Horizon 2020 Programme

METIS

Research and Innovation Action (RIA)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 945121

Start date : 2020-09-01 Duration : 48 Months

Verification & Validation of nonlinear mechanical model on SMART2013 campaign

Authors : Dr. David BOUHJITI (IRSN), Thomas LANGLADE, Thomas HEITZ, David BOUHJITI, Benjamin RICHARD

METIS - Contract Number: 945121

Project officer: Katerina PTACKOVA

Document title	Verification & Validation of nonlinear mechanical model on SMART2013 campaign
Author(s)	Dr. David BOUHJITI, Thomas LANGLADE, Thomas HEITZ, David BOUHJITI, Benjamin RICHARD
Number of pages	84
Document type	Deliverable
Work Package	WP6
Document number	D6.2
Issued by	IRSN
Date of completion	2024-04-16 16:25:04
Dissemination level	Public

Summary

Task 6.2 aims at verifying and validating the selected mechanical models and failure criteria identified and developed in T.6.1-1 based on a methodological comparative analysis with experimental analyses and test data. Thanks to the achieved work, IRSN developed a set of criteria to be fulfilled in order to ensure a relevant verification and validation process, both for detailed and reduced-order models. The retained methodology is based on a multi-fidelity and bias/variance reduction approach. Accordingly, the present document shows the verification and validation flowcharts proposed by IRSN and covers the mathematical tools for practical industrial applications in the case of linear and nonlinear mechanical models.

Approval

Date	By
2024-04-17 15:06:04	Mrs. Shadi FATHABADI (TUK)
2024-04-17 16:54:37	Dr. Irmela ZENTNER (EDF)

METIS

Seismic Risk Assessment
for Nuclear Safety

Research & Innovation Action

NFRP-2019-2020

Verification & Validation of nonlinear mechanical model on SMART2013 campaign

Deliverable D6.2

Version N°002

Author: Thomas LANGLADE
Radioprotection and Nuclear Safety Institute (IRSN)
Structural Performance Modeling and Analysis Laboratory (LMAPS)

Disclaimer

The content of this deliverable reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

Document Information

Grant agreement	945121
Project title	Methods And Tools Innovations For Seismic Risk Assessment
Project acronym	METIS
Project coordinator	Dr. Irmela Zentner, EDF
Project duration	1st September 2020 – 31st August 2024 (48 months)
Related work package	WP 6- Beyond Design Assessments and Fragility Analysis
Related task(s)	Task 6.2 - Verification and validation of mechanical nonlinear models using experimental analysis and tests
Lead organization	IRSN
Contributing partner(s)	
Due date	August 31st 2023
Submission date	
Dissemination level	Public

History

Version	Submitted by	Reviewed by	Date	Comments
N°001	Thomas Langlade	Thomas Heitz David Bouhjiti	30/08/2023	Verification
-	Thomas Heitz	Benjamin Richard	30/08/2023	Approval IRSN
N°002	Thomas Heitz	François Voldoire Abhinav Gupta	12/02/2024 12/02/2024	External review

Table of Contents

List of figures

List of tables

Abbreviations and Acronyms

Acronym	Description
WP	Work Package
VV	Verification and Validation
PDF	Probability Density Function
PGA	Peak Ground Acceleration
CPD	Cumulative Probability Distribution
LH	Latin Hypercube
MC	Monte-Carlo
EDP	Engineering Demand Parameter
UNUM	Numerical Uncertainties
UFORM	Model Formulation Uncertainties
UINP	Input Parameter Uncertainties
IM	Intensity Measure
OC	Overlapping Coefficient

Summary

Task 6.2 aims at verifying and validating the selected mechanical models and failure criteria identified and developed in T.6.1-1 based on a methodological comparative analysis with experimental analyses and test data. Thanks to the achieved work, IRSN developed a set of criteria to be fulfilled in order to ensure a relevant verification and validation process, both for detailed and reduced-order models. The retained methodology is based on a multi-fidelity and bias/variance reduction approach.

Accordingly, the present document shows the verification and validation flowcharts proposed by IRSN and covers the mathematical tools for practical industrial applications in the case of linear and nonlinear mechanical models.

.

Keywords

Verification, Validation, Flowchart, Uncertainties, Propagation, Reinforced-Concrete, Cumulative Probability Distributions, Conservatism, Extrapolation.

1.Introduction

1.1. Context

Task 6.2 aims at verifying and validating the selected mechanical models and failure criteria identified and developed in T.6.1-1 based on a methodological comparative analysis with experimental analyses and test data. Thanks to the achieved work, IRSN developed a set of criteria to be fulfilled in order to ensure a relevant verification and validation process, both for detailed and reduced-order models. The retained methodology is based on a multi-fidelity and bias/variance reduction approach.

Accordingly, the present document shows the verification and validation flowcharts proposed by IRSN and covers the mathematical tools for practical industrial applications in the case of linear and nonlinear mechanical models. In (IRSN, 2022), IRSN applied the method to the case of reinforced concrete beams of the IDEFIX experimental campaign. In the present report, IRSN considered a bigger structural scale following the SMART2013 experimental campaign involving an asymmetric reinforced concrete structure subjected to seismic input of increasing intensities.

1.2. Introduction to verification and validation

Validation and verification are two distinct actions that are essential to qualify simulation results as acceptable. This section below defines only shortly the concept of verification. Indeed, verification implies simply to check the software ability to yield accurate/expected outcomes. If this the case, one can move on to the validation phase; otherwise, one needs to correct the model's implementation until the tool is verified. On the contrary, the validation concept and its related flowchart will be more thoroughly presented in Section 2, before being applied to the SMART2013 experimental campaign in Section 4.

Figure 1: (ASN, 2017) verification and validation guidelines

1.3. Short introduction to verification concept

Verification aims at concluding if a numerical model reproduces correctly the theoretical (or conceptual) model from which it is originated. It is therefore not a question of judging the capacity of

this numerical model to reproduce reality, but rather to characterize the quality of the solution, often approximate, constructed numerically. In other words, the verification is the demonstration of the acceptability of the errors made due to a numerical transposition of the mechanical models.

It is worth mentioning that analytical mechanical models that have been transposed into the numerical software, object of the verification flowchart, have been anteriorly designed and subjected to a scientific consensus to corroborate an experimental evidence, in representative situations.

