Metrics and Guidelines for Rotorcraft Modeling Accuracy and their Relation to Model Uncertainty: A State-of-the Practice Survey

Marilena D. Pavel[1], Anna Aba[1,2], Pierluigi Capone[2]
[1]Delft University of Technology, The Netherlands
[2] ZURICH University of Applied Sciences

Abstract

The aim of this paper is to provide a survey on the validation metrics and criteria employed for validation of rotorcraft modeling accuracy, extending the present approaches to the uncertainty quantification (UQ) domain of validation. In this sense, the paper will review the current state-of-the-art metrics belonging to 1) the root mean square error method which are expressed as cost functions in frequency and time domain and 2) their enhancement towards the chi-square statistics consideration. The paper will extend afterwards the problem of validation towards uncertainty quantification (UQ) area with its distinction between aleatory and epistemic uncertainties. In UQ the emphasis is not on validation metrics but on the methods to validate the model, the choice of a metric depending on the problem under investigation. A simple example will be given for rotorcraft showing that the problem of flight controller modeling can become nondeterministic using different formulations for the equations of motion. A key message of the paper is that using UQ in the future can enhance our understanding of the simulation model fidelity.

1. INTRODUCTION

Computer-based mathematical models and simulations are developed and deployed to support aircraft/rotorcraft system analysis, design, test and evaluation, acquisition, pilot training and education and more recently certification. Examples of simulations goals to be reached are: ensure stability, design control systems, ensure aeroelastic stability, determine the fatigue as well as extreme loads in the critical components, study the failure in critical parts that could have a catastrophic effect upon the rotorcraft, etc. Simulations are run for a multitude of operating situations such as: normal or extreme windspeeds including various levels of turbulence, different control modes investigated during the design process, critical failure modes, parallel candidate design configurations as shown relevant by conceptual studies, critical failure modes, etc. The simulations are based on aerodynamic and dynamic models capable to predict the aircraft/rotorcraft behavior in different operating conditions. Clearly, the simulation-phase is in general a time-consuming stage. The models used need to be a representation of the real world, simplified, but designed such that the essential features that one is interested in, are focused on the model, and are correctly represented in the model. Also, the models used should be as simple as possible, as far as is consistent with the required accuracy and the specific case considered. This is especially true in the case of flight simulators. Due to their wide field of applications, it is essential that the simulation models used correctly represents the actual behavior of the aircraft.

Modeling and Simulation (M&S) of rotorcraft (helicopter, compound helicopter, tiltrotors, vertical take-off and landing (VTOL) aircraft configurations) dynamic response is known to be challenging. Creating a model that realistically predicts a rotorcraft characterized by its highly coupled, unstable, multi-input multi-output dynamic system is a difficult task and usually involves many scientists and years of research. Therefore, it is not surprising that ensuring modeling accuracy is a high cost and difficult problem. Part of the problem is that there is no guarantee that satisfying a given set of modeling accuracy criteria will result in a desired level of performance. The accepted way of assessing the model accuracy is to perform analysis-test correlation where the target model behavior is compared with corresponding flight test data. Verification and Validation (V&V) and Verification, Validation and Accreditation (VV&A) requirements, standards and tools have been developed to correlate modeling accuracy criteria with system reliability. A measure for M&S state complexity can be derived as the total numbers of inputs, outputs, and state variables spanning the whole state space of an M&S. This

Paper no. ERF2023_0078 presented at the 49th European Rotorcraft Forum, Sept. 4-7, 2023, Bückeburg, Germany

results in new challenges for model designs and simulation implementations as well as quite often to tremendous amount of data.

The aim of this paper is to provide a survey on the modelling validation metrics and criteria that rotorcraft M&S community can employ presently when developing rotorcraft models, extending the present approaches to the uncertainty quantification (UQ) domain of validation. In this sense, the paper aligns with the current need for quantification of uncertainties, especially in dynamic modeling of aircraft. Integrating uncertainties into M&S effort is a key element to increase aircraft safety.

2. REVIEW OF TERMINOLOGY AND PROCESSES IN MODELING AND SIMULATION

Modeling and Simulation (M&S) is an emerging science which, according to Wikipedia, it is based on developments in diverse computer science areas as well as influenced by developments in Systems Theory, Systems Engineering, Statistics, Probabilities, Software Engineering, Artificial Intelligence, and more. "This foundation is as diverse as that of engineering management and brings elements of art, engineering, and science together in a complex and unique way that requires domain experts to enable appropriate decisions… The diversity and application-oriented nature of this new discipline sometimes result in the challenge, that the supported application domains themselves already have vocabularies in place that are not necessarily aligned between disjunctive domains. " (https://en.wikipedia.org/wiki/Modeling_and_simulation) Indeed, M&S is a complex science with many applications. For engineers models are the core of a simulation system, the quantitative analyses of the phenomena occurring in many engineering applications being based on mathematical models which are then turned into computer codes for simulation.

Each model provide a representation of a real system which is dependent on several hypotheses and parameters. Model can be deterministic (non-probabilistic) (e.g. Newton's dynamic laws, Lagrange mechanics, Hamiltonian dynamics etc.) or probabilistic (stochastic) (e.g. the Poisson model for describing the occurrence of failures in a structural model). Nevertheless, in the world of M&S, two "stark realities" exist: on the one hand all models are approximations of their target phenomena; on the other hand, uncertainties exist, "due both to ignorance and poorly understood inherent processes, that corrupt our ability to state with certainty the level of approximation that is present in them." [Ghanem et. Al. (2008)]. This is why recommended practices have been established to fulfil the fundamentals of credibility of a model. These practices belong to the concept of VV&A and have been established by the IEEE 1278.4 standard "Recommended Practice Guide for Distributed Interactive Simulation – Verification, Validation and Accreditation" [IEEE, (1997)]. Later such practices have been applied to computational fluid dynamics [Roache (1998)] and simulation fidelity [Roza(2004)]. In the SISO guide GM-VV [SISO (2012-2013)] the following definitions are established [Lehmann & Wang (2019)]: **Verification:** The process of providing evidence justifying the M&S system's correctness. **Validation:** The process of providing evidence justifying the M&S system's validity. Validity is the property of an M&S system's representation of the simuland to correspond sufficiently enough with the referent for the intended use. The simuland is the system or process that is simulated by a simulation while the referent is "*… the codified body of knowledge about a thing being simulated*" (IEEE STD 1516.4, 2007). **Acceptance:** The process that ascertains whether an M&S system is fit for its intended use. **Accreditation:** The official certification that a model, simulation, or federation of models and simulations are acceptable for use for a specific purpose. **Tailoring:** The modification of V&V processes, V&V organization, and V&V products to fit agreed risks, resources, and implementation constraints (SISO-Guide-GM-VV, 2012–2013).

More succinctly said [Roache (1998)]: **Verification:** The process of determining that a model implementation accurately represents the developer's conceptual description of the model and the solution to the model. **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.

Finally, even more succinctly [Boehm (1981)] and [Blottner (1990)]: **Verification** means "*solving the equation right*" To solve "the equations right" means that the numerical model meets the requirements of consistency, stability, and convergence. **Validation** means "*solving the right equations*". To solve "the right equations" means that the selected approach is coherent with the physics of the problem that is going to be modelled. Therefore, verification is a condition on the numerical scheme and validation is a condition on its application to real-world problems [Roache (1998)].

In Computational modelling often the term of adequacy comes in place of accuracy [David et. Al. (2017)]: **Verification** for computational modelling is defined as checking the adequacy among conceptual models and computational models. Herein the meaning of adequately means that the inputs, outputs, and the mechanisms post-computationally modelled from the executable computational model are consistent with the ones specified through the pre-computational models. **Validation** is defined as ensuring that both conceptual and computational models are adequate representations of the target. The term “adequate” in this sense may stand for several epistemological perspectives.

Concluding, M&S with its VV&A procedures truly belongs to complex systems science where one needs to understand the processes on how one can scientifically approach the study of complex systems—physical, biological, and social [Siegenfeld & Bar-Jam (2020)].

