UNIVERSITY OF LIVERPOOL
School of Engineering

Developing quantitative validation metrics to assess quality of computational mechanics models relative to reality

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy
by
Ksenija Dvurecenska

May 2019

Abstract

Developing quantitative validation metrics to assess quality of computational mechanics models relative to reality

Computational models are widely used to assess and predict behaviour of engineering systems. It is desirable that simulation correctly represents the real behaviour of the system's intended use and thus establishment of the level of quality for obtained results through validation is vital. The focus of the research presented in this thesis is the development and implementation of a novel generic validation metric for quantifying the quality of simulation results.

The choice of the validation metric is governed by the data available, the outcome required and the model's range of use. Three categories of metrics have been identified: Hypothesis testing, Frequentist and Bayesian. In general, Frequentist methods, in comparison to Hypothesis testing, allow a better understanding of the quality of the current model, through quantifying the differences between the predicted and measured results; whereas Bayesian analysis is typically used for the model parameter calibration. The work presented in this thesis concentrates on developing a Frequentist validation metric, and two novel metrics are proposed, i.e. based on a Theil's inequality coefficient and a new relative error metric.

Previously in solid mechanics validation has been applied to data points obtained from strain gauge measurements; in the present work the application is extended to data fields obtained with the aid of optical measurement techniques, e.g. displacement fields. By incorporating orthogonal decomposition the dimensionality of data fields can be reduced to coefficients in the feature vector, while preserving the key information about the deformation of the entire surface, and equivalent measured and predicted data sets can be obtained, which is essential for validation purposes.

Utilising the feature vectors, both of the novel metrics provide a measure of quality of the model's predictions relative to reality. The outcome of the Theil's inequality coefficient is a value between 0 and 1, i.e. from excellent to poor correlation of predictions with measured data. Whereas the novel relative error metric combines the use of a threshold based on the uncertainty in the measurement data with a normalised relative error, and the quality of predictions is expressed in terms of a probability statement. Such outcome obtained with the new relative error metric is more quantitative and informative than the previous validation procedures but qualitatively equivalent. Three previously published case studies were successfully employed to demonstrate the efficacy of the novel methodologies.

Ksenija Dvurečenska
University of Liverpool

Acknowledgements

Whilst my name appears alone on the front page of this thesis, there are a number of individuals who have contributed to this work by providing support inside and outside of the academic environment.

I would like to express an unprecedented gratitude to my primary supervisor Eann Patterson for the continuous guidance and for being very supportive throughout the ups and downs of my project. I would also like to thank my secondary supervisor Edoardo Patelli for sharing his expertise and my industrial supervisor Steve Graham for providing his industrial insight. All together and individually, their valuable knowledge and advice has helped to progress my ideas and produce this thesis; including the friendly reminders to keep writing it. Thank you all for your patience and encouragement.

I really appreciate the support offered by colleagues at the University of Liverpool and beyond, including regular tea breaks. Also, I would like to thank George Lampeas, Vasilis Pasialis, Xiaoshan Lin, Luis Felipe-Sese, Xiaohua Tan and Weizhong Wang for providing access to their data for the case studies to test my research ideas.

This research project was supported by a studentship funded by the UK Engineering and Physical Sciences Research Council and by the National Nuclear Laboratory, and I am grateful for their financial support.

And last but not least, infinite gratitude goes to my family and many friends. Wherever you might be, thank you for your understanding and encouragement. The kind support I have received is worth more than I can express here.

☺

Contents

List of Figures

List of Tables

1 Introduction

Computational solid mechanics models are widely used to evaluate and predict behaviour of engineering systems. Due to the increasing computational capabilities, it is now possible to simulate a large variety of processes. Results from simulations allow a better understanding of physical phenomena and thus are often used to inform decisions that could potentially have socio-economic consequences. To avoid detrimental impact, it is desirable that a simulation does not just correctly compute the underlying mathematics used to model the physics but represents the real situation of the system's intended use. The provision of credibility for predicted results becomes vital and can be achieved through a Verification and Validation, (V&V) process.

The overall aim of the process is to assess and establish confidence that a numerical model is sufficiently accurate and reliable with respect to a specified intended use. Often the V&V term is used without distinguishing between the constituent parts, or verification and validation terms are used interchangeably; however, there is a distinct difference between the terms and it is important to clarify it prior to proceeding further. The ASME [1] guide [1] gives a modern succinct definition of these terms in solid mechanics and these will be used throughout the thesis when referring to Verification and Validation:

[1] American Society of Mechanical Engineers

Verification - 'the process of determining that a computational model accurately represents the underlying mathematical model and solution'.
Validation - 'the process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model'.

From these definitions it can be seen that verification of the model should precede validation. Nowadays, commercial software packages are used to build solid mechanics models and to run simulations based on them; it is usually assumed that these software packages have undergone the verification process to some standard, e.g. the NAFEMS [2] has developed number of verification benchmarks [2, 3]. In addition to using verified software packages, modellers undertake verification exercise to ensure that the underlying mathematics is being solved correctly. Once a level of accuracy defined for a specific application is reached, validation methodologies are applied to establish whether the model described by the set of mathematical concepts represents the reality of intended use. Along the process of building confidence in the model's predictions, a model calibration activities can also be undertaken, the aim of which is to adjust model parameters to improve agreement between the simulation results with a specified benchmark [4]. Further research on topics of verification and calibration of models describing novel behaviour is necessary as part of a V&V process [1], however, it is outside the scope of the research presented in this thesis. In this research it will be assumed that verification procedures have been performed to a certain level of confidence and thus a validation methodology can be applied.

A number of guides and approaches for validation in the area of mechanics are currently available (see e.g. [1, 2, 5, 6]). A similar template for the main

[2]National Agency for Finite Element Methods and Standards

processes is apparent throughout these guides; nonetheless, there is no single validation methodology that is widely accepted. Usually in solid mechanics, model validation is performed through a comparison between experimental and computational data sets consisting of point data, for example maxima or minima measured by a series of strain gauges and corresponding values extracted from a computational model; the experimental data is used as a referent, i.e. physical data representing the real world, and the outcome of the comparison leads to a conclusion on whether the model is valid for the intended application or not. This evaluation and a boolean valid or not valid outcome may be sufficient for some applications, but ultimately it is desired to utilise the entire field of data, e.g. displacement field on the entire surface of an aircraft wing, and to quantify the degree of quality of the model's predictions with respect to the validation criteria. In recent years, experimental data acquisition methods have advanced significantly and can provide an easy access to fields of data by means of optical measurement techniques such as Digital Image Correlation. Thus providing the desired quantity of data and presenting opportunities to develop novel validation methodologies. Another important aspect to consider is how to interpret obtained information to better support a decision-making process on the appropriateness of the model for the intended use. In these circumstances, a novel approach is required and research presented in this thesis addresses this issue by investigating statistical techniques of data comparison for validation purposes, utilisation of field data with these techniques and tools to represent validation outcomes.

Aim and objectives

Taking into account the significance of validation outcome on decisions that have socio-economic consequences and current lack of a robust methodology, the **aim** of the research presented in this thesis is to develop a reliable and transferable validation metric that can be used in different engineering applications, where modern sensor systems allow the acquisition of fields of measurement data.This is approached through working towards the following **objectives**:

a) To extend the application of validation metrics to field data;

b) To explore approaches to obtain a quantitative measure of quality of simulation outputs;

c) To consider ways of communicating validation outcomes to technical decision makers.

Thesis structure

The content of this thesis will be presented in *seven* main chapters. Following the introduction, in Chapter 2 the existing literature on validation and associated activities will be reviewed, including the topic of validation metrics. In Chapter 3 novel methodologies implemented and developed to achieve quantitative validation will be detailed, together with the description of case studies employed to demonstrate and compare the efficacy of the validation approaches. Subsequently, results based on the developed metrics will be presented in Chapter 4 and discussed in Chapter 5, including the overall implementation of the validation process. Chapter 6 will contain conclusions and summary of findings from the research, followed by the last part of the thesis, Chapter 7, containing an overview of potential further directions for this topic.

List of presentations and published work

The key findings described and discussed in this thesis have been presented at a number of international conferences. In addition, a manuscript based on the novel relative error metric and the three case studies (introduced in Chapter 3) has been published in a peer reviewed journal.

- Dvurecenska K, Graham SJ, Patelli E & Patterson EA, A probabilistic metric for the validation of computational models, *Royal Society Open Science*, vol. 5, no. 11, Nov 2018.

- Dvurecenska K, Patelli E & Patterson EA, What's the probability that a simulation agrees with your experiment?, Photomechanics 2018, March 19-22, 2018 [*extended abstract*]

- Dvurecenska K, Graham SJ, Patelli E & Patterson EA, Application of a frequentist metric for the validation of computational mechanics models, 12th Int. Conf. on Advances in Exptl. Mech., September 1-3, 2017 [*extended abstract*]

- Dvurecenska K, Patterson EA, Patelli E & Graham SJ, Preliminary evaluation of validation metrics for computational mechanics models, Proc. 10th Int. Conf. on Advances in Exptl. Mech., September 1-3, 2015 [*extended abstract*]

- Dvurecenska K, Patterson EA, Patelli E & Graham SJ, Preliminary evaluation of validation metrics for computational mechanics models, Universities Nuclear Technology Forum, March 31 - April 2, 2015 [*presentation*]

2 Literature review

This chapter contains an overview of previous work and important concepts that underpin the novel developments presented in this thesis.

2.1 Validation overview

Validation is part of the Verification & Validation (V&V) process, the overall aim of which is to assess and establish confidence that a model, i.e. a set of numerical equations and assumptions which capture the behaviour of a structure, behaves in accordance with the underlying equations and it produces realistic results with respect to a specified intended use. As mentioned in Chapter 1, the focus of this research has been on the validation process, its implementation and interpretation.

Initial discussions on validation can be found in the literature of the second half of 20th century, when validation became a significant topic amongst the simulation community. Papers concentrated on analysing the terminology, methods of performing the validation and the distinction between verification and validation. Fishman and Kiviat [7] and Van Horn [8], were amongst the first to discuss the idea of validation, and related questions. Their papers were based on economics science, but the ideas are relevant to simulations in different areas of science. The clarity in terminology to avoid the misuse, including the distinction between verification and validation, was widely discussed. Also, it was identified that a computational model is usually developed for particular objectives that reflect the intended use; consequently, the simulation results have to be evaluated against these objectives. Even though, Fishman and Kiviat [7] did not use the exact definition that eventually appeared in the ASME guide [1], i.e. their statements were concerned with whether the model reasonably approximates the real system, they did discuss that it is important that simulation results are compared to the particular representative parameter. Similarly, Van Horn [8] stressed that validation should be performed for a defined case. However, it was Sargent [9] who first included the intended use in the definition; if a model is valid for one set of objectives, it may not be valid for another set of objectives. It would be very costly, time consuming and often unnecessary to construct a model and run a simulation to obtain results for all possible conditions outside the intended use. Following the initial publications, papers summarising the general validation techniques started to appear around 1980 [10, 11, 12]. Throughout the past and present literature, the terminology with respect to validation has not been consistent such that some reports imply that a model is being validated and some that the predictions, however in both situations a model is being evaluated over a specific domain defined at the beginning of the validation process. In general, most of the ideas are similar and it is important to note the general principle is based on validation being performed by comparing model behaviour with the real system behaviour when both simulation and observation are conducted under identical conditions.

In general, techniques for comparison can be objective and subjective, with quantitative and qualitative aspects. For some models it might be sufficient to use historical data from experiments performed previously and included in the validation databases [13]. In these circumstances, it is assumed that experiments have been well-documented and performed according to certain standards or guidelines, which are widely accepted in the research community. However, there is a significant number of historical data sets for which some of the associated assumptions and uncertainties are unknown and are difficult to quantify. When the experimental results are used as the referent against which the computational results are compared, insufficient information on the accuracy of the experimental results does not allow to adequately build the confidence in the computational model [14]. This promotes the need for new experiments in order to obtain the necessary information.

In some cases, validation of simulation results can also be performed through comparison with previously validated models [15]. For instance, when a new model is designed, it is possible to validate it with another model that has been previously validated for the same conditions. In these circumstances, the epistemic value of models, i.e. the scope of knowledge, should be taken into account [16], such that only models of evolutionary designs with incremental differences in comparison to previous model can be evaluated [6]. The appropriateness and validity of the previous model is subject to interpretation but confidence can potentially be built by providing sufficient evidence. Nowadays many models are built for novel and complex designs, thus this technique might not often be applicable. Nevertheless, it is potentially useful for the validation of constituent components of a complex system [1], i.e. the lower levels of a hierarchy of models.

The techniques mentioned in the previous paragraphs can be categorised as a combination of both, objective and subjective, elements of comparison. In some circumstances it is not possible to obtain sufficient experimental data and to perform direct comparison with predictions, thus purely subjective evaluation is performed. Subjective evaluation can be achieved with the aid of expert opinion known as face validation, or information interpretation based on the application of the model, which is known as operational validation [15]. Operational validation evaluates the level of accuracy of the results in relation to the intended use. For example, the same model would require a different level of accuracy depending on the circumstances of the intended use and allocated costs. For more challenging applications, for example such as space exploration mission or nuclear power plant lifespan prediction, a more accurate model is required and higher costs will be involved. According to decision theory [17], the higher the importance or the risk of the decision, the higher level of evidence is required for validation. Face validation can be helpful at the initial stage of the validation, where an expert opinion can suggest whether the model and its behaviour are reasonable, based on the previous experience with similar systems. If the model is found to be unreasonable, it can be corrected and then subjected to a full validation. This could potentially save time in the overall validation process, as errors or inconsistencies in the model would be spotted early in the process. Nevertheless, an opinion can vary from person to person and cannot currently be effectively quantified in the scope of validation [18, 19]; however it can be related to the epistemic value of the model and interpreted from a philosophical perspective [20, 21, 22, 23].

2.1.1 Validation process

The ideas presented in the articles of the late 20th century have stimulated the development of guides on V&V in different areas of engineering. One of the first consensus guides was compiled by an accredited standards developer in the

USA AIAA[3], in 1998 for computational fluid dynamics simulations [5]. This has served as a foundation for the guide in computational solid mechanics [1] that was published in 2006 by the ASME. These guides provide concise definitions and a generalised methodology for performing the V&V, but neither include definitive step-by-step procedures. Other guidelines can be found too, which have attempted to include more detailed methodologies. For example during a series of European collaborative research projects publications on validation in the solid mechanics area, such as reference [6] or the CEN workshop agreement [24], were produced. In the solid mechanics, it is common to perform validation using measurements from single points, for example evaluating the maximum and minimum values of a response measured by strain gauges. However, in this recent work it was proposed to extend the application of validation process to field data, such as displacement fields on the entire surface of a specimen [24], where the experimental data is obtained by modern non-contact full-field optical measurement techniques [25], e.g. three-dimensional Digital Image Correlation. Some of the advantages of such techniques are that the data from the entire surface of the specimen can be acquired for in-plane and out-of-plane deformation, and they can be used at number of time-steps to evaluate the deformation in the spatio-temporal domain [26]. These techniques can be applied to components of different size, for example a full-scale car bonnet liner [26] or a small, i.e. 60x60x25mm, sample of a rubber block [27]. Further details of the validation methodology outlined in the CEN guideline [24] will be provided in the next section, Section 2.2.1. Nevertheless, there is currently no standardised, widely accepted and used validation methodology.

Typically, validation process is presented in a flowchart consisting of steps leading to the comparison between computational and physical outputs. The

[3] American Institute of Aeronautics and Astronautics

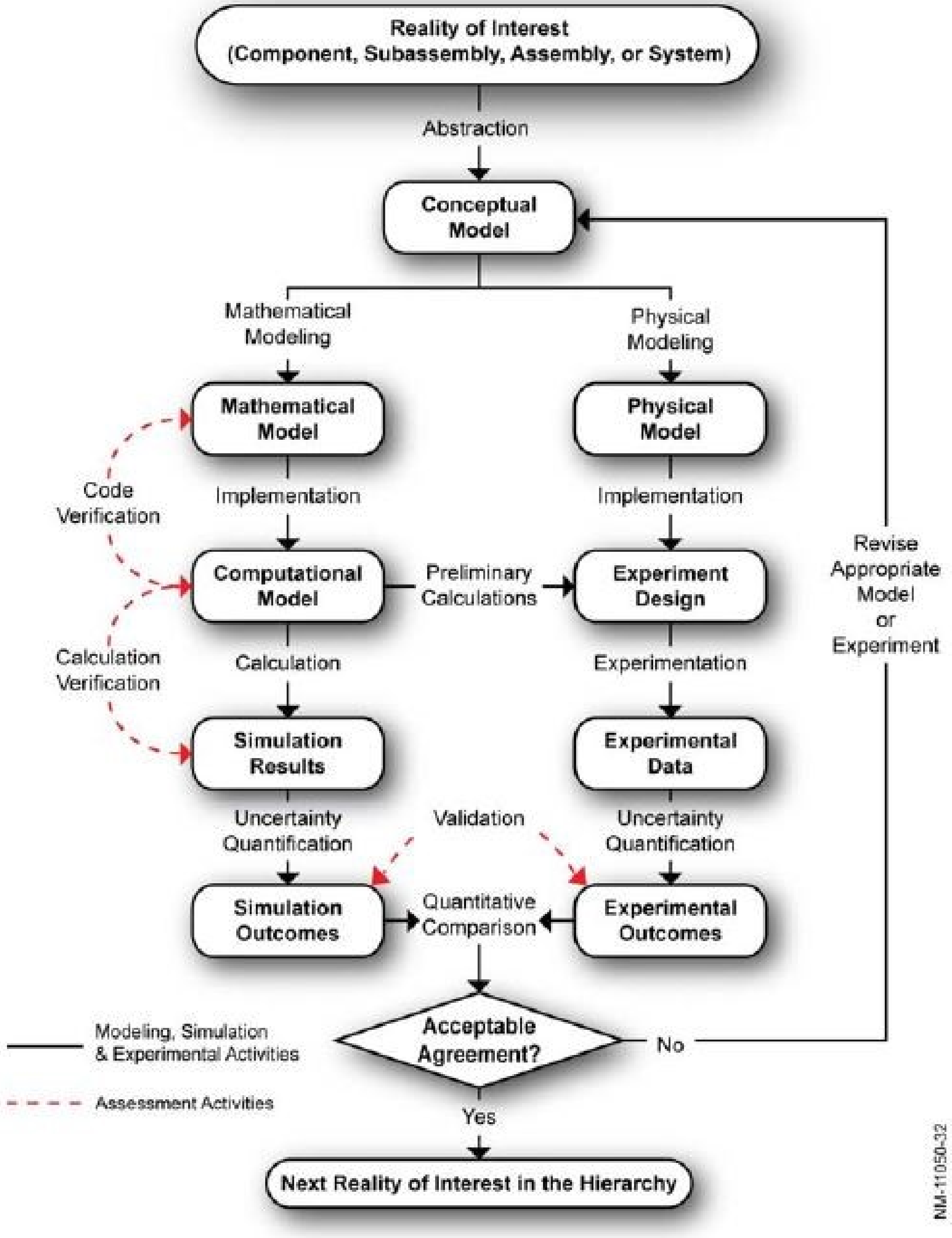


Figure 1: ASME Validation and Verification flowchart [1].

schematic in Figure 1 summarises a typical process that consists of two streams, i.e. computational and experimental data, incorporates data processing and comparison phases, and concludes with a decision box. The two streams are typically performed in parallel but with minimal interaction between the computational and experimental teams, and include activities such as identifying necessary parameters and boundary conditions to obtain the necessary data for the following steps in the validation process. Once the activities are completed and results assessed individually, outcomes are compared with the purpose to provide sufficient information for the subsequent decision on the model's validity for it's intended use. As a result of a comparison, if an acceptable agreement has been reached, then the model is stated to be valid. If the model has not been validated, usually the computational model or the experiment, or both, are reviewed [14]. The choice of the next action can, for example, depend on the available resources and the consequences of failure to validate the model. The process is repeated until the acceptable agreement is reached [24]. However, after several refinements some aspects of the computational model or the experiment will be changed, so they should be re-evaluated with respect to the initial conditions for the intended use, to make sure they are still appropriate for validation.

This approach can be applied to a system model or its components. However, it is currently arguable whether the validity of components can be used to draw a conclusion about the system model or whether new experimental data has to be obtained. In the ASME guide [1] it is suggested that the complexity of connections or behaviours between components is not always included in the individual component validation and thus a system model can only be validated when more information is acquired. Hierarchical validation of the system model is mentioned in some of the guidelines published in different disciplines; as yet no explicit methodology has been presented. The AIAA guide [5] suggests that a system model can be approximately divided into three levels: unit problems, benchmark cases and subsystem cases from simplest to complex and individual validation requirements are established at each level. A similar concept is mentioned in the ASME guide [1] and based on that an example of an aircraft wing was described in the subsequent ASME V&V 10.1-2012 guide [14], from which the example hierarchy is shown in the Figure 2.