The flowchart presented in Figure 2 summarizes the verification process from a literature review of recommendations from various institutions such as Institute of Electrical and Electronics Engineers (IEEE, 1984), American Nuclear Society (ANS, 1987), American Institute of Aeronautics and Astronautics (AIAA, 2002), American Society of Mechanical Engineers (ASME, 2007), International Atomic Energy Agency (IAEA, 1999), Autorité de Sûreté Nucléaire (ASN, 2017), and Sandia National Laboratories (Trucano et al., 2001).

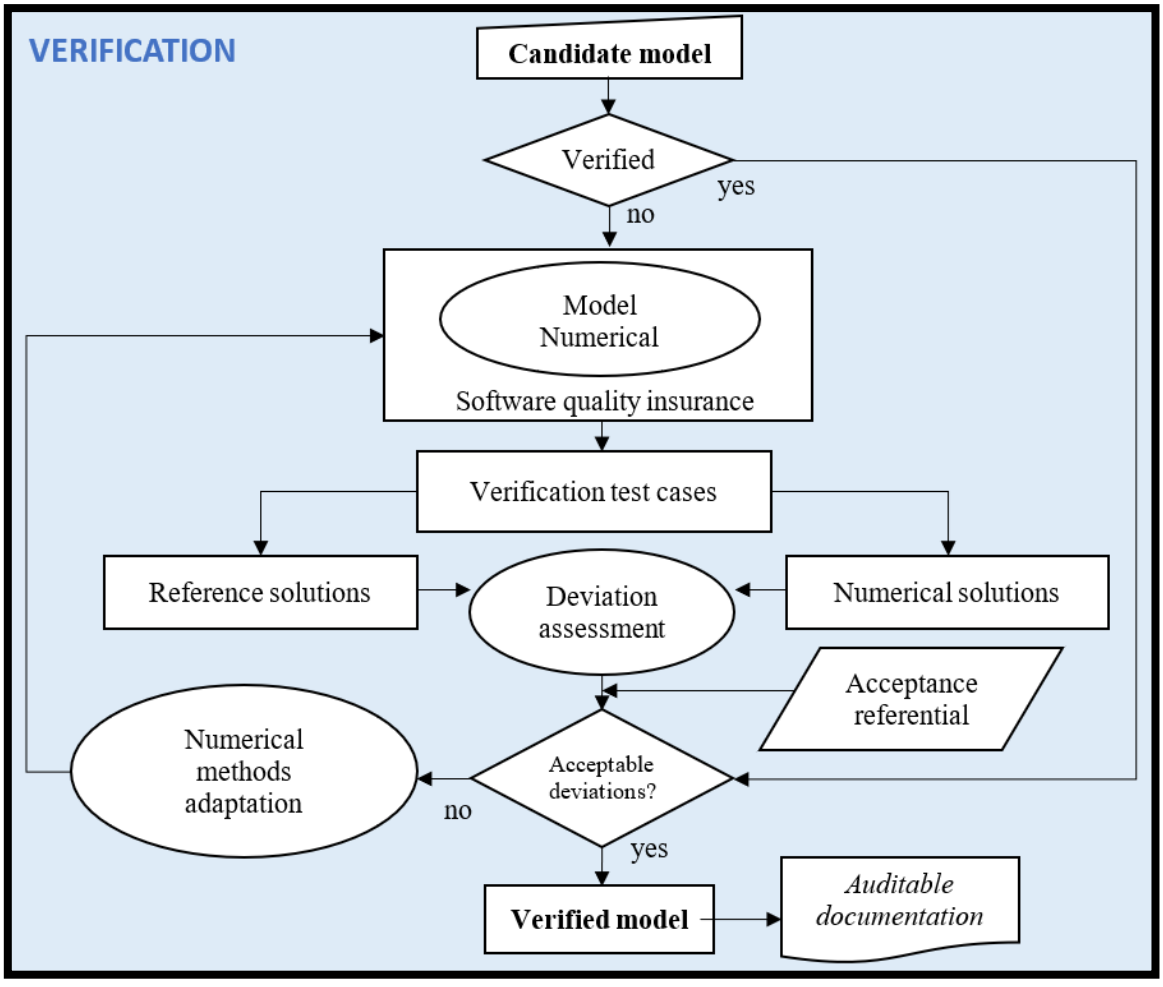


Figure 2: Verification flowchart

In the present document, the Cast3m (CEA, 2011) numerical software is used, considered verified, and the focus is directed entirely towards the Validation part.

1.4. Overview of the report

First, the concepts of validation are presented along with their respective diagrams used herein. The importance of accounting for and assessing the experimental and numerical uncertainties are also explained. Then, several propositions to improve the process are detailed, notably the model(s)

accuracy and behavior evaluation. Finally, the validation process is applied to the practical cases of the data extracted from the SMART2013 experimental campaign (DEN, 2014) (Richard, 2016).

2. Validation: definition and flowchart detailed

This whole section aims at introducing the main concepts and methodology that will be applied on practical cases in the following sections. First, the objective is to define the concept of validation and introducing its associated flowchart. This flowchart can be divided into three distinct sections. To understand concretely what happens in each of them, one must first explain the reason why uncertainties are accounted for in the validation process. Secondly, the various sources of uncertainties are presented with simple examples, as well as how they are propagated and combined. Thirdly, the concept of validation domain is explained as well as the need for methods and rules with which interpolate or extrapolate data. The method presented relies on quadratic regressions combined with a prediction interval computed with the Student's distribution law.

2.1. Definition

Validation aims to determine whether a numerical model (and therefore the associated physical hypotheses) correctly reproduces experimental observations. In a nuclear safety regulatory context, "correctly" can be read as "conservative and reliable" (see ASN Guide n° 28).

To do this, one could either use the output from a validated, reliable, and trustworthy software, or better, use experimental data. Validation can be viewed as a three-step process:

1. Quantification of the accuracy of the numerical model by comparing its responses with experimental measurements (which may be different from the application case of interest). Several models can be compared in the same process.