2.1. Validation for flight simulation: Is this only about the model being right or wrong?

Moving further from the field of M&S to its application of flight simulation and flight simulators (FSTD – flight simulation training device), the procedures of M&S can be seen as a concept of fidelity. Fidelity can be defined then as the quality of the simulator to replicate the real world. Whether a simulator is fit-for-purpose, or what level of complexity is needed, and which aspects are most important to let the pilot give the experience of realistic flight, is subject to the studies of simulator fidelity. Simulation/simulator fidelity is a science in itself and the goal of this paper is not to concentrate on this topic. A good review on the simulator fidelity characteristics has been done in [NATO AVT-296 (2021)] work. The goal of the paper is to concentrate more on the validation area of flight simulation. It is anyway worth mentioning that defining simulation fidelity is not an easy task. At least eight different definitions can be employed when discussing on simulation fidelity [Liu & Macchiarella (2008)]. Also, “*unfortunately simulation users often fail to be explicit about the fidelity required for simulation results to fully support the simulation’s intended use, and that accuracy is required for proper perspective on the uncertainty of simulation results. Appropriate acceptability criteria indicate what that accuracy is. Uncertainty and fidelity are inversely related. The more fidelity is required, the less uncertainty is acceptable.* “ [Pace(2012)]

In flight simulation. the most straightforward description of simulation fidelity is the one related to the system’s dynamics. Fidelity (in the viewpoint of verification and validation) of vehicle physics is achieved in a straight-forward manner – by directly comparing simulation predictions with acquired flight test data from the real aircraft. Herein, verification can be related to the question whether the model do what it intends to do. As opposed to verification, validation asks whether the model is correct [Mihram, (1972)]. The problem of validation might seem in this way relatively straightforward: one needs to take the prediction of the model, compare it to the flight test data, see whether and how they match and draw conclusions on the model fidelity. The answer could be that the model is good if it a little off the data or that the model is not good if the model is way off the data. In this sense validation is about checking whether a model is right or wrong. However, thinking of George Box aphorism that “All models are wrong, but some are useful” [Box (1976)], it follows that all models of any physical reality are wrong, at least in the narrow sense that they can never be perfect. This aphorism is a simple way to communicate the praxis of modelling: the usefulness of a model is not measured by the accuracy of representation but how well it supports the generation, testing and refinement of hypotheses. Therefore, validation is about assessing the accuracy of the model and assessing whether a model is good enough for some intended purpose. For a deterministic model as often is the case for the physical model , validation can be a fairly straightforward process: 1) the model produces a point estimate for its prediction about some quantity of interest (QoI) (also called system response quantity (SQR); 2) this prediction can be compared against one or more measurements about that quantity; and 3) finally the difference(s) can be understood as a measure of how accurate is the model. “*A model could be consistently inaccurate and yet close enough for its purpose. Likewise, even a highly accurate model might not be good enough if it is needed for some delicate, high-consequence decisions*.” [Ferson et. Al. (2008)] Herein Albert Einstein’s’ aphorism comes into play “Everything should be made as simple as possible, but no simpler” warning against too much simplicity in a physical model.

However, nowadays, the validation problem gets more complicated because: 1) one can start generating serious simulation models based on entire distributions rather than point estimates as their predictions of system response quantities. These distributions can characterize the stochastic variability of the quantities of interest (QoIs)and the model uncertainty made in predicting these quantities; 2) the data available for validation may not be directly relevant to the prediction of the QoI. In particular, the collected data might be under conditions that are similar, but not identical, to those for which a prediction is desired. Therefore, in order to allow the characterization of the model validity one is needs to extrapolate the validity of the model. "*In a strong sense, any forecast about future or general predictive capability is necessarily an extrapolation of some kind, even if directly relevant data are available, because such a forecast is always about hypothetical data that might be observed, rather than merely a summary of the consistency with past observations*." [Ferson et. Al (2008)]. A key insight for flight physics validation becomes then the point of not focusing on the model as an outcome, but rather considering the modeling process and simulated model predictions as "ways of thinking" about complex nonlinear dynamical systems. Looking at the flight physics from an uncertainty point of view means to have strategies to address several questions [Ferson et. Al (2008)]: 1) What if the prediction is a probability distribution rather than a point value? 2) What if there is experiment-to-experiment variability in the measured data? 3) What if there are statistical trends in the experimental data? 4) What if there are multiple predictions about different outputs to be assessed? 5) What if the predictions from the model are extremely expensive to compute? 6) What if the data were collected under conditions other than the intended application? 7) What if there is non-negligible experimental measurement uncertainty in the data?

Herein the validation of a model becomes a problem of uncertainty characterization and propagation to be discussed in this paper.

3. A PRACTICAL APPROACH TO MODEL VALIDATION FOR FLIGHT SIMULATION AND ITS METRICS

The most practical way to validate the model fidelity of rotorcraft flight dynamics model has been established by Tischler & Remple (2012) and later has been extensively applied by the NATO's STO AVT-296 Group led by Dr. Mark B. Tischler on "Rotorcraft Flight Simulation Model Fidelity Improvement and Assessment". [NATO, 2021)]. This group comprised research from 9 nations and 18 different organizations. The objective of the group was to assess the simulation model fidelity of different baseline models for rotorcraft and evaluate different sys id. methods for model improvement. The report of this group is interesting not only due to the different sys id methods used to improve the accuracy of the simulation model but also due to the metrics and criteria reviewed on assessing model fidelity. Next a short overview and the integration of the metrics used into statistical context will be given.

3.1. Root mean square (RMS) Method

For quantifying the accuracy of a flight dynamics model the standard metric to be used relates to the Root Mean Square Error (RMS or RMSE in the statistics literature) metric. RMSE is an established, widely accepted, and easy performance measure of the model accuracy/fidelity. It measures the **average difference** between model's predicted values and the actual values. Mathematically, it is the standard deviation of the residuals. Given a set *i* ($i = 1 \dots n$) data one can differentiate between of the y_i **measured data** (also called **real data, observation data)** which corresponds to the **flight test data** in this paper and the $\hat{y}_i$**simulated data (predicted data, estimated data). The residual** is defined as the difference between the real data y_i and the predicted data $\hat{y}_i$, i.e. [Chaterjee, (2011)], see :

$$\mathrm{Re}\,sidual = y_i - \hat{y}_i = y_{data,i} - y_i \qquad (1)$$

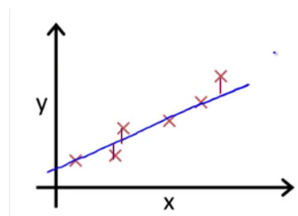


Figure 1: Calculation residuals

To evaluate the model accuracy, four of the more popular direct error measures are: mean squared error (MSE)) and its variants i.e. root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) defined as followings [Wang, 2011)]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \qquad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{3}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{4}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{5}$$

Minimizing these measures is usually the most essential criterion in comparing simulation models. These measures are most frequently used due to their mathematical convenience. For each of these measures, a smaller value indicates higher prediction accuracy. The RSME formula is often used for the assessment of the simulation fidelity in flight mechanics models. The RSME is essentially the standard deviation formula (this looks logic as the root mean square error is the standard deviation of the residuals). It measures the scatter of the observed values around the predicted values. In other words, RMSE quantifies how dispersed these residuals are, revealing how tightly the observed data clusters around the predicted values. When the number of measured data n differs from the number of predicted data p, the RSME formula for a sample becomes:

$$RMSE = \sqrt{\frac{1}{n \cdot p}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{6}$$

Finding the RMSE involves therefore: 1) calculating the residual for each observation (y – ŷ), 2) squaring it, 3) sum all the squared residuals, 4) divide that sum by the product of the number of observations and the number of estimated data in order to find the average squared residual and 5) finally, take the square root to find the RMSE.

Like any statistical measure, the RMSE has its strengths and weaknesses [Frost (2020)]. As advantage, one can mention the fact that RMSE has an intuitive Interpretation. It is a simple metric providing a straightforward interpretation of the model's overall error, making it accessible to engineers without a solid statistical background. It is an absolute measure of the average distance that the data points fall from the predicted values using the units of the dependent variable. Therefore, RMSE assesses prediction precision easily and directly. As disadvantages one can mention the fact that: 1) RMSE is Sensitive to Outliers. The squaring process in the root mean square error causes it to give a disproportionately higher weight to larger errors, making it more sensitive to outliers') RMSE is sensitive to Overfitting. The metric decreases every time one adds an independent variable to the model, even when it is just a chance correlation between variables. It never increases when adding a variable to the model. AS the goal is to have a small RMSE, one can create the appearance of a better model simply by adding more variables, even when they don't make sense. 3) RMSE is sensitive to scale, The metric is sensitive to the scale of the dependent variable. Consequently, interpreting RMSE it requires additional knowledge about the dependent variable (i.e. the variable that is measured by the user to see if this depends on the change in the independent variable.) and its scale. In some cases, the RMSE might not be comparable across different datasets or units of measurement.