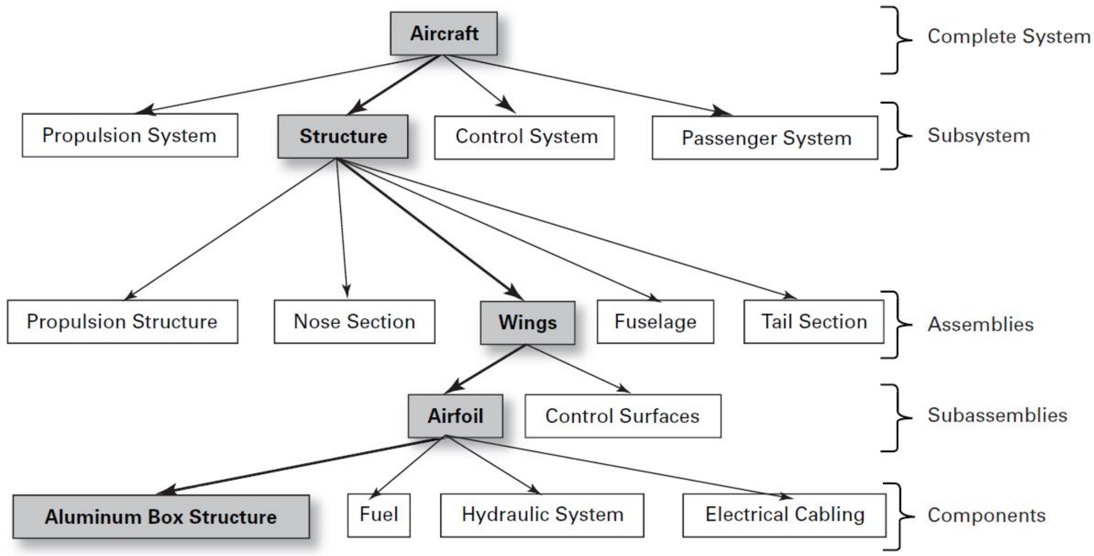


Figure 2: Hierarchy for the aircraft model validation [14].

One of the stages in the validation process is quantification of uncertainty, as can be seen in the flowchart presented in Figure 1. Uncertainty quantification activities are incorporated as part of the process in both streams, e.g. physical and computational, and are necessary to assess reliability and consequently build confidence in the obtained results.

There are two widely accepted categories of uncertainty: aleatory and epistemic [28]. It is a good practice to distinguish sources of uncertainties and assess each category individually. Aleatory, or irreducible and stochastic, uncertainty is associated with the variability and randomness of unknowns. Usually, it is not possible to reduce this uncertainty or inherent variation; however it can be described using traditional probabilistic approaches [29]. For example, a variability in material properties of a carbon fibre reinforced engineering components can be categorised as an irreducible uncertainty for a given manufacturing process, i.e. it can only be reduced by changing the manufacturing process. By taking measurements of individual components, information about the variability of the properties can be collated and represented by a probability density function, or, in the studies with limited data, by intervals. On the other hand, epistemic uncertainty is related to the parameters of the system that are not well-known but additional information can be obtained through measurements to reduce this uncertainty. An ill-defined conditions during a physical or virtual test, for example not taking into account the effect of damping on the component's response during a dynamic test, can contribute towards the epistemic uncertainty, but this uncertainty can be reduced by gaining more knowledge, through a detailed analysis of the system and component's interaction by relevant experts. In the above example it would be recognising the presence of damping and analysing it's effect on the component's response. Both categories described above are relevant to computational and experimental results.

To raise the confidence in results, uncertainty should be evaluated and quantified where possible [30]. In the simulation results, irreducible uncertainties can arise from, for example, ill-defined material properties or component interaction. Even though these cannot be reduced, they can be better quantified by more thorough analysis of the model [31]. Also, the initial uncertainty can be propagated through the simulation to evaluate the uncertainty in the final outcome using probabilistic analysis. This can be accompanied by a sensitivity analysis based on evaluating individual sets of inputs and their effect on the output uncertainty. The reducible uncertainty can be divided into statistical and model types. The statistical uncertainty can arise from the limited availability and poor comparison of the data, thus can be reduced by obtaining more samples or improving statistical method applied. The model type uncertainty is usually hard to quantify because it is related to the modelling assumptions associated with the parameters describing the system [32].

Experimental uncertainties can arise from a number of sources [33]. These can be associated with the accuracy of the experimental equipment, assumptions that the experimental procedure is based on and the environment. The key principle is that nothing can be measured precisely; some uncertainty will always be present but it can be reduced or better quantified through, for example, appropriate calibration of the equipment. For an adequate validation, as many sources of uncertainty as possible should ideally be identified for the experiment and simulation; then the uncertainties should be quantified to allow credible conclusions about the predictive capabilities of the computational model to be drawn [1].

2.2 Validation metrics

In the guidelines described earlier, validation was referred to as a single process. At a more detailed scale, it can be divided into two activities [1]:

- quantitative comparison between computational and experimental results
- assessment of the comparison outcome with respect to the accuracy requirements for the intended use of the model

This division into two activities is not always explicitly stated but both are performed when quantitatively validating a model [34]. Figure 3 illustrates interpretation of previously presented ASME flowchart with the two activities explicitly highlighted by Oberkampf and Barone [34]. It can be seen that first the difference is computed with the aid of statistical comparison between simulated and measured outcomes, through application of the validation metric, and then the outcome is evaluated in the context of the adequacy requirements.

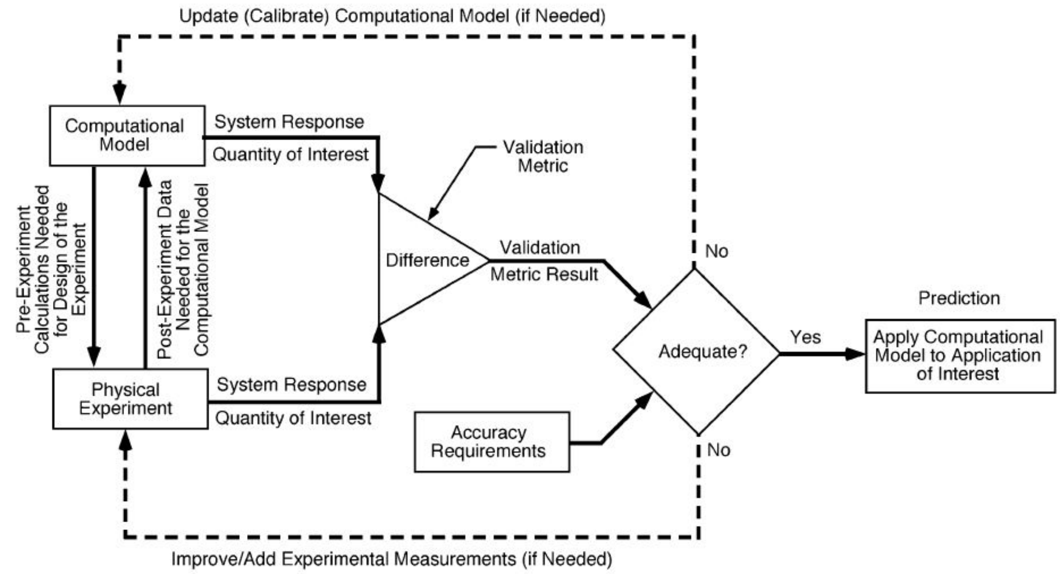


Figure 3: A schematic of the validation procedure that provides evidence to decision makers; a difference between simulated and measured outcomes is computed with the aid a validation metric, and the results of the comparison are evaluated against the adequacy requirements [34].

It is important to note that even though a qualitative validation is utilised in some research, in this research the emphasis is on quantitative validation. In mathematics and science a metric can be referred to as a function to determine a distance between any two points or elements within the given space or set; it can also refer to a value equivalent to the distance [32]. The first definition will be used in this thesis when referring to a validation metric and the value will be defined as an outcome or result of the metric. A validation metric is applied to measure the agreement between the computational results and the referent, in this case experimental results. The desired outcome of the validation metric is a quantity that can be easily interpreted to determine whether the model predictions are adequate for the intended use. To aid the interpretation during the decision-making process, the degree of confidence in the results representing the discrepancy or similarity between the data should be presented.

The validity of a model can be established through an adequacy assessment of the results obtained via application of a validation metric [34]. For example, for field data in solid mechanics the CEN guideline [24] suggests comparing experimental data against simulation results by plotting corresponding data pairs and evaluating the outcome against the minimum measurement uncertainty, as graphically illustrated in Figure 4; details of the methodology will be provided in the following subsection, Section 2.2.1. In general, the model can be stated to be valid if the simulation results fall within the accuracy limits, set beforehand, with respect to the physical data. Typically a single answer is produced, e.g. valid or invalid, however, this gives no indication of how good the model is, if it is valid, or how bad, if it is not. In some instances, particularly when the model has been found to be invalid, without this information, decision-makers could not perform an efficient trade-off for the next set of actions, apart from the general decision to refine the model [1]. The knowledge gap can potentially be closed by integrating a quantification of the agreement between data sets and the accuracy requirements into the second activity of validation. This will be discussed further in the next chapter, Chapter 3, where details of the novel methodologies addressing these aspects will be presented.

Throughout the literature, a number of suggestions can be found for the desired features of a validation metric [1, 34, 35, 36]. One of the main requirements is that a metric should be quantitative and objective, meaning that different researchers can obtain the same result for the same sets of data and requirements. Additionally, it has been stressed that uncertainties contained in the experimental data, i.e. measured or post-processed, and the computational results should ideally be considered. Another point worth noting is that, even though it would be beneficial to have a single validation metric that could be applied to various scenarios, most of the current validation metrics can only be applied with the assumption that sufficient experimental data is accessible for validation. Thus, it is also important to assess the effects of the extent of the experimental data available on the validation result [11, 37].

Validation metrics can be classified according to different criteria. Some can be differentiated by the assumptions made, for example, that the simulation output or input is deterministic or requires multivariate analysis, others by the type of the validation outcome. From a philosophical point of view, most of the approaches can be divided into two categories based on two opposite statistical philosophies, i.e. Frequentist and Bayesian [38]. For example, when assigning and evaluating a probability of a parameter or an event, Frequentist approaches are based on the frequency of the occurrence deduced only from past physical observations, whereas Bayesian approaches assign a degree of belief, e.g. credence, that can also be collated from a variety of sources. In this thesis in addition a third category of validation metrics has been identified as Hypothesis testing. Although there exist both Frequentist and Bayesian hypothesis testing, the information contained in the outcome of these statistical techniques is interpreted differently, and thus three categories of validation metrics, i.e. Hypothesis testing, Frequentist and Bayesian approaches, are discussed in the following subsections.

2.2.1 Hypothesis testing

In hypothesis testing the objective is to accept or reject a hypothesis or a set of hypotheses. For instance, during hypothesis testing two statistical hypotheses, the null hypothesis and the alternative hypothesis, are commonly stated prior to the statistical analysis, e.g. the predicted data belongs to an experimental distribution [39]. Such an approach usually can only reject the null hypothesis in favour of the alternative but not confirm it. For example, if the hypothesis that the predictions belong to the measured distribution has not been rejected, the predictions can be considered acceptable but one should assess factors that could influence the outcome before making the decision based on this model. Sparse experimental data and large uncertainty present in both sets of data, predicted and measured, could potentially lead to rejecting an acceptable model, i.e. a type I error, and accepting an inadequate model, i.e. a type II error, due to insufficient evidence to conclude otherwise [36]. It is possible to include the consideration of both types of errors in the analysis through computing the probability of making these errors based on significance level and the probability distribution of the test statistics associated with the alternative hypothesis [40]. In relation to the qualities of the validation metric discussed earlier, the hypothesis testing approach provides a quantitative comparison, however currently the outcome of the comparison does not provide a clear quantitative indication of a model's validity in terms of the degree to which a model is an accurate representation of the real world, as per the definition of the validation by the ASME [1].

One of the more recent examples of a metric in this category can be found in the CEN workshop agreement [24] for solid mechanics simulations. In the CEN guideline, it is suggested to compare model predictions with experimental measurements, obtained with the aid of a no-contact full-field optical measurement

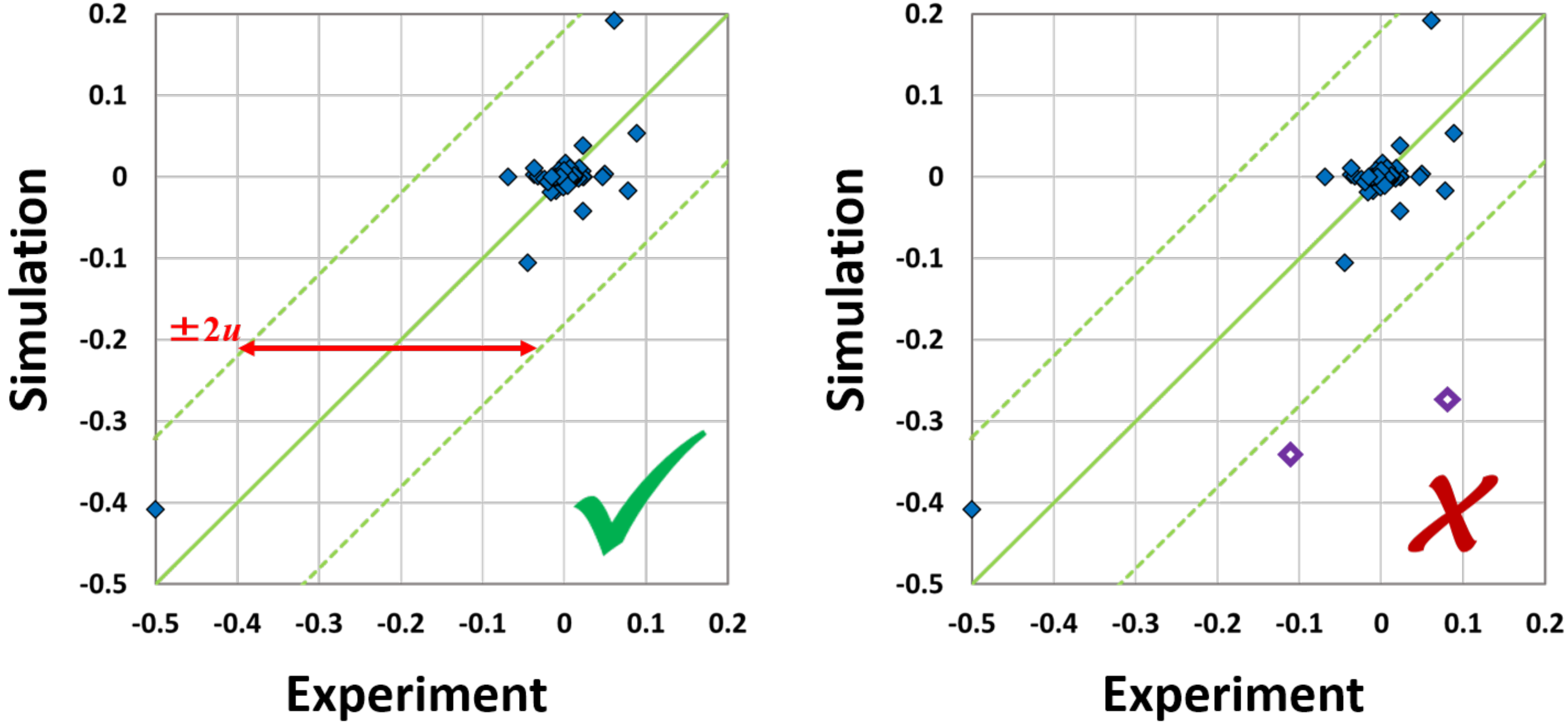


Figure 4: A graphical comparison of components of the feature vectors representing measured and predicted data fields by following approach recommended by the CEN guideline [24]. The predictions can be considered valid when all of the data falls within the zone bounded by the broken lines that are defined by equation (2.1); the central solid line represent a perfect match between the components of the feature vectors. Graph on the left demonstrates example of valid predictions and graph on the right demonstrates example of invalid predictions.

system, using an image decomposition technique. Predicted and measured data fields are treated as images and are decomposed using a set of polynomials to describe the essential features of the image; image decomposition and associated details will be covered in the next chapter, in Section 3.1. As a result, two equivalent feature vectors representing each data set, i.e. measured and predicted, are obtained, where the feature vectors contain the coefficients of the polynomials used to decompose the corresponding image. The metric proposed in the CEN guideline is based on a quantitative comparison of the feature vectors representing measured and predicted data sets, and includes a validation criterion based on the experimental uncertainty, u_{exp}, i.e. the uncertainty associated with the measured data. As illustrated in Figure 4, components of the feature vector obtained from the measured data set, S_M, are plotted against components of the feature vector from the predicted data set, S_P, and, if all of the points on the graph are within the uncertainty limits, i.e. zone bounded by the broken lines defined by

$$S_M = S_P \pm 2 \times u_{exp}, \tag{2.1}$$

a model can be considered valid. Such a statement of the validity is very common, but it only gives a yes/no answer, which might be unsatisfactory for certain applications and does not allow for interpretation of the models quality with respect to the validation criteria.

2.2.2 Frequentist approach

Methods based on a Frequentist approach have been implemented in different applications [38] and are based on the quantification of the difference between the two data sets, or, as defined in some literature, on a measure of error [34, 41]. The approach incorporates probability in some cases and can be summarised as mapping a discrepancy in the computational response relative to the referent, for example the mean or the distribution of the experimental response. Experimental results are assumed to be a golden standard and are used to compute the relative error of computational results. In reality, experimental data cannot be taken as true; uncertainties and errors are associated with the measurements and thus should be accounted for when evaluating the discrepancy between the data sets [35].

Most of the techniques falling into this category have been previously developed for analysis of time histories in structural dynamics. For validation purposes, Oberkampf and Barone [34] have used the approach to propose a technique that includes confidence intervals, i.e. an interval computed with the aid of the observed data that will contain a certain parameter value a specified proportion of the time, for example a mean of a population 95% of the time; in their work the confidence interval is based on the experimental uncertainty. The calculation of the metric proposed by Oberkampf and Barone [34] consists of three main steps:

a) Calculation of the average relative error;

b) Calculation of the maximum relative error; and

c) Confidence interval estimation over the range of the experimental data for both of the above.

Even though the outcome is a quantity that represents the degree of validity, this metric has not been widely used in the literature. The paper itself has been cited many times for its summary of validation procedures and definition of a metric, e.g. see references [42, 43, 44], but in the papers that have applied the metric, a simplified definition has been used or additional techniques were used to draw conclusions about the model's performance. For instance, Slaba et al [45], in their study of the space radiation doses, applied the metric in increments by dividing data into several subsets and evaluating trends between individual subsets, because the whole data set could not be averaged; or Fortunato et al [46] did not normalise the absolute average error to avoid division by zero, in their application of the metric to fluid velocity predictions during flameless combustion events. This could be explained by the drawbacks identified by the authors themselves, for example it is not appropriate for a system where a response quantity of interest cannot be time averaged or when the values used for calculations are close to zero, which will be the case in some applications. For example, in image decomposition depending on the content of an image the difference between the coefficients within the feature vector could be of orders of magnitude, with smallest coefficients being less than 10^{-2}.

Kat and Els [41] in their method, instead of calculating the average discrepancy like Oberkampf and Barone [34], computed an absolute percentage relative error for each pair of data points considered for validation of periodic signals, e.g. frequency response of a vibrating plate. By doing so, they highlighted an issue of drawing a conclusion about an overall data set that has a high variability of discrepancy over the quantity of interest. To overcome the issue, Kat and Els [41] evaluated the set of relative errors against the specified threshold, set by the assessment requirements, and consequently obtained the probability that the model is producing results at or below the threshold. However, the assumption is made that the data used is a deterministic quantity of interest and thus an uncertainty analysis is not included.

Another example of a Frequentist method is Theil's inequality coefficient and can be defined as a scaled version of the root mean square (RMS) prediction error, which is usually expressed as

$$RMSerror = \sqrt{\frac{\sum_{k=1}^{n}(P_k - M_k)^2}{n}} \tag{2.2}$$

and is a measure of a spread of the deviations between data sets, e.g. predicted, P, and measured, M, data, where a value of 0 as an outcome represents a perfect fit of the data. The application of the Theil's inequality coefficient can be found in articles not only on economic forecasting, for which it was initially developed, but also in many other fields. For example, Rowland and Holmes [37] have analysed Theil's inequality coefficient for an application with sparse data consisting of a serially correlated time-series for a ballistic missile simulation. They concentrated on proposing an unbiased estimate of the statistical components of the Theil's inequality coefficient associated with the sample mean, variance and covariance in order to modify the original definition of the coefficient. Similarly, Dorobantu et al [47] reviewed it for the validation of mathematical models of aircraft dynamics. There are two versions of the coefficient, proposed by Theil in 1961 [48] as a measure of accuracy, equation 2.3, and in 1966 [49] as a measure of quality, equation 2.4. The difference between the two formulations is the presence or absence of the predicted component in the denominator:

$$T_{C1} = \frac{\sqrt{\sum_{k=1}^{n}(P_k - A_k)^2}}{\sqrt{\sum_{k=1}^{n}(P_k)^2} + \sqrt{\sum_{k=1}^{n}(A_k)^2}} \tag{2.3}$$

$$T_{C2} = \frac{\sqrt{\sum_{k=1}^{n}(P_k - A_k)^2}}{\sqrt{\sum_{k=1}^{n}(A_k)^2}} \tag{2.4}$$

where P represents model's predictions and A represents physical measurements or observations, i.e. actual data of the corresponding quantity of interest.In the first case, the presence of the RMS predicted data in the denominator means that the coefficient is influenced by the computational model, because the coefficient becomes biased towards the prediction and is not determined uniquely by the mean square prediction error [49]. By removing this term, a more objective measure is obtained, so that usually when comparing the output for the same set of data $T_{C1} < T_{C2}$.