2. Interpolation or extrapolation of this model(s) output to the case of interest.

3. Determination of whether the accuracy requirements of the model(s) applied to the case of interest are met.

It is important to distinguish the validation case from the case of interest. Indeed, the need to have reference data, preferably experimental, often imposes a comparison of results to a case different from the one which is the object of the study. For example, a mechanical model can be validated on a structure A for which experimental data are available. If this model is well validated in this case, it is then necessary to evaluate the relevance of its application to the case of interest, which would be structure B, but for which little information is available. It is important to distinguish the problem linked to a lack of knowledge of the input parameters of a model from the problem linked to a bad choice of model to describe the studied physics.

In other words, Validation is the demonstration of the acceptability of the uncertainties inherent in the model chosen to describe the reality of a phenomenon. Figure 3 introduces the validation flowchart strongly inspired by (Trucano et al., 2001), (Thacker et al., 2004), (Oberkampf and Barone, 2006), and (Kats and Els, 2012). The step-by-step flowchart unfolding is strongly inspired by (Roy and Oberkampf, 2011), often cited in the present document.

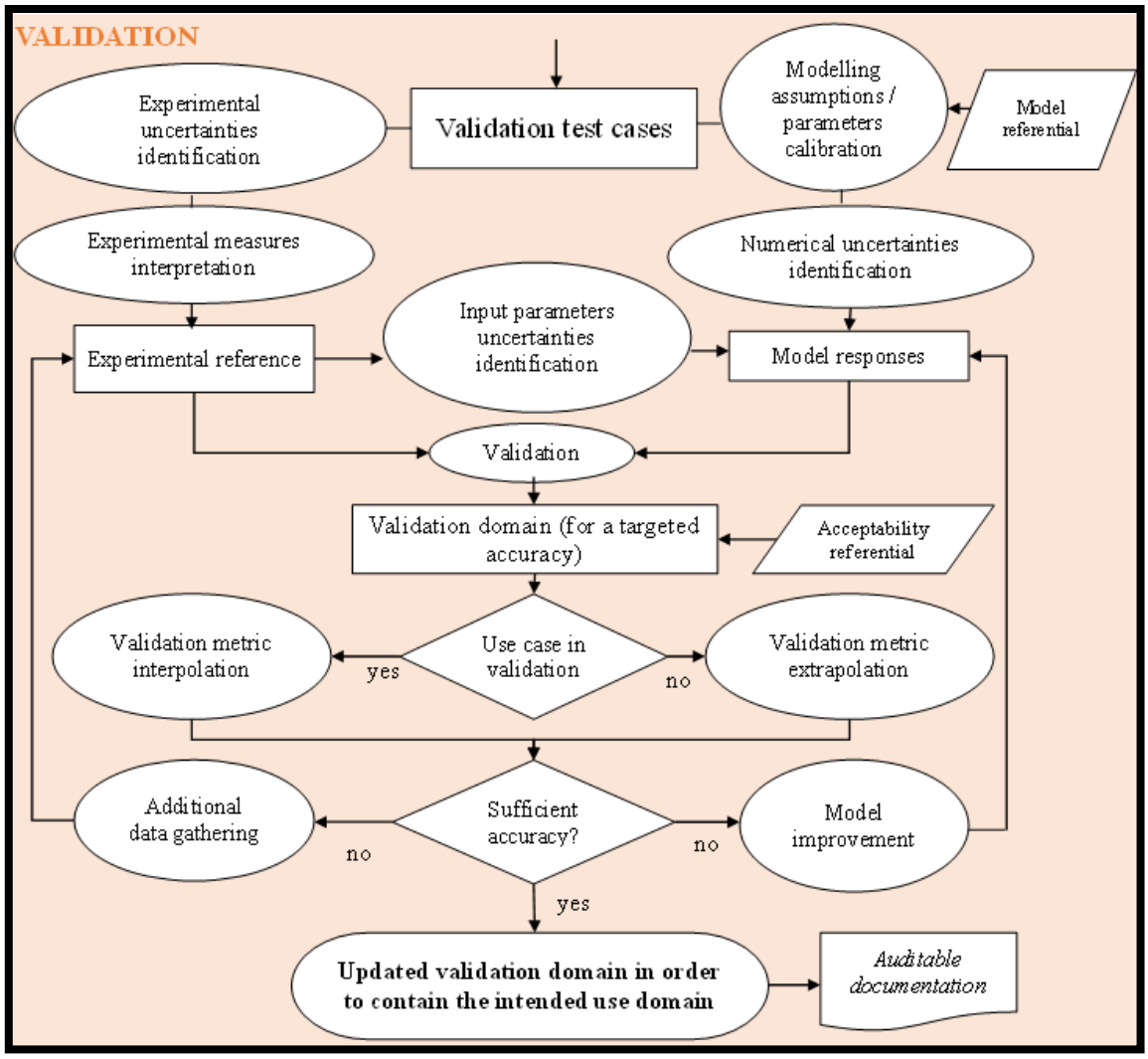


Figure 3: Validation flowchart

The flowchart can be divided into four parts that will be detailed below:

- Part 1: Analysis of the experimental tests, measurements, mock-up, hypothesis.
- Part 2: Numerical Analysis: Numerical modelling, parameter calibration, hypothesis, choice of the uncertainties to propagate.
- Part 3: Calculation of the metrics, e.g., the mismatch between the experimental measurements and the numerical output.
- Part 4: Extrapolation of the output of interest for a condition not achieved during the experimental tests.

But prior to that, it is first important to introduce the concepts of accounting for and assessing uncertainties.

2.2. Sources of uncertainties

This subsection objective aims at explaining why uncertainties are inherent to this validation process and details the main sources of uncertainties: numerical discretization, input parameters, and model formulation. Further subsection 2.2.1 will present into more details how they are concretely assessed,

and the subsection 2.2.2 will finally describe the above flowchart step by step. An illustration of the treatment of uncertainties is presented in section 4 in the case of SMART2013 mock-up.

2.2.1. Accounting for uncertainties

It is generally not expected to obtain an "exact" result from the numerical simulation of a mechanical problem concerning a real structure, the problem being far too complex, and the model formulation is most often partial. It is indeed impossible to know with precision the properties of the materials at all points, the boundary conditions, the imperfections, etc. Uncertainties are therefore inherent to modeling, especially when moving to a numerical model (or approximate solution). The acceptability of numerical results should be weighted based on the associated uncertainties. In this sense, the determination of the different types of uncertainties is an integral part of the V&V process.

