterministic vs. stochastic approaches', in S., K. and A., E. H. (eds) *Numerical Methods for Reliability and Safety Assessment*. Springer International Publishing, pp. 179–196. DOI: https://doi.org/10.1007/978-3-319-07167-1_5.

[174] Meier, H., Roy, R. and Seliger, G. (2010), 'Industrial Product-Service Systems - IPS2', *CIRP Annals - Manufacturing Technology*, 59 (2), pp. 607–627. DOI: 10.1016/j.cirp.2010.05.004.

[175] Roy, R., Stark, R., Tracht, K., Takata, S. and Mori, M. (2016), 'Continuous maintenance and the future – Foundations and technological challenges', *CIRP Annals - Manufacturing Technology*, 65 (2), pp. 667–688. DOI: 10.1016/j.cirp.2016.06.006.

[176] Tukker, A. (2004), 'Eight types of product–service system: eight ways to sustainability? Experiences from SusProNet', *Business Strategy and the Environment*, 13 (4), pp. 246–260. DOI: 10.1002/bse.414.

[177] Erkoyuncu, J. A., Roy, R., Shehab, E. and Wardle, P. (2009), 'Uncertainty challenges in service cost estimation for product-service systems in the aerospace and defence industries', in *Proceedings of the 1st CIRP Industrial Product-Service Systems (IPS2) Conference*, pp. 1–17. Available at: https://dspace.lib.cranfield.ac.uk/handle/1826/3830.

[178] Erkoyuncu, J. A., Durugbo, C., Shehab, E., Roy, R., Parker, R., Gath, A. and Howell, D. (2013), 'Uncertainty driven service cost estimation for decision support at the bidding stage', *International Journal of Production Research*, 51 (19), pp. 5771–5788. DOI: 10.1080/00207543.2013.794318.

[179] Saaty, T. L. (2006), *Fundamentals of decision making and priority theory with the Analytic Hierarchy Process, Vol. VI of the AHP series*. St Peterburgh.

[180] Aull-Hyde, R., Erdogan, S. and Duke, J. M. (2006), 'An experiment on the consistency of aggregated comparison matrices in AHP', *European Journal of Operational Research*, 171 (1), pp. 290–295. DOI: 10.1016/j.ejor.2004.06.037.

[181] Saaty, T. L. (2016), *The Analytic Hierarchy and Analytic Network Processes for the Measurement of Intangible Criteria and for Decision-Making*, *International Series in Operations Research and Management Science*. Springer Science & Business Media. DOI: 10.1007/978-1-4939-3094-4_29.

[182] Balioti, V., Tzimopoulos, C. and Evangelides, C. (2018), 'Multi-Criteria Decision Making Using TOPSIS Method Under Fuzzy Environment. Application in Spillway Selection', *Proceedings*, 2 (11), p. 637. DOI: 10.3390/proceedings2110637.

[183] Krohling, R. A. and Pacheco, A. G. C. (2015), 'A-TOPSIS - An approach based on TOPSIS for ranking evolutionary algorithms', *Procedia Computer Science*, 55 (Itqm), pp. 308–317. DOI: 10.1016/j.procs.2015.07.054.

[184] Wang, T. C. and Lee, H. Da (2009), 'Developing a fuzzy TOPSIS approach based on subjective weights and objective weights', *Expert Systems with Applications*, 36 (5), pp. 8980–8985. DOI: 10.1016/j.eswa.2008.11.035.

[185] Massam, H. (1988), 'Multi-Criteria (MCDM) Decision Making Techniques in Planning', *Progress in Planning*, 30 (Mcdm), pp. 1–84.

[186] Reiman, T. and Oedewald, P. (2004), 'Measuring maintenance culture and maintenance core task with CULTURE-questionnaire - A case study in the power industry', *Safety Science*, 42 (9), pp. 859–889. DOI: 10.1016/j.ssci.2004.04.001.

[187] Robson, C. (2016), *Real world research: a resource for users of social research methods in applied settings*. 4th edn. Wiley & Sons.

[188] Dixon, J. (1990), *The new performance challenge : measuring operations for world-class competition*. Homewood, Illinois: Business One Irwin.

[189] Charmaz, K. (2017), 'Constructivist grounded theory', *Journal of Positive Psychology*, 12 (3), pp. 299–300. DOI: 10.1080/17439760.2016.1262612.

[190] ACNS Support (2016), 'Building a Common Support Model for the Future', *Ministry of Defence, Brochure*.

[191] Grenyer, A., Addepalli, S., Zhao, Y., Oakey, L., Erkoyuncu, J. A. and Roy, R. (2018), 'Identifying challenges in quantifying uncertainty: case study in infrared thermography', *Procedia CIRP*, 73, pp. 108–113. DOI: 10.1016/j.procir.2018.03.301.

[192] Sceral, M., Erkoyuncu, J. A. and Shehab, E. (2018), 'Identifying information asymmetry challenges in the defence sector', in *Procedia Manufacturing: Proceedings of the 6th International Conference in Through-life Engineering Services*. Bremen, Germany, pp. 127–134. DOI: 10.1016/j.promfg.2018.01.018.

[193] Zhang, X., Han, X., Liu, X., Liu, R. and Leng, J. (2015), 'The pricing of product and value-added service under information asymmetry: A product life cycle perspective', *International Journal of Production Research*, 53 (1), pp. 25–40. DOI: 10.1080/00207543.2014.922707.

[194] Lind, H. and Muyingo, H. (2012), 'Building maintenance strategies: planning under uncertainty', *Property Management*, 30 (1), pp. 14–28. DOI: 10.1108/02637471211198152.

[195] Greenough, R. M. and Grubic, T. (2011), 'Modelling condition-based maintenance to deliver a service to machine tool users', *The International Journal of Advanced Manufacturing Technology*, 52 (9–12), pp. 1117–1132. DOI: 10.1007/s00170-010-2760-x.

[196] Grenyer, A., Schwabe, O., Erkoyuncu, J. A. and Zhao, Y. (2021), 'Dynamic multistep uncertainty prediction in spatial geometry', *Procedia CIRP*, 96, pp. 74–79. DOI: 10.1016/j.procir.2021.01.055.

[197] Ben-Haim, Y. and Hemez, F. M. (2012), 'Robustness, fidelity and prediction-looseness of models', *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 468 (2137), pp. 227–244. DOI: 10.1098/rspa.2011.0050.

[198] Putnik, G. D., Manupati, V. K., Pabba, S. K., Varela, L. and Ferreira, F. (2021), 'Semi-Double-loop machine learning based CPS approach for predictive maintenance in manufacturing system based on machine status indications', *CIRP Annals*, 70 (1), pp. 365–368. DOI: 10.1016/j.cirp.2021.04.046.

[199] Booth, A. and Brice, A. (2005), 'Evidence-based Practice for Information Professionals: A Handbook', *Journal of Documentation*, 61 (6), pp. 803–805. DOI: 10.1108/00220410510632095.

[200] LoBiondo-Wood, G. and Haber, J. (2002), *Critical reading strategies: overview of the research process*. 5th edn, *Nursing Research: Methods, Critical Appraisal, and Utilization*. 5th edn. St. Louis: Mosby.

[201] Schmid, K. M., Kugler, R., Nalabothu, P., Bosch, C. and Verna, C. (2018), 'The effect of pacifier sucking on orofacial structures: a systematic literature review', *Progress in Orthodontics*, 19 (1), p. 8. DOI: 10.1186/s40510-018-0206-4.

APPENDICES

Appendix A. Literature review methodology: Search, appraisal, and synthesis

A.1 Literature search

The literature search detailed the formulation of the search string entered in online databases with applied filters (article type and publication year), inclusion of previously cited and recommended papers (hand search), along with publications cited in highly relevant sources [94]. The resulting string and search results are illustrated below and in Table A.1:

Search string: *("Uncertainty quantification" AND (“aggregation” OR "industrial maintenance" OR “forecasting” OR "challenges" OR "complex engineering systems”))*

Table A.1. Database search results

Database	Search fields	Date	Documents found
Google Scholar	Title	07/07/2020	59
IEEE Xplore	Title-Abs-Key	07/07/2020	79
Science Direct	Title-Abs-Key	07/07/2020	275
Scopus (open access)	Title-Abs-Key	07/07/2020	218
		Total	**631**

From the database search, 148 files were downloaded on a basis of accessibility, format, title and date. The hand search sourced 119 papers, while 24 were sourced from citations within sourced papers. This resulted in a pool of documentation to assess in the appraisal phase. Inclusion and exclusion criteria are required to refine the results, as well as a structured data extraction methodology, defined in the following sections.

A.2 Appraisal of identified literature

A.2.1 Quality assessment

It is necessary to refine the number of publications obtained to appropriately satisfy the RQs and assess the evidence base. To do this, a critical assessment of relevance and

quality was conducted. The broad selection process in Figure A.1, adapted from Booth [53], was implemented considering the PICOC framework in Section 2.3, as well as other review examples and author experience. Specific inclusion and exclusion criteria, based on the PICOC framework, are identified in Table A.2 [53,54,199].

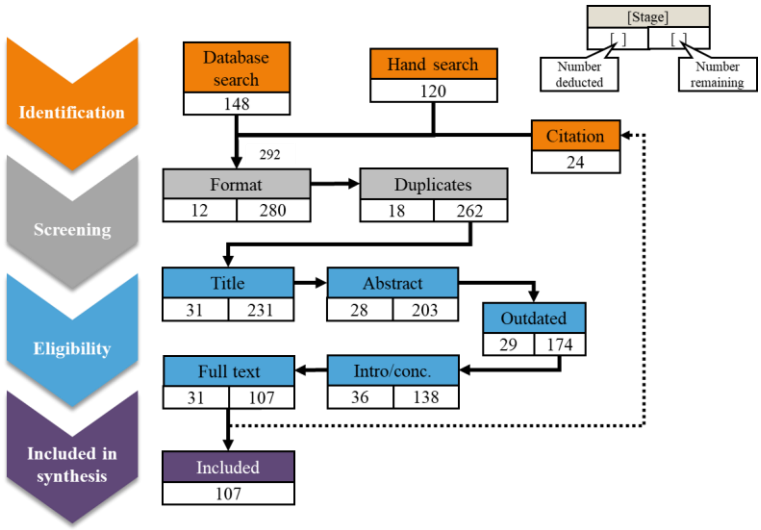


Figure A.1. Appraisal: Publication selection process [53,201]

Table A.2. Appraisal: Inclusion and exclusion criteria

Inclusion	Exclusion
• UQ theory, uncertainty prediction and analysis, contextual application, compound uncertainty • Uncertainty propagation and forecasting • Industrial maintenance applications • Clear techniques & referred sources for validity & additional searching	• Full source not accessible • Not written in English • Limited in-text citations or references to verify findings

A.2.2 Data extraction

A data extraction table was composed in MS Excel (Appendix A.2) to manage the literature and assess the evidence base, allowing different studies to be appraised in a consistent manner [53]. This included a record of:

- Publication details: Source folder, filename, publication title, author, year, type (journal, book, etc.), source method (database search, citation search, recommended) and author keywords

- Study details: Context, aims/objectives, methodologies/theory adopted, data collection strategies
- Results: Author's conclusions, outcome / findings, strengths, limitations

Publication details were recorded for all sources that passed the screening stage in Figure A.1. Eligibility was established in four main sifting stages: title, abstract, introduction/conclusion and full-text reading. If deemed eligible based on title, a preliminary understanding of study details and results was obtained from the abstract to gain familiarity and identify key information. Publications considered relevant were then looked over in more detail to gain a comprehensive understanding in the next two stages. This allowed papers to be summarised into categories and identify relationships for synthesis [200]. Cited publications within papers that could enhance the research picture were searched for directly and fed back into the start of the process. A total of 185 papers were eliminated in the process, based on the sifting stages illustrated in Figure A.1 and Figure A.2.

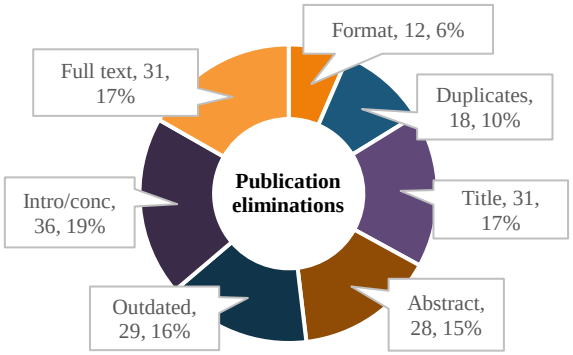


Figure A.2. Appraisal: Quality assessment – Publication eliminations

A.3 Synthesis of extracted data

The synthesis phase of the SALSA framework overlaps with the search and analysis phases to produce a breakdown of extracted data, comparing similarities and differences within each category [53]. This phase will identify what the literature says. The analysis identifies what it means.