RMSE was extensively applied in rotorcraft system identification especially by the NATO's STO AVT-296 Group. More precisely, **cost functions based on RMSE** are proposed in Chap 4 of the NATO report both in time and frequency domain and used in Chaps. 5&7 for quantitative model validation. The metrics introduced in this report are summarized as follows.

Table 1: Cost functions metrics and their guidelines for rotorcraft model fidelity validation

	Time domain	Frequency domain
RMSE Metric for model validation	$J_{RMS} = \sqrt{\frac{1}{n \cdot p}\sum_{i=1}^{n}\left(y_{data,i} - y_i\right)^T\left(y_{data,i} - y_i\right)}$ (7) $\boldsymbol{y}_{\text{data}}$= time-history measurement vector from the flight data $\boldsymbol{y}$ = simulation model time-history response vector	For a single-input/single-output (SISO) system: $J_{SISO} = \frac{20}{n_\omega}\sum_{\omega_1}^{\omega_{n\omega}} W_\gamma \left[W_g\left(\lvert\hat{T}_c\rvert - \lvert T\rvert\right) + W_p\left(\langle\hat{T}_c - \langle T\right)\right]^2$ (12) T=frequency response simulation model $\hat{T}_c$=frequency response flight-test *data*

<table>
<tr>
<td></td>
<td>n = number of time-history points in the model (training data)
p = number of measurements

Units of the measurement vector (SI units):

deg, deg/sec, m/sec, m/sec^2 (8)

Units of the measurement vector (English units):

deg, deg/sec, ft/sec, ft/s^2 (9)

Eq. (7) can contain a Weight matrix W allowing the weighting of the output errors (residuals):

$$J_{RMS} = \sqrt{\frac{1}{n \cdot p} \sum_{i=1}^{n} \left(y_{data,i} - y_i \right)^T W \left(y_{data,i} - y_i \right)}$$
(10)</td>
<td>$|\;|$□□ magnitude (dB) at each frequency ω
∠□□phase (deg) at each frequency ω

n_ω□□number of frequency points (typically selected as $n_\omega = 20$

ω_1 and ω_{n_ω}□□ starting and ending frequencies of fidelity assessment (typically covering 1–2 decades).
$\mathrm{Mag}_{\mathrm{err}}(f) = (|T| - |\hat{T}_\mathrm{c}|)$ magnitude error (dB) response,

$\mathrm{Phase}_{\mathrm{err}}(f) = (\angle T - \angle \hat{T}_\mathrm{c})$phase error (deg) response.

Weightening factors are according to USAF MIL-STD-1797B (2006) meaning that a 1-dB magnitude error is comparable with a 7.57-deg phase error.

$W_\mathrm{g} = 1.0$ and $W_\mathrm{p} = 0.01745$ (13)

A function W_γ can be introduced to weigthen more heavily when the flight data are reliable:

$$W_\gamma = \left[1.58 \left(1 - e^{-\gamma_{xy}^2} \right) \right]^2 \quad (14)$$

γ_{xy}^2 coherence function at each frequency ω.
A coherence of $\gamma_{xy}^2 = 0.6$, $W_\gamma$reduces the weight on the squared errors by 50%.

J_SISO was generalized to multi-input/multi-output (MIMO):

$$J_{MIMO} = \sum_{l=1}^{n_{TF}} \left\{ \frac{20}{n_\omega} \sum_{\omega_1}^{\omega_{n\omega}} W_\gamma \left[W_g \left(|\hat{T}_c| - |T| \right) + W_p \left(\langle \hat{T}_c - \langle T \right) \right]^2 \right\}$$
(15)

T_l, l = 1, 2, 3, …, n_TF frequency-response pairs in the frequency-range (ω_1,…, ω_{n_ω})

J_MIMO accuracy is best characterized by an average overall cost function or integrated cost function:

$$J_{ave} = \frac{J}{n_{TF}} = \frac{1}{n_{TF}} \sum_{l=1}^{n_{TF}} \left\{ \frac{20}{n_\omega} \sum_{\omega_1}^{\omega_{n\omega}} W_\gamma \left[W_g \left(|\hat{T}_c| - |T| \right) + W_p \left(\langle \hat{T}_c - \langle T \right) \right]^2 \right\}$$
(16)</td>
</tr>
<tr>
<td>Guideline RMSE metric</td>
<td>Based on experience, J_RMS values are in the range of:

$Guideline: J_\mathrm{RMS} \leq 1.0 \text{ to } 2.0$ (11)

Reflecting acceptable level of accuracy (adequate fidelity).

For good accuracy J_rms should be smaller than 1.</td>
<td>Based on experience, J_SISO values are in the range of:

$Guideline: J_{SISO} \leq 100$ (17)

reflecting acceptable level of accuracy for rotorcraft flight-dynamics modelling. J_SISO smaller than 50 produces a model nearly indistinguishable from the flight data in the frequency/time domain.

For the MIMO system:

$Guideline: J_{ave} \leq 100$ (17)</td>
</tr>
</table>

		reflects acceptable level of accuracy. $Guideline: J_{ave\ off-axis} \leq 150\ to\ 200$ (18) reflects acceptable accuracy flight-dynamics modelling. Some of the individual cost functions, especially for the **off-axis** responses, can use the guideline of without resulting in a noticeable loss of overall predictive accuracy.

NATO has extensively applied accuracy J_{rms} and J_{ave} in the project for model fidelity (Pavel. et. al. (2021)]. However, care should be taken and do not consider these metrics as an absolute statement on model validity. For example, Scepanovic and Dorinc (2021) applied these metrics in the case of an Airbus EC-135 helicopter in hover and forward flight. Their example showed that even a simple constant offset of 1 m/s in model velocity w.r.t the flight test data (this offset corresponded to the conversion of m/s to ft⁄s) and a perfect trend tracking, would lead to $J_{\text{rms}} = 3.28$ (which is well above acceptable model quality). Therefore one cannot use J_{rms} as a definitive statement about the predictive ability of a simulation model. The authors accompanied model accuracy by the Qualification Test Guide (QTG) performance standard [EASA, 2012]. This standard is used by simulator manufacturers and certification authorities to validate a simulator and define the acceptable tolerances for simulator model fidelity in the form of QTGs.

3.2. From cost functions guidelines to statistics guidelines: Chi Square Statistics considerations

Instead of using guidelines based on engineering experience to develop accuracy criteria for simulation modeling, one can think of applying statistical tools to measure model predictability. If this is not the case, a statement on the validity of the model needs to be calculated, the so-called “**goodness-of-fit**". In statistics this is called the “**test of hypothesis, hypothesis testing or significance testing**”. Hypothesis testing is a well-developed statistical method of deciding which of two contradictory claims about a model, or a parameter, is correct. In hypothesis testing the validation assessment is formulated as a ‘‘decision problem’’ to determine whether or not the computational model is consistent with the experimental data. The level of consistency between the model and the experiment is stated as a probability, based on what has been observed in comparing system response quantities (SRQs) from the model and the experiment. In other words, hypothesis testing it tells how probable it is that the data have occurred if the assumed model were true. Most often **least-squares statistics** are applied, e.g., one of the most well-known **χ^2 methods.** The classical statistics is based on the χ^2 density which is defined as

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(y_i - \hat{y}_i\right)^2}{\sigma_i^2} \tag{19}$$

where $y_i - \hat{y}_i$ have the same significance as in Eq. (1) and they are normally (Gauss) distributed with the known variance σ_i^2 (the variance in Gauss distribution is an indication of the spread of the data). In the χ^2 statistics one can perform **Chi-squared statistic test** (pronounced with a hard ch as in “character”) [Snedecor and Cochran, (1989)] in order to determine the model confidence. A statistic is a measure of the consistency of the model and the experimental data.