2.2.2.Sources of uncertainties

There are three main sources of uncertainties. Their naming and calculations are strongly inspired by (Roy and Oberkampf, 2011).

1. The first source of uncertainty concerns the numerical approximations, further named **UNUM**, and can be divided into four categories:

- the discretization error, related to the spatial discretization (finite element mesh) and temporal discretization of the problem.
- the convergence error, linked among others, in the case of an iterative resolution process, to the respect of a stopping criterion generally reached before obtaining the exact solution.
- the rounding error, linked to the limit of numbers after the decimal point according to the type of variable used.
- the programming error, generally of human origin, but which can most of the time be avoided by the application of good practices, for instance by applying a verification process as displayed by Figure 2.

UNUM is often relatively small as software capabilities have grown in power over the years.

2. The second source of uncertainty concerns the input data of the model, namely the characteristics of the structure (geometry, initial conditions, material properties), its environment (boundary conditions, stresses), or any numerical or theoretical model parameter. It also includes uncertainties related to the measurement techniques and instruments used. Subsequently, it is named **UINP**. Subsection 2.3.2 will introduce the natures of input uncertainties, and the way they are propagated to obtain an estimation of the Engineer Demand Parameter (EDP) of interest (example: building interstorey-drift, crack opening, peak floor response spectra, …).

3. The third source of uncertainty, named **UFORM** in (Roy and Oberkampf, 2011), is the remaining difference between the simulations output accounting for input uncertainties **UINP** and the experimental measurements. **UFORM** reflects the lack of knowledge linked to the choice of the model, or the inherent limits of the model itself that may not be able to reproduce reality, i.e., the experimental measurements, accurately enough.

The following sub-section will present into more details and with simple examples how to compute successively **UNUM** and **UINP**. **UFORM** will be presented more thoroughly in the next section in the step by step reading of the validation flowchart.

2.3. Uncertainties assessment

This subsection objective aims at presenting in more details and with simple examples how the three abovementioned uncertainties are assessed for the practical case of Section 4.

2.3.1. Numerical uncertainty UNUM

Due to the spatial discretization of the problem, approximation errors appear. These errors tend to decrease when the step, whether it is spatial or temporal, is refined at the cost of an increase in the overall calculation time. It is necessary to evaluate the discretization error in order not to unnecessarily increase the computation time: this error may already be satisfactory. The numerical uncertainty **UNUM** can be described, in a conservative manner where all uncertainties add up, as follows:

$$\boldsymbol{U_{NUM} = U_{DE} + U_{IT} + U_{RO}}$$

Where U_{DE}, U_{IT}, and U_{RO} are respectively the uncertainty linked to the discretization error in space, the uncertainty linked to convergence errors, and the uncertainty linked to rounding numbers. The works by (Roach, 1994), (Roy, 2005), and (Roy, 2010) present them into more details.

Rounding uncertainties U_{RO}

Rounding errors are due to the storage of variables in the computer memory. Double precision floating point variables can store up to 16 decimals which is sufficient most of the time. With today's machines, it is normally no longer necessary to limit the number of decimals by using, for example, a simple precision (7 decimals). U_{RO} is not a priori a problem and is neglected in the following studies.

Convergence uncertainty U_{IT}

Convergence errors appear in the case of an implicit scheme when the solution at each time step is obtained by an iterative process. A stopping criterion must be defined. It is most often a threshold for the relative rate of change of a quantity between two successive iterations: when this rate of change falls below the threshold, the solution is considered as converged. In some cases, a criterion on the maximum number of iterations is defined in addition, it is then necessary to consider the fact that the error can exceed the previously defined threshold when this number is reached. The convergence of the calculations is in any case a necessary condition for the validation and verification of the model: it is therefore considered to be reached a priori. The choice of a sufficiently low threshold for the stopping criterion allows to guarantee that the convergence error is negligible compared to the values of the results.

Discretization uncertainty U_{DE}

Different methods exist to assess the discretization uncertainties, especially Richardson's extrapolation which consists in monitoring the quantities of interest for several levels of discretization and known levels of refinement as detailed in (Roach, 1994), (Roy, 2010), and (NCARP, 2022). It is nevertheless important that these evaluations are done in the asymptotic convergence domain of the mesh refinement, i.e., when the mesh is fine enough not to distort the physics of the phenomena involved. Let is consider a quantity of interest $\hat{y}$ approximated by three simulations for increasing levels of mesh fineness $\{y_1, y_2, y_3\}$ and assume that the element size between two successive simulations is divided by r each time.

Richardson's extrapolation estimates the approximation error δ_{RE} associated with the result of the i calculation by a polynomial:

$$\boldsymbol{\delta_{RE} = y_i + y = \alpha h_i^p}$$

$$\hat{p} = \frac{ln\frac{y_1 - y_2}{y_2 - y_3}}{ln\,(r)}$$

where h_i is the characteristic size of the mesh elements of the calculation, p is the formal order of accuracy of the method (the error decreases proportionally to the power p of h_i), and p is the observed order of accuracy. If $p \approx \hat{p}$, then the Richardson extrapolation can be applied. The size h_i does not matter if one knows the ratio r between successive mesh sizes. Richardson then showed that the calculation converges to a value $\bar{y}$ as the size h_i decreases:

$$\bar{y} = y_3 + \frac{y_1 - y_2}{r^p - 1}$$

Thus, the discretization error for any mesh size, at least as fine as mesh 1, can be estimated:

$$\delta_{DE} \approx \bar{y} - y$$

The transition from discretization error to discretization uncertainty can be done via the use of a safety factor F_s and an error measure δ_{DE} :

$$U_{DE} = F_s \cdot |\delta_{DE}|$$

In the case where the solution lies in the asymptotic mesh fineness convergence domain, the value F_s=1.25 is recommended according to (Roy and Oberkampf, 2011). In fact, the previous method can also be applied with only 2 calculations giving the values y_1 and y_2. In this case, the formula for calculating p should be adapted and a safety factor of F_s=3 should be taken because the confidence in the value of δ_{RE} is lower.