Data extracted from the papers was categorised through a thematic synthesis. This is a well-validated method for synthesising qualitative data [53,54,56,58,201]. Key themes were established according to the research scope defined in the PICOC framework (Table

2.3), RQs, discussions with academic supervisors and the author's understanding of the topic. The thematic categorisation involved the generation of several categories for each established theme. This was achieved through a repetitive word counting process whereby the most frequently used words in the full text of each included paper were cross-referenced with the proposed category names using VLOOKUP functions in Excel (snapshot in Figure A.4). The most recurring words were more likely to be identified as categories that could be applied to the themes. Variations of words e.g., 'predict' and 'prediction' were included to account for word stemming and acronyms. This process required several iterations to combine and refine categories within a larger area and eliminate less frequent or irrelevant words; identified by the same method as key themes.

The category term most frequently recurring for each paper was highlighted. For each category, the number of highlighted cells over the 107 papers was added to the number of papers that contain that category term. The resulting 'score' was then used to identify the most relevant categories in each theme, combining similar terms. The resulting themes and categories are defined in Table A.3. Where applicable, the pros and cons of these categories are discussed in the Analysis phase. Theme and category definitions were determined through the author's interpretation of occurrences in literature as well as dictionary definitions. An example of the thematic synthesis data extraction for 3 papers is illustrated in Table A.4.

Table A.3. Synthesis: Definition of data extraction themes and categories

Theme	Definition	Category	Definition
Contextual application	The field in which the proposed framework or study is applied	Aerospace & defence	Includes defence and commercial or military aerospace sectors
		Emissions, energy & environment	Includes oil & gas, energy & power and greenhouse gas cases
		Manufacturing & maintenance	Includes general maintenance and manufacturing processes in miscellaneous applications
		Theory	Qualitative or quantitative theory and frameworks with no applied context
Analysis type	Type of analysis carried out according to the nature of uncertainty sources	Quantitative	Type A, considering purely statistical data sources
		Qualitative	Type B, considering purely heuristic data sources
		Compound	Combination of Type A and B data sources
Propagation & simulation techniques	Most prominent techniques used to propagate uncertainty in the analysis process	Bayesian	Expresses the probability of an event occurring given that a prior event has occurred
		Confidence	Probability that true parameter value lies within a specified range
		Correlation	Level of interdependence between 2 or more variables
		Degrees of freedom	Amount of information in a sample relevant to the estimation of a parameter
		Expertise / assumption	Derivation of a parameter through opinion-based, non-statistical means
		Fuzzy set theory	Function assigns a grade between 0 and 1 to each input parameter of a set, as opposed to Boolean which is 0 or 1
		Monte Carlo	Highly effective and flexible simulation technique to generate random variables about specified input parameters for multiple distribution types
		Neural network	Network of cooperating processing elements to give an output. This is applied to a model and 'trained' to give an optimum output
		Pedigree matrix	Scores results of qualitative expert judgement or assumptions according to predefined criteria to allow for quantitative assessment
		Sensitivity analysis	Identifies key input parameters for uncertainty analysis. Quantifies how

			changes in input value alter that of the outcome
		Survey / interview	Qualitative data collection method for expert or general population opinion on a given topic
		Other	Methods not used in many papers
Probability distributions	Type of distribution function (PDF) used to represent uncertainty about a given range in the analysis process	Beta	(See Table 2.7)
		Lognormal	“
		Normal	“
		Poisson	“
		Triangular	“
		Uniform	“
		Weibull	“
Uncertainty assessment and forecasting	Most prominent terms and qualities used to predict and forecast uncertainty	Challenges	Hinders, adds complexity or prevents action towards a given entity
		Deep learning	Use of artificial networks to learn from existing data to predict or optimise future results
		Forecasting	Predicting future trends based on past and present data
		Life cycle	A series of stages or developments that take place over the useful lifetime of a given product or service
		Optimisation	Finding the best or most effective use of a situation or resource
		Over time	Measurable progress of past, present, and future events
		Prediction	Estimate that something will happen or will be a consequence of something else – Synonym for forecasting

Table A.4. Synthesis: Thematic data extraction example for 3 papers

Theme	Simmons et al. [167]	Baek et al. [103]	Erkoyuncu et al. [13]
Contextual application	Emissions, energy & environment	Emissions, energy & environment	Manufacturing & maintenance
Analysis type	Quantitative	Compound	Compound
Propagation & simulation techniques	Bayesian, Confidence, Monte Carlo, Sensitivity analysis	Confidence, Correlation. Monte Carlo, Pedigree matrix, Sensitivity analysis, Survey / interview	Confidence, Sensitivity analysis
Probability distributions	Normal, Uniform	Normal, Uniform, Triangular	Normal, Triangular, Poisson, Weibull
Uncertainty assessment and forecasting	Estimation, Optimisation, Prediction	Estimation, Life cycle	Estimation, Life cycle, Prediction

Publication details												
Title	**Author**	**Year**	**Publication**	**Publication**	**Author Keywords**							
A framework for considering uncertainty in quantitative life cycle	Goh, Y., Newnes, L., McMahon, C., Mileham, A., Paredis, C.	2010	Conference	Proceedings of the ASME								
A framework for geometric quantification and forecasting of	Schwabe, O., Shehab, E., Erkoyuncu, J. A.	2016	Journal	Progress in Aerospace	Cost estim	Geom	Symm	Topol	Uncer			
A GEOMETRICAL FRAMEWORK FOR FORECASTING COST UNCERTAINTY	Schwabe,O.	2018	Thesis	PhD thesis	Cost estim	Cost uncer	Geom	Scarc				
A new uncertainty importance measure	Borgonovo, E.	2007	Journal	Reliability Engineering	Globa l	Impor	Proba	Uncer	Uncer			
A novel standard for the experimental estimation of the	Lanza, G., Viering, B.	2011	Journal	CIRP Annals -								
A procedure for the evaluation of measurement uncertainty based on	Willink, R	2005	Journal	Metrologia								
Aleatoric or Epistemic? Does it matter?	Kiureghian, A., Ditlevsen, O.	2009	Conference	Special Workshop								
Approaches to mitigate the Impact of Uncertainty in Development	Chalupnik, M., Wynn, D., Clarkson, P.	2009	Conference	16th Internationa	flexibi lity	produ ct	reliab ility	robus tness	uncer tainty	versat ility		
Assessment of uncertainty in structured surfaces using	MacAulay, G., Giusca, C.	2016	Journal	CIRP Annals -	Struct ured	Surfac	Uncer					
DEALING WITH UNCERTAINTY: A SURVEY OF THEORIES AND	Li, Y., Chen, J., Feng, L.	2013	Journal	IEEE Transaction	demp ster-	uncer tainty	fuzzy datab	fuzzy theor	info- gap	proba bilisti	proba bility	
Empirically based uncertainty factors for the pedigree matrix in ecoinvent	Ciroth, A., Muller, S., Weidema, B., Lesage, P.	2016	Journal	Internationa l Journal of	Data qualit	Ecoin	Empir	Pedig	Uncer	Uncer		
Estimating and Combining Uncertainties	Castrup, Howard	2004	Conference	8th Annual ITEA								
Extracting value from uncertainty: A methodology for engineering	Cardin, M, Neufville, R., Dahlgren, J.	2007	Conference	17th Symposium								
Predicting Human Interest: An Application of Artificial Intelligence	Ahmed, T., Srivastava, A.	2016	Journal	Uncertainty Analysis	artific ial	data engin	intere st	intere st	uncer tainty			
Quantification of margins and uncertainties: Alternative	Helton, J., Johnson, J.	2011	Journal	Reliability Engineering	aleat ory	episte mic	evide nce	interv al	possi bility	proba bility	quant ificati	uncer tainty
Quantifying Uncertainty in Analytical Measurement	Ellison, S., Williams, A.	2012	Journal	EURACHEM / CITAC								
Quantitative Inventory Uncertainty	Ciroth, A.	2013	Journal	Greenhouse gas protocol								

Figure A.3. Snapshot of data extraction table for publication details, study details and results of selected papers

Study details					Results			
Category sp	Context	Aims	Methodology	Data collection strategy	Author conclusions	Outcome/findings	Strengths	Limitations
Life-cycle costing	Uncertainty in LCC has significant influence on	Presents a framework for considering	Overview of literature - qualitative	Overview of literatur Development framework	Presented a framework for considering	Presented a framework for considering	Novel framework to validate LCC for	
Cost uncertainty	Quantification and forecasting of cost	Develop a framework based on the principles	Qualitative - literature review, expert	Qualitative - literature review, expert	Uncertainty of a cost estimate at any future	Cost variance can be quantified through	Novel combination of methods, clear	Some figures are unclear, only relates to
Life-cycle costing	Oliver Schwabe thesis Increasing competition	The aim of the investigation was to	Standard lit review Research methodology:	Various qualitative and quantitative	Aims & objective achieved, hypothesis	Key contributions to knowledge:		
Sensitivity &	Uncertainty in parameters leads to	Discuss proposed global sensitivity indicator with	Propose Global SA indicator, derive	Other methods - quantitative	Literature refined and established theoretical	New method compared with others show that	Novel method that improves on previous	Only focuses on variance on the input,
Type A estimation	Uncertainty estimation in micro gear	Present a novel standard for micro gear	Use of cylinders to represent involute tooth	Quantitative analysis of proposed cylindrical	Presents 2 approaches for the experimental	Process for estimating measurement	Novel, robust process, well cited	Purely addresses measurement (Type A),
Type B estimation	Method to evaluate measurement	Propose a practical, transparent and	Quantitative theory to combine and estimate	Quantitative theory to combine and estimate	Allows treatment of correlated input	Measurand distribution is approximated by the	Strong quantitative theory and novel	How can it be applied?
Guide/defin itions	Many sources of uncertainty are	Discuss whether it is necessary to distinguish	Qualitative discussions with literature sources	Qualitative - literature	The distinction between aleatory and epistemic	The distinction between aleatory and epistemic	Strong comparison and discussions	
Mitigation &	Complexity and pressure in the product	Mitigate the likely impact of process risks	Qualitative examination of literature in 2	Qualitative literature - current methodologies	Introduced conceptual framework to clarify	Conceptual framework based on system	Robust framework developed, thoroughly	
Type A estimation	Uncertainty in metrological properties	Presents a method uncertainty in structured	Quantitative data such as a flatness map are	Quantitative measurements and	Presented an approach to calculate	Presented an approach to calculate	Good example of combining measurement	Only focuses on measurement
Methodolog ies	Examination of theories and practices for	Give a structured overview of existing	Qualitative - from literature, overview of	Qualitative - from literature, overview of	Knowledge of uncertainty is necessary	Knowledge of uncertainty is necessary	Structured overview and comparison of	
Type B estimation	Ecoinvent - method for estimation of standard	Presents: 1. A procedure to obtain an empirical	Considers uncertainty as a geometric standard	Ecoinvent database - Quantitative	Gives an approach to obtain an empirical	Derived empirically based uncertainty factor	Useful methodology for pedigree matrix, well	Only considers lognormal distribution,
Uncertainty derivation	Guidance to estimate and combine statistical	Present a general uncertainty	Theoretical derivation of formulas - Quantitative	Qualitative - literature review	Measurement errors are random variables that	Defined axioms (basic theory)	Strong mathematical theory	All theory - no contextual application
Mitigation &	Enabling design flexibility by examining	Present a design methodology that adds	Design process developed from	Qualitative from literature	Approach allows designers to consider	Approach allows designers to consider	Robust framework developed, thoroughly	Cases only focus on military applications
Automation	Numerical estimation of human interest towards	Propose a novel approach to estimate	Use of Bayesian inference to estimate	Use of Stack Overflow to measure daily changes	Model for quantifying level of interest was	Model for quantifying level of interest was	Rigorous approach, well tested, well cited	Initial model, needs improving
Aleatory and	Nuclear weapons risk and reliability	Illustrate alternatives to possibility theory for the	Quantitative - Considers mathematical structure	Quantitative theory from literature	Uncertainty analysis for complex systems	Uncertainty analysis for complex systems	Comprehensive analysis of theories and	Challenges: Aggregation of information from
Guide/defin itions	Guide: Evaluation and expression of	Provide a practical, understandable and	Gives guidance based on existing methods	Guide - Quantitative		Gives guidance based on existing methods	Application examples given	
Guide/defin itions	Guidance on going beyond qualitative	Guidance on going beyond qualitative	Use of pedigree matrix for qualitative data,	Example data from Greenhouse gas	Shows the importance of providing uncertainty	Results of example	Clear use of existing methods, well cited	No clear author or date