Along with the formulation, Theil [48] set criteria for evaluation of the coefficient:
$T_C < 0.3$ as a good correlation;
$0.3 < T_C < 0.6$ as a medium correlation;

The closer to zero, the better the prediction is relative to the actual data, with $T_C = 0$ corresponding to two data sets being the same. Any value beyond 0.6 is considered to correspond to a low quality prediction. There is no upper limit for T_{C2} though values rarely exceed 1; T_{C1} is bounded between 0 and 1.

In addition to Theil's comments [49], a number of papers have discussed the differences between the two formulations above, and concluded that T_{C2} provides a more meaningful result. In 1973 Granger and Newbold [50] were the first, after Theil, to criticise T_{C1} in relation to its adequate evaluation of the predictive model. They showed the influence of the variance of the predictor data on the outcome and concluded that data with a high variance is more likely to result in a low coefficient. In the same year, Bliemel [51] published a note to clarify the difference and also mentioned the effect of the variance. He concluded that T_{C2} is a simpler version with a more meaningful outcome, as it evaluates the deviation between the data sets against the measured data only and thus is not biased towards the predictions. Kloek [52] in 2001 in his summary on contributions by Theil devoted a subsection to the inequality coefficient. He stressed that often authors neglect the differences and do not specify which formulation they used, making it hard to understand their results. Following this criticism, one would expect the first formulation to fade into the past, nevertheless most of the literature still cites T_{C1}. Dorobantu et al [47] identified the above mentioned drawbacks of the T_{C1} and later proposed a modification of this formula to produce a gap metric that was similar to T_{C2}. In addition, it was noted that a common misinterpretation is not only in the use of T_{C1} instead of T_{C2}, but also some authors take P as actual data and A as predicted, e.g. Li et al [53] when using T_{C1} for image quality evaluation and Kanayath et al [54] for validation of aerodynamic coefficients. This would not make a difference in the results for T_{C1} but would make T_{C2} a completely inaccurate representation of the prediction quality due to the predicted data being taken as a referent instead of the experimental data.

2.2.3 Bayesian analysis

The third class of metrics can be distinguished as being based on Bayesian analysis. In Bayesian analysis, initial information about the quantity of interest, e.g. a model parameter, is described by a probability distribution, known as a prior distribution, and is updated using additional data described as a likelihood, to produce a new or updated probability distribution describing the quantity of interest, known as the posterior distribution. Thus, Bayes' formula can be expressed as [55]:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{\int_\theta P(D|\theta)P(\theta)d\theta} \propto P(D|\theta)P(\theta) \qquad (2.5)$$

where $P(\theta)$ is a prior, $P(D|\theta)$ is a likelihood and $P(\theta|D)$ is a posterior, with $\boldsymbol{\theta}$ representing a parameter or parameters of interest and D is the additional data, e.g. observations such as physical measurements or expert knowledge. Through this procedure one can see how the prior data considered changes with the evidence that is presented as a likelihood. The analysis of the probability is appropriate for the purpose of validation as referred to in this project, because the data itself is not of the main interest but the similarity of the data to the reality, given the experimental evidence and associated uncertainties. Although, in the literature Bayesian analysis is usually associated with model calibration, [4], i.e. adjusting model parameters to improve agreement between the simulation results with a specified benchmark, and updating, [56], rather than validation of a model, i.e. assessing the extent to which a model is an accurate and reliable representation of the reality of interest, and is common in work on engineering design [43].

From the perspective of the validation, Bayesian analysis does not directly give an indication of the extent to which a model is a good or bad representation of reality for it's intended use [44]. Instead, the main focus of these approaches has been the definition of model parameters. For instance, in the Sandia validation challenge [44] none of the participants used Bayesian approaches for validation. Instead some authors concentrated on uncertainty quantification and model parameter calibration [57, 58]. These are important steps and can be appropriate for a sensitivity analysis before performing a validation. Bayesian techniques have also been used to aid model selection, however the purpose of model selection is to distinguish between the competing models and does not necessarily involve validation, hence has not been included in this review. The utility of a Bayesian analysis as a validation metric has been actively debated but has been pursued by some researchers.

Wang et al [43] identify a Bayesian approach as a category of validation metric; however in their examples they only used the analysis to update model parameters and then used a root mean square percentage error to evaluate the discrepancy. Some may argue that validation can be performed through the use of the Bayes factor, the ratio between the prior and the posterior, and a confidence index derived from the Bayes factor. For example, Liu et al [36] identify the Bayes factor and associated confidence index as a metric, and Rebba and Mahadevan [59] discuss a reliability measure based on the Bayes factor as an indication of validity. Bayarri et al [42] built a validation framework that is based on Bayesian statistics, although they concentrate on the analysis of the data rather than the validation. Brynjarsdottir and O'Hagan [60] do not claim to discuss validation but they investigate model discrepancy defined as a difference between the reality and the model output. In their paper, model discrepancy is part of the model calibration, but it could be used to aid establishing the validity of the model by providing more information and as such more evidence to be included in the validation process [42].

Considering the Bayes' formula, the prior allows to state the initial belief about the model's parameters or its predictions for the intended use in terms of probability distribution; the distribution of the prior, e.g. Gaussian or binomial, can be assumed in the beginning. Though, it is important to know that this decision will influence the properties of the posterior distribution, unless there is a lot of observations and this additional data, i.e. the likelihood, provides compelling evidence. The likelihood function makes it possible to incorporate the information such as, for example, the uncertainties associated with experimental and simulation sets of data to draw the conclusion about the model's validity. This term updates the initial statement of the probability, i.e. prior stated above, and potentially gives a more realistic outcome, than just relying on the prior. It should help to avoid the over or under prediction of the models validity if carefully defined [60]. However, the concern is that the choice of the likelihood and of what information it includes is entirely the user's choice and thus is subjective, which contradicts the aim for the objective validation metric. Following on the Brynjarsdottir and O'Hagan [60] idea of model discrepancy, we would like to update the initial statement of the validity using a more detailed uncertainty analysis. For a better estimate of the validity, uncertainty in the model should be included and the representativeness of the reality in the experiment needs to be considered, in addition to the errors due to the instrumentation. The last term of the formula, the posterior, is the updated prior distribution. The interpretation of the outcome strongly depends on the data and assumptions in the right hand side of the Bayes formula. Currently, if a model or its parameters are considered as a prior, posterior is a new updated data, i.e. a model or its parameters.

At the moment, there is no practical validation metric based on the Bayesian analysis. The methods presented in the literature follow a similar trend of concentrating on improving a model rather than evaluating its present performance. Hence, Bayesian analysis is more appropriate for model calibration and sensitivity studies which are important and should be performed before starting a validation process. Model discrepancy has been mentioned in a number of papers and is thought to aid the evaluation of validity, though for complex models the overall analysis requires demanding computational iterations. A few papers, for example see references [38, 61], have considered the topic of integration between Frequentist and Bayesian ideas including from the philosophical perspective; however these limited studies have not yet led to a comprehensive metric, nor are applicable to fields of data that are the focus of this thesis.

2.3 Summary of the review

Model validation is part of the Verification & Validation process and is relevant across all areas of science where models are used to make predictions. In the second half of 20th century, its definition and underlying ideas were clarified, and the first publications on validation techniques started to appear. This has stimulated the development of V&V guides where validation is presented as a process with number of activities on data acquisition, processing and comparison. Ultimately, the outcomes of these activities feed into a decision stage (Figure 3) by providing sufficient information to evaluate the extent to which model's predictions are representative of the real world. One of the key steps in this process is the application of a validation metric, which is a quantitative comparison of experimental and computational results.

A qualitative comparison of field data such as displacement maps, for example through graphical representations, is rarely sufficient to validate the model and conclude the extent to which the computational results represent reality, whereas the implementation of quantitative measures could improve the usefulness of the validation outcome. Three categories of validation metrics have been identified, namely Hypothesis testing, Frequentist and Bayesian approaches, however neither methodology falling in these categories has fulfilled the desired criteria for the validation metric nor has been extensively implemented in industry. Also, there has not been a lot of effort on utilising data obtained with the aid of optical measurement systems, apart from a recent European collaborative research project that led to the CEN guideline [24]. This guideline is the most up-to-date for a validation of full-field data in solid mechanics, although further development is required. For example, the methodology suggested in the guideline does not provide a quantitative measure of the model's predictions quality, beyond a boolean answer such as *'valid'* or *'not valid'*. Other methodologies falling within the Frequentist category have number of short-falls, e.g. with respect to the range of data within the data set evaluated or difficulty to interpret the outcome, but most importantly have not been applied to field-data. Whereas it was concluded that currently there are no comprehensive validation metrics based on the Bayesian analysis.

3 Methodologies

In this chapter validation metrics developed to meet the desired criteria identified in the earlier chapter are presented. The desired criteria include:

- objective
- quantitative
- take into consideration uncertainties associated with data sets
- produce quantitative outcome
- easy to communicate

Examples of the approaches used to obtain and process data fields are also included.

3.1 Data processing

Simulation outcomes are usually graphical and it is beneficial to obtain similar output from the experiments, for example colour maps of deformation. Instead of a visual qualitative comparison of such deformation data, it has been previously proposed to utilise an image decomposition, or orthogonal decomposition, technique to process the data from both the experiment and simulation for the model updating [62, 63] and later for the validation [6, 64]. Image decomposition techniques are based on a principle of fitting a set of selected polynomials to an image and are commonly used to compress or capture shape features of the images for applications such as object tracking, face and natural structures recognition [65]. More recently, it was suggested that field data, i.e. displacement or strain fields, can also be treated as an image where the level of, for example, displacement is represented by a colour or grey level values [63, 64]. An image is a two-dimensional matrix of data, in this scenario a displacement field, and with the aid of image decomposition it can be condensed from the order of 10^6 pixels or data points to a one-dimensional feature vector consisting of only a hundred or less shape descriptors. Shape descriptors are coefficients of polynomials, or also referred to in the literature as moments, and are obtained by fitting the polynomials to the image.

Table 1: Example of polynomials used for image decomposition, defined on continuous and discrete domains, and distinguished by the sensitivity to local or global features present in the image (based on table from Wang and Mottershead [66]).

		Global	**Local**
Continuous		Fourier Zernike	Continuous wavelets
Discrete	Uniform lattice	Chebyshev	Discrete wavelets Krawtchouk
	Non-uniform lattice	Racah	

A number of different polynomials are described in the literature on decomposition techniques; for example, Table 1, adapted from Wang and Mottershead [66] , lists some common functions used for image processing. For validation purposes Chebyshev polynomials, Zernike polynomials and Advanced Geometric Moment

Descriptors (AGMD) were applied in recent publications [67, 68, 26]. These can be classified as continuous or discrete, orthogonal or geometric moments, and can be applied to rectangular, circular or irregular surfaces respectively [69]. For instance, Chebyshev and Zernike polynomials are both orthogonal, but Chebyshev polynomials are discrete and defined in Cartesian coordinate system, which is suitable to process an image with rectangular shape, whereas Zernike polynomials are continuous and defined in polar coordinate system, which is suitable to process circular shape. Typically, discrete orthogonal moments are computationally easier to implement by comparison to geometric moments, and they overcome the accumulation of the discretization error of continuous moments, which leads to minimal information redundancy [70, 71]. Decomposition using polynomials not only reduces the dimensionality of the data, i.e. from an image that is a two-dimensional matrix to a one-dimensional feature vector, but also provides a unique way of translating the data from different sources and in different co-ordinate systems into the same format [63, 72], i.e. invariant and feature preserving shape descriptors can be obtained. The magnitudes of the shape descriptors are insensitive to the difference in scale, rotation and translation between predicted and measured images, and each of the descriptors corresponds to a specific feature of the image. For example, in the Chebyshev polynomial the first coefficient represent the magnitude of the data captured in the image and the other coefficients represent different shapes of the deformation, as shown in Figure 5. Coefficients have the same units as the data in the image and the magnitude of each coefficient corresponds to the strength of that particular feature, so that coefficients with higher values represent more dominant features. The polynomial used for the decomposition and the corresponding coefficients obtained can be stored and used to reconstruct the original image when necessary. The CEN guideline [24] is based on the idea that measured and predicted data fields can be treated as an image, and it was followed in the current research to perform decomposition. The choice of a specific polynomial was based on the individual cases studies.

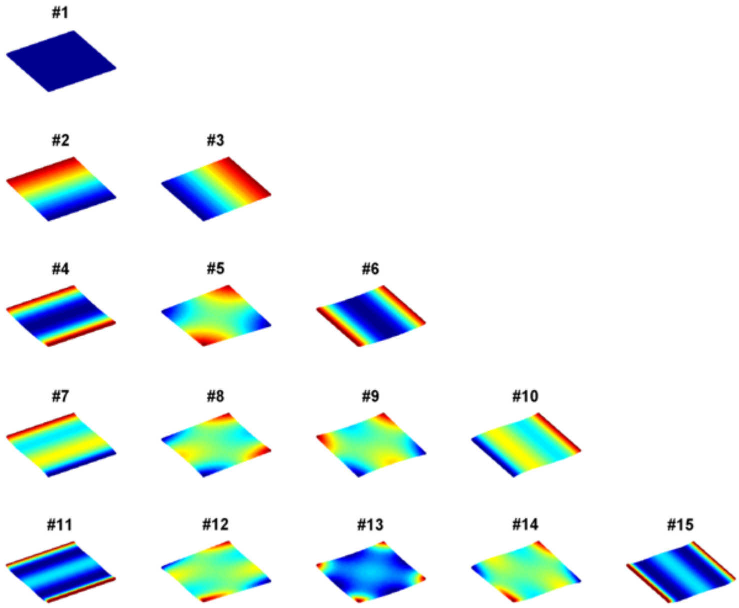


Figure 5: A visual representation of first 15 Chebyshev coefficients as surface maps (reproduced with permission from Berke et al [67]).

It is important to ensure that a feature vector is a good representation of the original data before utilising it in a validation procedure. As suggested in a recent publication by the CEN [24], data from experiment and simulation can be decomposed with the same order of polynomials, and the accuracy of the representation can be evaluated through a reconstruction procedure. Following the guidelines, if the data from the original decomposed image is expressed as $I(i,j)$ and for the reconstructed image as $\hat{I}(i,j)$, then the average squared residual u obtained from N data points is

$$u^2 = \frac{1}{N}\sum_{i,j}^{N}(\hat{I}(i,j) - I(i,j))^2 \tag{3.1}$$

The average squared residual u represents the uncertainty introduced by decomposing the image and N corresponds to the number of data points in the image. Ideally, $u < u_{cal}$ should be satisfied [24], where u_{cal} is the minimum measurement uncertainty obtained using a calibration procedure for the measurement apparatus; for example Patterson et al [73] provide guidance on a calibration material and procedure for optical measurement systems. In addition, there should be no region with a cluster, i.e. a region of adjacent pixels, $> 0.3\%$ of the total region of interest with a residual $> 3u$. These two criteria, namely $u < u_{cal}$ and no cluster $> 0.3\%$ with a residual $> 3u$, are used to assure minimal loss of the information during the decomposition; if these criteria are not met, a higher order polynomials should be used to decompose images. The total uncertainty associated with the moments of the experimental data can be obtained using

$$u_{exp} = \sqrt{u_{cal}^2 + u_{deco}^2} \tag{3.2}$$

where u_{deco}^2 is the average squared reconstruction residual of the image from the experiment calculated using equation (3.1) with N number of data points in the image [24].

3.2 Theil's Inequality Coefficient

Theil's inequality coefficient has been previously identified as a useful metric for validation purposes, so it was decided to modify this method to be applicable with full-field data. In the current work, the metric was evaluated based on the definition of the T_{C2} such that:

$$T_{C2} = \frac{\sqrt{\sum_{k=1}^{n}(S_{P,k} - S_{M,k})^2}}{\sqrt{\sum_{k=1}^{n}(S_{M,k})^2}} \tag{3.3}$$

where $S_{P,k}$ and $S_{M,k}$ are moments, or coefficients of the feature vectors, obtained from decomposing predicted and measured data respectively. Corresponding pairs of orthogonal moments, e.g. the first for experimental data with the first for simulation data when k=1, were evaluated to obtain the final coefficient. The result for the metric is a single value and the outcome is evaluated with respect to the criteria originally set by Theil [48]: given that a value of the coefficient close to zero would represent a model that produces results similar to experimental results and 0 would indicate identical data sets.

3.3 Relative error

As the name suggests, this metric is based on calculating a relative error. It is applied to feature vectors containing coefficients from orthogonal decomposition to validate the model's predictions [74]. Initial steps to establish the validity consist of:

- Computing a normalised relative error for each pair of data in the feature vector;
- Computing a weight for each error; and
- Defining a threshold.

In the first step, the relative errors are calculated by normalising the absolute error of each data pair, i.e. using the pair of first coefficients in the feature vectors, by the coefficient with the largest absolute value from the experimental feature vector such that the error is given by:

$$e_k = \left| \frac{(S_{P,k} - S_{M,k})}{max|S_{M,k}|} \right| \tag{3.4}$$

As noted previously in Sections 2.2.2 and 3.1, the values of the coefficients in a feature vector can vary in the orders of magnitude, which is governed by the prominence of a particular shape in the image such that coefficients with higher values represent more dominant features. Consequently, to avoid division by very small numbers, i.e. $< 10^-2$, or zero, all absolute errors are normalised by the largest absolute value from the measured feature vector.

The second step is to calculate the weight of each error, defined as its percentage of the sum of all errors, thus

$$w_k = \frac{e_k}{\sum_{k=1}^{n} e_k} \times 100 \tag{3.5}$$

where n is the number of components in each vector. This provides a vector of values that represents the quality of the model relative to the experimental data. The following piece of information required for the validation is a threshold and it is defined by combining the concept of pre-specified threshold from Kat and Els [41] with the uncertainty limits from the CEN guideline [24]. The threshold is based on the experimental uncertainty, namely the total experimental uncertainty u_{exp} from equation (3.2), as suggested in the CEN guideline [24], and is given by

$$e_{unc} = \frac{2u_{exp}}{max|S_M|} \times 100 \tag{3.6}$$

By doing so, the evaluation becomes more objective and incorporates uncertainty in the measured data. The last step to establish the quality of a model's predictions consists of comparing the vector of the error weights, w_i against the error threshold, e_{unc} and the sum of those errors below the threshold yield the validation metric outcome, thus

$$VM = \sum_{k=1}^{n} w_k \qquad \text{for} \qquad e_k < e_{unc} \tag{3.7}$$

Following the interpretation of Kat and Els [41], this sum corresponds to the probability of the normalised errors being at or below the experimental uncertainty. However, here it is proposed to interpret this outcome in terms of the probability that the model is representative of reality. By overcoming the drawbacks of the Frequentist approaches reviewed in Section 2.2.2, this new relative error metric is capable of evaluating data with a naturally high variance between the individual values in the data set, including very small values close to zero, and it takes into account uncertainties in the measurement data. Most importantly, the validity of the model is expressed as a numerical quantity and a clear statement can now be created to reflect the definition of the validity by the ASME [1]. The statement being proposed includes:

- The probability of the model being representative of reality;
- The intended use or loading conditions considered; and
- The quality of the data used as a referent.

3.4 Case studies

To explore the features and possible outcomes from the metrics described in Sections 3.2 and 3.3, three case studies were utilised:

- An I-beam subject to three-point bending,
- A rubber block subject to indentation, and

- A bonnet liner response to impact.

These case studies were selected from the available data due to their diversity of sample geometry, material and mechanical behaviour. The experimental data in all three case studies was obtained with the aid of a stereoscopic digital image correlation (DIC). This is a non-contact full-field optical measurement technique and is commonly used in solid mechanics to analyse in-plane and out-of-plane surface deformation by tracking the relative difference between a sequence of images captured during the experiment [25]. Each image is subdivided into evenly spaced array where each sub-image, or a facet, has a unique signature typically achieved by applying a random speckle patter on the surface of the test specimen; these unique signatures allow tracking of individual sub-images in the sequence of images. By correlating the relative displacement of each sub-image, an array of displacement vectors is generated that describes two or three-dimensional deformation of the entire surface in the field of view and can be plotted as a displacement map. In all three case studies, the test specimens were treated with white paint, and black paint was used to create a speckle pattern. Further details of the three case studies are summarised in the following subsections.

3.4.1 I-beam subject to three-point bending

The data for this case study was taken from an earlier study [68] of the efficacy of the validation methodology described in the CEN guide [24], and key details of the model and experiment are included here. A half metre length of aluminium I-section with cross-section dimensions 42x65mm was subject to static bending by a central load while supported symmetrically by two 50mm diameter solid rods of circular cross-section that were 450mm apart. The thickness of the web and flange was 2.5mm and a series of four 35mm diameter circular holes penetrated the web at 100mm intervals along its length, as shown in Figure 6. In the

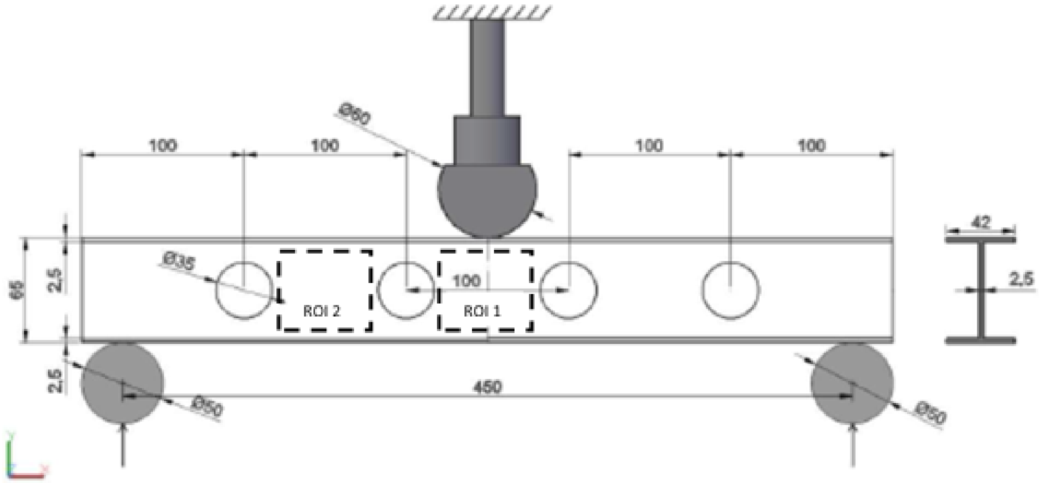


Figure 6: A diagram of the I-beam subject to three-point bending showing the regions of data used in case study 1 (reproduced with permission from Lampeas et al [68]).

experiment, a stereoscopic digital image correlation system was used to acquire displacement data at 15 frames per second and the minimum measurement uncertainty was established as $10\mu m$ and $30\mu\epsilon$ for displacement and strain using the calibration procedure described in [75]. A finite element model was created using 23,135 shell elements with the Ansys software package and employing an elasto-plastic material model with kinematic hardening. In response to the loading, the beam bent elastically and stress concentrations were observed at discontinuities; a sample of a measured full field displacement map is illustrated in Figure 7.