However, these adaptations are at the discretion of the expert and there does not appear to be any objective criteria today. As indicated above, should p be significantly different than $\hat{p}$, the Richardson extrapolation would not be applicable. One could then take U_{DE} equal to $F_S \,.\, |y_3 - y_1|$ with F_S equal at least to three.

2.3.2. Input uncertainties UINP: types and propagation

Types of input uncertainties

Two types of uncertainty are often distinguished in the literature: random uncertainty and epistemic uncertainty.

Random uncertainty, also called irreducible, stochastic, **aleatory uncertainty** or **variability**, is inherent to the variations observed within a population of realizations, generally close but not identical. It is characterized by a Probability Density Function (PDF) of the value of a parameter. This is for example the case of material properties such as the compressive strength of concrete: on a set of standardized cylindrical specimens, the compressive strength value measured by a press will be different from one test to another. If one is often satisfied with an average value, it is more interesting, if one has enough data, to represent this resistance by a stochastic variable to which is associated a probability density.

Epistemic uncertainty, also called **reducible uncertainty**, or **uncertainty of ignorance**, is characterized by an interval in which no probabilistic information is available. It is possible to have a variable degree of confidence in this interval. Measurement accuracies are generally associated with epistemic uncertainties. An office ruler graduated to the millimeter will only give information about a measurement within ± 0.5 mm, without any specific value being more likely within this interval.

Propagation of uncertainties to obtain UINP

UINP is obtained by propagating the main uncertainties through the model. The uncertainties propagation can be summarized by these successive steps:

1. The input parameters with the most influence on the EDP of interest must be identified. This can be done with some expert recommendations and/or based on sensitivity analyses methods such as Morris (Herman, 2013), FAST (Mara, 2009), or Sobol methods (Dimov et al., 2010). To keep the computation cost as low as possible, it is advised to limit the number of parameters to the most influential and to calibrate the remaining ones (deterministic).

2. The input uncertainties must be assigned each a representative and well-defined random distributions.

3. Then, a random sampling method (Monte Carlo or quasi Monte Carlo methods) is applied to the epistemic and aleatoric uncertainties. The number of random realizations depend on how sensitive the EDPs are, how fast the computations are, and how accurate the results are expected to be.

4. Finally, for each set of samples (where aleatory realizations are considered given a fixed epistemic random value), the Cumulative Probability Distribution (CPD) of the EDP of interest is computed. One gets as many CPDs as the number of random epistemic values. Figure 4 illustrates this step where five colored CPDs (plotted as Probability of Exceedance) are plotted altogether and **UINP** is obtained by taking the overall black envelop, called the probability-box, or p-box. In the provided example, Latin Hypercube Sampling method was used to generate 5 random epistemic values and Crude Monte Carlo Method was used to generate 30 random aleatory values.

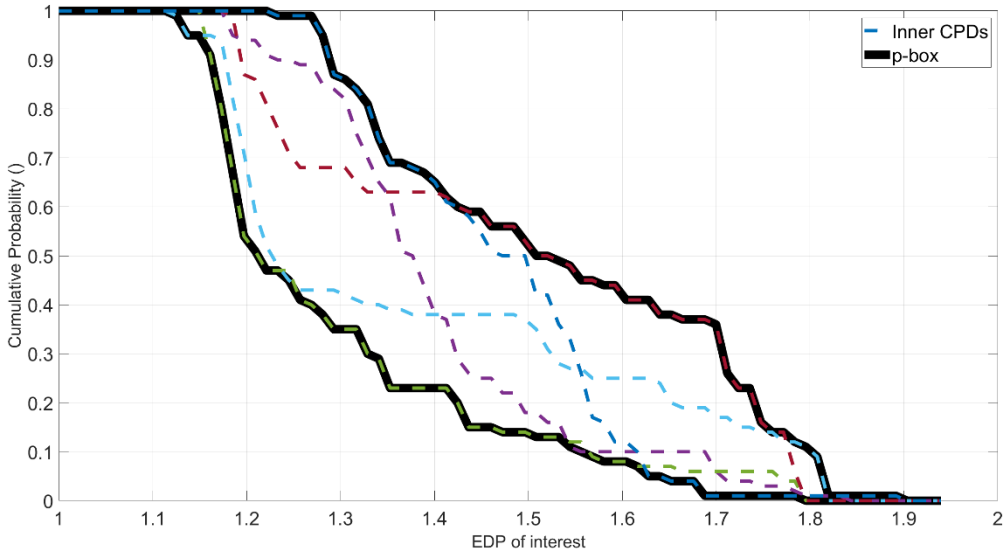


Figure 4: Scheme example of five CPDs combination shaping a probability-box

The p-box can also be used to encompass experimental uncertainties in terms of measurement.

2.4. Validation flowchart step by step

Now that the importance and unfolding of the uncertainties calculation have been introduced, it is possible to present into more details the validation flowchart. It can be divided into three parts: the upper-left and upper-right parts (respectively Experimental analysis and Numerical analysis recalled in Figure 5), and the lower part (Accuracy estimation inside/outside the validation domain, Figure 6).

2.4.1. Upper-Left part: Experimental analysis

The upper-left part of the validation flowchart is highlighted by the yellow square of Figure 5. The successive steps could be summarized as following:

1. The validation test cases are analyzed and interpreted. Only the trustworthy data are kept, with the potential uncertainties linked to their measurement. For instance, if a measure is too large to be realistic, its presence in the validation process should be questioned and discarding such unrealistic value should be considered.

2. The main sources of uncertainties are identified. This will feed with information the numerical analysis part to plan the simulation strategy.

3. The experimental reference box consists in setting and detailing the hypothesis of the whole system before running the full numerical analysis. For instance, let us take the example of a structure mock-up subjected to 20 seismic runs, all desired to be only directed in its longitudinal axis. However, among the 20 tests, 10 of them were achieved with the dynamic loading inducing undesired transversal motion. The following questions arise: should these 10 tests be kept despite these experimental missteps, or should they be discarded to keep more reliable data, but in less quantity, or should the numerical model be adapted to account for potential out-of-plane motions? The answers to these example questions depend on the user's final objective and the foreseen domain of validation.