The Group for Aeronautical Research and Technology in Europe (GARTEUR) GARTEUR HC/AG-06 on “*Mathematical Modelling for the Prediction of Helicopter Flying Qualities*” [Padfield et al. (1996)] and GARTEUR HC/AG-09 on “*Mathematical Modelling For The Prediction Of Helicopter Flying Qualities*” [Haverdings, (2000)] (https://garteur.org/technical-reports/) explored the Chi-squared statistics boundaries for judging the goodness of simulation models as compared to the flight test data. More exactly, *ValCrit-T* and *ValCrit-F* metrics in the time domain and respectively frequency domain were developed within GARTEUR HC/AG-09 to detect statistically the differences between simulation model and flight test data that might be missed by engineering judgment alone. Looking at Eq. (19), one can define the χ^2 density of the simulation model. GARTEUR HC/AG-06 defined a the χ^2 density when the simulation model and data are known in the time-domain as follows [Haverdings, (2000)]:

Table 2: Chi-statistics metrics and their guidelines for rotorcraft model fidelity validation

<table>
<tr><th></th><th>Time domain</th><th>Frequency domain</th></tr>
<tr>
<td>Chi-statistics metric for model validation</td>
<td>

$$ValCrit-T = J_T = \frac{1}{n \cdot p}\sum_{i=1}^{N}\left(y_{data,i}-y_i\right)^T Y^{-1}\left(y_{data,i}-y_i\right) \quad (20)$$

$Y = diag\left[\sigma_{y,i}^2\right]$ a weighting matrix representing the error covariance.

$\boldsymbol{y}_{\text{data}}$= time-history measurement vector from the flight data
$\boldsymbol{y}$ = simulation model time-history response vector
n = number of time-history points in the model (training data)
p = number of measurements

$J_T pn$ is the simulation error statistic to be tested against a Chi-squared χ^2_{pn} distribution. The p-value of this function is the probability under the null hypothesis of getting the observed values of the test statistic or something even larger. It can be found using Chi-square tables.

<table>
<tr><th></th><th>P</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th></tr>
<tr><td>DF</td><td>0.995</td><td>0.975</td><td>0.20</td><td>0.10</td><td>0.05</td><td>0.025</td><td>0.02</td><td>0.01</td><td>0.005</td><td>0.002</td><td>0.001</td></tr>
<tr><td>1</td><td>0.0000393</td><td>0.000982</td><td>1.642</td><td>2.706</td><td>3.841</td><td>5.024</td><td>5.412</td><td>6.635</td><td>7.879</td><td>9.550</td><td>10.828</td></tr>
<tr><td>2</td><td>0.0100</td><td>0.0506</td><td>3.219</td><td>4.605</td><td>5.991</td><td>7.378</td><td>7.824</td><td>9.210</td><td>10.597</td><td>12.429</td><td>13.816</td></tr>
<tr><td>3</td><td>0.0717</td><td>0.216</td><td>4.642</td><td>6.251</td><td>7.815</td><td>9.348</td><td>9.837</td><td>11.345</td><td>12.838</td><td>14.796</td><td>16.266</td></tr>
<tr><td>4</td><td>0.207</td><td>0.484</td><td>5.989</td><td>7.779</td><td>9.488</td><td>11.143</td><td>11.668</td><td>13.277</td><td>14.860</td><td>16.924</td><td>18.467</td></tr>
<tr><td>5</td><td>0.412</td><td>0.831</td><td>7.289</td><td>9.236</td><td>11.070</td><td>12.833</td><td>13.388</td><td>15.086</td><td>16.750</td><td>18.907</td><td>20.515</td></tr>
<tr><td>6</td><td>0.676</td><td>1.237</td><td>8.558</td><td>10.645</td><td>12.592</td><td>14.449</td><td>15.033</td><td>16.812</td><td>18.548</td><td>20.791</td><td>22.458</td></tr>
<tr><td>7</td><td>0.989</td><td>1.690</td><td>9.803</td><td>12.017</td><td>14.067</td><td>16.013</td><td>16.622</td><td>18.475</td><td>20.278</td><td>22.601</td><td>24.322</td></tr>
<tr><td>8</td><td>1.344</td><td>2.180</td><td>11.030</td><td>13.362</td><td>15.507</td><td>17.535</td><td>18.168</td><td>20.090</td><td>21.955</td><td>24.352</td><td>26.124</td></tr>
<tr><td>9</td><td>1.735</td><td>2.700</td><td>12.242</td><td>14.684</td><td>16.919</td><td>19.023</td><td>19.679</td><td>21.666</td><td>23.589</td><td>26.056</td><td>27.877</td></tr>
<tr><td>10</td><td>2.156</td><td>3.247</td><td>13.442</td><td>15.987</td><td>18.307</td><td>20.483</td><td>21.161</td><td>23.209</td><td>25.188</td><td>27.722</td><td>29.588</td></tr>
</table>

</td>
<td>

$$ValCrit-F = J_F = \frac{1}{N}\sum_{i=1}^{N} Z_F^2\left(\omega_i\right) \quad (21)$$

where the variable to be summed in the frequency domain is:

$$Z_F{}^2(\omega_i) = \frac{1}{2p}\left[\vec{g}(\omega_i)-\vec{g}_{\text{model}}(\omega_i)-\vec{g}_{bias}\right]^T \cdot X_g{}^{-1}(\omega_i)\cdot\left[\vec{g}(\omega_i)-\vec{g}_{\text{model}}(\omega_i)-\vec{g}_{bias}\right]+ \frac{1}{2p}\left[\vec{\phi}(\omega_i)-\vec{\phi}_{\text{model}}(\omega_i)-\vec{\phi}_{bias}\right]^T \cdot X_\phi{}^{-1}(\omega_i)\cdot\left[\vec{\phi}(\omega_i)-\vec{\phi}_{\text{model}}(\omega_i)-\vec{\phi}_{bias}\right]$$

p-dimensional state vector for magnitude and phase (measurements)

output state vector in magnitude and phase from the model prediction

bias correction for magnitude and phase that may be applied

Matrix allowing for the weighting of the difference between model and measurements magnitude and phase

$X_i = diag\left\lfloor \sigma_{xi}{}^2 \right\rfloor$

X is set equal to the measurement error covariance

(22)

J_F exactly as J_T relates to the Chi-squared statistic, i.e. $Z_F^2\left(\omega_i\right) = \chi_F^2\left(\omega_i\right)$:

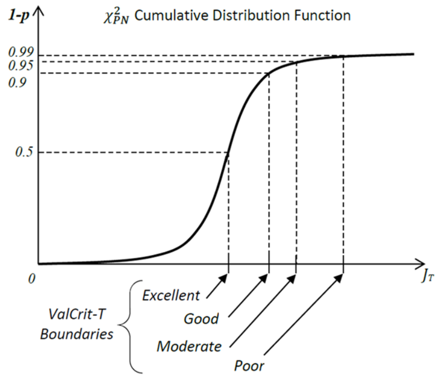
$$ValCrit-F = J_F = \frac{1}{\nu}\sum_{i=1}^{N}\chi_F^2\left(\omega_i\right) \quad (23)$$

where $\nu = pN$ is the number of degrees of freedom.

The values for Z_F are based on a 2-tailed probability, as Z_F may be both positive and negative, i.e. the null-hypothesis tested is:

$$H_0\left\{Z_F > g \; or \; Z_F < -g\right\} = H_0\left\{\left|Z_F\right| > g\right\}$$
$$H_0\left\{Z_F > \phi \; or \; Z_F < -\phi\right\} = H_0\left\{\left|Z_F\right| > \phi\right\} \quad (24)$$

</td>
</tr>
<tr>
<td>Guideline ranges Chi-statistics metric</td>
<td>The values of p for four qualitative levels of simulation error as recommended in GARTEUR HC/AG-09 [Haverdings et al. (2000)] are:</td>
<td>GARTEUR HC/AG-09 used for each frequency a maximum allowable error of 10 percent (95th percentile) of the gain and 20 degrees error (also 95th percentile) in phase angle. The factor of 1.645 in eq. (25) stems from conversion of the 95th percentile value to standard deviation.</td>
</tr>
</table>



p-value	Model quality
0.5	Excellent (insignificant error)
0.1	Good
0.05	Moderate
0.01	Poor (significant error)

$$X_{g_i}(\omega) = (0.1\times 1.645)\, diag\left\lfloor \vec{g}_{model}(\omega_i)\,\vec{g}^{\,T}_{model}(\omega_i)\right\rfloor$$
$$X_{\phi}(\omega) = 3.183 \tag{25}$$

Discussing on the application of Chi-statistics to the flight dynamics models in GARTEUR AG-09, one important conclusion of this work was that: 1) model type had no significant effect on the ValCrit-T, although the effect was nearly becoming weakly significant (p<.10) with regard to the off-axis response 2) control input-type affected significantly ValCrit T (doublet & step w.r.t. 3-2-1-1 and pulse inputs. This is thought to be due to differences in frequency content between the control inputs; 3) The lateral control axis had a much worse mean ValCrit-T value for the off-axis response than that for the longitudinal control axis. With the BO-105's higher bandwidth in the lateral response (i.e. roll) than in the longitudinal response, any small deviation between model and aircraft immediately resulted in a "large" error (large ValCrit-T).