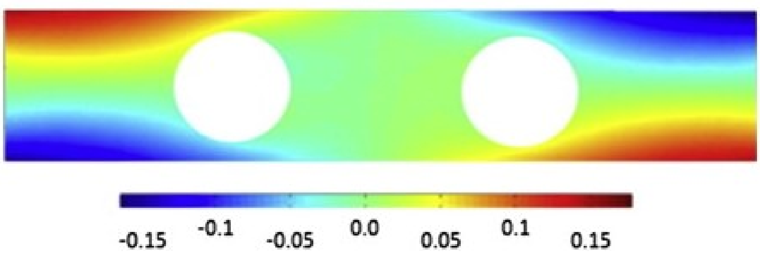


Figure 7: A measured longitudinal displacement field in mm around two middle holes, obtained using DIC system in case study 1: I-beam subject to three-point bending (reproduced with permission from Lampeas et al [68]).

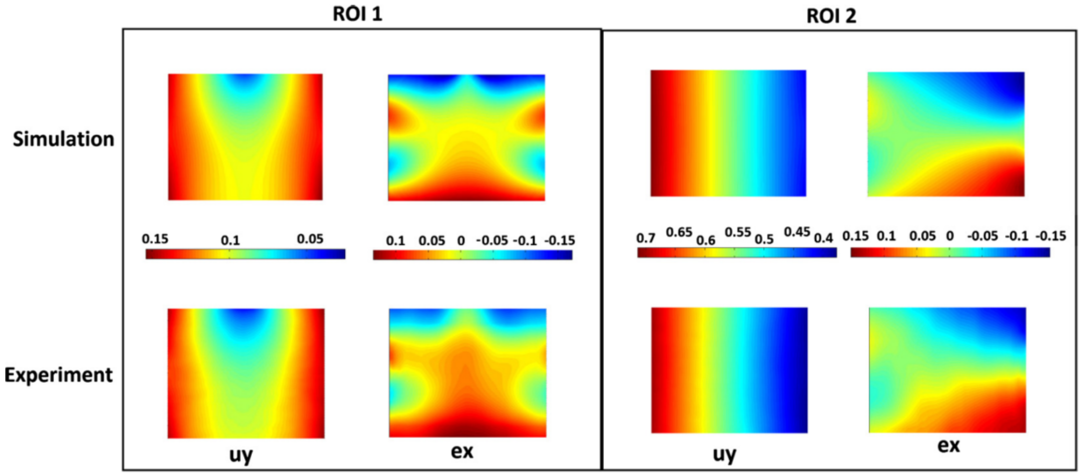


Figure 8: Predicted and measured data fields of transverse displacement, uy in mm, and longitudinal strain, ex in %, obtained from Region 1 and 2 as highlighted in Figure 6 (reproduced with permission from Lampeas et al [68]).

In this case study, the validity of the predictions of the transverse displacement of the web and the longitudinal strain in regions 1 and 2 in the Figure 6 were evaluated. Each region was 60 x 50 mm, and for post-processing the facet size and pitch, i.e. the grid spacing, were set to 46 and 11 pixels; images of displacement and strain fields from region 1 and 2 are shown in Figure 8. These predicted and measured data fields were decomposed using Zernike polynomials, and only the significant coefficients were included in the validation. When selecting polynomials to use for the image decomposition, Lampeas et al [68] followed the original work by Wang at el [76], where Zernike polynomials were successfully applied to process images of strain maps on the surface of a square plate with a circular hole. Although, Lampeas et al [68] suggested an alternative method to achieving an optimal feature vector to represent measured and predicted data, consisting of significant coefficients or moments. In general, significant coefficients can be selected after decomposing images with large number of coefficients, e.g. 200, and applying a threshold to remove coefficients with relatively small magnitude, i.e. coefficients whose values are below certain percentage of the largest coefficient in the feature vector; the remaining coefficients should still lead to a reasonable reconstruction as described in Section 3.1. In this case study, Lampeas et al [68] used 400 Zernike coefficients to decompose data fields and then applied a threshold of $< 5\%$ to obtain the final feature vectors. After applying a specified threshold, it is not always possible to obtain equivalent feature vectors representing experimental and simulation results, e.g. different number or order of coefficients might be retained, thus extra coefficients must be appropriately added to create matching feature vectors before proceeding with validation process.

3.4.2 Rubber block subject to indentation

The indentation of a 60x60x25mm rubber block by a rigid wedge has been investigated previously by experiment and modelled analytically [27] and computationally [77]. Consequently only a brief outline is provided here. Deformation data for the rubber block was acquired using a stereoscopic digital image correlation system when a compressive displacement of 2mm was applied across the entire 30mm thickness of the block by an aluminium alloy wedge of external angle 73.45 degrees and tip radius 1.68mm (see Figure 9). A stereoscopic digital image correlation system was used and calibrated to provide minimum measurement uncertainties of $3.2\mu m$ and $23.8\mu m$ for the in-plane [33] and out-of-plane [78] displacements respectively. Predictions of the x-, y- and z-direction displacements were obtained from a refined original finite element model simulated in the Abaqus 6.11 software package using 49,920 three-dimensional eight-noded linear elements for the block and 2,870 three-dimensional four-noded bilinear quadrilateral elements for the wedge. The model is illustrated in Figure 10. The material of the wedge was assumed to be rigid while the rubber was modelled as a hyperelastic material defined by the Mooney-Rivlin relationship with the constants taking the following values: C_{10}=0.9 and C_{01}=0.3 with a bulk modulus, J=20.

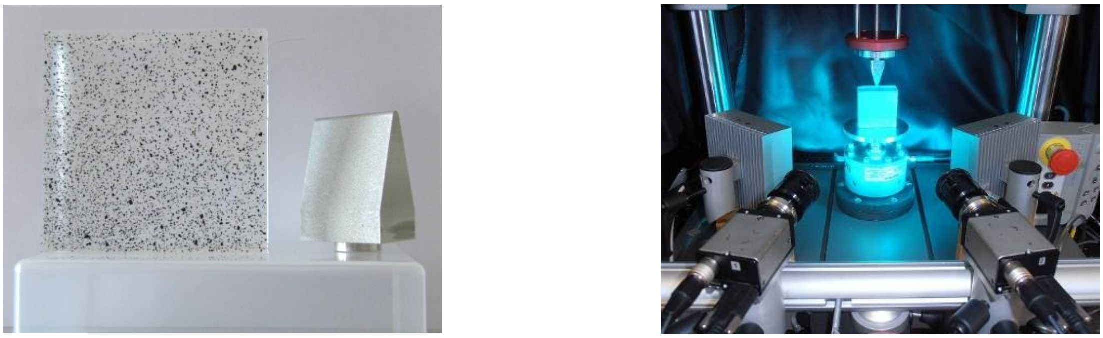


Figure 9: The specimen and experimental set up: rubber block (60x60x25mm) and the aluminium indenter with the 1.68mm radius tip and 73.45 degrees included angle (reproduced with permission from Tan et al [27]).

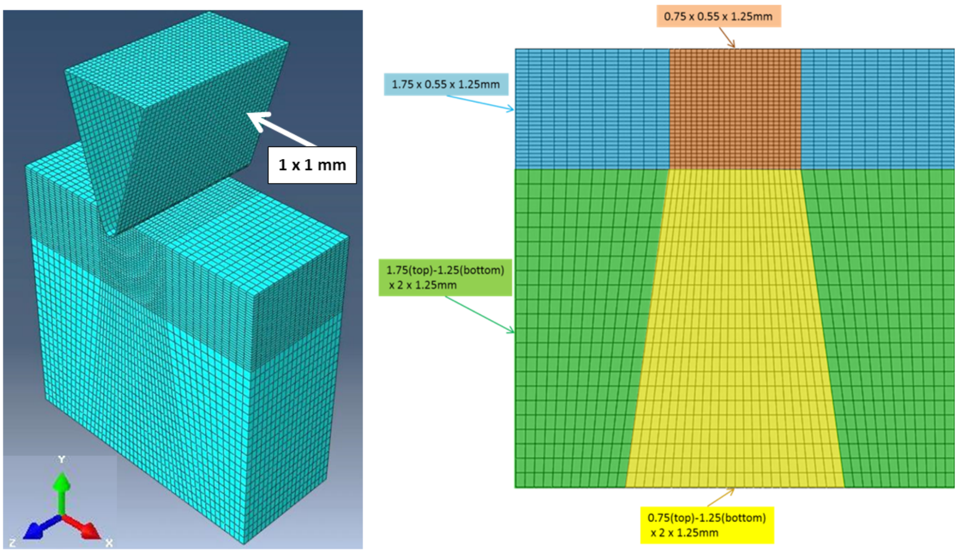


Figure 10: Finite element model of the indenter and the rubber block (Figure 9) with the mesh size for different sections.

The measured and predicted displacement fields are shown in Figure 11 and were decomposed using Chebyshev moments, due to the rectangular shape of the validation region and to avoid the accumulation of the discretization error as mentioned in Section 3.1. 170, 210 and 15 coefficients were computed for the displacement in x-, y- and z- directions on the surface respectively by achieving average reconstruction residuals that are just below the minimum measurement uncertainty.

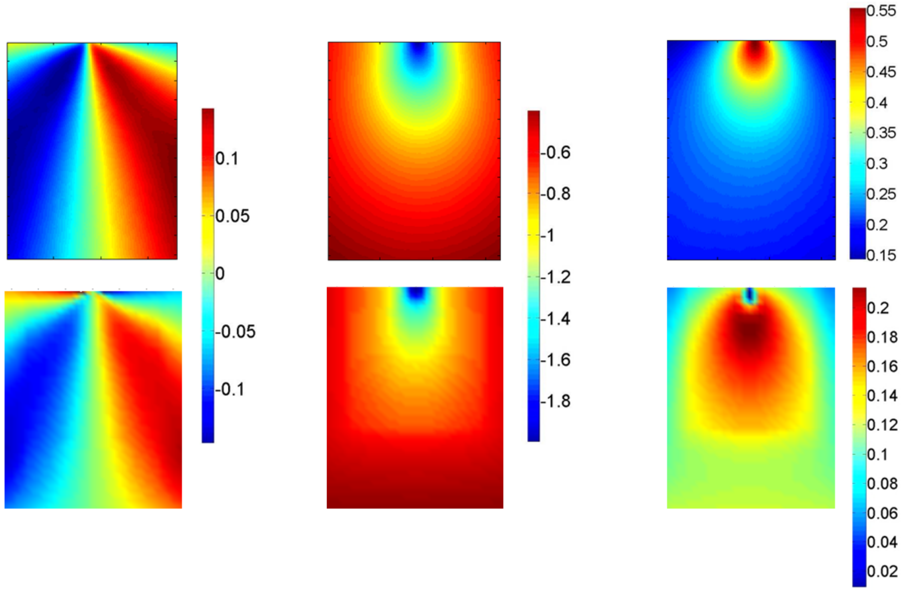


Figure 11: Measured (top) and predicted (bottom) x-direction (left), y-direction (middle) and z-direction (right) displacement fields for a 28.5x23mm area of the rubber block shown in figure 9 when it was subject to 2mm displacement load by the wedge in the y-direction. The centre of the top edge of each data area corresponds the location of contact by the wedge and the units are millimetres (based on data from Tan et al [27]).

3.4.3 Bonnet liner impact

Burguete et al [26] have described the analysis of the displacement fields for an automotive composite liner for a bonnet or hood and so only an outline of the data acquisition and processing will be given here. The composite liner, which had overall dimensions of approximately 1.5x0.65x0.03m, was subject to a high velocity (70m/s), low energy (<300J) impact by a 50-mm diameter projectile with a spherical head as shown in Figure 12. A high-speed stereoscopic digital image correlation system was used to obtain maps of out-of-plane displacements at 0.2ms increments for 100ms. The minimum measurement uncertainty was based on a previous calibration study performed by Sebastian and Patterson [75] following a methodology prescribed in Patterson et al [73]; it was established to be $14\mu\varepsilon$ at $290\mu\varepsilon$ rising to $29\mu\varepsilon$ at $2110\mu\varepsilon$ [26].The finite element code Ansys-LS-Dyna was employed to model the bonnet liner following impact using an elastic-plastic material model with isotropic damage and four-noded elements based on a Belytschko-Tsay formulation. Typical fields of predicted and measured fields of out-of-plane displacements are shown in Figure 13 and were decomposed using adaptive geometric moment descriptors (AGMD) specifically tailored for the complex geometry of the liner. Burguete et al [26] compared the data fields from the model and experiment for 100ms following impact by plotting the absolute difference between pairs of corresponding AGMDs as shown in Figure 14. They concluded that when any of the absolute differences were greater than the uncertainty in the experiments, indicated by the broken lines in Figure 14, then the model was not valid. In this study, the probability of the model predictions being representative of reality was assessed using the error threshold in equation 3.6 for each increment of time for which a displacement field was measured up to 100ms into the impact event.

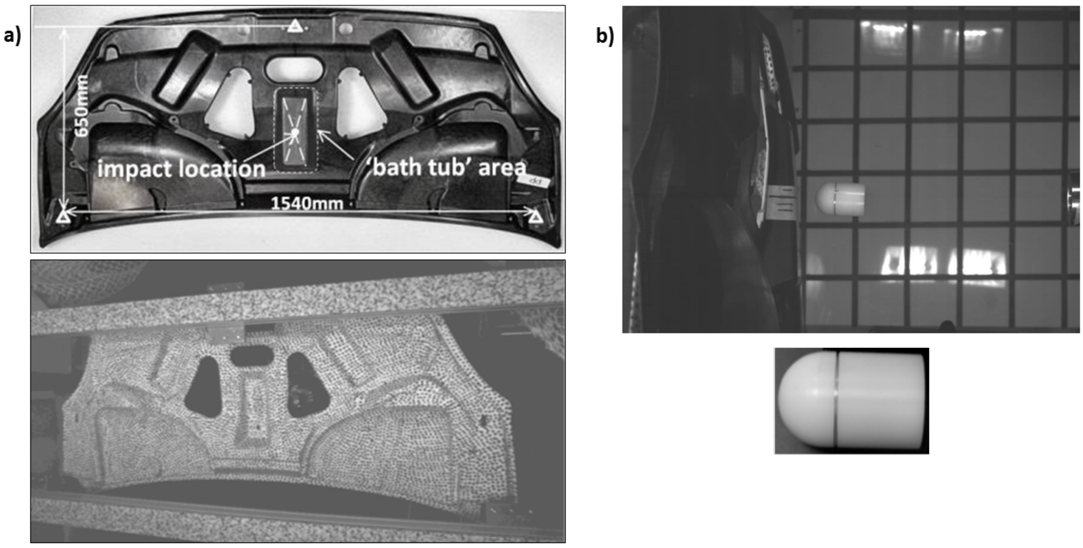


Figure 12: a) Specimen and b) test configuration: car bonnet liner (left) and a projectile (right) (reproduced with permission from Burguete et al [26]).

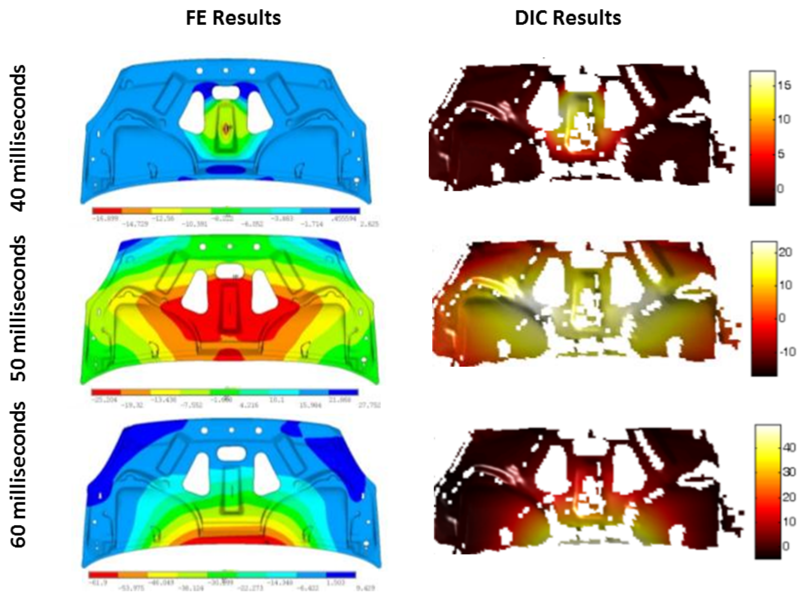


Figure 13: Predicted (left) and measured (right) out-of-plane displacement fields for the car bonnet liner shown in figure 12 at 40, 50 and 60 ms after a high-speed, low-energy impact by a projectile (reproduced with permission from Burguete et al [26]).

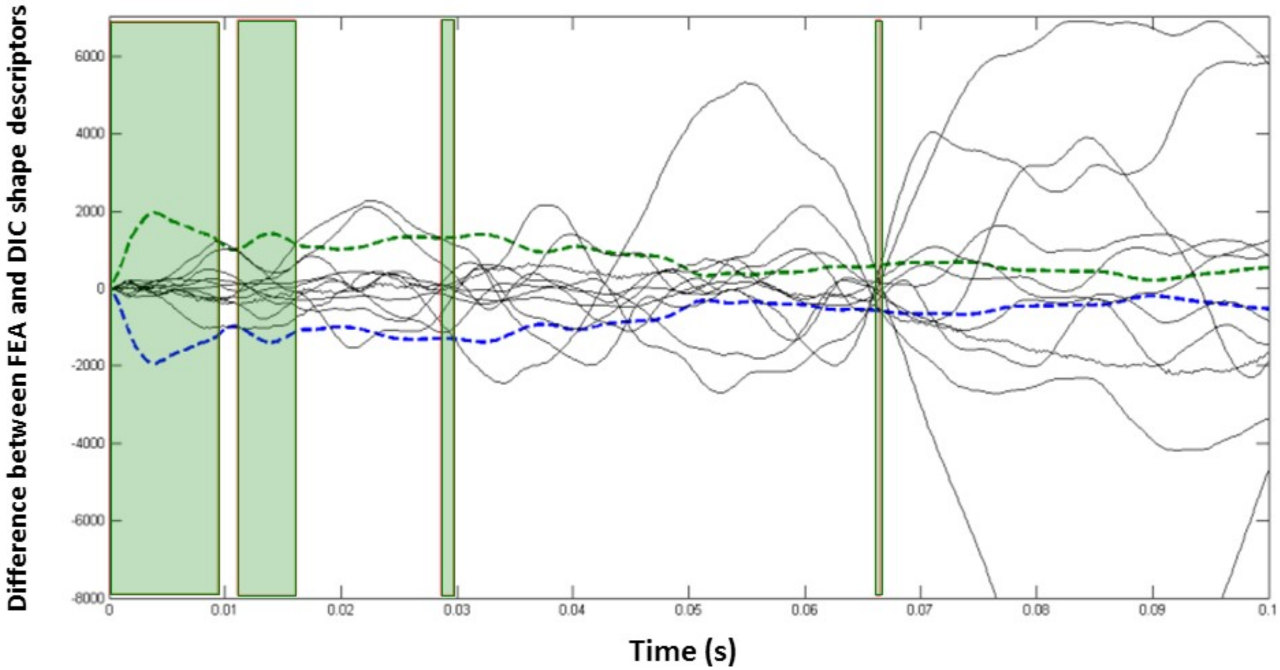


Figure 14: Absolute difference between the first twelve corresponding adaptive geometric moment descriptors (AGMD) representing predicted and measured out-of-plane displacement fields for the car bonnet liner (Figure 12-13) during the 100ms following the impact (reproduced with permission from Burguete et al [26]).

4 Results

Results obtained through the application of the novel methodologies described in the previous chapter are presented in the following sections. For the three case studies described in Section 3.4, i.e. I-beam, rubber block and bonnet liner, key outcomes are summarised and are supported by tables and figures at the end of this chapter. These results demonstrate the applicability of the novel validation metrics to data-fields and the depth of the information that the outcome offers. More detailed discussion of these results and implications within the validation process are covered in the next chapter, Chapter 5.