Figure A.3 (cont'd). Snapshot of data extraction table for publication details, study details and results of selected papers

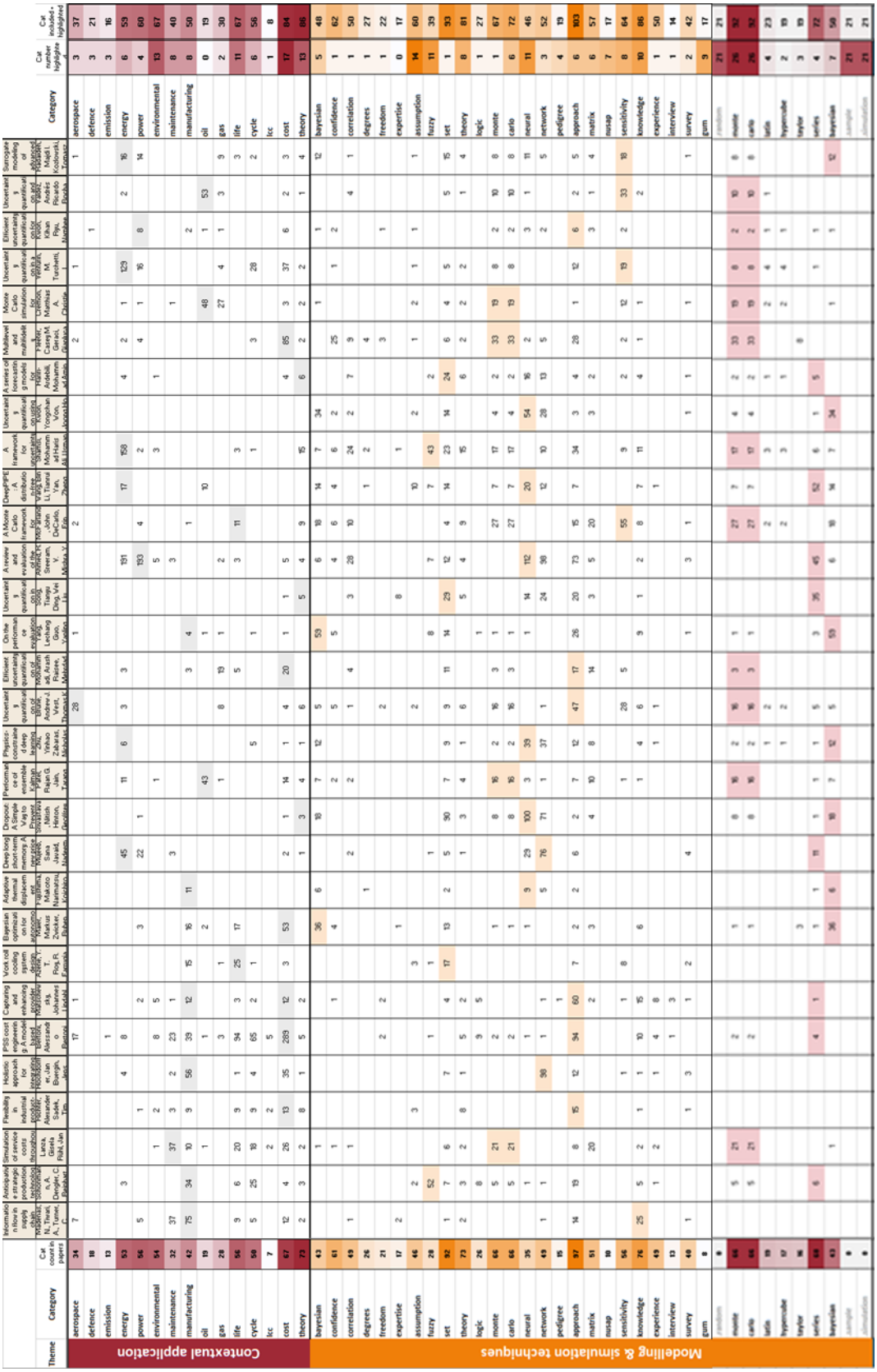
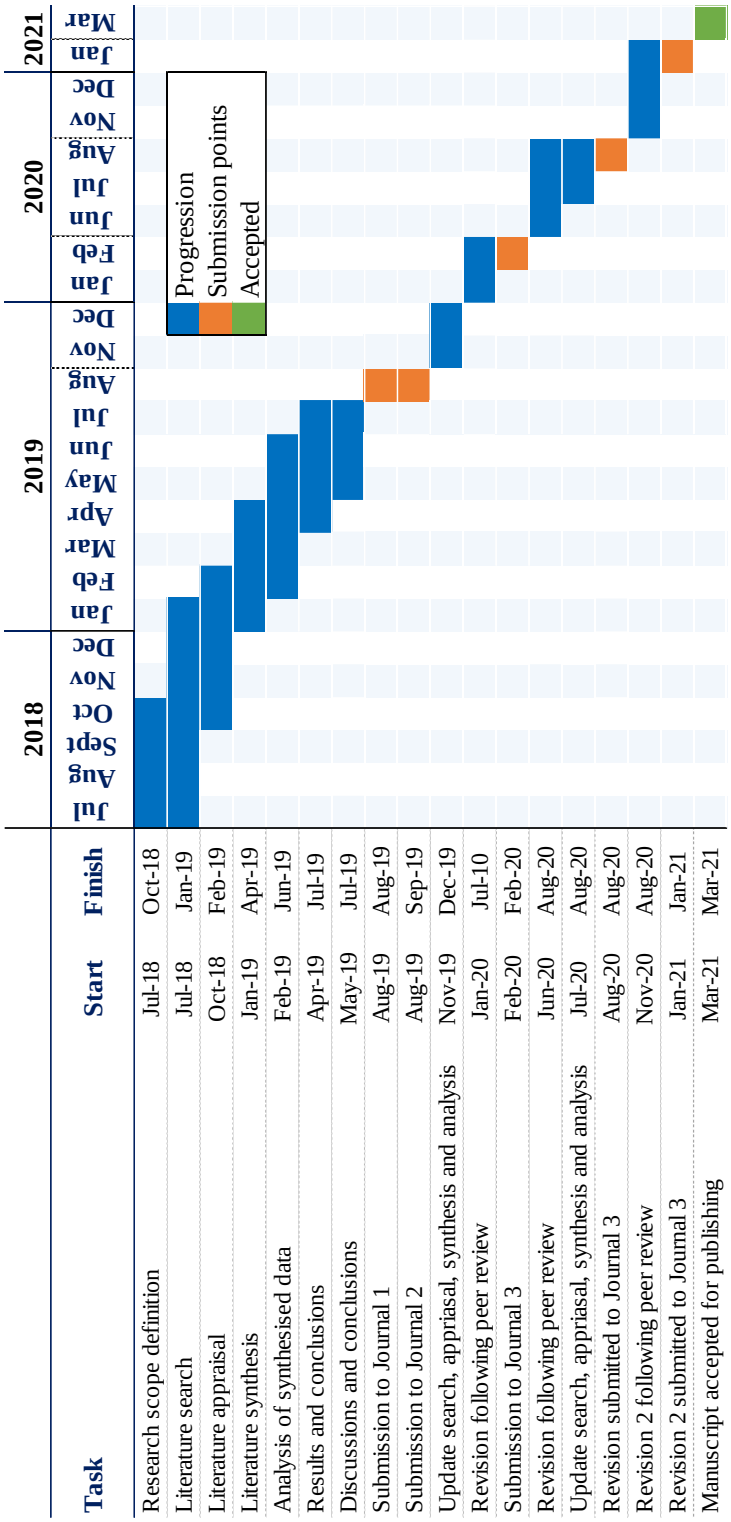


Figure A.4. Snapshot of word frequency count matrix for the thematic synthesis using Excel VLOOKUP functions

A.4 Review timeline



Task	Start	Finish
Research scope definition	Jul-18	Oct-18
Literature search	Jul-18	Jan-19
Literature appraisal	Oct-18	Feb-19
Literature synthesis	Jan-19	Apr-19
Analysis of synthesised data	Feb-19	Jun-19
Results and conclusions	Apr-19	Jul-19
Discussions and conclusions	May-19	Jul-19
Submission to Journal 1	Aug-19	Aug-19
Submission to Journal 2	Aug-19	Sep-19
Update search, appriasal, synthesis and analysis	Nov-19	Dec-19
Revision following peer review	Jan-20	Jul-10
Submission to Journal 3	Feb-20	Feb-20
Revision following peer review	Jun-20	Aug-20
Update search, appriasal, synthesis and analysis	Jul-20	Aug-20
Revision submitted to Journal 3	Aug-20	Aug-20
Revision 2 following peer review	Nov-20	Aug-20
Revision 2 submitted to Journal 3	Jan-21	Jan-21
Manuscript accepted for publishing	Mar-21	Mar-21

Figure A.5. Review timeline

A.5 Research methods validity and neutrality

The validity of research methods is distinguished here as the extent to which they achieve the objectives. Neutrality is the measure to avoid bias and increase transparency and replicability of the research. The following points examine these traits for the frameworks and methods adopted in this review.

- **Systematic review procedure:** The SALSA framework was adopted to carry out the review procedure due to its contextual flexibility and validity, as well as its successful implementation in other systematic reviews [53,54,56,58]
- **Scoping framework:** The PICOC framework (Table 2.3) was adopted to scope the research and define the aim, objectives and research questions as it provides a transparent and duplicable identification of key concepts to be implemented in the SALSA framework.
- **Literature search:** The PICOC framework was used to construct, refine and enhance the search string (Table A.1). Literature deemed to encapsulate the scope of the research criteria was selected to assess in the appraisal phase.
- **Appraisal:** Inclusion and exclusion criteria were defined through the research scope and PICOC framework, as well as examples in literature [53,54,199]. Publications were eliminated on a basis of format (accessibility), duplication, title, abstract, date, introduction/conclusion and full-text reading (Figure A.1) according to these criteria via the author's interpretation of their relevance. The remaining papers were deemed most relevant to answer the research questions. Data management was upheld using the data extraction table described in Appendix A.2.2.
- **Synthesis:** Themes and categories were established through the repetitive word counting process described in Appendix A.3. This reproducible process was validated and refined by comparison with other reviews and academic feedback [56,58,200,201]
- **Analysis:** A combination of thematic, narrative, tabular and graphical approaches were adopted to examine the literature and answer the research questions. Types of uncertainty were discussed in Section 2.2 to provide context for the research scope.

Appendix B. Methodology selection – TOPSIS

B.1 TOPSIS process

The six steps of the TOPSIS process are as follows:

Step 1: Define decision matrix, D and criteria weights, W for each criterion, C_i compared to the approaches, A_j

$$D = \begin{matrix} \\ A_1 \\ A_2 \\ \vdots \\ A_j \end{matrix} \begin{matrix} \begin{matrix} C_1 & C_2 & \dots & C_i \end{matrix} \\ \begin{bmatrix} X_{1\,1} & X_{1\,2} & \cdots & X_{1\,i} \\ X_{2\,1} & X_{2\,2} & \cdots & X_{2\,i} \\ \vdots & \vdots & \ddots & \vdots \\ X_{j\,1} & X_{j\,2} & \cdots & X_{j\,i} \end{bmatrix} \end{matrix}$$

$$W = [W_1 \quad W_2 \quad \dots \quad W_i]$$

Step 2: Calculate normalised decision matrix, $\bar{X}_{ij}$

$$\bar{X}_{ij} = \frac{X_{ij}}{\sqrt{\sum_{i=1}^{n} X_{ij}^2}} \tag{B-1}$$

Step 3: Calculate weighted normalised matrix, V_{ij}

$$V_{ij} = \bar{X}_{ij} \times W_j \tag{B-2}$$

Step 4: Calculate ideal best, V^+ and worst, V^- value

$$\begin{aligned} V^+ &= (\tilde{V}_1^+, \tilde{V}_2^+, \dots, \tilde{V}_j^+) \\ V^- &= (\tilde{V}_1^-, \tilde{V}_2^-, \dots, \tilde{V}_j^-) \end{aligned} \tag{B-3}$$

Step 5: Calculate Euclidean distance from ideal best, S_i^+ and worst, S_i^- value

$$\begin{aligned} S_i^+ &= \sqrt{\sum_{j=1}^{m} \left(V_{ij} - V_j^+\right)^2} \\ S_i^- &= \sqrt{\sum_{j=1}^{m} \left(V_{ij} - V_j^-\right)^2} \end{aligned} \tag{B-4}$$

Step 6: Calculate performance score, P_i

$$P_i = \frac{S_i^-}{S_i^+ + S_i^-} \tag{B-5}$$

B.2 TOPSIS application

Scoring criteria for the identified approaches are adapted from inclusion and exclusion criteria used in the systematic review in Chapter 2 [97] and challenges recognised from interviews with industry in Chapter 3 [10]. These are shown in Table B.1 along with an indication of whether they are a benefit or cost to the ranking process.