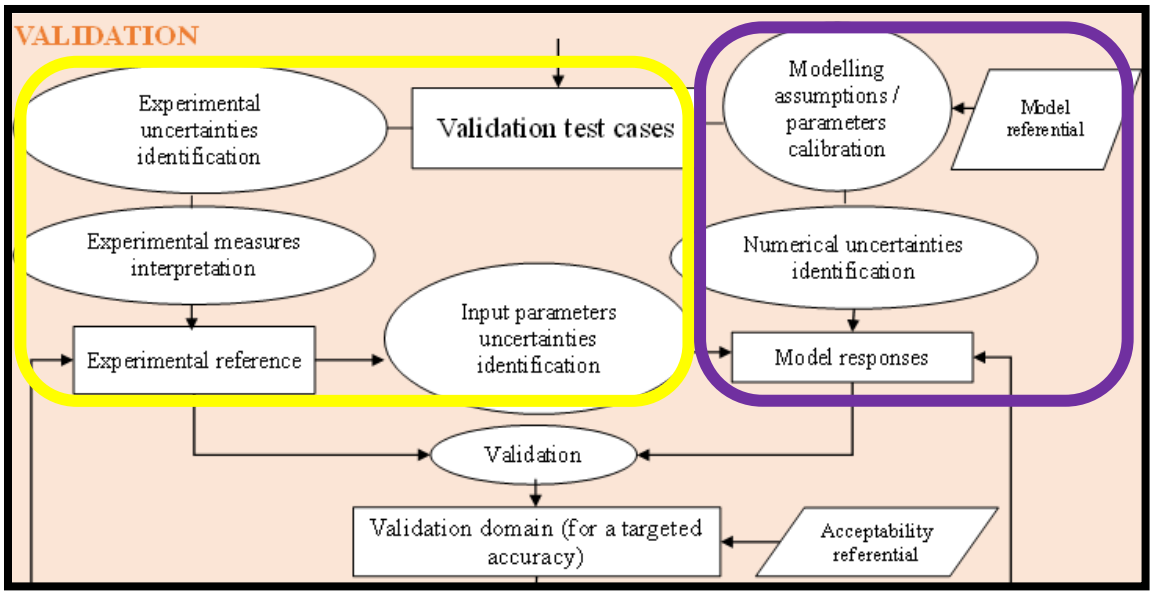


Figure 5: Experimental and Numerical analyses of the validation flowchart

2.4.2. Upper-right part: Numerical analysis

The upper-right part of the validation flowchart is highlighted by the purple square of Figure 5. The successive steps could be summarized as this:

1. The experimental analysis yielded notably the main input uncertainties identification along with the experimental reference framework. This is where the decision of the nature of the model should be defined, and a first "draft" built. At his discretion, the engineer should aim first at a model with a reasonable computational cost.

2. A first step of parameter calibration must be executed to narrow their range of uncertainties, especially when dealing with complex constitutive behavior laws. The validation test cases to the least (i) help to check if the output of interest has realistic value, (ii) ensure that no convergence issues show up, and (iii) allow for having a first evaluation of the computation runtimes.

3. After this first set of analyses, if the model nature and design are considered satisfying enough, the numerical uncertainties **UNUM** must be assessed. Various versions of the model with increasing degrees of discretization should be computed.

Despite perhaps not having a clear view of the input parameters values, or having input parameters with large variabilities, they should all be set to a single value. Each model should be subjected to solicitations of various intensities. This will help to spot, (i) any computational convergence issues, (ii) if the results between each model remain realistic for different ranges of intensities, (iii) if the results remain realistic, if not it would help to correct the parameter calibration. **UNUM** can then be assessed as presented in Section 2.3.1.

4. Once the model nature and discretization have been chosen, a rigorous way to check the input parameters assumption issued from the experimental analysis would be to run sensitivity analyzes such as Morris (Herman, 2013), FAST (Mara, 2009), or Sobol methods (Dimov et al., 2010). Such methods come with a computation cost (lower for Morris, larger for Sobol) and may not always be available and applicable in time for the engineers. This initial cost can be beneficial if it allows to classify the input parameters and thus reduce the uncertainties from the less to the most influential.

   This step is not mandatory if engineers already trust enough their knowledge on the model behavior and the experimental data. Otherwise, it could be an elegant way to justify the choice of the input parameter uncertainties, or to help identifying one between many.

   One must point out that the studied EDPs might not have the same sensitivity to the input parameter, which may lead to difficult decision-makings. For instance, in (Crozet et al., 2017), the authors showed after running a Sobol method that the contact element stiffness k_c has very small influence on the maximum structural displacement but has much larger importance on the floor response spectra at high frequency. So the list of influential inputs is dependent on the considered EDP.

5. Once the input uncertainties are chosen, and given a probability density function, the method presented in Section 2.3.2 can be applied to compute and obtain **UINP** for each validation case kept from the experimental analysis.

2.4.3. The validation domain

The validation domain is the domain inside which the experimental data are available. It can include the system boundary and environmental conditions (e.g., insulated, embedded, temperature, pressure, …) during the experiments. In (Roy and Oberkampf, 2011), authors studied the flow of air through a wind tunnel. The tunnel section output temperature under three different stagnation pressures, 7 MPa, 10 MPa, and 12 MPa, is the output of interest. The validation domain corresponds here to these three distinct pressures. The numerical model corresponds here to the analytical Euler equation and is only validated inside this domain where the experimental data is available. The validity of running simulations with a stagnation pressure of 11 MPa or 20 MPa is to be demonstrated, as no experimental test was performed at 11 MPa or 20 MPa. It is consequently up to the expert and engineer opinions to decide whether the numerical model is valid or not at these conditions of interest.

The same reasoning can be applied within civil engineering domain. For instance, let us assume a structure mock-up subjected to four seismic records of PGAs = [0.05, 0.1, 0.15, and 0.2] g. The numerical model is chosen, and the numerical and input uncertainties are assessed as presented above, showing satisfying accuracy to the project leaders, with very little difference between experimental and numerical EDPs. One could then assume that the model is as well validated for a seismic record of PGA 0.11 g as it is relatively close to the validation domain, even if technically outside of it. Moreover, 0.11 g is located between the points of the validation domain, where interpolation brings generally less uncertainty than extrapolation (depending on the distance between the points). The assumption is more difficult if a ground motion of PGA 0.5 g were to be applied, as it is much further away from the validation domain, especially considering the presence of nonlinearities at this level of loading.