4.1 I-beam validation

As described in Section 3.4, two regions on the surface of the beam were considered for validation: Region 1 in the middle, around the area of applied loading between the two holes, and Region 2 on the side (Figure 6). Feature vectors representing displacement and strain data fields (Figure 8) in both regions are presented in Figure 15 and the number of significant coefficients summarised in Table 2. Results from applying Theil's inequality coefficient and a new relative error metrics are collated in Table 3, and graphically presented in Figure 17. For Region 1 it was found that the probability of predicting the displacement in the y-direction is 100% and for the strain in the x-direction is 48%. This correlates well with the outcomes in Figure 16 from Lampeas et al [68], where for

Table 2: Number of retained significant coefficients in feature vectors representing longitudinal strain and transverse displacement in Region 1 and 2 as highlighted in Figure (6) together with corresponding total measurement uncertainty as calculate using equation (3.2).

Region of Interest	**Number of coefficients**	**Total measurement uncertainty %**
Region 1 area uy	10	2.69
Region 1 area ex	50	3.57
Region 2 area uy	3	2.73
Region 2 area ex	41	3.97

Table 3: Results obtained from validation metrics for I-beam case study.

Region of Interest	**Theil's Inequality Coefficient**	**Error threshold %**	**VM %**
ROI 1 uy	0.07	24.15	100
ROI 1 ex	0.45	15.11	48
ROI 2 uy	0.05	4.61	100
ROI 2 ex	0.12	11.53	100

the displacement data the model was found to be valid, as all the data points in the graph based on the CEN guideline [24] were inside the uncertainty bounds, and for the strain it was found to be invalid, as in the graph there were some data points outside the uncertainty bounds. For the data in the Region 2 the probability using the new relative error metric was found to be 100% for both responses, i.e. strain and displacement, which also correlates well with the conclusions of Lampeas et al [68] who obtained concordance coefficient of 0.99 for the displacement prediction in the Region 2 that indicated a very good model.

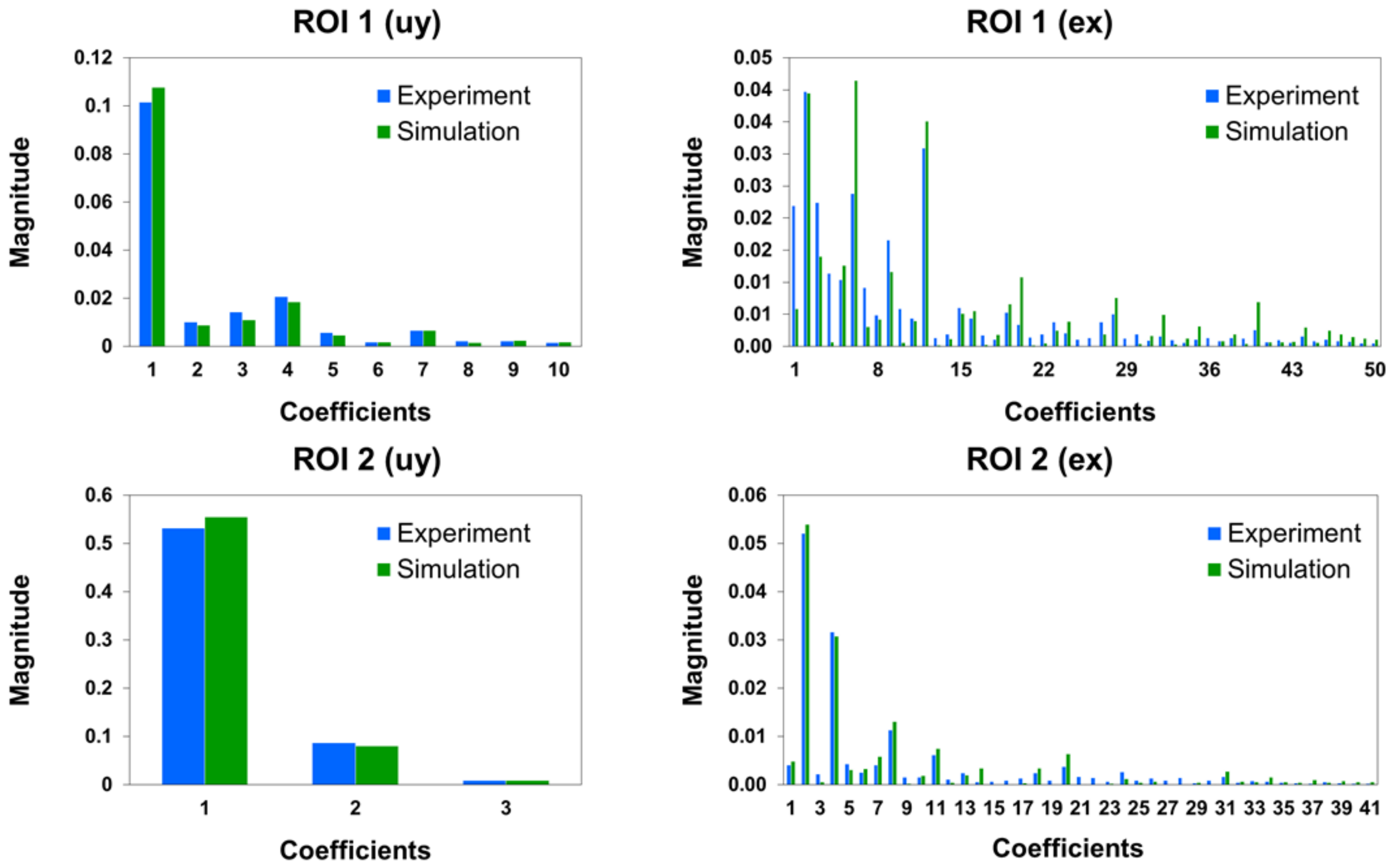


Figure 15: Retained significant Zernike coefficients obtained by decomposing measured and predicted fields in regions 1 (top) and 2 (bottom) of the I-beam subject to three-point bending shown in Figure 6. Transverse displacement is presented on the left and longitudinal strain on the right of the figure.(based on Lampeas at el [68])

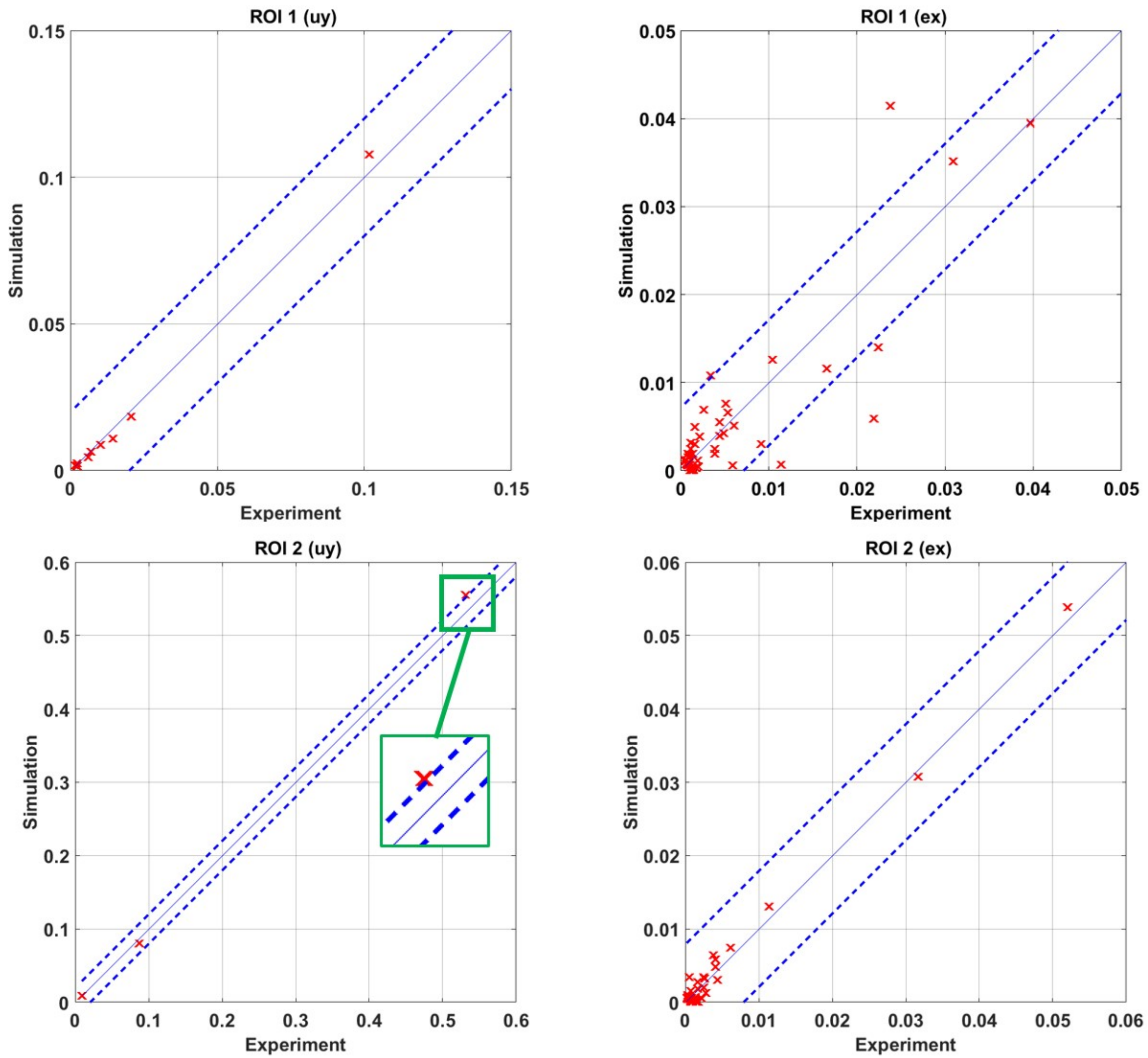


Figure 16: Graphical comparisons, using the approach recommended by the CEN guideline [24] for evaluating the validity of model predictions, of the Zernike coefficients representing the predicted (y-axis) and measured (x-axis) transverse displacement (left) and longitudinal strain (right) in regions 1 (top) and 2 (bottom) of the I-beam subject to three-point bending shown in Figure 6. (based on Lampeas at el [68])

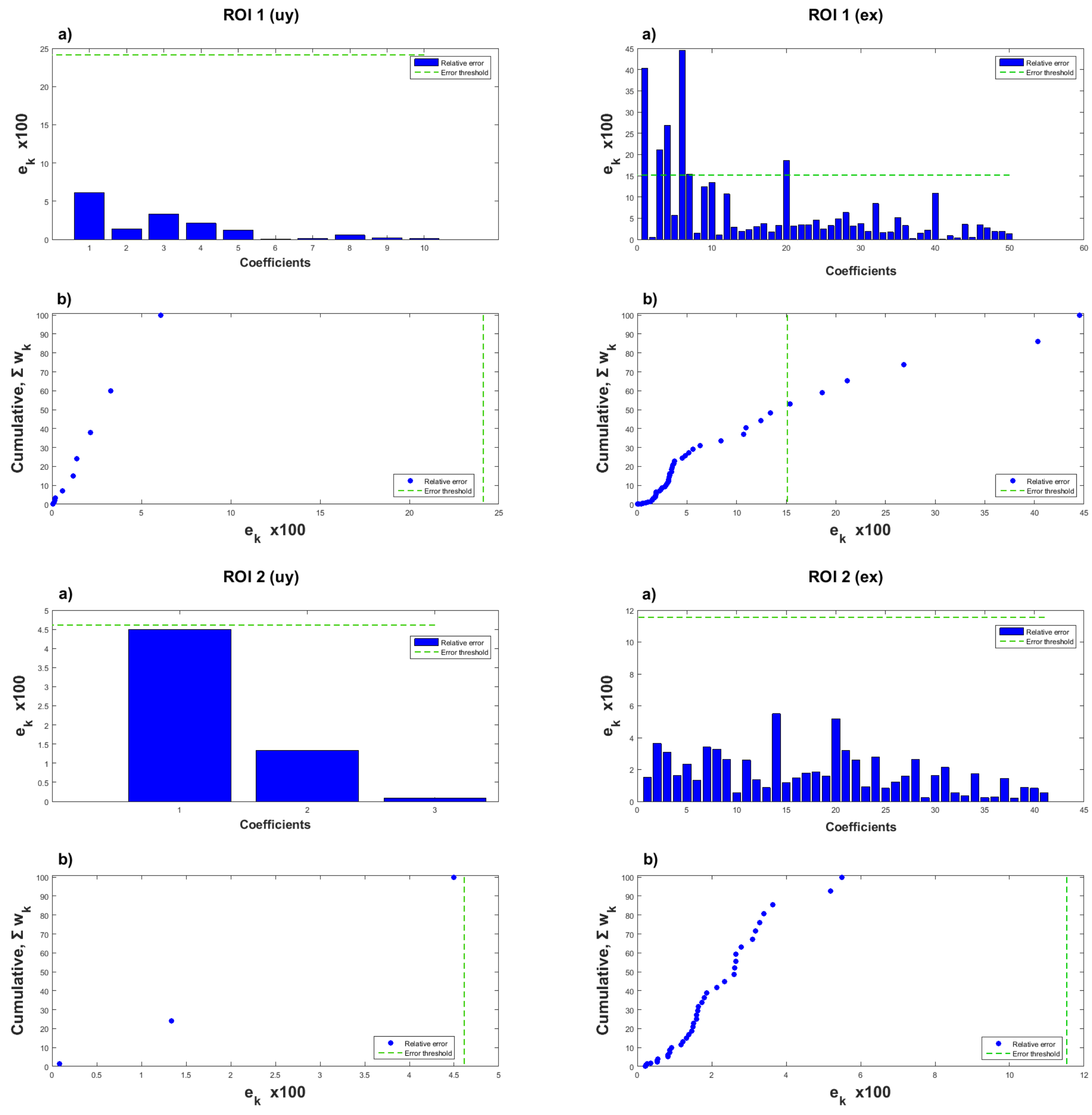


Figure 17: **a)** A bar charts of the normalised absolute errors computed using equation (3.4) with a superimposed threshold, e_{unc}, for the transverse displacement, uy (left) and longitudinal strain, ex (right) fields in region 1 and 2 of the I-beam (Figure 6). **b)** The sum of the weighted errors below the threshold represents the probability of the prediction being representative of experimental measurements. Corresponding final outcomes of the metric are summarised in Table 3.

4.2 Rubber block validation

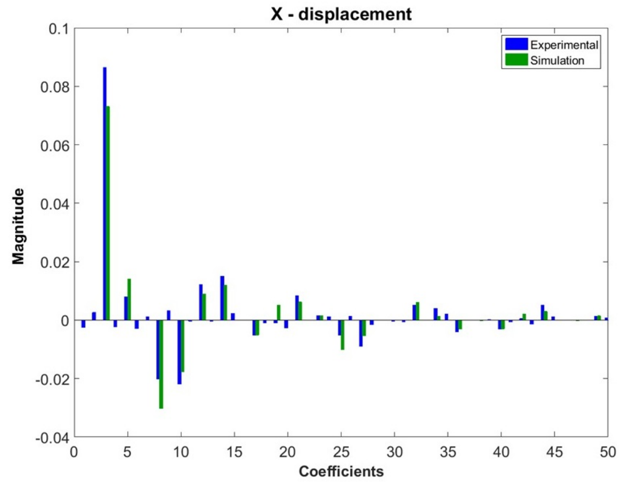
Results for the case study on the rubber block are summarised in Table 5, and graphically presented in Figure 20. The probability of the model's predictions being representative of reality was found to be 83%, 62% and 34% for x-, y- and z-direction displacements respectively, and Theil's inequality coefficient outcome also corresponds well with these results. The relative uncertainty used in the new relative error metric calculations was 10%, 1.2% and 19.4% for the individual data sets. It is evident that the outcome of the metrics has captured the discrepancies between the simulation and the experiment. As was expected from a visual comparison of the data fields, the model is poor at predicting z-direction displacement and this is reflected in the metric's outcome, i.e the probability is very low in comparison to the rest of the results even given the high uncertainty in the measured data. At the same time, the validation outcomes for the new relative error metric for the other two sets of data fields, i.e. x- and y-direction displacements, successfully reflect and quantify the minor differences, mostly due to the shape of the deformation. This can be deduced by examining relative errors for the individual pairs of coefficients in the left and middle sub-figures in Figure 20. Both of the relative errors for the first pair of coefficients are smaller than the relative uncertainty, meaning that they contribute towards the probability of the predictions representing reality. Whereas the largest relative errors, above the uncertainty limit, come from the coefficients representing the shape of the deformation and they cause the probability to decrease.

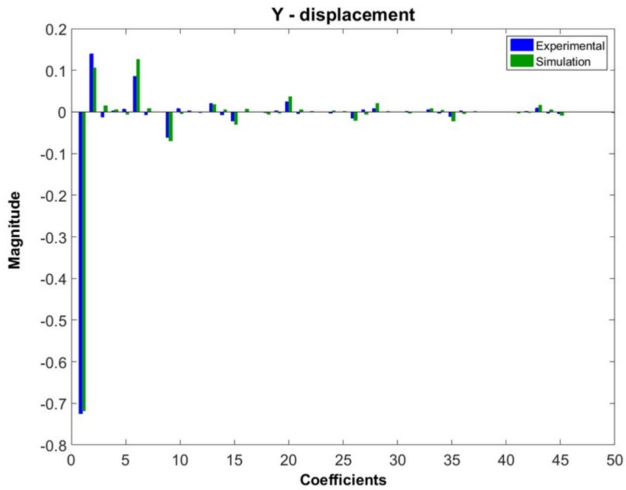
Table 4: Number of coefficients in feature vectors representing displacement fields for the 28.5x23mm region on the surface of the rubber block (Figure 11) together with corresponding total measurement uncertainty as calculated using equation (3.2).

Quantity of Interest	**Number of coefficients**	**Total measurement uncertainty** μm
X-displacement	171	4.1
Y-displacement	210	4.3
Z-displacement	15	23.9

Table 5: Results obtained from validation metrics for rubber block case study.

Quantity of Interest	**Theil's Inequality Coefficient**	**Error threshold %**	**VM %**
X-displacement	0.24	9.95	82.5
Y-displacement	0.11	1.2	62.4
Z-displacement	0.58	19.43	34.3





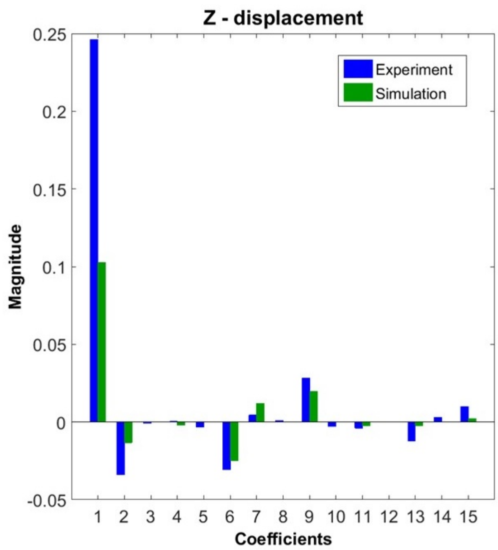


Figure 18: Feature vectors consisting of Chebyshev coefficients and representing the predicted and measured x-direction (top), y-direction (middle) and z-direction (bottom) displacements on the region of the surface of the rubber block shown in Figure 9. Only the first fifty coefficients for the x- and the y-direction are presented above.

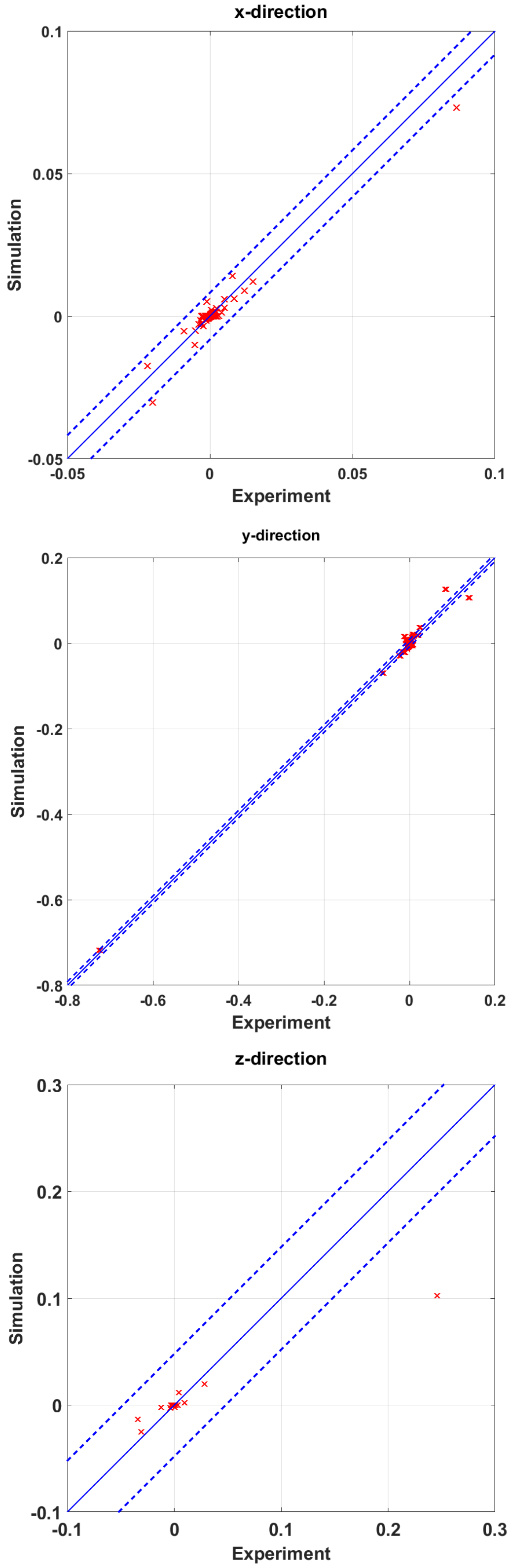


Figure 19: Graphical comparisons, using the approach recommended by the CEN guideline [24] for evaluating the validity of model predictions, of the Chebyshev coefficients representing the predicted (y-axis) and measured (x-axis) x-direction (top), y-direction (middle) and z-direction (bottom) displacements on the region of the surface of the rubber block shown in Figure 9.