Table B.1. Criteria definition with cost or benefit clarification

Number	Criteria	Cost or benefit
1	Applications in industrial maintenance for CES or PSS	Benefit
2	Presents flexible approach to quantify qualitative uncertainties	Benefit
3	Allows for consideration of human and environmental factors	Benefit
4	Demonstrates methods to combine uncertainties	Benefit
5	Considers methods for multiple types of PDF	Benefit
6	Enables interpolation of gaps in data	Benefit
7	Capability to forecast uncertainty	Benefit
8	Complexity of method	Cost
9	Referred sources for validity	Benefit

The decision matrix for each criterion, scored on a 5-point Likert scale defined in Table B.2, against the identified approaches in Table B.3. The criteria are numbered in the table heading. These scores are based on the author's understanding of the approaches following the systematic review, summarised by contrasting matrices in Appendix B.3, and conclusions drawn by the authors that presented the approaches in literature.

Table B.2. 5-point Likert scale definition for scoring

Score	Definition
5	Fulfils criterion fully to solve problem with implementation examples and is well cited/validate
4	Fulfils criterion in theory but is not widely validated
3	Fulfils criterion but is challenged by other sources in its accuracy and rigour
2	Does not satisfactorily fulfil criterion
1	Not related to criterion

Table B.3. Decision matrix for defined criteria against identified approaches, scored on 5-point Likert scale

Criteria	1	2	3	4	5	6	7	8	9
GUM	3	2	3	4	4	1	1	3	5
NUSAP	4	5	5	4	2	1	1	2	4
Muller	3	5	5	5	5	1	1	2	3
Willink	2	4	3	4	3	1	1	4	3
Nordtest	3	3	3	3	2	1	1	3	4
BPN	4	1	1	1	2	5	4	3	5
DUQ	4	3	1	3	2	4	5	3	3
Dropout	3	2	1	1	2	4	5	4	3
Spatial geo	5	3	1	2	3	4	5	4	4

The comparison matrices are scored on a 1-3 scale according to whether one method improves on, is neutral to, or not improved by the other. The sum product of these scores gives an initial indication of the best method, shown in Table B.4 and Table B.5.

Table B.4. Sum-product of scoring for RQ1 and RQ2

	GUM method	NUSAP	Muller method	Willink method	Nordtest approach
Green (3)	0	3	2	2	0
Orange (2)	1	1	2	1	1
Red (1)	3	0	0	1	3
Total	**5**	**10**	**10**	**9**	**5**

Table B.5. Sum-product of scoring for RQ3

	BPN	DUQ	Dropout	Spatial geometry
Green (3)	0	2	0	1
Orange (2)	3	1	2	2
Red (1)	0	0	1	0
Total	**6**	**8**	**5**	**7**

The subjective weights applied for the level of importance of each criterion to RQ1-3 are given in Table B.6. These were determined by the author on an interval of 0.2 between 0 and 1. Criteria that are not applicable to a problem are given a 0 weighting. These weights, as well as the scores in Table B.2, are subjective and naturally have a direct influence on the resulting rankings from the TOPSIS method. For a larger decision problem, additional insights from decision-makers would be necessary to take a mean or mode of each criterion and relative weight [183,184]. However, for this application, the scores are deemed sufficient.

Table B.6. Problem weightings for defined criteria

Criteria	1	2	3	4	5	6	7	8	9
RQ1 weight	0.6	0.4	0.2	1.0	1.0	0.0	0.0	0.6	0.6
RQ2 weight	0.6	1.0	1.0	0.4	0.4	0.0	0.0	0.6	0.6
RQ3 weight	0.6	0.2	0.0	0.4	0.4	0.8	1.0	0.8	0.6

The resulting rankings for each method to RQ1-3 are shown in Table B.7 (1 is ranked top, 5 is ranked last). Muller's [34] semi-quantitative approach is deemed the most suitable for RQ1 and 2, using the coefficient of variation to combine quantitative and qualitative inputs represented by different PDFs. Qualitative inputs are defined through the Pedigree method, as they are for NUSAP [52,128]. These two approaches may be amalgamated to optimise the use of Pedigree for RQ2.

Table B.7. Ranking of identified approaches

Rank	Methods for RQ1	Methods for RQ2	Methods for RQ3
1	Muller	Muller	Spatial geometry
2	GUM	NUSAP	DUQ
3	NUSAP	Willink	BPN
4	Willink	GUM	Dropout
5	Nordtest	Nordtest	

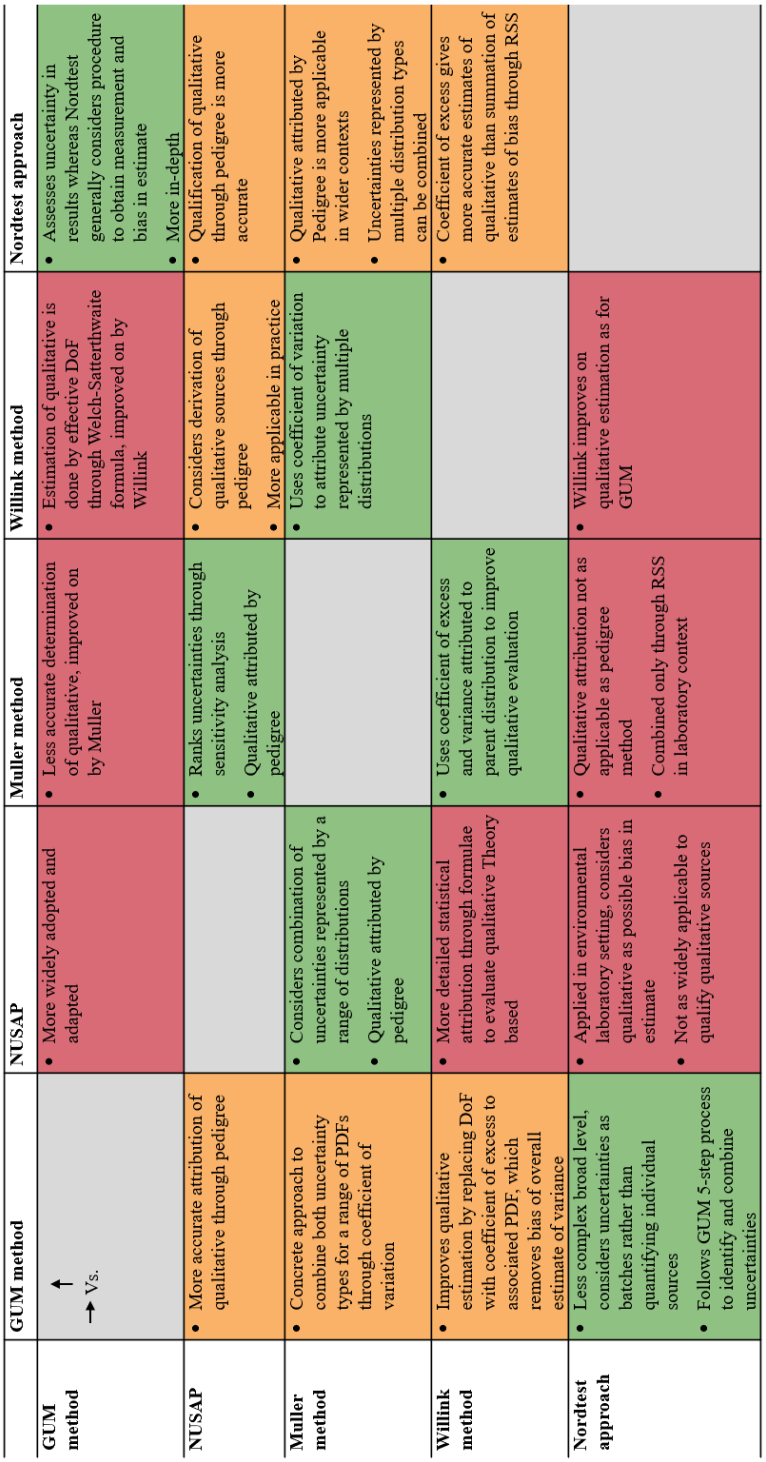
Schwabe's [28] spatial geometry is ranked best of the identified approaches for RQ3. While focused on cost uncertainty, the approach to determine geometric symmetry of data variance for a given point may be extended to consider uncertainties from a range of

inputs. Once qualitative inputs are converted through the pedigree approach, combined with quantitative types and equated through the coefficient of variation, Schwabe's approach can be applied to forecast how they may change over time. As with Wang's [149] and Gal's [15] approaches, spatial geometry can make use of dropout training to approximate Bayesian inference in Gaussian processes to update probability as more evidence becomes available.

B.3 Contrasting matrices of identified approaches

3	Left method improves on or is favourable to top
2	Methods can benefit each other / no clear positive or negative of using either
1	left method has been improved on by top

Approaches suited to RQ1 & RQ2:



	GUM method ← / → Vs.	NUSAP	Muller method	Willink method	Nordtest approach
GUM method		• More widely adopted and adapted	• Less accurate determination of qualitative, improved on by Muller	• Estimation of qualitative is done by effective DoF through Welch-Satterthwaite formula, improved on by Willink	• Assesses uncertainty in results whereas Nordtest generally considers procedure to obtain measurement and bias in estimate • More in-depth
NUSAP	• More accurate attribution of qualitative through pedigree		• Ranks uncertainties through sensitivity analysis • Qualitative attributed by pedigree	• Considers derivation of qualitative sources through pedigree • More applicable in practice	• Qualification of qualitative through pedigree is more accurate
Muller method	• Concrete approach to combine both uncertainty types for a range of PDFs through coefficient of variation	• Considers combination of uncertainties represented by a range of distributions • Qualitative attributed by pedigree		• Uses coefficient of variation to attribute uncertainty represented by multiple distributions	• Qualitative attributed by Pedigree is more applicable in wider contexts • Uncertainties represented by multiple distribution types can be combined
Willink method	• Improves qualitative estimation by replacing DoF with coefficient of excess to associated PDF, which removes bias of overall estimate of variance	• More detailed statistical attribution through formulae to evaluate qualitative Theory based	• Uses coefficient of excess and variance attributed to parent distribution to improve qualitative evaluation		• Coefficient of excess gives more accurate estimates of qualitative than summation of estimates of bias through RSS
Nordtest approach	• Less complex broad level, considers uncertainties as batches rather than quantifying individual sources • Follows GUM 5-step process to identify and combine uncertainties	• Applied in environmental laboratory setting, considers qualitative as possible bias in estimate • Not as widely applicable to qualify qualitative sources	• Qualitative attribution not as applicable as pedigree method • Combined only through RSS in laboratory context	• Willink improves on qualitative estimation as for GUM	

Approaches suited to RQ3:

	Backpropagation neural network (BPN)	Deep uncertainty quantification (DUQ)	Dropout as Bayesian approximation	Spatial geometry
Backpropagation neural network (BPN)	← Vs. →	• Backpropagation is implemented in DUQ to optimise the network	• Can be used together to optimise approximations	• Optimisation not applied but could be considered in future work
Deep uncertainty quantification (DUQ)	• Predictive network trained by negative log-likelihood error and optimised by backpropagation		• Combines deep learning and UQ to forecast multi-step meteorological time series • Fuses information from multiple sources and provides multiple output prediction • Improves generalisation compared to MSE and MAE	• Point estimation forecasting is not suitably applicable for human decisions • Training by negative log-likelihood error (NLE) significantly improves point estimation generalisation compared to MSE and MAE
Dropout as Bayesian approximation	• Compared against probabilistic backpropagation – found to give better performance against this and RMSE • Backpropagated with sampled Bernoulli random variables	• Uses dropout training as approximate Bayesian inference in Gaussian processes • Improves on predictive log-likelihood and RMSE		• Used to obtain uncertainty in the model itself • Framework continuously learns from new and past data and updates probability accordingly • Applicable in alternative domain to spatial geometry
Spatial geometry	• Able to forecast the rise in cost variance over time under conditions of small data to aid decision making • Neural networks not used so backpropagation not necessary in cost estimation context	• Describes how cost estimate uncertainty changes over time compared to setting a single point figure to be considered valid over the product life cycle • Focus is solely on quantitative cost uncertainty estimation – can qualitative factors be considered?	• Applicable where data available is not sufficient to fulfil the central limit theorem • Framework is used to obtain cost uncertainty estimates	

Appendix C. Current practice and challenges

C.1 Interview questions and discussion points

Part 1 – Influencing factors

- Discuss and highlight important factors from questionnaire results – Which ones can be filtered out?
- General discussion of scores given for each section – Mean score, highest effect on quality of maintenance
- How can this impact maintenance times/costs?
    - E.g. having to redo a task because it wasn't carried out effectively
- What can be improved?