3.3. Discussion on the use of Chi-statistics for flight simulation

The χ^2 methods and with it the Chi-statistics used previously are only valid for Gauss (normal) distributed data [Stoneking & Hartog (1997)]. One can argue the fact that the flight test data measurements are more Poisson distributed (i.e. discrete Gauss). Poisson distribution differs significantly from the Gauss (normal) distribution, especially if the total number of counts per measurement is very small. Therefore the form of the Chi-statistic as expressed in Val-Crit-T and Val-Crit-F should be carefully considered. Research performed by Hauschild & Jentschel (2001) on nuclear applications comparing different statistics formulations for the Chi-Square statistics concluded that if goodness-of-fit estimations are to be done, Pearson's chi-square should be used. It is the only statistic that leads to the **correct expectation value for chi-square**. Person's Chi-square statistics is weighting w.r.t. the measured data [Cox, (2002)] It is discussed that the chi-square per degree of freedom is not well suited for judging the consistency of a model and the data. One should know well in advance which Chi-square form should be used for which case.

A different way of system identification modeling and updating for model accuracy is to fit the flight flight test data according to the statistics tools for parameter estimation. This procedure brings the question of fitting a model function to experimental data and estimate their parameters. The model function is called `**"fit function**" and the result of the "fit" is the best value for the parameter(s). One of the used techniques for fitting a function are in the **maximum likelihood estimation (MLE)**. These statistics maximise the likelihood that the experimental data is a random sample, that could have been drawn from the model. MLE approaches are opposed to chi-square techniques. It could be shown that the MLE derived for Poisson distributed data (Poisson MLE) produces the best statistic to estimate parameters. MLE approaches are numerically more costly than χ^2 methods. With modern computers, however, this difference becomes meaningless.

4. FROM STATISTICS QUANTIFICATION TO UNCERTAINTY QUANTIFICATION

While the previous sections used statistical background to determine model accuracy, one can look at the system from a probabilistic point of view bringing the problem of validation to a problem of uncertainty quantification as discussed in the introduction. The term ‘uncertainty’ is imbued with many meanings (exactly as the terms ‘error’, ‘deviation’ or ‘residual’). The term may refer to a variety of different things: for instance, models, data, tests, practices, forecast, can be ‘uncertain’. In many of these cases, the term is used in a loose manner. A typical example of a definition of ‘uncertainty’ is the following: Uncertainty can be defined as a lack of precise knowledge as to what the truth is, whether qualitative or quantitative [National Research Council, 1994)].

Consider a simple example in order to understand the importance of uncertainty in model prediction, see Figure 2. The example concerns the choice between two competing models, with that choice based upon simulation results. The system is a helicopter model where the pitch rate accuracy during a pull up maneuver needs to be determined in the simulator. In Figure 2a the flight test data would require a desired capability D with a region of acceptable tolerances according to the certifiable simulator tolerances. Consider S1 and S2 predictions of respectively model S1 and model S2. Figure 2a shows D, S1, and S2 without consideration of uncertainty in their location, and indicates the acceptable deviation from D by an oval about it. As can be seen, S1 model is the better choice because its predicted capability, S1, is within the area of acceptable performance, while the capability predicted for S2 is not. Figure 2b indicates that uncertainty in the location of was introduced resulting in a new region of acceptable deviation. Now the predicted capabilities of both models, S1 and S2, are within the region of acceptable deviation. In this situation, it is no longer so obvious that S1 is a better choice as it was when uncertainty was ignored. Figure 2c indicates that uncertainty in predicted model capabilities, S1 and S2, was introduced. When the uncertainty of S1 and S2 locations are considered as well as the uncertainty in D, it appears that model S2 might be the better choice because a greater portion of its potential locations are within the acceptable deviation region than is the case for model S1. This simple example makes two important points: 1) if uncertainty in modelling is not considered, wrong conclusions can be drawn from simulation results. This makes uncertainty consideration in both the data and the models important. This is shown clearly in this example. Model S1 was a clear choice when no uncertainties were considered. When some uncertainty was considered, it was not so clear that S1 was the best choice. When more uncertainties were considered, it appears that S2 might be the better choice. 2) uncertainty should be considered comprehensively, otherwise, wrong conclusions can be drawn from simulation results. In this case, more sophisticated methods of analysis may be necessary to be employed than when uncertainty is neglected. In this example, without consideration of uncertainty, simple tolerances from the desired capability were considered. When uncertainties are considered, probabilistic calculations need to be considered.

Simply to ignore the uncertainty in any process is almost sure to leave critical parts of the process incompletely examined, and hence to increase the probability of generating a risk estimate that is incorrect, incomplete, or misleading. (Pace, 2012). This is also the case for flight dynamics modelling. Typically, these models offer what may be called ‘‘point prediction’’ in the sense that the models, which are themselves deterministic (non-probabilistic), yield deterministic response when acted upon by deterministic inputs. Sometimes stochastic inputs can be applied (ex. wind gusts) that yield stochastic response. In this case one can discuss about model uncertainty. One should note that uncertainty can be added to a model in a deterministic or probabilistic way.

An illustration of deterministic uncertainty added to the model is in the area of robust control [Balas et. Al. (2001)]. The idea behind this approach, which is often referred to as μ analysis, is to separate the known parts of the system from the uncertain parts and pass information between them in a feedback loop. Bounds on the possible values of the uncertain elements are either assumed or based on existing data. The system can be assumed to be robustly stable if it can be shown to be stable for variations within the specified uncertainty bounds.

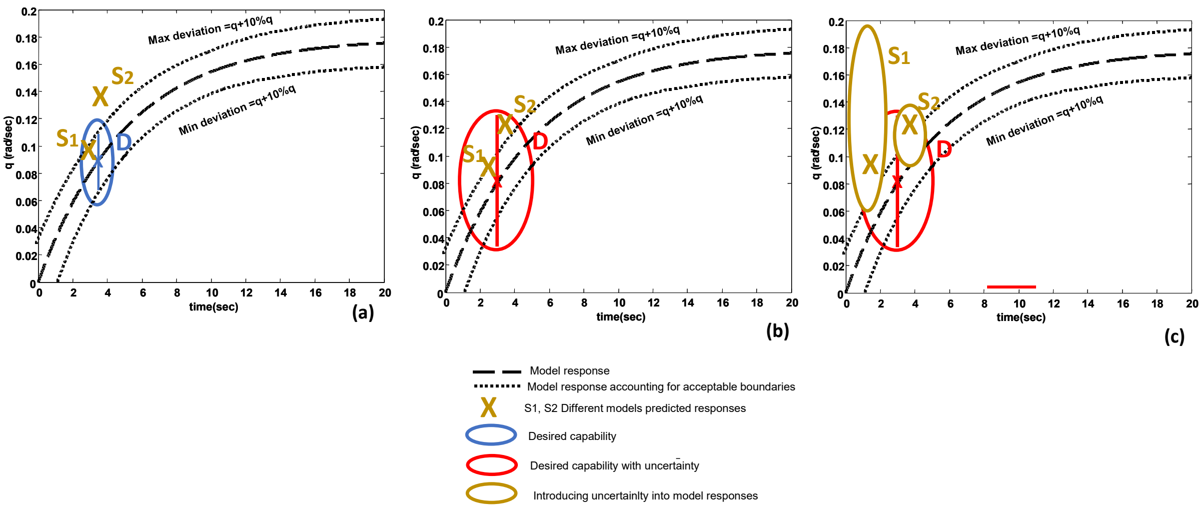


Figure 2: Illustration on how uncertainty can impact model validity.

Examples of sources of uncertainty that can be contained into the flight mechanics models are [Kennedy & O'Hagan (2001):

- parameter uncertainty: it relates to uncertainty associated with the values of model inputs;
- residual variability: it relates to the variation of a particular process outcome even when the conditions of that process are fully specified;
- parametric variability: it results when certain inputs require more detail than is desired (or possible) and are thus left unspecified in the model;
- observation error: it involves the use of actual observations in a model calibration process;
- code uncertainty: it results when a code is so complex or computationally involved that it may not be possible to execute the code at every possible input configuration of interest, thus there is some additional uncertainty related to regions of the input space that have not been interrogated;
- model discrepancy: it and model discrepancy relates to the fact that no model is perfect, and thus some aspects of reality may have been omitted, improperly modeled, or contain unrealistic assumptions.