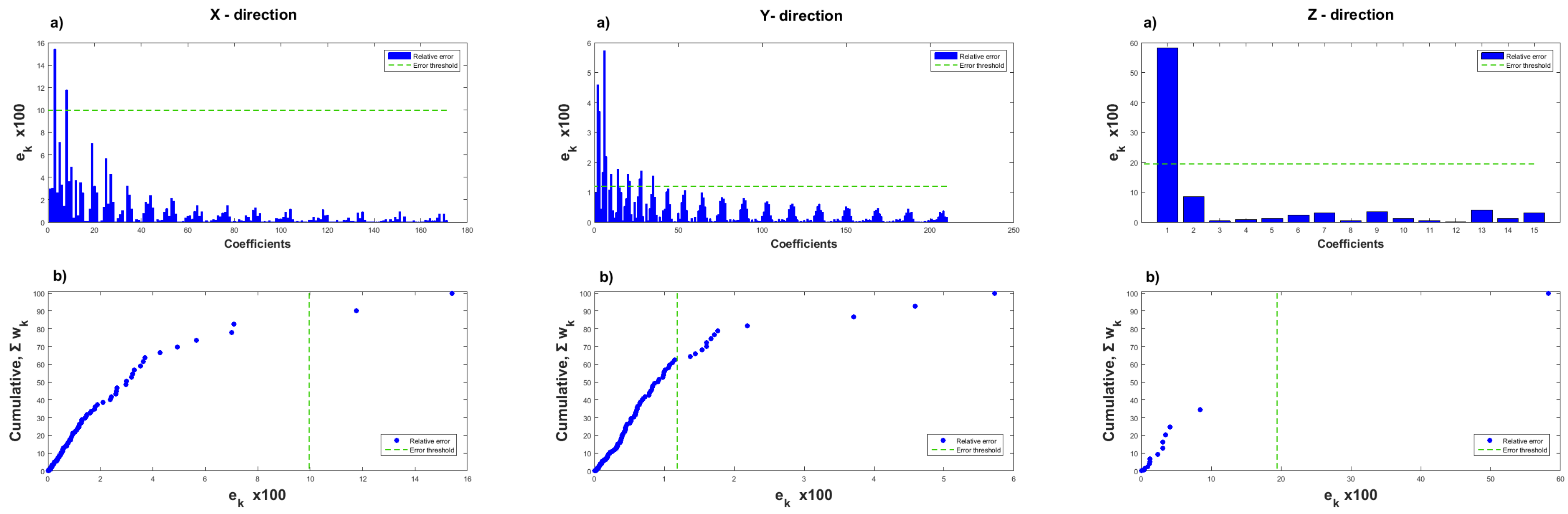


Figure 20: **a)** A bar charts of the normalised absolute errors computed using equation (3.4) with a superimposed threshold, e_{unc}, for the x-, y- and z-direction displacement on the surface of the rubber block shown in Figure 11. **b)** The sum of the weighted errors below the threshold. Corresponding final outcomes of the metric are summarised in Table 5.

4.3 Bonnet liner validation

The last set of results is associated with the car bonnet liner study and validation was performed for the data sets obtained at the time steps after the impact, see Figure 21. The trend of the relative error outcome with time is similar to the trend reported by Burguete et al [26] as can be observed in Figure 14, where the authors claimed that the model can satisfactory predict the displacement in z-direction up to 0.035s following the impact. After this time, the probability steeply decreases and continues to oscillate until the final decrease, at around 0.07s, when probability stays below 10%, apart from the last time-steps. It was noted by Burguete et al [26] that the validation of the second half of the period after the impact is no longer appropriate, because the model no longer corresponds to the experimental conditions; a crack has propagated in the region close to the impact, which was not simulated.

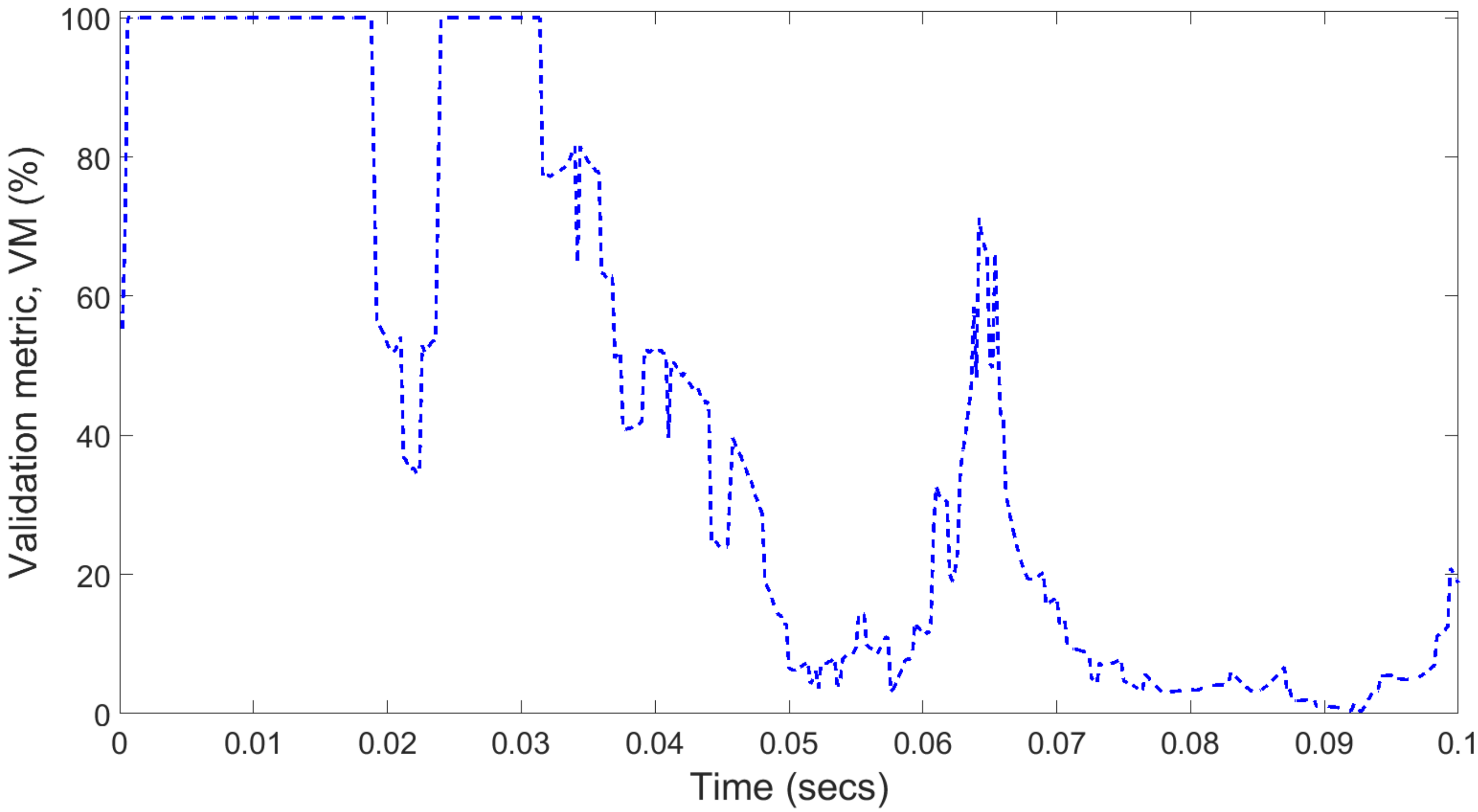


Figure 21: Graphical representation of the validation outcome obtained through the new relative error metric to evaluate out-of-plane displacement predictions for the car bonnet liner (Figures 12-13). Probability of predictions being acceptable was evaluated at 2ms time increments over 100ms following the impact.

5 Discussion

In the this chapter the application of validation methodologies to field data, corresponding validation outcomes and their interpretation as part of a validation process are discussed. This is first put into context of the whole validation process. Then followed by a critical analysis of methodologies described in Chapter 3 and corresponding results from Chapter 4, including comparison with methodologies mentioned in Section 2.2 and the communication of validation outputs. Strengths and weaknesses are discussed with respect to the objectives of this research and the desired criteria for a reliable and transferable validation metric.

5.1 Validation process

A validation metric is a tool incorporated in the validation process and applied to obtain quantitative information about the quality of a computational model with respect to the real world for a specified intended use. As described in Chapter 2, a validation process encompasses number of activities associated with evaluating and comparing computational predictions with experimental measurements, amongst other activities. Figures 1 and 3 consist of diagrams found in the literature and visualise some of the steps in different amounts of detail. When combined with findings from this research, the validation process can be represented by a schematic diagram in Figure 22. This diagram in Figure 22 emphasises a number of aspects highlighted in this research and discussed below.

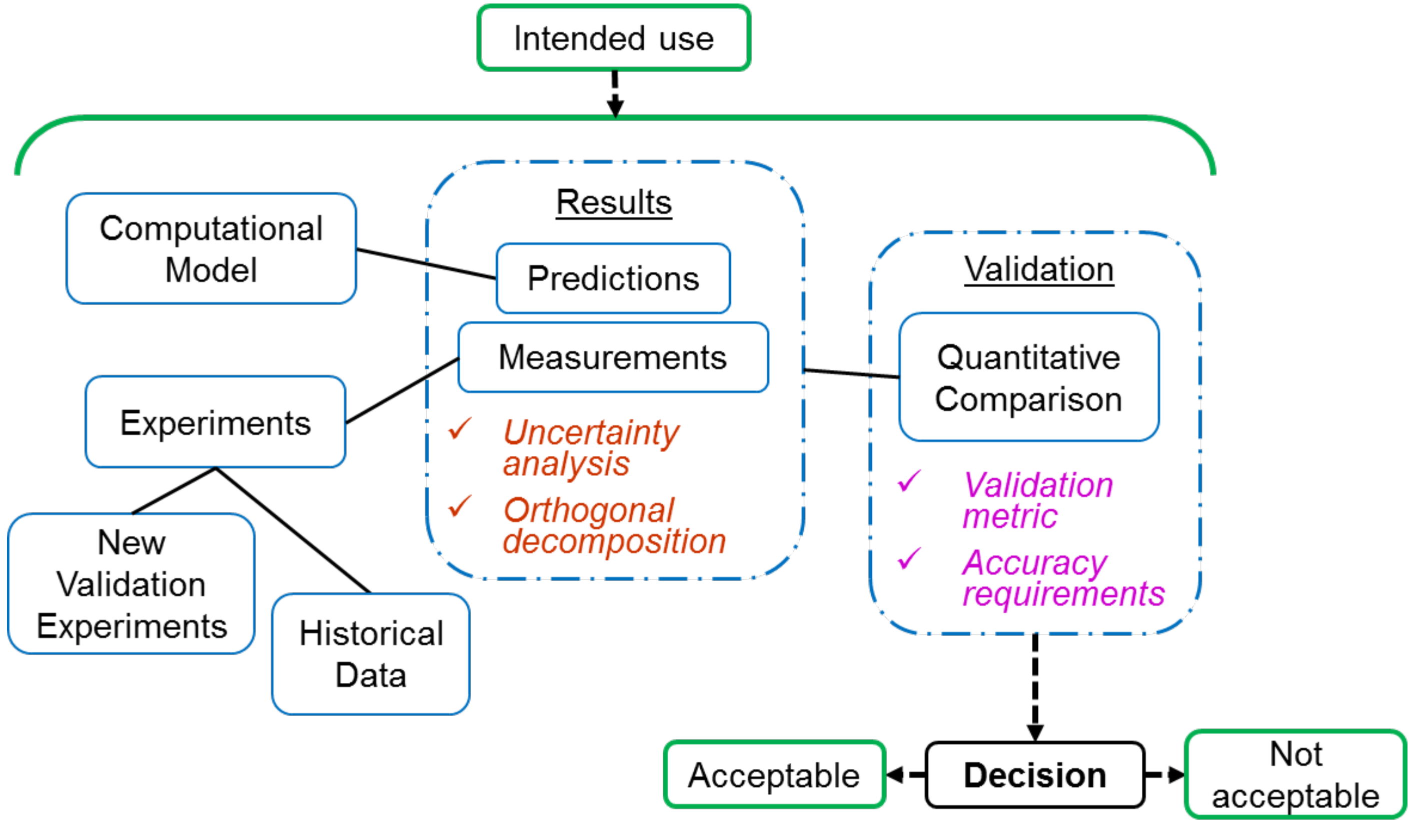


Figure 22: A schematic diagram of the validation process with dashed boxes highlighting some of the novel aspect presented in this thesis. a) Results: evaluation of predicted and measured results with the aid of orthogonal decomposition techniques, and evaluation of associated uncertainties, b) Validation: quantitative comparison of results through the application of the validation metric, taking into consideration all outcomes from Results' box.

Intended use and acquiring data

At the top of the diagram the objective for the validation activity is identified, e.g. the intended use such as predict buckling of an aircraft panel, and is then carried forward to aid the validation process until the decision stage. The left side of the diagram encompasses the development of the computational model and the experiment, which includes identifying parameters and boundary conditions to obtain predicted and measured data that is in line with the objective of the validation process. At this step, previously obtained experimental data, e.g. historical data, can be used, if it satisfies the requirements, or a validation experiment should be performed. This will vary between engineering disciplines. For example in the nuclear industry, due to the nature of operating conditions in nuclear power reactors, only limited access to physical data is available, which leads to sparse data sets and almost exclusive use of historical data for the validation purposes. Whereas in the aerospace industry, physical testing is an integral part of any aircraft certification activity and it is a normal practice to request a validation experiment to assess quality of computational model predictions. However, in the interest of reducing costs and time, the aerospace industry is considering ways to optimise the process of building confidence in computational models with a reduced number of experiments. To move towards a minimum amount of physical testing, validation of evolutionary and revolutionary types of computational model needs be considered, where evolutionary represents applications of known physics and behaviour, whereas revolutionary encompass studies of materials and structures for which only limited knowledge about their mechanical behaviour is available [79].

Predicted and measured results

The next step in Figure 22 is collating and analysing predicted and measured results, as indicated by a dashed box 'Results' on the diagram. At this step any necessary data processing techniques are applied in order to obtain equivalent data sets representing predicted and measured results. Uncertainties associated with predicted and measured results, and with any data processing applied to data also have to be computed. This step is crucial in order to successfully apply a validation metric in the next step and to obtain a meaningful outcome in the end of the validation process. In the current research, field data was used instead of sparse data points typically obtained by strain gauges in solid mechanics. The displacement and strain fields were treated as images, thus allowing the application of decomposition techniques to condense data and produce equivalent data sets, i.e. feature vectors. Recently, Balcaen at el [80] have proposed a technique to process a deformed FE[4] mesh as a DIC[5] grid at different loading steps followed by the smoothing of both images, i.e. predicted and measured fields, to obtain equivalent outputs. However, from the validation perspective this method is less efficient in comparison to methodology described in this thesis and it does not allow a straightforward application of validation metrics to quantify the quality of predictions, because the images still need to be further processed. It may also introduce additional sources of uncertainty.

Quantitative comparison

Once all information about the results is collated, a quantitative comparison is performed. This is highlighted by the second dashed box 'Validation' in the Figure 22. As described in Section 2.2, to proceed with the validation process and

[4]finite element
[5]digital image correlation

establish the quality of a model's predictions, a quantitative comparison should be undertaken with the aid of a validation metric and evaluated in the context of the adequacy requirements. Ideally, an independent party would apply a validation metric to assure an objective evaluation of the predicted results; and having a clear and a robust metric would make this activity much more streamlined and less prone to the influence of subjective judgement. Yet, there is no validation methodology in the current literature that fulfils the desired criteria for the metric, listed in Chapter 3. As a consequence, there is a lack of widely accepted validation metrics across engineering industries. The research in this thesis has concentrated on the development of a validation metric that overcomes the pitfalls of previous methodologies and has the potential to significantly improve implementation of the validation process, in particular the opportunity to utilise field data. The novel methodologies were presented in detail in Chapter 3, with a clearly identified sequence of actions to achieve the desired objective quantitative information necessary for the subsequent decision stage in the validation process. Results presented in Section 4 have successfully demonstrated the applicability of the novel metrics to different case studies in solid mechanics and details are discussed in the following section, Section 5.2.

Decision stage

The ultimate outcome of the validation process is the decision on whether a model's predictions are acceptable for the intended use, as shown in the diagram in Figure 22. In the case of a positive outcome, the model's predictions are used to inform further actions; for example, a successful certification of a machine, a continuous use of a safety critical component or an increase in the current state of the knowledge about a scientific phenomenon. In the opposite case, when a model is found to be unacceptable for the intended use, a request for model or experiment refinement can be requested. This decision will change case by case. For example, the historical data used might be found to be not sufficient and thus a request for a new validation experiment is necessary. In the situation when it is decided to refine the model and associated simulation there are further details that can be extracted from the novel metric and help to decide the next best action.

As described in the Section 3.1, the magnitude of a shape descriptor, or a coefficient, corresponds to the strength of a particular feature in the image or, as presented in this work, in the deformation map, thus individual normalised errors can indicate which features in the predicted deformation map do not correspond to physical measurements. The contribution of individual coefficients was initially studied by Wang at el [62] in the scope of mode shape analysis, and they have demonstrated a correlation between individual dominant coefficients and mode shape patterns of a vibrating disk. Later, Berke at el [67] also employed the analysis of shape descriptors, and demonstrated the correlation between Chebyshev coefficients and mode shapes of a rectangular plate. In the current work, this can be illustrated by looking at the results for the x- and the z-direction displacements in the rubber block case study. For the x-direction displacement, in the top bar chart in Figure 18 a number of dominant coefficients can be distinguished, e.g. #3, 5, 8, 10 and 12, and the visual representation of these Chebyshev coefficients, Figure 5, can be directly related to the shape of the displacement fields in Figure 11. As for the z-direction displacements, as mentioned in Section 4.2 and can be observed in the Figure 11, the predicted and measured displacement fields are evidently different; it was not even possible to apply the same colour scale to visualise both images in the Figure 11 without loosing features in one of them. After applying the validation metric, the probability of the model's prediction being representative of reality was found to be 34.3%, given 19.4% relative uncertainty in the measured data, which can be considered a low probability and a high measurement uncertainty for this particular application in engineering. Depending on the intended use of the model and the accuracy requirements, this outcome can potentially lead to a conclusion that the model is not acceptable. Further investigation of individual normalised errors, e_k in Figure 20 indicates that the largest contribution to the discrepancy comes from the first pair of coefficients, which correspond to the magnitude of the deformation, and the rest of the normalised errors, which correspond to the shape of the deformation map, are an order of magnitude less. Such information about the contribution of a specific error can help to identify the source of the discrepancy. In this case, it can be speculated that it is due to incorrect material properties and thus corresponding model input parameters should be refined. It is important to note here that it was not the intent of this research to build a reliable model that can be used with confidence, and as such no further refinement of the model was undertaken. Similar analysis can be followed for other data fields represented by sets of coefficients obtained from orthogonal decomposition, where visual representation of individual coefficients resembles modes of deformation, e.g. Chebyshev (Figure 5) or Zernike polynomials. This is the great advantage of integrating orthogonal decomposition techniques into the validation process and in combination with the novel validation metric it allows much more useful information to be obtained than previously was possible during the validation activities.

5.2 Validation metrics

Two novel Frequentist metrics were presented and investigated in the main body of this thesis, namely the methodologies based on the Theil's inequality coefficient and the new relative error metric. These were also compared to the methodology suggested by the CEN guidelines [24], the most recent guidelines on the validation in solid mechanics. The robustness of the novel metrics was evaluated based on three case studies (Section 3.4), which represent diverse applications in spatial and temporal domains.

The CEN methodology [24] was originally developed for use in solid mechanics and for use with feature vectors obtained from orthogonal decomposition. It was relatively straightforward to apply and the outcome was easy to interpret, since with the aid of the graphical representation, it was possible to establish whether the model was a reasonable representation of the real world for all three case studies. However it was not possible to determine, or to quantify, the extent of model's predictive quality. Overall, this methodology allows a simple interpretation of the outcome without complex analysis and, in comparison to the desired qualities of the metric mentioned in the Section 3, it does take into consideration uncertainties associated with the experimental data, yet it lacks the quantification of the level of the model's quality with respect to reality.

The Theil's inequality coefficient, Section 3.2, originates from econometrics, and has not been previously applied to field data for the validation of the solid mechanics models. The original formula was successfully modified and applied with the components of feature vectors obtained from decomposing deformation maps. The validation outcomes were successfully obtained for two case studies, the I-beam and the rubber block, and were complementary to the outcomes of the CEN methodology. For example, in the rubber block case study the outcome showed an acceptable agreement for the x- and y-displacement fields, i.e. $T_{C2} < 0.3$, whereas for z-displacement $T_{C2} = 0.58$, which indicates a poor agreement between predicted and measured data fields. The Theil's inequality coefficient proved to be a reliable measure of the model's quality, although it does not consider the uncertainty in the data, which in turn is a disadvantage when referring to the desired criteria of the validation metric.