2.4.4. Lower part: Accuracy inside/outside the validation domain

The center-lower part of the validation flowchart is highlighted by Figure 6. The possible steps could be summarized as this:

1. A metric is a measure of a difference between two entities, whether they are scalar, vectors, areas, etc. For instance, the relative error in percentages between two scalars is a metric. The metric, chosen by the engineer, is calculated for each of the validation cases available. This allows then to plot the metric versus the conditions of each test and analyze if/where the model tends to overestimate/underestimate reality. Engineers can already decide (i) whether the model is accurate enough or not, (ii) if the **UINP** p-box is too wide and needs further investigation of the input parameters, (iii) if more experimental tests are needed.

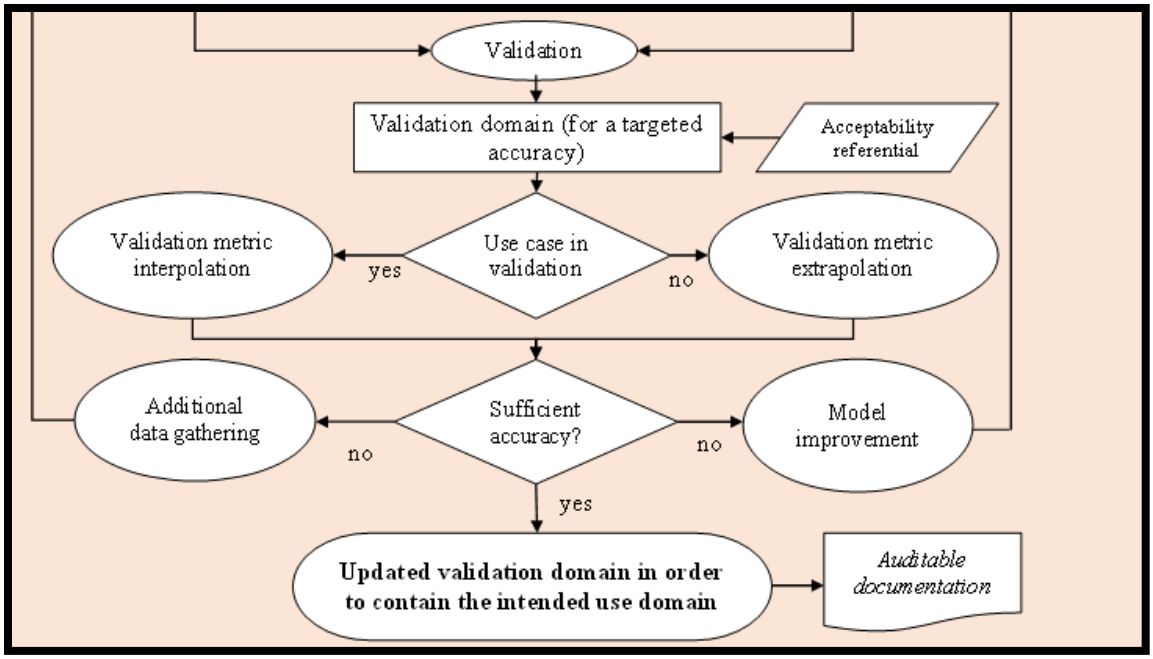


Figure 6: Lower part of the validation flowchart

2. One can check if unexpected or undesired phenomenon appear in the simulations, and if so, if it is detrimental towards the model acceptability. For instance, in reinforced concrete simulation where rebars are considered linear elastic in first assumptions, to check if no significant irreversible deformation has occurred.

3. If the model accuracy is satisfying, then simulations under conditions outside the validation domain may be applied if deemed reasonable and justified by the engineers. The main question could be: What would be the probability that the structure inter-storey drift overpasses 1% if subjected to an earthquake with a PGA record namely Outside the Validation Domain, namely PGA_{ovd} ?

   If so, then a relevant PGA_{ovd} record is selected.

   a. **UNUM** is kept as before or re-assessed at this given intensity as described in Section 2.3.1.

   b. **UINP** is computed for this ground motion as in Section 2.3.2.

   c. Finally, the metric is extrapolated (or interpolated) at the condition PGA_{ovd} , it is named Model Form Uncertainty, **UFORM**, and will be more detailed in the following subsection.

d. Then, the total combination of uncertainties can be executed and shape the final probability-box by a direct summation of the above uncertainties **UTOT** = **UNUM** + **UINP** + **UFORM** at PGA_{ovd}.

Next section presents how (Roy and Oberkampf, 2011) performed the calculation of **UFORM** using the area validation metric.

2.5. UFORM calculated with (Roy and Oberkampf, 2011) method

The model formulation uncertainty **UFORM** is basically the error between the experimental measurements' distribution and the numerical output accounting for the input and numerical uncertainties. Its detailed explanation implies beforehand to define what the validation domain is. Then, because numerical output and measurements may be represented by CPDs, the errors cannot be calculated with simple mathematical differences, so the area validation metric is introduced. Afterwards, a proposition of improvement is presented. Then, in case **UFORM** needs to be assessed outside the validation domain, two types of extrapolation methods and prediction intervals are introduced.

2.5.1. UFORM inside the validation domain

When comparing the numerical and experimental output inside the validation domain, **UFORM** corresponds to their inferred model error. It corresponds simply to how much apart the numerical outcomes are from the "real" observed response.

Example: the maximum displacements on a mock-up subjected to four seismic records of [0.0520, 0.124, 0.1530, 0.402] g Peak Ground Acceleration (PGA) are measured "exactly" and respectively equal to [0.3, 1.0, 1.2, 1.9] cm. Four computations at these conditions are then ran, yielding the following displacements, respectively [0.35, 0.7, 1.3, 1.5] cm. The discrepancies **UFORM** between the experimental data and numerical outcomes can be calculated with metrics of any kind depending on the engineer choice:

A) If the metric corresponds to the absolute error, then **UFORM** equals respectively [0.05, -0.3, 0.1, -0.4] cm.

B) If the metric corresponds to the relative error, then **UFORM** equals respectively [17, 30, 8.3, 21] %.