Padfield et. al. (2023) in project RoCS gave examples on the uncertainties characterizing rotorcraft models: blade torsion dynamics, inertias, flap-pitch coupling, wake distorsion and decay, rotor-fuselage interference, rotor wake interference on tail rotor, dynamic stall, yawed flow, and reverse flow effects.

With uncertainty in modeling the field of uncertainty quantification (UQ) was born. UQ is studying how uncertainties in parameters and models affect the uncertainty in quantities of interest. It provides tools and methods to include statistics and randomness into model predictions. It covers many different topics, such as sensitivity analysis, optimization, parameter estimation and calibration, model reduction, and data assimilation. A main distinction should be made between uncertainty characterization and uncertainty propagation.

- **uncertainty characterization** is determining the nature of the uncertainties. These can be categorized as input (parametric) uncertainties and model-form uncertainties. Other characterization belongs to aleatory (irreducible, stochastic) and epistemic (reducible, systematic, due to a lack of knowledge) uncertainties.
    - aleatory uncertainty corresponds to the inherent variation associated with the physical system or the environment under consideration, e.g. the variation of geometric and or material properties due to manufacturing process It is stochastic and irreducible below a certain threshold.
    - epistemic uncertainty refers to the potential inaccuracy in any phase of the modelling process that is due to a lack of knowledge or to intentional approximations applied by the engineer. It is potentially reducible by model improvements or by a better measuring technique employed to assess model parameters.
- both aleatory and epistemic uncertainties get better with an increase in knowledge. Aleatory uncertainty can become epistemic, sometimes the user

being able to decide to which category an uncertainty belongs. For example, the atmospheric wind field, it can be be treated as aleatory uncertainty; however, when the wind field is modelled with meteorological models, the uncertainty in wind speed prediction can be seen as epistemic. Therefore, it is quite important to understand in the problem analyzed if one deals with aleatoric or epistemic uncertainty. [Kiureghian & Ditlevsen (2009)].

- **uncertainty propagation** is related to predictive accuracy. Herein, uncertainties in inputs and models are related to uncertainties in output quantities, the quantities of interest (QoI) (also called system response quantity (SQR). In other words, uncertainty propagation relates to the accuracy of the model under conditions for which it has not been tested. A model ''tuned'' to match a particular set of data it might not be good in predicting vehicle behavior under different circumstances (e.g., different operating conditions). In other words, a tuned model does not necessarily imply an improved model, although it might. "*The proof of the pudding, so to speak, is how well the model predicts the future, not how well it postdicts the past.*" (Hasselman, 2001) An important part of uncertainty propagation is the analysis of sensitivities: which input quantities have most influence on the QoI? For example, an outcome of a sensitivity analysis in helicopter flight mechanics modelling could be that there is a large uncertainty in the variation of the dynamic inflow, but that its effect on the hovering flight is marginal.

4.1. A Short Review on Uncertainty Quantification Methods

Discussing on the methods that can be used for uncertainty characterization and propagation in a model, a myriad of approaches are found in the literature of specialty. Uncertainty propagation can be studied using intrusive and non-intrusive methods.

Non-intrusive methods rely on 'black box' modelling approach wherein the uncertain input parameters are passed through the black box without understanding what happened in the model. The black box models can be computationally very expensive (e.g. CFD codes), and therefore the art is to smartly sample the input parameters and get accurate output predictions with as few samples as possible. As non-intrusive methods, the sampling-based methods (MonteCarlo, polynomial-based methods (ex polynomial chaos expansions), and Gaussian process regression (also called Kriging)) are often applied in the aerodynamics and structural dynamics modelling.

Intrusive methods rely on changes in the mathematical model. Depending on the model properties, this can lead to efficient methods, but it may require considerable modelling effort. For example, sensitivity analysis is a central topic in the field of uncertainty quantification. It can be used, for example, to determine what parameters have a large influence on the model output and what parameters have little effect and can potentially be fixed at their nominal values in uncertainty propagation studies. Within the sensitivity analysis, one can distinct between local and global sensitivities. Local sensitivities are usually defined via partial derivatives (ex. change of the model input and its related change in the outputs). Global sensitivity analysis is related to determining which part of the output can be ascribed to the model input, over the entire range of input values. A widely used technique in global sensitivity analysis is based on an analysis of variance (ANOVA) decomposition of the model. This decomposition naturally leads to the definition of sensitivity indices, the so-called Sobol' indices [Sobol(2001)]. The first order indices measure which part of the variance of the model output is related to a certain input, while higher order indices also measure interaction effects. For models that can be written as a sum of sub-models, each depending on a single Gaussian random variable, the ANOVA decomposition yields the same results as summation of variances [Sullivan (2015)].

While the above-described methods can be seen as direct problems (uncertainties introduced in model inputs without the need of using experimental data data), one can think also at the backwards problem (i.e. calibrate the models using experimental data) to obtain more accurate predictions under uncertainties. Calibration can be seen as the inverse problem of UQ. Basically, calibration is the process of revising code parameters to bring model predictions into better agreement with test data. This revision process of calibration is also commonly known as model updating, model tuning, parameter calibration and parameter estimation. A common way to solve calibration problems is by using regression analysis, for example with least squares methods [Sullivan (2015)]. An alternative approach is to use Bayesian model calibration or data assimilation. Bayesian model calibration also called Bayesian analysis has been applied often in uncertainty quantification problems with parametric uncertainties [Sullivan (2015)]. The Bayesian framework models uncertainties (both aleatoric and epistemic) in a probabilistic way. Although the method is rather complicated, it involves basically three steps: 1) assume a probability distribution for each input quantity that is chosen to be a random variable; 2) update the previously chosen probability models for the input quantities based on comparison of the computational and experimental results. 3) compare new computational results with the existing experimental data or any new experimental data that might have been obtained. Therefore, the following ingredients are necessary in a Bayesian approach: 1) prior knowledge, 2) measurement data, 3) a computational model, 4) a statistical model, and 5) Bayes' formula. Using these ingredients, the parameters of the model can be calibrated as random variables, resulting in a posterior which can be used to assess the model uncertainty and make future predictions. The advantage of the Bayesian framework is that prior knowledge and the statistical model are naturally included in the formulation.

Within Bayesian framework, data assimilation can be seen as the integration of mathematical-physical models with measurement data, typically to be performed in real-time. The classical example of a data assimilation problem is the weather prediction, where weather simulation models are updated once new observations are made. The most basic data assimilation method, the linear Kálmán filter, is the best unbiased estimator assuming a linear model and linear observation operator, and Gaussian distributions for the errors in both the model and the data. In the extended Kálmán filter the restriction of linearity is relaxed by a linearization of the model and the observation operator. In general, Kálmán filters aim to reduce the computational cost of Kálmán filtering (associated with inverting the covariance operator) by Monte Carlo sampling.

Finally, as UQ methods one should mention the input parametrization and model decomposition approach as applied in structural dynamics models. This is the case when the number of random variables is very large. For example, a turbulent inflow field can be modelled as a high-dimensional correlated random field with a certain power spectrum and covariance function. In a UQ context, such correlated random fields are typically approximated with a lower dimensional expansion, such as a Karhunen-Loève (KL) expansion. This expansion is determined by the eigenvalues and eigenvectors of the covariance function. The coefficients in the expansion constitute the random variables that can be used for uncertainty quantification studies.

Discussing on the methods to be used for describing aleatory or epistemic uncertainties, consensus exists on the use of probability theory (probability density functions) for aleatory uncertainties. For epistemic uncertainties several approaches are being used: interval analysis [Jaulin et. Al. (2001)] which is used when only lower and upper bounds are available, Bayesian theory [Faber(2005)]; evidence theory [Helton&Johnson(2011), Bae&Grandhi (2004)] which is a generalization of classical probability theory and uses belief and plausibility as measures; possibility theory (Dubois & Prade, 1986); fuzzy theory [Beer et. Al., 2015]; probability box (P-box) [Ferson & Ginzburg (1996)], etc. The major differences among these theories are the way that the incomplete knowledge is interpreted and mathematically described. [Zhang et al. (2017)]. For example, applying Bayesian theory implies that one can represent our incomplete knowledge as prior probability distributions, evidence theory expresses incomplete knowledge by identifying basic probability assignments (BPA), possibility theory relies on possibility distributions (or membership functions) describing the state of knowledge, probability box (P-box) expresses the incomplete knowledge based on intervals or bounds of probability distributions .