The new validation metric proposed in Section 3.3 is based on a relative error metric but, through the application of appropriate normalisation of the relative error and the error threshold, the drawbacks of the previous Frequentist approaches are avoided. This means that, unlike previous metrics, the proposed metric is capable of evaluating data with a naturally high variance between the individual values in the data set, including very small values close to zero. It also takes into account uncertainties in the measurement data. The result is a value for the probability that predictions from a model are a reliable representation of the measurements based on the uncertainty in the measurements used in the comparison. This metric was applied to all three case studies and the quantitative validation outcomes agreed with previous findings from other authors, i.e. Lampeas et al [68] and Burguete et al [26]. The new validation metric has been described in generic terms and the case studies illustrated its application to information-rich spatial data fields for a variety of conditions. However, the vectors, S_P and S_M describing the predicted and measured data could be constructed from many types of data, providing that there is correspondence between the components of the vectors. The generic nature of the approach should allow its application in a wide variety of industries such as aerospace, mechanical and nuclear engineering.

There are also other statistical techniques, that can be generally grouped under statistical distance metrics. Some authors have attempted to develop validation metrics based on these methods, however neither provide a desired and informative validation outcome that can be easily interpreted. For example, Ringuest

[81] used Chi-squared statistics, i.e. a comparison of the variance between the predicted and measured results, to evaluate quality of a dynamic control system response, but did not consider the effect of the uncertainties on the measured and predicted results. Xi et al [82] proposed to use Bhattacharya distances, i.e. a measure of similarity between predicted and measured probability distributions that ranges from 0 for no overlap between the two distributions to 1 for a complete overlap, to evaluate the predicted response of a dynamic system, where a reference Bhattacharya distance is compared with the distances representing the discrepancy between predicted and measured results for different scenarios. Nonetheless, the main effort was concentrated on calculating a model bias, i.e. uncertainty in simulated results, and the final outcome did not give a clear indication of the model's quality. Zhao et al [83] applied Mahalanobis distances, i.e. a special case of Bhattacharya distance that measures a distance between a sample and a distribution, to multivariate models, but, similarly to Xi et al [82], no indication of how to interpret the outcome in terms of validation was provided and rather a comparison of competing models was performed. Overall, these statistical techniques have been found useful for some applications, for example uncertainty qualification or model calibration, and can be a useful tool at the early stages of the validation process. However from the perspective of the quantitative comparison for the purpose of establishing quality of the model's predictions, outcomes of these techniques do not currently provide a sufficiently clear statement.

5.2.1 Communicating validation outcomes

The approach to the validation process described in the ASME V&V guide [1] implies that it should be an interactive effort between those responsible for the model and those developing and conducting the experiments required to generate measurement data. However, as mentioned earlier it is unlikely that either group will be responsible for making decisions based on the predictions from the model and hence the credibility of the model becomes a critical factor. Model credibility is the willingness of others to make decisions supported by the predictions from the model [39]. Thus, it is important to present the outcomes from the validation process in a manner that can be readily appreciated by decision-makers who may not be familiar with principles embedded in the model or the approach taken to validation, including the techniques used to acquire the measurement data used in the validation process. The information about the model's predictions obtained by applying the novel validation metric can be expressed in a clear quantitative statement that reflects the complete definition of the validation process. Such a statement includes the following three components:

- the probability of the model's predictions being representative of reality
- for the stated intended use and conditions considered, and
- based on the quality of the measured data defined by its relative uncertainty.

For example, one of the validation outcomes for the rubber block case study can be expressed as: *'there is an 83% probability that the model is representative of reality, when simulating x-direction displacements induced by a 2mm indentation, based on experimental data with a 10% relative uncertainty'*. The first part of the statement, which refers to the probability, presents the quantitative validation outcome, i.e. the degree to which a model is an accurate representation of reality of interest based on the ASME [1] philosophy; the second part of the statement summarises the validation case, e.g. displacement induced by indentation, and the last part states the quality of the measured data, used in the validation process, as an indication of the level of confidence in the validation outcome. All three parts must be communicated together to avoid a misinterpretation of the validation results during the decision stage. A value of 100% for the probability in the statement above indicates that there is no deviation between predicted and measured results larger than the uncertainty in the measured data, whereas values $< 100\%$ can be interpreted as the chance that the relative error of the predictions is less than the measurement uncertainty. A similar statement can be found in other disciplines; for example, in weather forecasting it is common to make a statement about precipitation in an area, e.g. there is an 82% probability of rain tomorrow in Liverpool. The implementation of this type of statement in solid mechanics would represent a significant advance on current practice and could be interpreted relatively straightforwardly by decision-makers. It allows the decision-maker, e.g. customer or stakeholder, to make the final judgement based on the evidence from the validation and their required or desired level of quality.

Brynjarsdottir and O'Hagan [60] have discussed the issue that experiments and simulations both mimic reality so that both have a certain level of approximation which has to be accounted for during a validation process. They concentrated on the concept of model discrepancy, i.e. a difference between the reality and the model output; however it is also important to recognize that the process of experiment design results in a representation of the real-life situation based on our current understanding and that the resultant measurements should not be regarded as the absolute truth. Hence, it is not enough to compare a simulation with an experiment, but also it is necessary to consider the relation of the experiment to reality [84]. As a consequence, some caution needs to be exercised in employing the type of statement expressed above in italics, nevertheless it represents an improvement on current practice in terms of its specificity.

5.2.2 Bayesian updating

As stated in Section 2.2.3, currently, there is no practical validation metric based on Bayesian analysis that is widely used and implemented in industry. Most of the effort is concentrated on evaluating a model's parameters or on improving the model's behaviour. To overcome this drawback, the idea of combining two statistical approaches, Frequentist and Bayesian, has been suggested previously but has not yet led to a successfully applied validation metric [38, 61]. Here, it is proposed that the combined concept can help to explicitly incorporate different sources of evidence as well as errors from experiments and simulations, and effectively assess the validity by increasing or decreasing the confidence in model's predictions. This would require to change the objective of the evaluation to the quality of the predictions for the intended use by using the initial estimate of the validity and updating it when new evidence are available.

Referring to the Bayes' formula in equation 2.5, the initial estimate about a model's validity can be defined as a prior and then updated by the likelihood containing new evidence, for example based on the outcome of the Frequentist metrics, when it becomes available. The concept is summarised in Figure 23, where prior is calculated in *Step 1*, then *Step 2* is introduced when additional experimental data is obtained or more extensive uncertainty analysis has been performed, and in *Step 3* the two sets of outcomes from previous steps are combined through the Bayes formula. The posterior obtained in *Step 3* is an updated value of the model's quality with respect to reality. By combining different sets of experimental data a more informed measure of the validity could be achieved. This could potentially help to avoid the over or under prediction of the model's validity if carefully defined [60].

By incorporating previous findings, the prior can represent the initial belief about the model's performance for the intended use. The value of the probability of a model being representative of reality will always be between 0% and 100%, or can be expressed as between 0 and 1. Such an interval can be defined by a Normal distribution, although assumptions are required on the width of the distribution due to only having information about one parameter, i.e. the model's validity as a mean of the distribution. A Beta distribution would be more appropriate to describe the interval of interest as the distribution is always between 0 and 1. Another advantage of using the Beta distribution is the availability of an analytical solution that does not require a numerical integration, and thus is simple to compute and can help to implement and verify the methodology. The methodology suggested above starts by defining a distribution to describe the validation data. First, the assumption is made that there is no knowledge about the validity of the model and this is expressed by a uniform Beta distribution $\beta(1, 1)$. This is the initial prior. Then, the new validation outcome from a Frequentist metric, i.e. the relative error metric described in this thesis, is added as a likelihood term. Different options to express the data as the likelihood are available and the choice will depend on a specific case; for instance, selecting a Binomial will lead to a posterior as another Beta distribution. As a result, a posterior is obtained, which corresponds to the updated statement of the validity. Every subsequent new piece of information is defined as a likelihood and is used to update the Beta distribution.

Overall, this concept is a more practical solution in comparison to previous work on Bayesian validation methodologies, as it concentrates on evaluating the quality

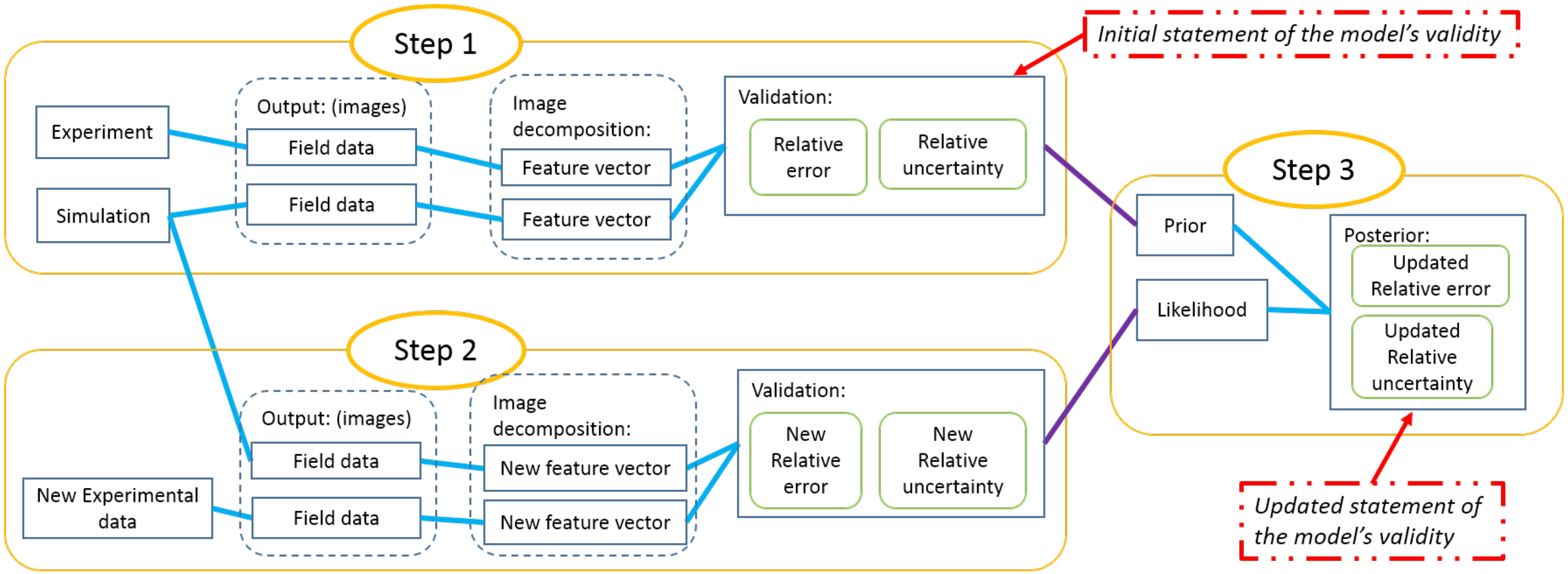


Figure 23: A diagram of a proposed methodology based on Bayesian Updating principles.

of the model's predictions rather than on calibrating it's parameters. It allows the Frequentist and Bayesian approaches to be combined, thus adding together the advantages of the two categories of validation metrics and creating a robust framework. There are some aspects that need further investigation; for example how to integrate and propagate measurement uncertainty within the updating process or how to extract informative outcome from a posterior. However, once these are resolved, such methodology could potentially be applied to sparse data sets and used to combine information from different experimental sources or to evaluate time-varying phenomena.

5.3 The effect of data on validation outcome

5.3.1 Significant coefficients

The image decomposition technique was employed in this research to reduce the dimensionality of the data used for the analysis. It was successfully applied to matrices of data using orthogonal and geometric polynomials, and the corresponding feature vectors were used for validation. It was also shown earlier, specifically in the I-beam case study in Section 3.4, that the data sets can be further condensed by removing coefficients with smaller magnitudes and retaining only the significant coefficients. This approach was further explored to evaluate the sensitivity of the metrics to the amount and magnitude of data.

In the case of the Theil's inequality coefficient, because it is based on the root-mean-square error, the larger absolute differences are given a larger weight. For example, in the Figure 19 it can be observed that the differences are larger for the larger valued coefficients and hence these have a stronger influence on the outcome. This would mean that applying a threshold to only retain significant coefficients, as described in Section 3.4, should not have a strong effect on the outcome of the validation metric, i.e. Theil's inequality coefficient is expected to be insensitive to size of the data set. As for the relative error metric, a slight difference in the outcome is expected, due to the change in weight of individual normalised errors, w_k. Although, errors with smaller values close to zero do not greatly contribute to the sum of all the errors, $\Sigma_{k=1}^{n} e_k$ they could potentially reduce the cumulative proportion below the error threshold, i.e. the probability VM.

The rubber block case study was used for this analysis. In Figure 11 it can be observed that none of the displacement maps have complicated or small features present, whereas a large number of coefficients with low values is present in Figure 18. This suggests that coefficients of the feature vector with very low values can be considered to not represent the main features of the deformation map. Consequently, the original set of coefficients in Table 4 obtained following the CEN guidelines [24] can be easily reduced by following the approach described by Lampeas at el [68] and retaining only significant coefficients. The significant coefficients were obtained from the original sets for the x-, y- and z- displacements on the surface by applying a threshold of $> 1\%$ of the value of the largest coefficient. As a result, it was possible to reduce by two-thirds the number of coefficient required to describe x- and y- data sets, as seen in Table 6.

Comparing validation results in Tables 5 and 6, it is clear that there is no change in the Theil's inequality coefficient and thus it can be concluded that it adequately reflects the difference between the main features of the data sets in question. As expected, the results for the novel relative error metric have

Table 6: Number of retained significant coefficients in feature vectors representing displacement fields for the 28.5x23mm region on the surface of the rubber block (Figure 11) and corresponding validation results.

	Number of significant coefficients	**Theil's Inequality Coefficient**	**VM %**
x-displacement	59	0.24	79.15
y-displacement	69	0.11	54.1
z-displacement	13	0.58	33.6

changed, although comparing the outcomes in Tables 5 and 6 the differences are small, i.e. 3%, 8% and 1% for the three displacement fields respectively. Bearing in mind that the feature vectors were significantly reduced, this change in the outcome can be considered not significant. In turn, this leads to believe that the novel metric is robust against the noise in feature vectors, i.e. discrepancies in coefficients that do not represent the main features of the deformation maps.

5.3.2 Data quantity

Currently, the amount of data used for the comparison is not reflected in the outcome of the validation metric, apart from defining a region of interest for which the validation is performed. It is evident that it is not always possible to generate measurement data at all points in the region of interest, such as when optical access is obstructed, only a small number of point sensors can be employed or the system is inaccessible. In these circumstances the validation metric cannot be calculated for all of the predictions and this shortfall should be reflected in the statement about the outcome of the validation process, i.e. it would be appropriate to state what percentage of the predictions were used in constructing the validation metric and how well the position of these data values covered the region of interest. The interpretation of this additional information will be specific to the intended use of the model and hence, further research is necessary before a conclusive prescription can be provided.

From another point of view, a statistically significant or a minimum number of points are required to apply a statistical method. With respect to the new relative error metric and the validation of field data, a minimum number of points are required to define the cumulative distribution, for example as shown in Figure 17, in order for the metric to yield a reliable outcome. It is impossible to define this minimum number of points for an unknown distribution; however, if it is assumed to be a simplest non-linear curve, i.e. a conic, then at least five points are required according to Pascal's theorem [85] on geometric properties and construction of a conic, assuming there is no uncertainty associated with the points. Hence, it is reasonable to propose that at least five coefficients in the feature vector need to be used to describe the data fields in order for the validation metric to yield reliable results. At the same time, when implementing orthogonal decomposition, care will need to be taken to ensure deformations associated with both the x and y axes are included. In Figure 5, it is evident that Chebyshev coefficients seven to fifteen, and beyond, are variations of the first six coefficients; this also holds true for some of the other polynomials used for decomposition, e.g. Zernike. In combination with an earlier statement this indicates that at least six coefficients in the feature vector are required, i.e. so that ideally $N_{min} = 6$.

With respect to the case studies in Section 3.4, most of the feature vectors consist of a significant number of coefficients, apart from the data representing the displacement field in Region 2 of the I-beam, where the shape of the deformation is relatively simple, as can be seen in Figure 8, and only three coefficients are required to represent the deformation. In such circumstances, i.e. the data field can be described by the first three coefficients, it can be argued that it is not reasonable to perform a statistical comparison and as such to apply a validation metric. As noted in Section 2.2, where the concept and overview of the validation metric was presented, even though it might be desired to have a single metric that could be applied in a wide variety of scenarios, it is important to understand the practical limitations of achieving this objective. The novel relative error metric was successfully applied to all of the case studies, however more care should be taken at the preceding stages of the validation process to assure that data is appropriate for the purpose of the quantitative validation.

5.4 Summary of discussion

The results show that through the application of the new metrics proposed in Chapter 3 it is possible to quantify the quality of the model's predictions and express it in terms of a probability. Overcoming the disadvantages of the previous validation metrics, the novel relative error methodology provides a quantitative outcome that is computed based on the given level of the experimental uncertainty. The metric was successfully applied to field data representing displacement or strain. The quality of the model was expressed as a probability of it's predictions being representative of the real world relative to the specified uncertainty limit. The outcome was found to depend to some extent on the size of the data set used for the validation, i.e. a statistically significant quantity of data is required to apply the metric.

Further research is required and some aspects are discussed in the last chapter, Chapter 7, but at this stage the novel metric fulfils the desired criteria from the Section 2.2 and meets the objectives of this research.

6 Conclusions

Novel validation metrics based on Frequentist approaches have been proposed and successfully applied to fields of data, such as displacement and strain fields. With the aid of the orthogonal decomposition techniques, measured and predicted data fields were condensed and converted into equivalent format, i.e. feature vectors. Using the feature vectors, the novel validation metrics produce a quantitative measure of the quality of the model's predictions and allow to obtain an informed conclusion on whether the model is acceptable for the intended use or not.

Three case studies have demonstrated the use of the novel metrics in computational mechanics for a linear elastic planar static analysis, for a large deformation elastic static analysis and for a non-linear elasto-plastic time-varying analysis. In the latter case, the validation metric was applied as the analysis stepped forward in time. Theil's inequality coefficient was applied to the first two case studies and a quantitative measure of the model's quality was obtained, however the outcome did not take into consideration the uncertainties in the data. The novel relative error metric was applied to all three cases studies, and the outcomes obtained were more quantitative and informative than the previous validation procedures, but qualitatively equivalent.

The advantages of the relative error metric are that it can handle data sets with large amplitude variations in data values as well as close to zero values and that the uncertainty in the measured data can also be included. The metric provides a statement of the probability that a model's predictions are representative of the real world for the intended use, based on the uncertainty in the measurements. Although the case studies included in this thesis relate to structural analysis, the principles illustrated are applicable to validation in a wide range of disciplines where modelling and simulation plays a pivotal role. For example, surface temperature maps from an environmental science studies can be used instead of displacement fields; a measured and predicted temperature maps can be compared using the image decomposition technique and the validation metric to aid understanding of the climate changes.

Evaluating results for the three case studies, the probability of model's predictions being representative of reality was consistent with the previously published results, i.e. high probability when the model was previously found valid, although it gave much more insight into the model's performance. For example, by analysing individual pairs of coefficients in the feature vectors, it is possible to recognise whether the model predicts correctly the magnitude or the shape of the deformation. This information can be used to identify specific aspects of the model to be improved.

In addition to quantifying the quality of model's predictions, a clear validity statement was proposed that can be used to inform decision and policy-makers. It contains the three core sets of information: the quality of a model's predictions relative to the real world, the intended use or loading conditions for which the evaluation was performed and the quality of the measured data, expressed in terms of the measurement uncertainty, used for the validation. Based on this, a set of next actions can be devised; for example whether the quality of the data should be improved or the probability of the model's predictions being representative of reality is satisfactory for the intended use. The statement of the probability of the validity defined here is a significant advancement in the current interpretation of the validation outcome, including of the field data predictions.

Major contributions to knowledge presented in this thesis have been disseminated at a number of international conferences at different stages throughout the project. Furthermore, a journal paper based on the advancements of the novel relative error metric has been submitted and is currently under review.

7 Further work

The novel validation metrics presented in this thesis have allowed a further understanding of the validation outcomes of computational models and were successfully applied to field-data. A significant advance has been made, nevertheless, there is room for further development and some of aspects have been discussed in Chapter 5. Further directions are proposed below.

It was previously mentioned that some applications have data rich sets available for validation, whereas in other circumstances only data sparse sets can be obtained. Currently, most of the validation process is followed with the assumption that a sufficient amount of data is available to perform the validation, and not much effort has been concentrated on considering **the effects of data sparsity** on the validation process and subsequent outcomes. In order to apply a statistical analysis to evaluate the quality of a model's predictions, a statistically significant amount of data is required, so a question to explore further would be what constitutes a significant amount of data for a variety of engineering applications in the industrial context. Different ways of combining information from various sources [86] to combat data sparsity should also be explored. This was briefly discussed in Section 5.2.2, although, at this stage, further research is required to represent the terms of the Bayes' formula (equation 2.5); for example, a likelihood expressed as a normal distribution with the mean as a probability of the model's predictions being representative of reality, VM and the variance as an experimental uncertainty is a potential solution. It would also have to be assured that the quantity of experimental data used in estimating the validation outcome is reflected in the analysis. The topic of data sparsity could also be approached from the perspective of using sufficient data in the validation process.