Part 2 – Current practice

- Maintenance regimes – contractor vs client – effect on measurement accuracy
- Equipment quality
- Complex system context – change in one system maintained by another shareholder has an unknown impact on another, increasing uncertainty
- Discussion of scores given in survey

Methods or systems used to record maintenance data

- Are systems effective and fully utilised?
- If not, how could they be improved?
- How well is recorded data used to assess maintenance?
- How is data from previous projects used to influence decisions on new projects?

Risk and Uncertainty

- Key points that influence uncertainties in their field?
- How are these mitigated to reduce uncertainty?
- Are methods in place to measure changes in uncertainty over time?
- Are uncertainties categorised? (E.g. people, processes, equipment)

Summary of discussions

Appendix D. Compound uncertainty quantification and aggregation

D.1 Heat exchanger test rig

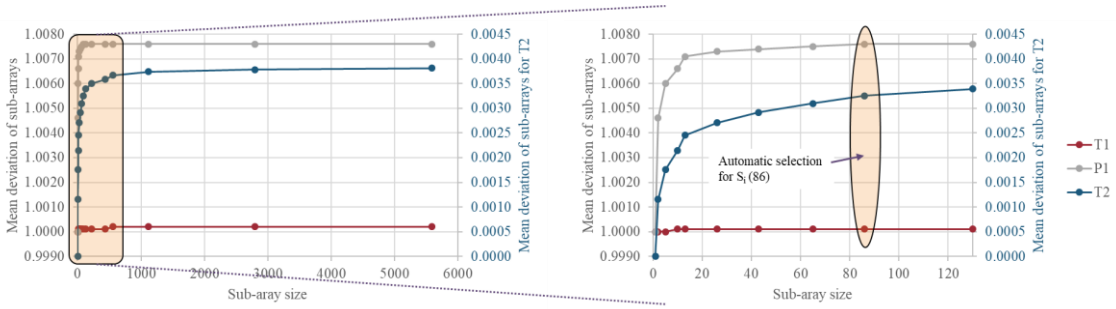


Figure D.1. Heat exchanger test rig: Increasing deviation (uncertainty) with sub-array size

D.2 C-MAPSS turbofan engine

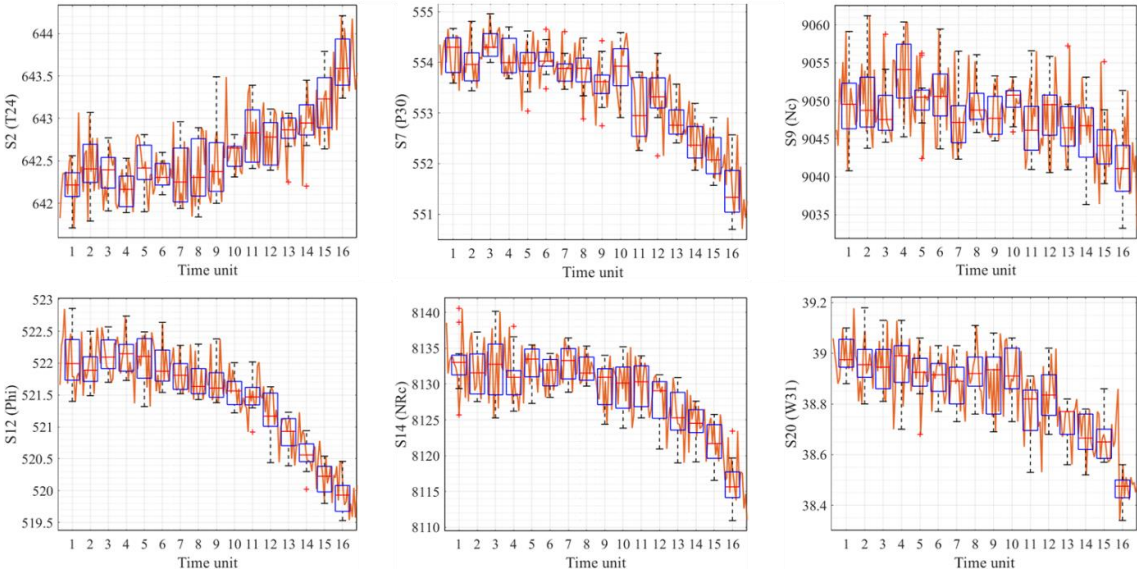


Figure D.2. C-MAPSS turbofan engine: Sub-array boxplots over time-series data (example for six input parameters)

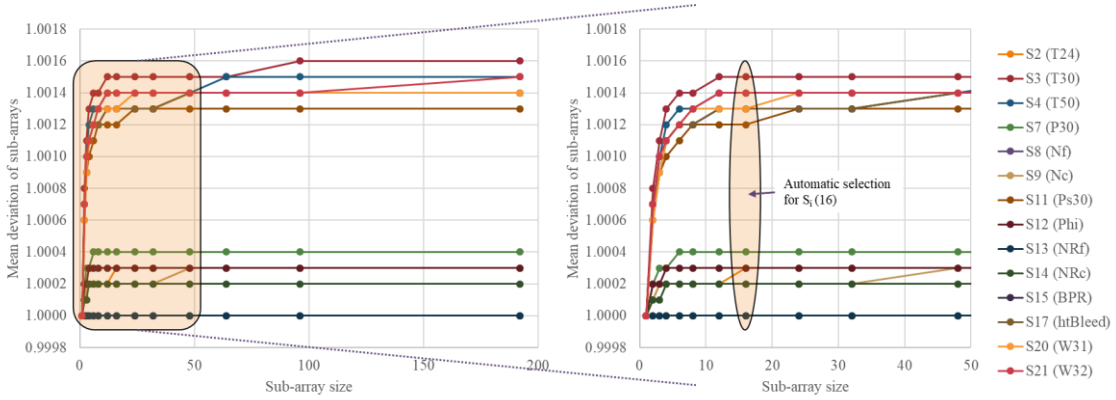


Figure D.3. C-MAPSS turbofan engine: Increasing deviation (uncertainty) with sub-array size

Appendix E. UPLD framework interview questions

Semi-structured interviews were held with key personnel from a leading defence company. This included data analysts and modellers, some of whom participated in the initial studies in Chapter 3. The interview questions and summary of answers is given below:

1. What methods are used?
    - LSTMs in prognostics and health management (PHM) – autoencoders used to reconstruct signals from assets
    - This is used to produce a health index – subject of patent, cannot discuss
    - Knowledge of reconstruction quality and operational defects is used to predict when maintenance will be needed
    - Question of how long components have been failing for
    - Scoring is a big issue – which method is best for different failure modes?
2. What time scales are forecasts made on and what factors are forecast or predicted? (Costs, turnaround times, required equipment, measured data)
    - There is a time window in which maintenance could be performed
    - An anomaly score is assigned for measured data
    - Depends on quality of reconstruction – if error is constant that's ok
3. How much data is typically available on which to base forecasts?
    - Highly variable on number of signals and sampling rate
    - Can have unpredictable gaps in the data
    - Uncertain if all failure modes are being captured
4. For what level are forecasts made? (Whole system units or individual components)
    - Binary classification if it's healthy or not – system level
    - Work on the level of a machine, not e.g., a faulty pipe
    - There is not enough data to define a specific fingerprint for specific faults

Appendix F. Implementations of the work: Perspectives on the effectiveness of the final framework

The first stage of the CUQA app is to load in the data and split it into sub-arrays to quantify the uncertainty. The automatic sub-array size is used as default (as defined in Chapter 4) along with the corresponding number of boxes, controlled by the left-hand panel in Figure F.1. When switched to manual, the user can select any other possible size using the 'Sub array size' drop-down component and view the corresponding plots.

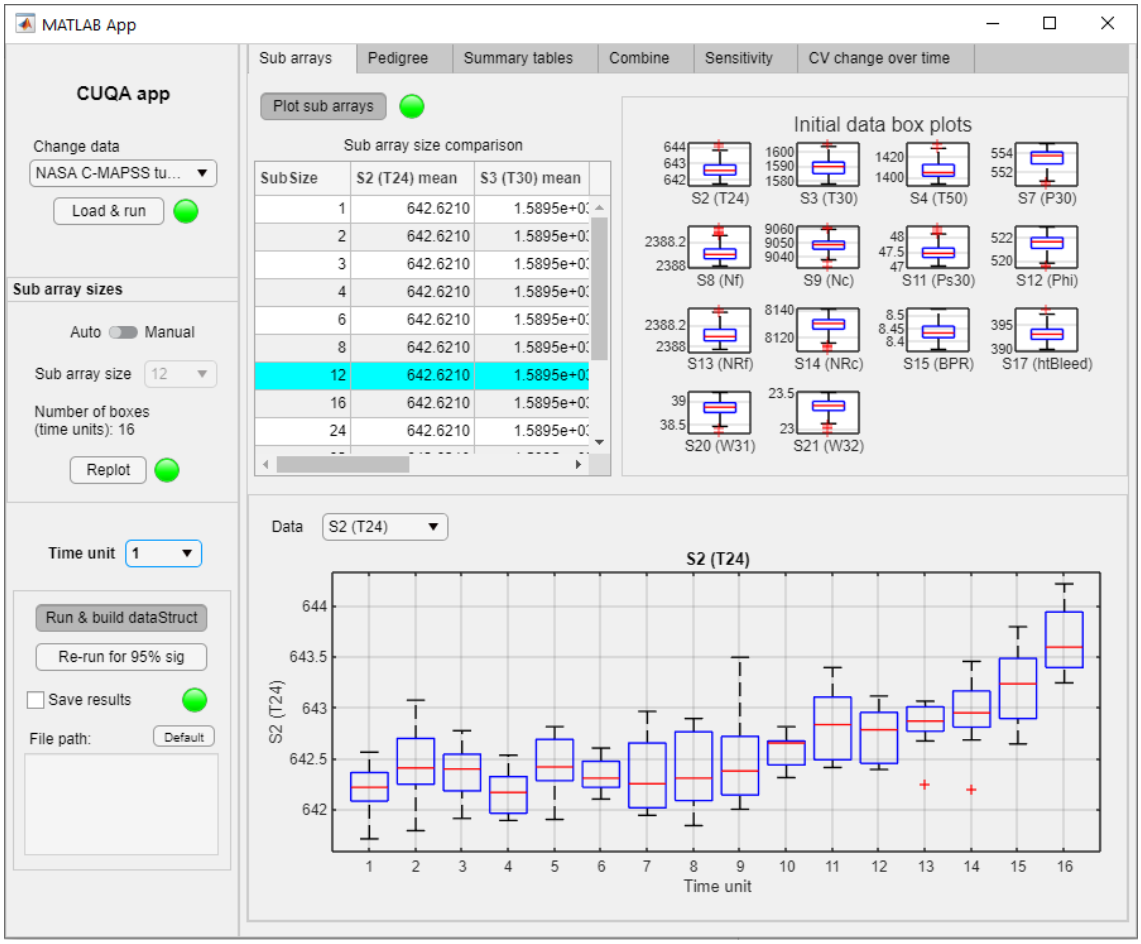


Figure F.1. MCDUQ app: CUQA sub-array tab

The displayed input data is changed using the 'Data' drop-down component. The examples in the figures use the C-MAPSS turbofan engine degradation dataset as described in Chapter 4 and Chapter 5.

The corresponding pedigree factors are plotted in the second tab shown in Figure F.2. The user can define uncertainty factors for increasing pedigree scores, displayed in the line plot, as well as the defined score for each factor.

When the ‘factor score type’ is set to ‘Random’, the app plots scores over the sub-array time period ±1 of the set score in individual bar charts. When set to ‘Constant’, the defined scores are held constant over the time period. This mirrors the CUQA framework defined in Chapter 4.