4.2. On the Validation Metrics for Uncertainty Quantification

In UQ the emphasis is not on validation metrics but on the methods to validate the computational model. However, discussing on the validity metrics for uncertainty quantification, a myriad of validation metrics exist, the choice of validation metric depending on the problem under investigation. The purpose of the validation metrics is to assess the quality of agreement between the model predictions and experimental data, the comparison being carried out in relation the specific quantity of interest (QoI). [Oberkampf et. al., (2001), Oberkampf (2002), Oberkampf & Barone (2006), Ghanem et. al. (2008), Lehmann et. al (2019)]. As validation metrics one can use qualitative or quantitative metrics. Example of qualitative metric is the visual metric where the engineer visually compares the graphs of computed response with those of experimental observations. Such a metric is purely qualitative and cannot properly accommodate the effects of uncertainty. Quantitative metrics are for example hypothesis testing-based validation metrics based on classical hypothesis testing or Bayesian hypothesis testing-based which generally accept or reject the model based on some error statistics. Discussing on the recommended features of the validation metrics in UQ, Oberkampf & Barone (2006) concluded that the main metric characteristics should be:

1) a metric should be objective such that no subjective input in terms of, e.g., calibration is required;

2) the metric should be deterministic, i.e. its predictive capability should remain unchanged if identical inputs are repeated;

3) the metric should be kept general such that it is applicable for any units of QoI and any dimension input space. However, Ferson et. al., (1996) argue that the unit of the metric should be the same as in simulated and measured QoI. This property needs to be weighed against its general characteristics;

4) a metric should incorporate, or include in some explicit way, an estimate of the measurement errors in the experimental data for the QoI. Usually, measurement errors are commonly divided two types: bias (systematic) errors and precision (random) errors. At a minimum a validation metric should include an estimate of precision errors, and, to the extent possible, the metric should also include an estimate of bias errors.

4.3. Example of Uncertainty Quantification Application to Rotorcraft

Consider the numerical example of a helicopter with the following characteristics: mass m=2200 kg, rotor radius R=7.32 m, rotor tip speed ΩR =200 m/sec, Lock number γ =6, moment of inertia I_y =10625 kg m^2; distance h=1 meter, semi-rigid rotor with spring constant K_b=460000 Nm and consider ist pitch motion, see Figure 3.

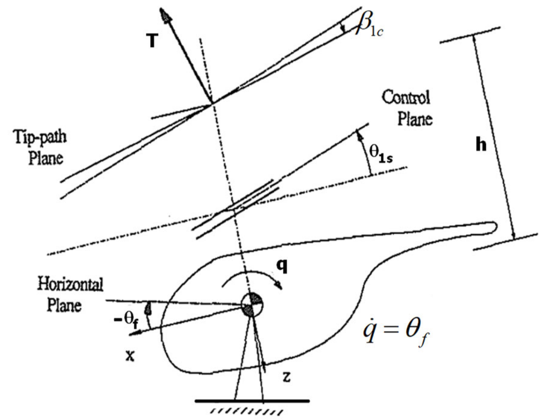


Figure 3: A pitch motion on a helicopter model

The equation of motion of pitch coupled to the flapping dynamics can be written in two ways:

Approach 1:	Approach 2:
$\dot{q} = -K(\theta_{1s} - a_1)$	$\dot{q} = -K(\beta_{1c})$
$\tau_\beta \dot{a}_1 + a_1 = -\dfrac{16}{\gamma}\dfrac{q}{\Omega}$	$\tau_\beta \dot{\beta}_{1c} + \beta_{1c} = \theta_{1s} + \dfrac{16}{\gamma}\dfrac{q}{\Omega}$

where θ_{1s} the longitudinal tilt of swashplate (cyclic stick displacement), a_1 is the longitudinal disc tilt w.r.t. plane of no-feathering plane, β_{1c} is longitudinal disc tilt w.r.t. plane of disc, τ_β a time constant for flapping dynamics.

Consider that a flight controller tracking controller is developed for a reference pitch rate qref, see Figure 4. Looking at this figure one can see that using Approach 2 in writing the eqns. of motion (i.e. expressing the rotor dynamics into the disc plane) results in an oscillatory controller. This was not the case for Approach 1 where the pitch reference was tracked more precisely. This example shows that although the model fidelity was not varied, the Controller became non-deterministic.

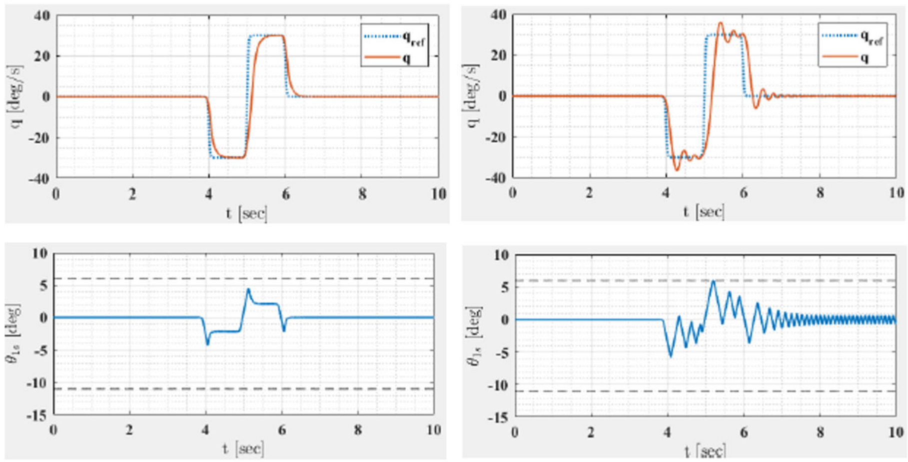
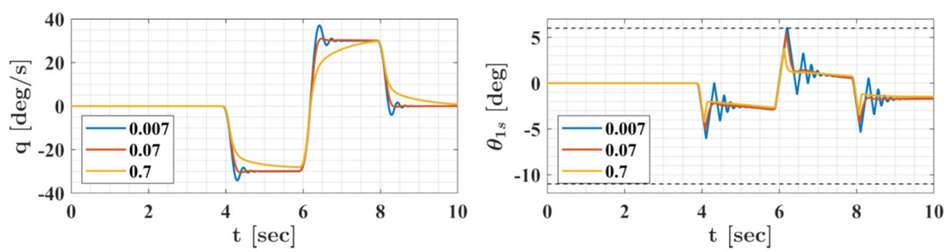


Figure 4: Helicopter pitch response acquires a nondeterministic character although its epistemic uncertainty is not varying

Next, including epistemic uncertainty in the controller (variation in the time constant τ_β) results in model uncertainty variations.



(a) Pitch rate for varying τ_β

(b) Control input for varying τ_β

Figure 5: Helicopter pitch response variation with model uncertainty

Looking at Figure 6 one can understand how the model uncertainty variations propagate on the tracking performance oft he flight controller. Therefore, one can conclude that model uncertainities affect muchly the performance indicators of a flight controller.

5. CONCLUSIONS

The present paper surveyed the state-of-the-practice validation metrics and criteria employed for rotorcraft modeling accuracy, extending these approaches to uncertainty quantification (UQ) domain of validation.

First, the cost function metrics J_{rms} and J_{ave} belonging tot he RMSE method and their guidelines used extensively by NATO AST-296 group show to give a very practical approach in assessing model fidelity. However, care should be taken in considering these metrics as an absolute statement on model validity. This is true especially in validation of models for flight simulators where small constant offset in model result in guidelines below acceptable of model quality.

Next the metrics defined in the Chi-square statistics and applied by the GARTEUR AG-09 are different from the cost function metrics J_{rms} and J_{ave} because of the weighting used which belongs to the error covariance. When applied to flight dynamics modeling for rotorcraft, these metrics show in some cases that small deviation between the model and aircraft data result in large error belonging to unnaceptable model quality. A word of caution when using these validation metrics is that there exist different Chi-square statistics depending on the test data distributions and one should know well in advance which Chi-square form should be used.

Extending the model validation to the uncertainty quantification is important, otherwise, wrong conclusions can be drawn from simulation results. However, uncertainty should be considered comprehensively, otherwise, wrong conclusions can be drawn from simulation results. There are a myriad of methods and metrics to be used for UQ and care should be taken to which quantity of interest (QoI) are they applied.