Another concept that is important to investigate in the future is validation of **true predictions**, as opposed to retrodictions, where the analysis is already known. As described by Patterson and Whelan [79] computational models can be differentiated depending on the level of physical evidence available to test the model's predictions in the real world. This implies that there are some models that predict events or phenomena that have not yet been observed in a physical world. A lifetime prediction of a stable nuclear plant operation [87, 88] or a climate changes [16] fall within this category, and in both circumstances the credibility and the level of acceptance of these models is crucial in accepting the predictive results. Extending the statistical methodologies to consider such cases of true predictions and how this relates to the traditional definition of validation, i.e. comparison with the real world, is an interesting and important topic to investigate in the future.

This thesis has concentrated on data obtained using the full field measurement systems that provide surface information. There are also volumetric imaging systems, for example X-ray Computed Tomography, that allow **volumes of data** to be obtained and are widely used to study the microstructure and the bulk material response of various materials, including ceramic matrix composites [89, 90] and nuclear graphite [91, 92], with the aid of a Digital Volume Correlation, as opposed to a Digital Image Correlation. It would be a substantial advancement to develop decomposition techniques to condense and represent volumes of data, and integrating it with the novel validation metric to allow a wider spectrum of applications.

Nomenclature

A	actual data that represents reality as defined by Theil [49]
e_k	normalised relative error
e_{unc}	error threshold
(i, j)	co-ordinates of general point in image
$I(i, j)$	strain / displacement value in the image at point (i, j)
$\hat{I}(i, j)$	reconstruction of I(i, j)
k	index of the component in the feature vector
M, P	measured and predicted data
N	number of data points in an image
S_M, S_P	feature vector describing data from experiment and model
T_{C1}, T_{C2}	Theil's inequality coefficient
u	average reconstruction residual
u_{cal}	minimum measurement uncertainty
u_{deco}	decomposition uncertainty
u_{exp}	total uncertainty associated with experimental data
VM	probability of model's predictions being representative of reality
w_k	weighted relative error

References

[1] ASME, "Guide for verification and validation in computational solid mechanics," *American Society of Mechanical Engineers*, no. ASME V&V 10-2006, 2006.

[2] W. L. Oberkampf and T. G. Trucano, "Verification and validation benchmarks," *Nuclear Engineering and Design*, vol. 238, no. 3, pp. 716–743, 2008.

[3] NAFEMS, "Quality Assurance," pp. 500–517, 2014. [Online]. Available: http://www.nature.com/doifinder/10.1038/labinvest.2014.37

[4] M. C. Kennedy and A. O'Hagan, "Bayesian calibration of mathematical models' by M. C. Kennedy and A. O'Hagan." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001. [Online]. Available: http://dx.doi.org/10.1111/1467-9868.00294

[5] W. L. Oberkampf, M. Sindir, and A. T. Conlisk, "Guide for the verification and validation of computational fluid dynamics simulations," *American Institute of Aeronautics and Astronautics*, no. AIAA G-077-1998, 1998.

[6] C. Sebastian, E. Hack, and E. A. Patterson, "An approach to the validation of computational solid mechanics models for strain analysis," *The Journal of Strain Analysis for Engineering Design*, vol. 48, no. 1, pp. 36–47, 2013.

[7] G. S. Fishman and P. J. Kiviat, "The statistics of discrete-event simulation," *Simulation*, vol. 10, no. 4, pp. 185–195, 1968.

[8] R. L. Van Horn, "Validation of Simulation Results," *Management Science*, vol. 17, no. 5, pp. 247–258, 1971.

[9] R. G. Sargent, "Validation of simulation models," in *Proceedings of the 11th conference on Winter Simulation-Volume 2*. IEEE Press, 1979, pp. 479–503.

[10] S. I. Gass, "Model accreditation: A rationale and process for determining a numerical rating," *European Journal of Operational Research*, vol. 66, no. 2, pp. 250–258, 1993.

[11] O. Balci, "How to assess the acceptability and credibility of simulation results," in *Proceedings of the 21st conference on Winter simulation*. IEEE Press, 1989, pp. 62–71.

[12] R. E. Shannon, "Tests for the verification and validation of computer simulation models," in *Proceedings of the 13th conference on Winter Simulation -Volume 2*. IEEE Press, 1981, pp. 573–577.

[13] W. L. Oberkampf and T. G. Trucano, "Verification and Validation in Computational Fluid Dynamics," *Progress in Aerospace Sciences*, vol. 38, no. 3, pp. 210–270, 2002.

[14] ASME, "An Illustration of the Concepts of Verification and Validation in Computational Solid Mechanics," *American Society of Mechanical Engineers*, no. ASME V&V 10.1-2012, 2012.

[15] R. G. Sargent, "An introduction to verification and validation of simulation models," *Proceedings of the 2013 Winter Simulation Conference*, pp. 321–327, 2013.

[16] J. Biddle and E. Winsberg, "Value judgement and the estimation of uncertainty in climate modeling," in *New Waves in Philosophy of Science*. Palgrave Macmillan, 2010, pp. 172–197.

[17] S. Groesser and M. Schwaninger, "Contributions to model validation: hierarchy, process, and cessation," *System Dynamics Review*, vol. 28, no. 2, pp. 157–181, 2012.

[18] D. Sornette, A. B. Davis, K. Ide, K. R. Vixie, V. Pisarenko, and J. R. Kamm, "Algorithm for Model Validation: Theory and Applications," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 16, pp. 6562–6567, 2007.

[19] K. T. Hu, A. Urbina, and J. Mullins, "A perspective on the integration of verification and validation into the decision making process," in *Conference Proceedings of the Society for Experimental Mechanics Series*, vol. 3, 2015, pp. 265–273.

[20] E. Winsberg, "Computer Simulation and the Philosophy of Science," *Philosophy Compass*, vol. 4, no. 5, pp. 835–845, 2009.

[21] G. B. Kleindorfer, L. O'Neill, and R. Ganeshan, "Validation in Simulation: Various Positions in the Philosophy of Science," *Management Science*, vol. 44, no. 8, pp. 1087–1099, 1998.

[22] Y. Barlas and S. Carpenter, "Philosophical roots of model validation: two paradigms," *System Dynamics Review*, vol. 6, no. 2, pp. 148–166, 1990.

[23] T. H. Naylor, J. M. Finger, J. L. Mckenney, W. E. Schrank, and C. C. Holt, "Verification of Computer Simulation Models," *Management Science*, vol. 14, no. 2, pp. 92–101, 1967.

[24] CEN, "Validation of computational solid mechanics models," *Comite Europeen de Normalisation (CEN)*, no. CWA16799(2014), 2014.

[25] F. Hild and S. Roux, *Optical Methods for Solid Mechanics: A Full-Field Approach*, P. K. Rastogi and E. Hack, Eds. John Wiley & Sons, 2012.

[26] R. L. Burguete, G. Lampeas, J. E. Mottershead, E. A. Patterson, A. Pipino, T. Siebert, and W. J. Wang, "Analysis of displacement fields from a high-speed impact using shape descriptors," *Journal of Strain Analysis for Engineering Design*, vol. 49, no. 4, pp. 212–223, 2014.

[27] X. Tan, Y. Kang, and E. Patterson, "An experimental study of the contact of a rounded rigid indenter with a soft material block," *Journal of Strain Analysis for Engineering Design*, vol. 49, no. 2, pp. 112–121, 2014.

[28] B. M. Ayyub and G. J. Klir, *Uncertainty Modeling and Analysis in Engineering and the Sciences.* CRC Press, 2006.

[29] A. C. Cullen and C. H. Frey, *Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs.* Springer Science & Business Media, 1999.

[30] W. L. Oberkampf and C. J. Roy, *Verification and Validation in Scientific Computing.* Cambridge University Press, 2010.

[31] C. J. Roy and W. L. Oberkampf, "A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing," *Computer Methods in Applied Mechanics and Engineering*, vol. 200, no. 25-28, pp. 2131–2144, 2011.

[32] G. Upton and I. Cook, *A Dictionary of Statistics*, 2nd ed. Oxford: Oxford University Press, 2008.

[33] E. Hack, X. Lin, E. A. Patterson, and C. M. Sebastian, "A reference material for establishing uncertainties in full-field displacement measurements," *Measurement Science and Technology*, vol. 26, no. 7, p. 075004, 2015.

[34] W. L. Oberkampf and M. F. Barone, “Measures of agreement between computation and experiment: Validation metrics,” *Journal of Computational Physics*, vol. 217, no. 1, pp. 5–36, 2006.

[35] B. M. Rutherford, “Computational modeling issues and methods for the 'regulatory problem' in engineering - Solution to the thermal problem,” *Computer Methods in Applied Mechanics and Engineering*, vol. 197, no. 29-32, pp. 2480–2489, 2008.

[36] Y. Liu, W. Chen, P. Arendt, and H.-Z. Huang, “Toward a better understanding of model validation metrics,” *Journal of Mechanical Design*, vol. 133, no. 7, pp. 071 005–071 005–13, 2011.

[37] J. R. Rowland and W. M. Holmes, “Simulation validation with sparse random data,” *Computers and Electrical Engineering*, vol. 5, no. 1, pp. 37–49, 1978.

[38] J. O. Berger and M. J. Bayarri, “The interplay of bayesian and frequentist analysis,” *Statistical Science*, vol. 19, no. 1, pp. 58–80, 2004.

[39] L. W. Schruben, “Establishing the credibility of simulations,” *Simulation*, vol. 34, pp. 101–105, 1980.

[40] Y. Ling and S. Mahadevan, “Quantitative model validation techniques: New insights,” *Reliability Engineering & System Safety*, vol. 111, pp. 217–231, mar 2013.

[41] C. J. Kat and P. S. Els, “Validation metric based on relative error,” *Mathematical and Computer Modelling of Dynamical Systems*, vol. 18, no. 5, pp. 487–520, 2012.

[42] M. J. Bayarri, R. Paulo, J. O. Berger, J. Sacks, J. A. Cafeo, J. Cavendish,

C. H. Lin, and J. Tu, "A framework for validation of computer models," *Technometrics*, vol. 49, no. 2, pp. 138–154, 2007.

[43] S. Wang, K. L. Tsui, and W. Chen, "Bayesian validation of computer models," *Technometrics*, vol. 51, no. 4, pp. 439–451, 2009.

[44] R. G. Hills, K. J. Dowding, and L. Swiler, "Thermal challenge problem: Summary," *Computer Methods in Applied Mechanics and Engineering*, vol. 197, pp. 2490–2495, 2008.

[45] T. C. Slaba, S. R. Blattnig, B. Reddell, A. Bahadori, R. B. Norman, and F. F. Badavi, "Pion and electromagnetic contribution to dose: Comparisons of HZETRN to Monte Carlo results and ISS data," *Advances in Space Research*, vol. 52, no. 1, pp. 62–78, 2013.

[46] V. Fortunato, C. Galletti, L. Tognotti, and A. Parente, "Influence of modelling and scenario uncertainties on the numerical simulation of a semi-industrial flameless furnace," *Applied Thermal Engineering*, vol. 76, pp. 324–334, feb 2015.

[47] A. Dorobantu, G. J. Balas, and T. T. Georgiou, "Validating Aircraft Models in the Gap Metric," *Journal of Aircraft*, vol. 51, no. 6, pp. 1665–1672, 2014.

[48] H. Theil, *Economic forecasts and policy*, 2nd ed. Amsterdam: North-Holland Publishing Company, 1961.

[49] ——, *Applied economic forecasting*. Amsterdam: North-Holland Publishing Company, 1966.

[50] C. W. J. Granger and P. Newbold, "Some comments on the evaluation of economic forecasts," *Applied Economics*, vol. 5, no. 1, pp. 35–47, 1973.

[51] F. Bliemel, “Theil' s Forecast Accuracy Coefficient: A Clarification,” *Journal of Marketing Research*, vol. X, no. November, pp. 444–446, 1973.

[52] T. Kloek, “Obituary : Henri Theil , 1924 2000,” *Statistica Neerlandica*, vol. 55, no. 3, pp. 263–269, 2001.

[53] D. Vanhoenacker and P. Carrette, “Wavelet transform,” in *Resonance*, vol. 9, no. 3, 2009, pp. 411–416.

[54] S. Kanayath, M. Jayakumar, and G. R. Bindu, “Estimation of longitudinal aerodynamic coefficients of a technology demonstrator aircraft using modified maximum likelihood algorithm,” *2012 IEEE International Conference on Computational Intelligence and Computing Research*, pp. 1–8, 2012.

[55] P. M. Lee, *Bayesian statistics: an introduction*, 4th ed. John Wiley & Sons, 2012.

[56] E. Patelli, Y. Govers, M. Broggi, H. M. Gomes, M. Link, and J. E. Mottershead, “Sensitivity or Bayesian model updating: a comparison of techniques using the DLR AIRMOD test data,” *Archive of Applied Mechanics*, vol. 87, no. 5, pp. 905–925, 2017.

[57] D. Higdon, C. Nakhleh, J. Gattiker, and B. Williams, “A Bayesian calibration approach to the thermal problem,” *Computer Methods in Applied Mechanics and Engineering*, vol. 197, no. 29-32, pp. 2431–2441, 2008.

[58] F. Liu, M. J. Bayarri, J. O. Berger, R. Paulo, and J. Sacks, “A Bayesian analysis of the thermal challenge problem,” *Computer Methods in Applied Mechanics and Engineering*, vol. 197, no. 29-32, pp. 2457–2466, 2008.

[59] R. Rebba and S. Mahadevan, “Computational methods for model reliability assessment,” *Reliability Engineering and System Safety*, vol. 93, no. 8, pp. 1197–1207, 2008.

[60] J. Brynjarsdóttir and A. O'Hagan, "Learning about physical parameters: The importance of model discrepancy," *Inverse Problems*, vol. 30, no. 11, p. 801812, 2014.

[61] A. Raue, C. Kreutz, F. J. Theis, and J. Timmer, "Joining forces of Bayesian and frequentist methodology: a study for inference in the presence of non-identifiability," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1984, dec 2013.

[62] W. Wang, J. E. Mottershead, and C. Mares, "Mode-shape recognition and finite element model updating using the Zernike moment descriptor," *Mechanical Systems and Signal Processing*, vol. 23, no. 7, pp. 2088–2112, oct 2009.

[63] W. Z. Wang, J. E. Mottershead, A. Patki, and E. Patterson, "Construction of Shape Features for the Representation of Full-Field Displacement/Strain Data," *Applied Mechanics and Materials*, vol. 24-25, pp. 365–370, 2010.

[64] A. Patki, W. Wang, J. E. Mottershead, and E. A. Patterson, "Image decomposition as a tool for validating stress analysis models," in *EPJ Web of Conferences*, F. Bremand, Ed., vol. 6, 2010, p. 46005.

[65] J. Flusser, T. Suk, and B. Zitova, *2D and 3D image analysis by moments*. John Wiley & Sons, 2016.

[66] W. Wang and J. E. Mottershead, "Adaptive moment descriptors for full-field strain and displacement measurements," *Journal of Strain Analysis for Engineering Design*, vol. 48, no. 1, pp. 16–35, 2013.

[67] R. B. Berke, C. M. Sebastian, R. Chona, E. A. Patterson, and J. Lambros, "High Temperature Vibratory Response of Hastelloy-X: Stereo-DIC Mea-

surements and Image Decomposition Analysis," *Experimental Mechanics*, vol. 56, no. 2, pp. 231–243, 2016.

[68] G. Lampeas, V. Pasialis, X. Lin, and E. A. Patterson, "On the validation of solid mechanics models using optical measurements and data decomposition," *Simulation Modelling Practice and Theory*, vol. 52, pp. 92–107, 2015.

[69] S. Huazhong, L. Limin, and J.-L. Coatrieux, "Moment-Based Approaches in Imaging. Part 1. Basic Features," *IEEE Engineering in Medicine and Biology Magazine*, vol. 26, no. 5, pp. 70–74, 2007.

[70] R. Mukundan, S. H. Ong, and P. A. Lee, "Image analysis by Tchebichef moments," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1357–1364, 2001.

[71] M. R. Teague, "Image analysis via the general theory of moments," *J. Opt. Soc. Am.*, vol. 70, no. 8, pp. 920–930, 1980.

[72] A. S. Patki and E. A. Patterson, "Decomposing Strain Maps Using Fourier-Zernike Shape Descriptors," *Experimental Mechanics*, vol. 52, no. 8, pp. 1137–1149, 2012.

[73] E. A. Patterson, E. Hack, P. Brailly, R. L. Burguete, Q. Saleem, T. Siebert, R. A. Tomlinson, and M. P. Whelan, "Calibration and evaluation of optical systems for full-field strain measurement," *Optics and Lasers in Engineering*, vol. 45, no. 5, pp. 550–564, 2007.

[74] K. Dvurecenska, S. Graham, E. Patelli, and E. A. Patterson, "A probabilistic metric for the validation of computational models," *Royal Society Open Science*, vol. 5, no. 11, nov 2018.

[75] C. Sebastian and E. A. Patterson, "Calibration of a digital image correlation system," *Experimental Techniques*, vol. 39, no. 1, pp. 21–29, 2015.

[76] W. Wang, J. E. Mottershead, C. M. Sebastian, and E. A. Patterson, "Shape features and finite element model updating from full-field strain data," *International Journal of Solids and Structures*, vol. 48, no. 11-12, pp. 1644–1657, 2011.

[77] E. Hack, G. Lampeas, and E. A. Patterson, "An evaluation of a protocol for the validation of computational solid mechanics models," *Journal of Strain Analysis for Engineering Design*, vol. 51, no. 1, pp. 5–13, 2016.

[78] L. Felipe-Sesé, P. Siegmann, F. A. Díaz, and E. A. Patterson, "Simultaneous in-and-out-of-plane displacement measurements using fringe projection and digital image correlation," *Optics and Lasers in Engineering*, vol. 52, no. 1, pp. 66–74, 2014.

[79] E. A. Patterson and M. P. Whelan, "A framework to establish credibility of computational models in biology," *Progress in Biophysics and Molecular Biology*, vol. 129, pp. 13–19, 2017.

[80] R. Balcaen, L. Wittevrongel, P. L. Reu, P. Lava, and D. Debruyne, "Stereo-DIC calibration and speckle image generator based on FE formulations," *Experimental Mechanics*, vol. 57, no. 5, pp. 703–718, 2017.

[81] J. L. Ringuest, "A chi-square statistic for validating simulation-generated responses," *Computers and Operations Research*, vol. 13, no. 4, pp. 379–385, 1986.

[82] Z. Xi, Y. Fu, and R.-J. Yang, "A model validation approach for various design configurations with insufficient experimental data for model accuracy check," *SAE Technical Papers*, 2012.

[83] L. Zhao, Z. Lu, W. Yun, and W. Wang, "Validation metric based on Maha-

lanobis distance for models with multiple correlated responses," *Reliability Engineering and System Safety*, vol. 159, pp. 80–89, 2017.

[84] H. B. Henninger, S. P. Reese, A. E. Anderson, and J. A. Weiss, "Validation of computational models in biomechanics," *J. Engineering in Medicine*, vol. 224, no. 7, pp. 801–812, 2010.

[85] W. Hamilton, "On a proof of Pascal's theorem by means of quaternions; and on some other connected subjects," *Proceedings of the Royal Irish Academy*, vol. 3, pp. 273–292, 1847.

[86] F. Hemez, H. S. Atamturktur, and C. Unal, "Defining predictive maturity for validated numerical simulations," *Computers and Structures*, vol. 88, no. 7-8, pp. 497–505, 2010.

[87] C. Unal, B. Williams, F. Hemez, S. H. Atamturktur, and P. McClure, "Improved best estimate plus uncertainty methodology, including advanced validation concepts, to license evolving nuclear reactors," *Nuclear Engineering and Design*, vol. 241, no. 5, pp. 1813–1833, 2011.

[88] I. Aho-Mantila, O. Cronvall, U. Ehrnstén, H. Keinänen, R. Rintamaa, A. Saarenheimo, K. Simola, and E. Vesikari, "Lifetime prediction techniques for nuclear power plant systems," in *Nuclear Corrosion Science and Engineering*, D. Feron, Ed. Woodhead Publishing Limited, 2012, ch. 14, pp. 449–470.

[89] T. Whitlow, E. Jones, and C. Przybyla, "In-situ damage monitoring of a SiC/SiC ceramic matrix composite using acoustic emission and digital image correlation," *Composite Structures*, vol. 158, pp. 245–251, 2016.

[90] L. Saucedo-Mora, T. Lowe, S. Zhao, P. D. Lee, P. M. Mummery, and T. J.

Marrow, "In situ observation of mechanical damage within a SiC-SiC ceramic matrix composite," *Journal of Nuclear Materials*, vol. 481, pp. 13–23, 2016.

[91] M. Mostafavi, S. A. McDonald, P. M. Mummery, and T. J. Marrow, "Observation and quantification of three-dimensional crack propagation in polygranular graphite," *Engineering Fracture Mechanics*, vol. 110, pp. 410–420, 2013.

[92] T. J. Marrow, D. Liu, S. M. Barhli, L. Saucedo Mora, Y. Vertyagina, D. M. Collins, C. Reinhard, S. Kabra, P. E. Flewitt, and D. J. Smith, "In situ measurement of the strains within a mechanically loaded polygranular graphite," *Carbon*, vol. 96, pp. 285–302, 2016.