Figure F.2. MCDUQ app: CUQA pedigree factors tab

The summarised quantitative, qualitative and correlation data is plotted in the summary tables tab (Figure F.3). The correlation matrices are plotted in a separate app window for all factors or only significant factors. The ‘Time unit’ drop-down component in the left panel controls the time unit for which data is summarised. The individual and aggregated

uncertainty of the selected time unit is then plotted along with the breakdown of distribution types in Figure F.4.

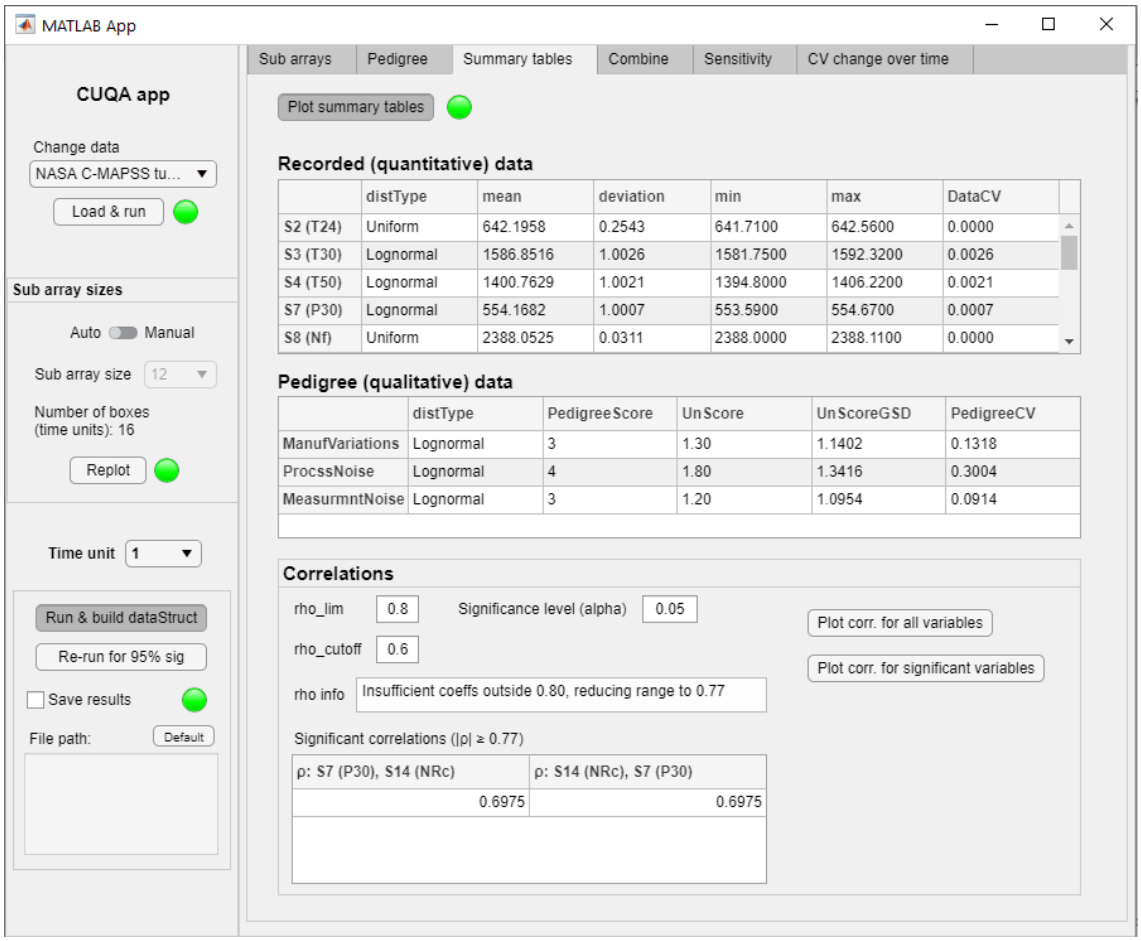


Figure F.3. MCDUQ app: CUQA summary tables tab

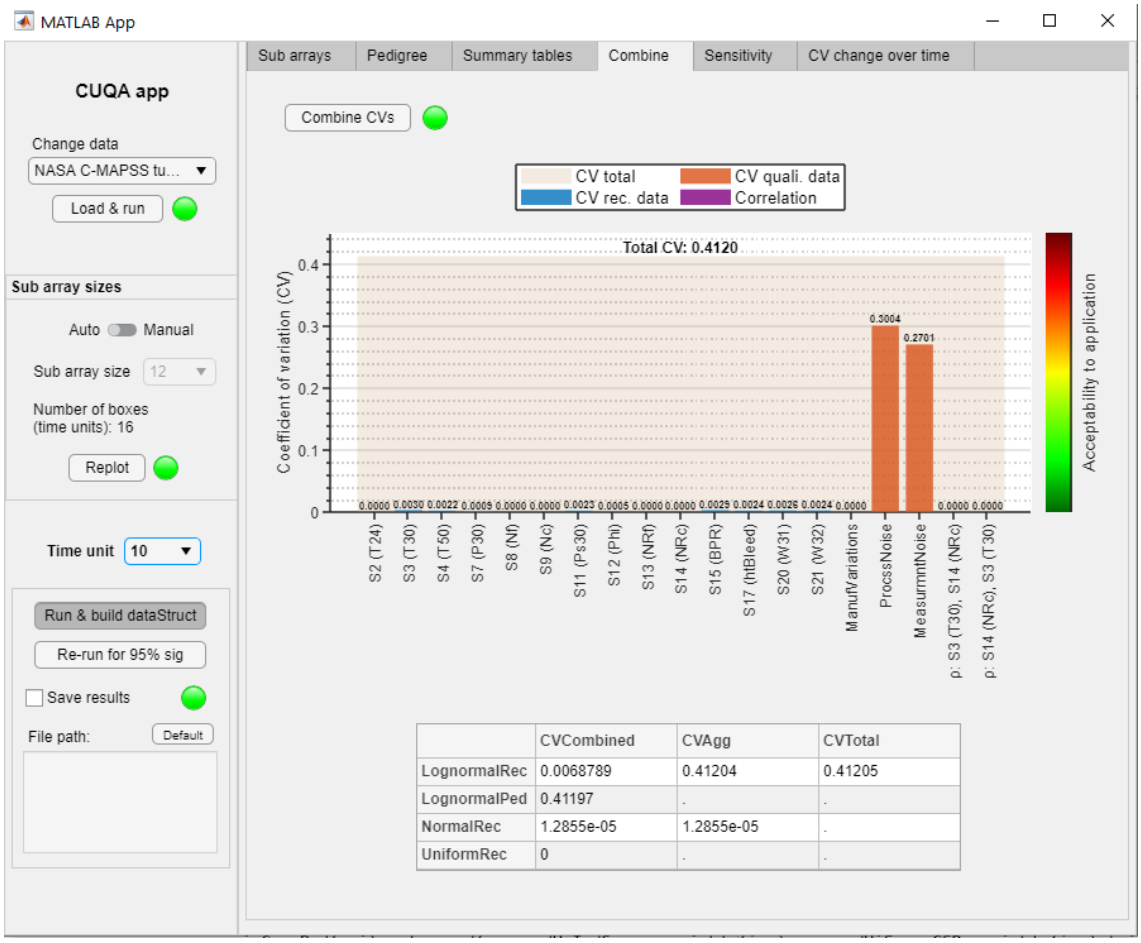


Figure F.4. MCDUQ app: CUQA combine CV tab

The 3D bar plots for the results of the global sensitivity analysis are plotted in the sensitivity tab (Figure F.5) [40]. This is given for all variables along with a table denoting the sensitivity indices. Factors within the 95% significance are highlighted red, while non-significant factors are grey. Using the panel in the bottom left, the CUQA assessment is looped for all time units and results saved in a structure array. The most significant factors (95%) are then reassessed, as in Chapter 4, and plotted in the right figure for comparison. These results are saved in a separate structure array, both of which can be optionally saved to a defined file path location using the 'Save results' check box.

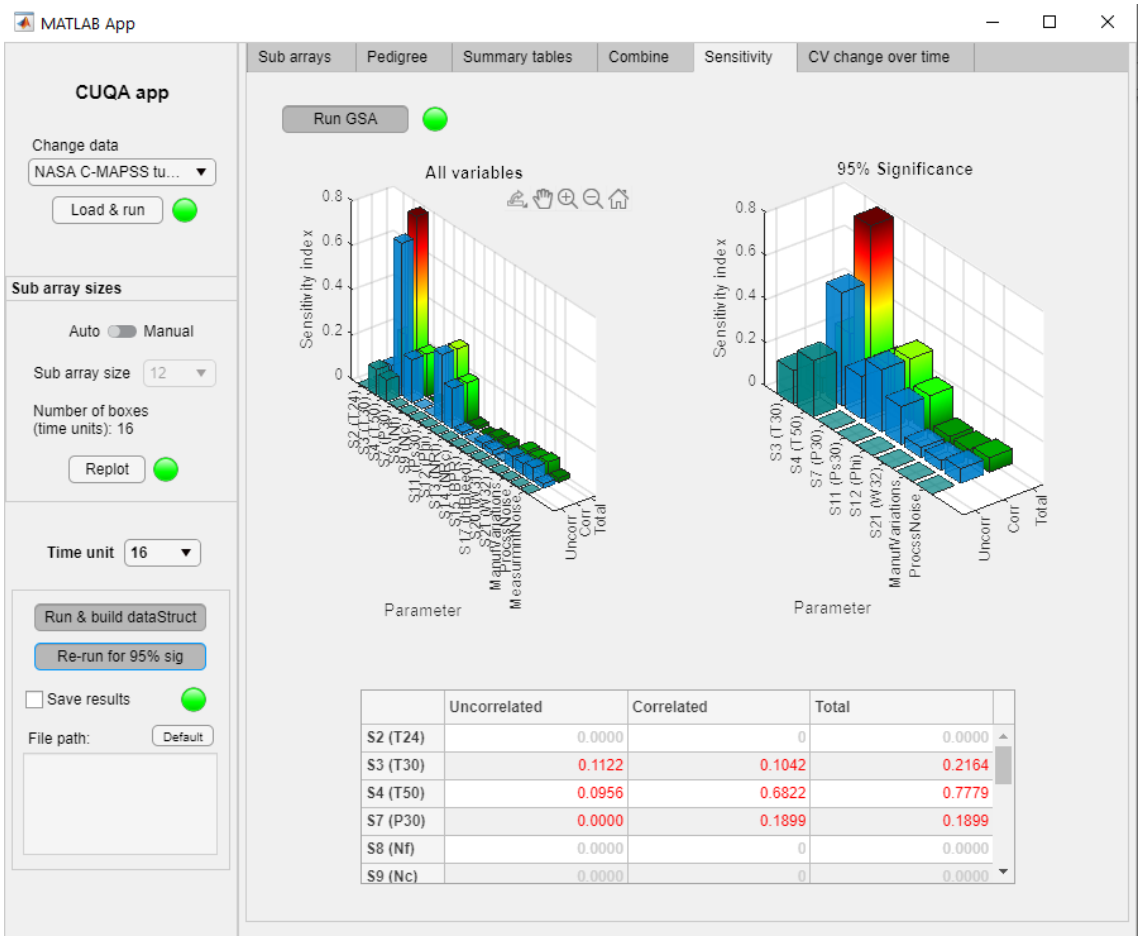


Figure F.5. MCDUQ app: CUQA sensitivity tab

Once the uncertainty is quantified, aggregated and the most influential inputs and dependencies are identified, the most significant uncertainties are forecast according to the framework denoted in Chapter 5. The UPLD application tool also features a fixed panel on the left side and a range of tabs displaying different plots and information on the right side. The fixed panel allows the user to:

- Select different datasets
- Adjust the partition split between training and test data
- Train the networks
- View how many input dimensions are allocated to each network
- Open separate apps to view the 3D visualisation and run hyperparameter tests for the network architecture

The tabs on the right side are described in Table F.1. To aid usability, red and green lamps were added next to each button to indicate when changes are being processed and plots updated.