A simple example was presented for rotorcraft model validation showing that the problem of flight controller modeling can become nondeterministic depending on the formulation of the equations of motion. A key message of the paper is that using UQ in the future can enhance our understanding of the simulation model fidelity.

Author contact: Marilena Pavel m.d. pavel@tudelft.nl

6. REFERENCES

Bae, H-R., Grandhi, RV., Canfield, RA., An approximation approach for uncertainty quantification using evidence theory, Reliab. Eng. Syst. Saf. 86 (3) (2004) 215–225.

Balas, G. J., Doyle, J. C., Glover, K., Packard, A., and Smith, R., µ-Analysis and Synthesis Toolbox, MathWorks, Natick, MA, 2001

Beer, M., Ferson, S., Kreinovich, V., Imprecise probabilities in engineering analyses, Mech. Syst. Signal Process. 37 (1–2) (2013) 4–29.

Boehm, B. W., Software engineering economics. Englewood Cliffs, NJ: Prentice-Hall, 1981

Blottner, F. G., Accurate Navier-Stokes results for the hypersonic flow over a spherical nosetip. Journal of Spacecraft and Rockets, 27(2), 113–122., 1990

Box George, Science and statistics. J American Stat Assoc 71(356), 791–799, 1976

Chatterjee, S. (2011). Residuals. In: Lovric, M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. 2011, https://doi.org/10.1007/978-3-642-04898-2_491

Cox, D.R. (2002). Karl Pearson and the Chi-Squared Test. In: Huber-Carol, C., Balakrishnan, N., Nikulin, M.S., Mesbah, M. (eds) Goodness-of-Fit Tests and Model Validity. Statistics for Industry and Technology. Birkhäuser, Boston, MA. https://doi.org/10.1007/978-1-4612-0103-8_1]

Dubois, D., Prade, H., Possibility Theory, Plenum, New York, 1986.

EASA: Certification Specifications for Helicopter Flight Simulation Training—CS-FSTD(H). Initial Issue (2012) UPDATE THIS

Faber, M.H., On the treatment of uncertainties and probabilities in engineering decision analysis, J. Offshore Mech. Arct. Eng. 127 (3) (2005) 243–248.

Frost, Jim, Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries (2020), Statistics by Jim Publishing, 2020

Ferson, S., Oberkampf, W.B., Ginzburg, L. Model validation and predictive capability for the thermal challenge problem, Comput. Methods Appl. Mech. Engrg. 197,2008, pp. 2408–2430

Ferson, S., Ginzburg, LR., Different methods are needed to propagate ignorance and variability, Reliab. Eng. Syst. Saf. 54(2–3),1996, pp.133–144.

GARTEUR HC/AG-21 on "Rotorcraft Simulation Fidelity Assessment Predicted and Perceived Measures of Fidelity" [White et al. (2020)] applied ValCrit-T parameter to the problem of helicopter simulator validation

Ghanem, R.G., Doostan, A., Red-Horse, J., A probabilistic construction of model validation, Comput. Methods Appl. Mech. Engrg. 197 (2008) 2585–2595

Hasselman, T. "Quantification Of Uncertainty In Structural Dynamic Models" Journal Of Aerospace Engineering, Vol. 14, No. 4, October, October 2001

Hauschild J., and Jentschel M., Comparison of maximum likelihood estimation and chi-square statistics applied to counting experiments, Nuclear Instruments and Methods in Physics Research A 457 (2001) 384}401

Haverdings H, Dequin AM, Basset PM, von Grunhagen W, Kampa K, Massey C, McCallum A, Meerwijk EL, Mathematical modeling for the prediction of helicopter flying qualities within GARTEUR -phase 3, ERF 2000

Helton, JC., Johnson, JD., Quantification of margins and uncertainties: alternative representations of epistemic uncertainty, Reliab. Eng. Syst. Saf. 96 (9) (2011) 1034–1052.

IEEE, Recommended Practice Guide for Distributed Interactive Simulation – Verification, Validation and Accreditation, IEEE, USA, pp 2-8, 1997

Jaulin, L., Kieffer, M., Didrit, O, Walter, E., "Applied Interval Analysis, With Examples in Parameter and State Estimation, Robust Control and Robotics", Springer Nature, 2001

Kennedy, M. and O'Hagan, A., Bayesian calibration of computer models, J.R. Statist. Soc. B, 63(3):425–464, 2001

Kiureghian, A.D., Ditlevsen, O., Aleatory or epistemic? Does it matter? Struct. Saf. 31 (2) (2009) 105–112.

Lehmann, A. W., Wang, Z. "Verification, Validation, and Accreditation (VV&A)—Requirements, Standards, and Trends, Model Engineering for Simulation, 101-121, Academic Press, 2019

Liu, D., Macchiarella, N.D., and Vincenzi, D.A., Simulation Fidelity. CRC Press, 2008, ch. 4, pp. 61–73.

Mihram, G. Some Practical Aspects of the Verification and Validation of Simulation Models. J Oper Res Soc 23, 17–29 (1972). https://doi.org/10.1057/jors.1972.3

National Research Council, Science and Judgment in Risk Assessment, Washington, dc: Taylor & Francis, 1994

NATO STO: Rotorcraft flight simulation model fidelity improvement and assessment (2021). https:// doi. org/ 10. 14339/STO- TR- AVT- 296

Oberkampf, WL., Helton, JC., and Sentz. K., Mathematical representation of uncertainty. American Institute of Aeronautics and Astronautics Non-Deterministic Approaches Forum, (April): 1–23, 2001. doi:http://dx.doi.org/10.2514/6.2001-1645.

Oberkampf, WL., et.al., Error and Uncertainty in Modeling and Simulation, Reliability Engineering and System Safety, 75, pp.333-357, 2002

Oberkampf, WF., Barone, MF., Measures of agreement between computation and experiment: validation metrics, J. Comp. Phys. 217(1) (2006) 5–36.

Pace, Dale, "Comprehensive consideration of uncertainty in simulation use" Journal of Defense Modeling and Simulation: Applications, Methodology, Technology, 10(4) 367–380

Roache, P.J., Verification of Codes and Calculations, AIAA Journal, 36(5), 696-7023, 1998

Roza, Z.C., Simulation fidelity: theory and Practice, Dissertation, Delft University Press, 2004

Padfield, G.D., vant' Hoff, S., Lu, L., Podzus, P.; White, M., Quaranta, G., RoCS- Preliminary Guidelines for the Rotorcraft Certification by Simulation Process, update no. 1, March 2023 https://dib2.aero.polimi.it/index.php/s/RdrMLL2dcZZwgNb.

Pavel, M.D., et. al., Simulation Fidelity Assessment for Rotorcraft–Methods and Metrics, Sketches from the Work of NATO AVT-296, 77th Annual Forum & Technology Display of the Vertical Flight Society, May 10-14, 2021, Virtual.

Scepanovic P., and Doring F, Improving a real-time helicopter simulator model with linear input Filters, CEAS Aeronautical Journal, Vol 12, pp.605–619, 2021 https://doi.org/10.1007/s13272-021-00518-6

SISO-Guide-GM-VV, 2012–2013. Generic Methodology for Verification and Validation (GM-VV), Volume 1.1, Volume 1.2, Volume 1.3. Simulation Interoperability Standardization Organization.

Siegenfeld, A & Bar-Jam, Y., An Introduction to Complex Systems Science and Its Applications, Complexity, 2020, ID 6105872, https://doi.org/10.1155/2020/610587

Stoneking, MR., Den Hartog, DG., Maximum-likelihood fitting of data dominated by Poisson statistical uncertainties, Review of Scientific Instruments 68 (1), pp 914-917, 1997).

Sobol, M.. "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates", Mathematics and Computers in Simulation, 55:271–280, 2001.

Sullivan, T.G., Introduction to Uncertainty Quantification, Springer, 2015

Tischler, M.B., Remple, R.K.: Aircraft and Rotorcraft System Identification: Engineering Methods with Flight Test Examples, 2nd edn. American Institute of Aeronautics and Astronautics, Reston, 2012

USAF MIL-STD-1797B, Flying Qualities of Piloted Aircraft, 2006

Wang, S., Chaovalitwongse, W. A., Evaluating and Comparing Forecasting Models, Wiley Encyclopedia of Operations Research and Management Science, 2011

Zhang, Q. et. Al, "Probability box as a tool to model and control the effect of epistemic uncertainty in multiple dependent competing failure processes", Applied Soft Computing, Vol 56, 570-579, 2017