Table F.1. MCDUQ app: UPLD tabs description

Tab name	Description
Input data	Plots for initial and updated forecasts with live-updating partition line
Input data summary	Displays summary statistics as bar charts for each input and has an option to use alternative deterministic measures for network allocation
LSTM architecture	Allows the user to adjust the initial structure and training options for each network
Initial and updated forecasts	Plots respective individual forecasts against observed data
Initial and updated error	Plots respective individual forecast errors as MAPE or RMSE
SG: Symmetry	• Summary table for shape areas, symmetry and cumulative variables for each time unit • Plot of symmetry over time • Plot actual and reference areas over time • Plot cumulative variables
SG: Shape comparison	Plots reference shape dimensions and compares actual and reference shape areas over time
SG: Correlation factor	Plots percent change of cumulative actual area and symmetry over time with comparison of correlation factors using linear, lognormal and exponential trend line equations
SG: Linear regression	Compares relationships between all input dimensions, displays summary table with linear trend line equations and R^2 correlation, plots top 5 strongest relationships

The ‘Input data’ tab is shown in Figure F.6. Three plots are displayed: the full-scaled dataset, initial forecast given by the trained networks and the updated multistep prediction. The partition slider is directly linked to the live partition line in the three plots to visualise the training and test split. The forecasts are automatically updated using the trained networks and given training data. Predicted data is plotted as dashed coloured lines (matching the colour of the training data) against the dashed grey observed (test) data. Prediction error is given by individual plots of predicted against observed data. These plots also update when the partition slider value is changed.

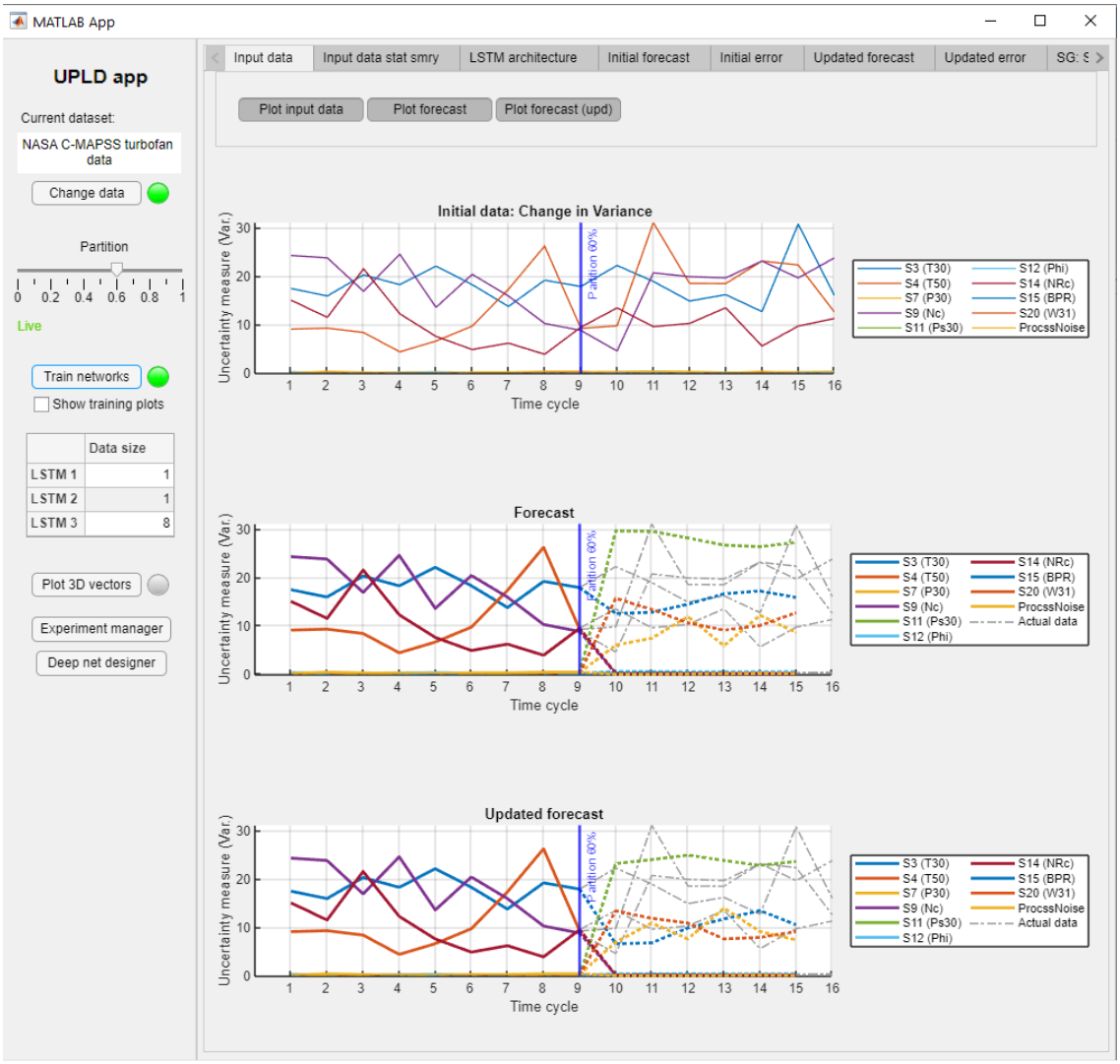


Figure F.6. MCDUQ app: UPLD input data tab

The ‘SG: symmetry’ tab is shown in Figure F.7. This is the main page for the results of the spatial geometry calculations. The table summarises the areas, symmetry and cumulative values for each time unit, which are then visualised in the three plots. Before the networks are trained, only the values for the initial dataset are displayed. Once trained, the ‘Use predicted data’ button is enabled. This recalculates the areas and symmetry using the predicted data in place of the observed data. When active, the same partition lines are layered over the plots to show the initial and predicted data split. This also updates then the partition slider value is changed.

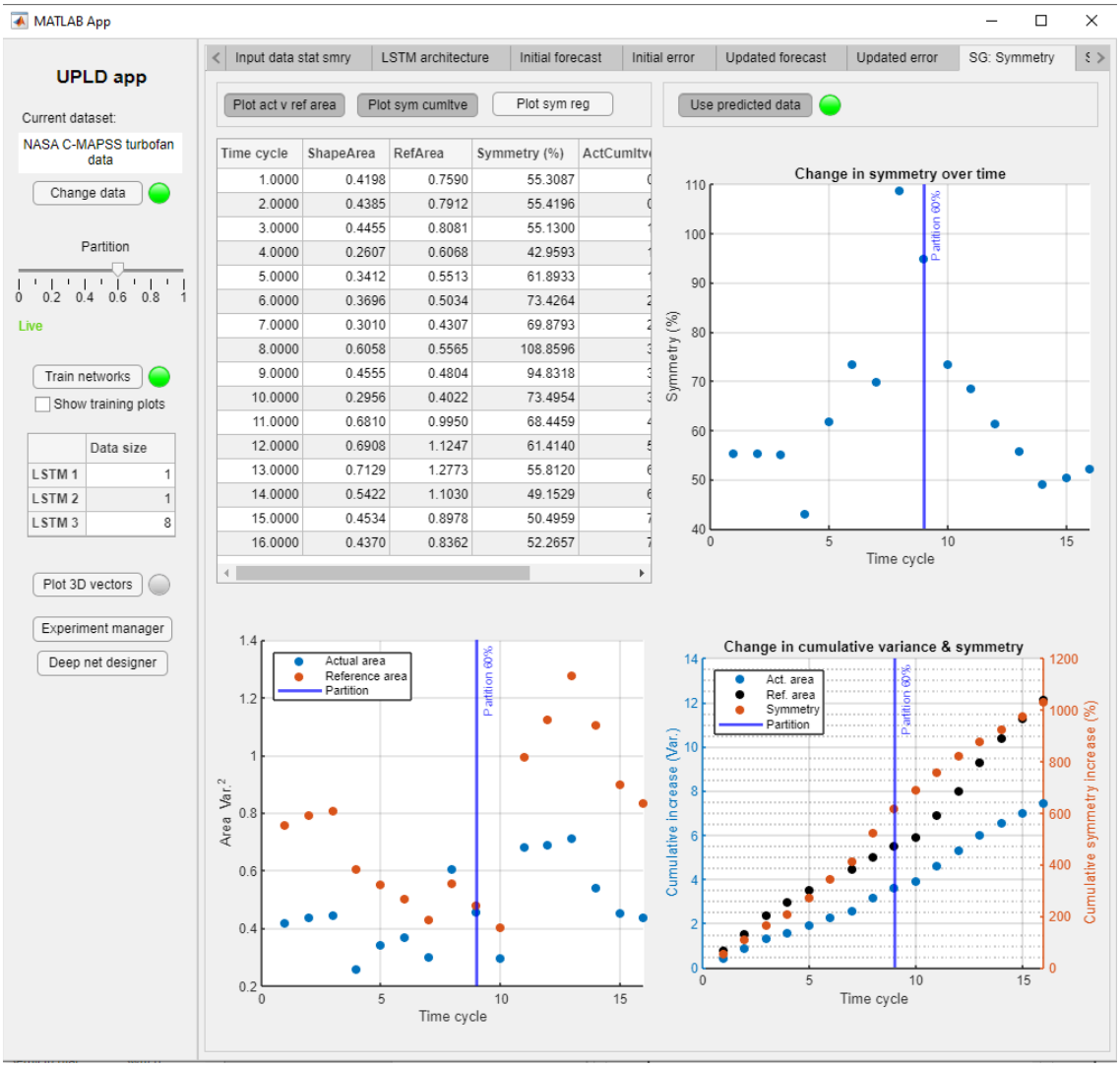


Figure F.7. MCDUQ app: UPLD spatial geometry symmetry tab

The ‘SG: Linear regression’ tab in Figure F.8 displays the relationships between all input dimensions. The table gives linear trend line equations and R^2 correlation, sorted to show the strongest correlation first. This indicates dependencies between variables. The top 5 strongest relationships are plotted below the table.

The ‘Plot 3D vectors’ button calls a separate app that displays the 3D plotting functionality, shown in Figure F.9. This is linked directly into the forecasting app so that when the partition slider value is changed, the subsequent training and test data split and updated prediction are reflected in the 3D plot. The app has two additional tabs for a 2D plot with each web overlayed on a single axis and the uncertainty range given by a 3D bar plot and line plot of the triangular distribution.

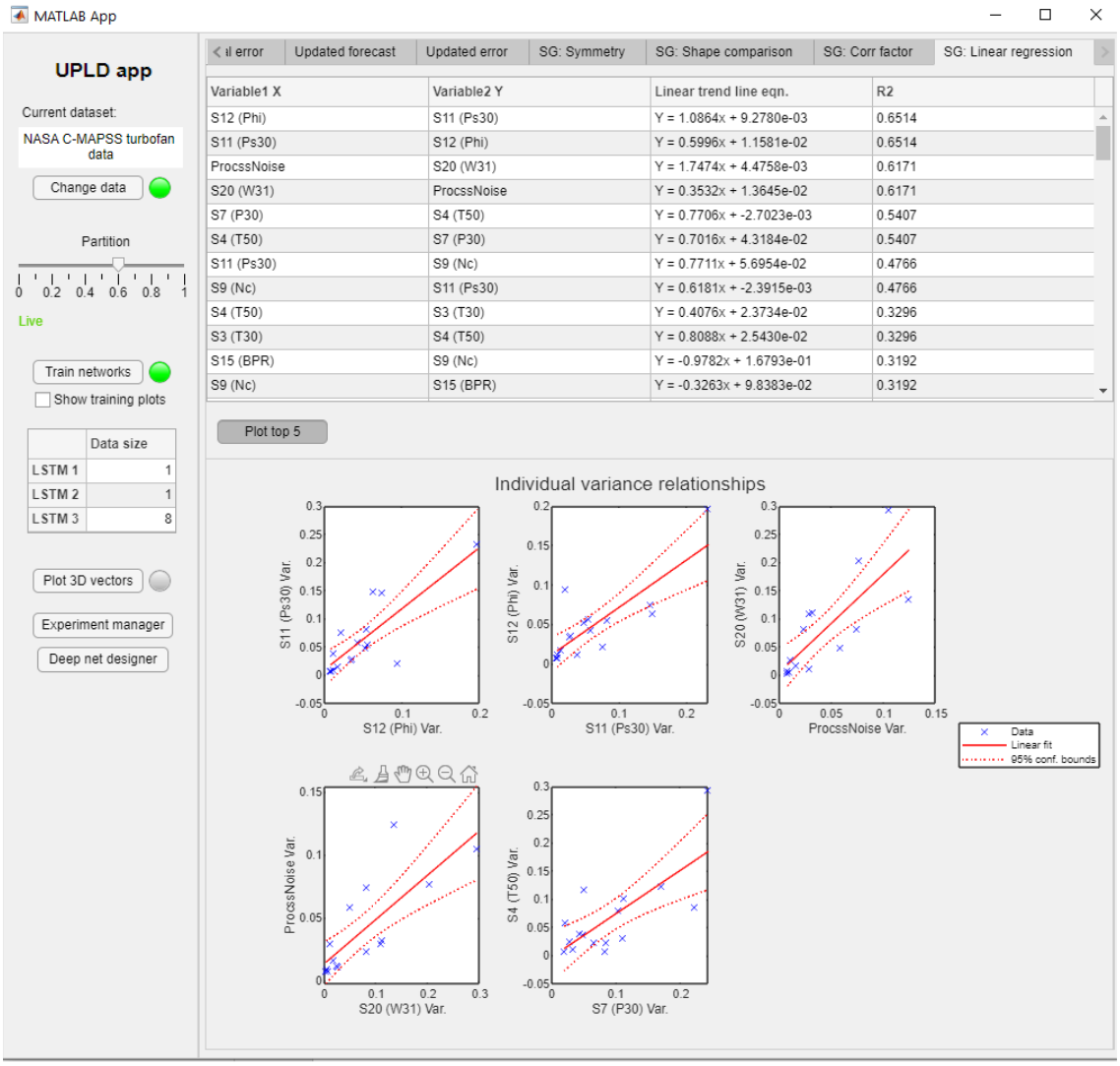


Figure F.8. MCDUQ app: UPLD spatial geometry regression tab

The ‘3D vectors’ tab has a series of checkboxes on the left side that allow the user to switch on and off each element within the plot. The web shape fill transparency can also be adjusted by a slider for the initial, observed and predicted data individually. The colours of the end vector coordinate points are given by default. The predicted (dashed) lines match the colour of the initial (solid) lines over time. the observed lines are shown in grey. For the visualisation in Figure F.9, only the predicted lines are shown. To boost visibility through the web fill and against the colours of the initial and predicted data lines, the colour and scale of the aggregate vector end markers can be adjusted by the user. Colours can be given as a 3-element RGB vector or by single letters as set by MATLAB. Linking lines between the observed and predicted data points and aggregate

vectors can also be shown to visualise the prediction error, as well as a text box annotation. The figure has the links between the aggregate vectors switched on. The plot legend automatically updates to allow identification of each selected element.

These features cover some of the issues highlighted in Chapter 5 for the useability of the 3D visualisation. Further developments may see these functionalities translated to a truly immersive AR visualisation. The user would then be able to expand on each data point and see additional information surrounding it, such as mitigation strategies and highly correlated variables.

The schematics in Figure F.10 and Figure F.11 illustrate the model flow and links between functions for the two phases of the MCDUQ app. The green diamond represents the linking point for the resulting uncertainty data from the CUQA phase to the input topology evaluation in the UPLD phase. Precalculated variance data can also be fed directly into the UPLD phase.

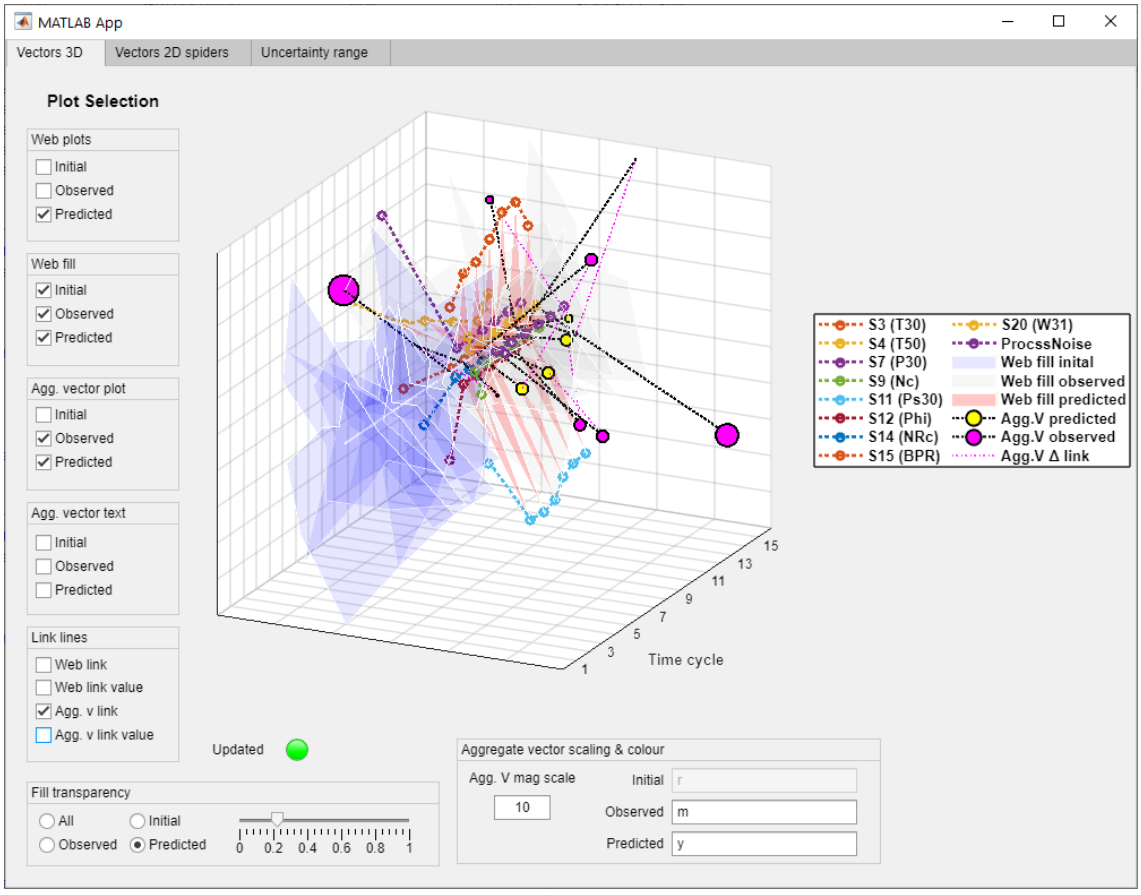


Figure F.9. MCDUQ app: UPLD 3D visualisation plot with optional perspectives

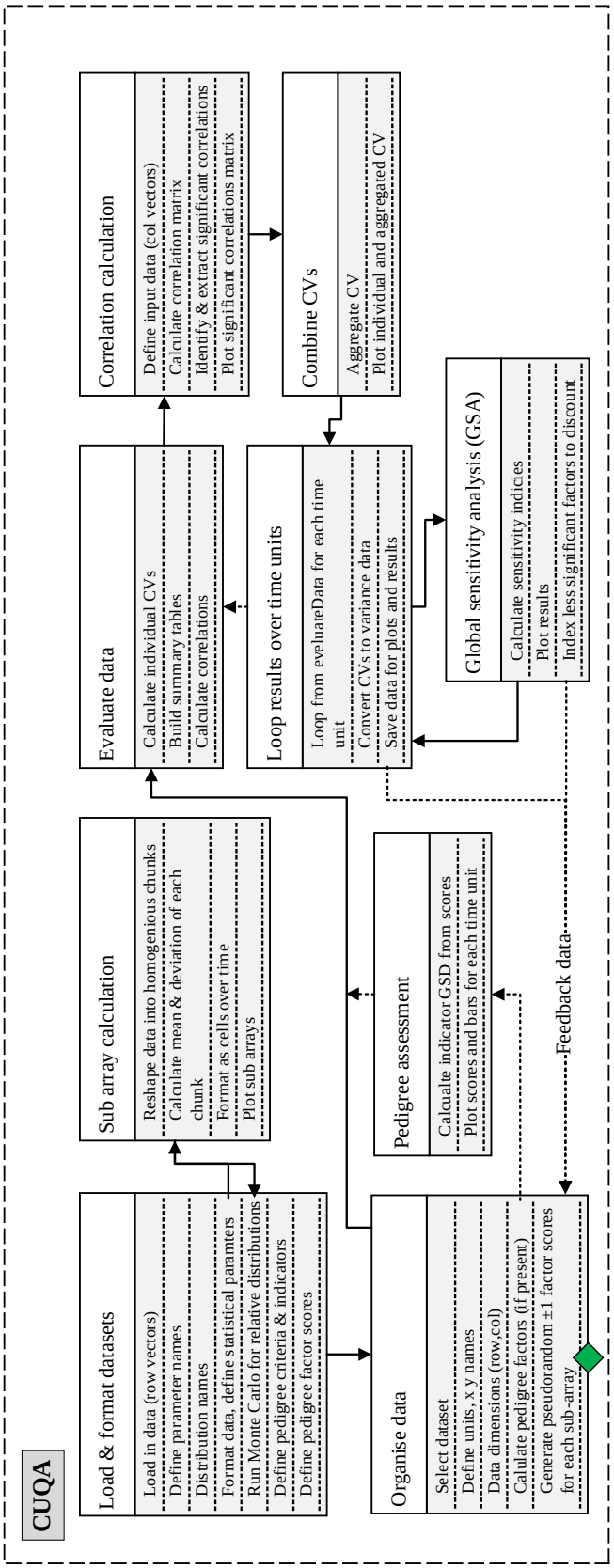


Figure F.10. MCDUQ app: Schematic for CUQA phase

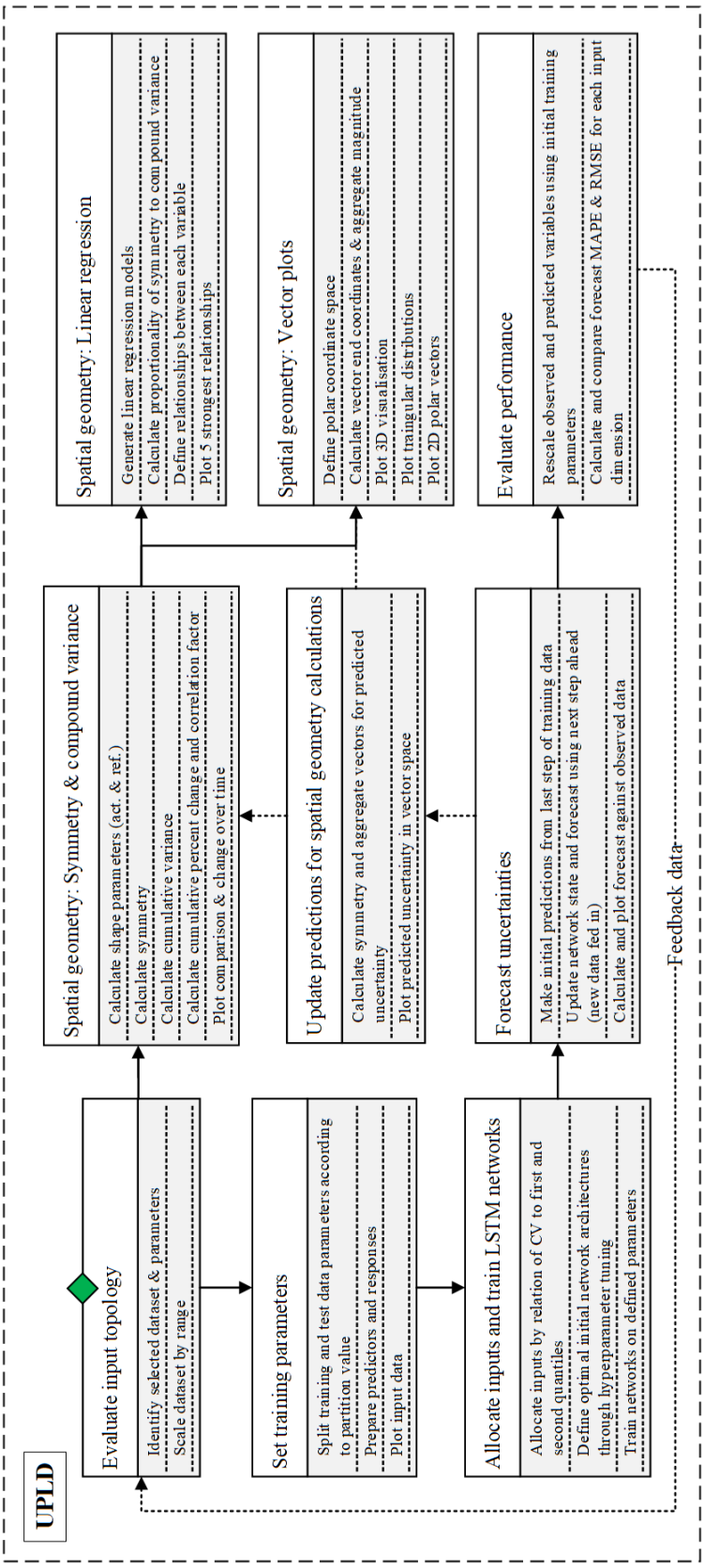


Figure F.11. MCDUQ app: Schematic for UPLD phase