It seems that the model both overestimates and underestimates the experimental data over the validation domain with larger errors when underestimating. The following subsection presents the area validation metric $\boldsymbol{d_0}$.

2.5.2. Area Validation metric d_0 to calculate UFORM as in (Roy and Oberkampf, 2011)

In the simple example above, no input parameters uncertainties were considered, all the parameters were considered calibrated to a single value, yielding one numerical output for each computation. It makes the experimental/numerical comparison easy to calculate each **UFORM** as a simple difference of two scalars.

Nevertheless, accounting for input uncertainties yields a **UINP** probability-box, as presented in subsections 2.3.2 above. In such cases, one needs a way to distinguish the discrepancies brought by the variability of the input parameters, and the one related to the model formulation itself.

The area-validation metric, namely $\boldsymbol{d_0}$, allows such operation. It corresponds to the absolute sum of the areas between the numerical values and the experimental data, as it is used in (Roy and Oberkampf, 2011) work.

$$\boldsymbol{d_0(Num, Exp)} = \int_{-\infty}^{+\infty} \left| \boldsymbol{CPD_{Num}(x) - CPD_{Exp}(x)} \right| \cdot \boldsymbol{dx}$$

With x the system response quantity of interest, i.e. the EDP in this context, CPD_{Num} the CPD of the simulation results obtained by the input uncertainties propagation, and CPD_{Exp} the CPD of the experimental test. Figure 7 displays a scheme of the area validation metric (green area) drawn from the numerical external CPD (black line) of the probability box of numerical output (grey area) and the experimental measurement defined equal deterministically to 1.65 here (red line). In this example, $\boldsymbol{d_0}$ ≈ 0.3.

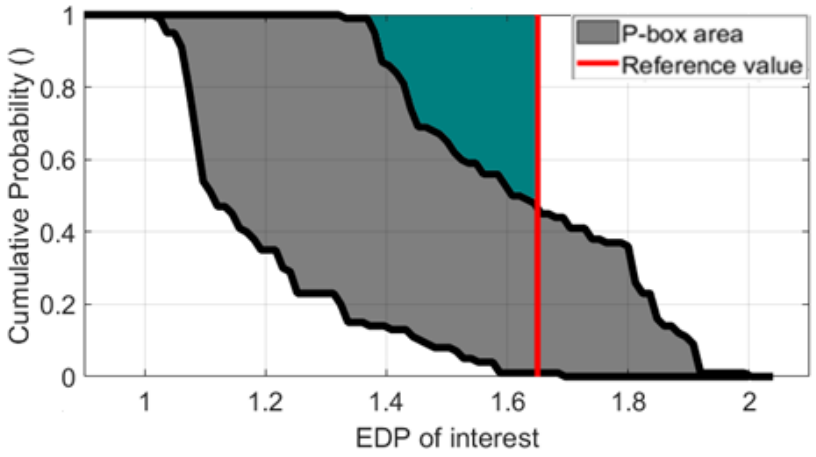


Figure 7: Scheme example of the area validation metric $d_0 = 0.3$

The area metric quantifies the mismatch between predictions and observations that are not explained by inputs or numerical uncertainties. Model acceptance is case-specific and is determined according to a prescribed accuracy requirement.

2.5.3. Linear Regression and Student's 95% prediction interval

Using the above example, uncertainty arises with the following question: how does the model behave outside the validation domain, let us say at 0.75g PGA intensity record? At such intensity, there is no experimental data available to help to assess the model accuracy. Nevertheless, one must quantify this accuracy beyond the validation domain.

UFORM is therefore extrapolated at PGA 0.75 g (outside the validation domain) by using the $\boldsymbol{d_0}$ metrics values formerly calculated from inside the validation domain by using any type of quadratic regression or conservative assumption the engineers deem justified. (Roy and Oberkampf, 2011) uses the following method: for a polynomial regression (or other regression formula) $\widehat{\boldsymbol{d}}$ formulated from the four values of model uncertainty $\{\boldsymbol{d_1}, \boldsymbol{d_2}, \boldsymbol{d_3}, \boldsymbol{d_4}\}$ calculated for the four experimental conditions $\{\boldsymbol{C_1}, \boldsymbol{C_2}, \boldsymbol{C_3}, \boldsymbol{C_4}\}$ the authors relied on Student's law (generalization of the standard normal distribution), see (Devore, 2015), to estimate the extra model uncertainty related to this extrapolation. This relies on the assumption that the data follows such distribution.

$$\widehat{\boldsymbol{d}} \pm \boldsymbol{t_{\frac{\alpha}{2},N-p}} \cdot \boldsymbol{s} \cdot \sqrt{\boldsymbol{1} + \frac{\boldsymbol{1}}{\boldsymbol{N}} + \frac{\boldsymbol{N} \cdot (\boldsymbol{C} - \overline{\boldsymbol{C}})}{\boldsymbol{N} \cdot \left(\sum_{i=1}^{N} \boldsymbol{C_i^2}\right) - \left(\sum_{i=1}^{N} \boldsymbol{C_i}\right)^2}}$$

where:

$\boldsymbol{t_{\frac{\alpha}{2},N-p}}$ is Student's law of parameters $\frac{\alpha}{2}$, N, and p ;

$\boldsymbol{\alpha}$ is the sought confidence level (e.g. 95%);

$\boldsymbol{N}$ is the number of different conditions for which measurements are available (e.g. four for $\{\boldsymbol{C_1}, \boldsymbol{C_2}, \boldsymbol{C_3}, \boldsymbol{C_4}\}$, that could correspond to four values of PGA) ;

$\boldsymbol{p}$ is the number of parameters of the polynomial regression (e.g. 2 if it is a linear regression);

$\boldsymbol{s}$ is the root mean square error of the polynomial regression with respect to the available experimental values (e.g. here comparing the values of the regression $\hat{d}$ with respect to the calculated values $\{d_1, d_2, d_3, d_4\}$) ;
$\boldsymbol{C}$ is the value of the operating condition for which one wishes to estimate the model uncertainty (e.g. a high intensity ground motion of PGA 1.78 g like in Section 4.3.8);
$\overline{\boldsymbol{C}}$ is the average of the operating conditions for which measurements are available (e.g. the average of the four different PGAs);

Figure 8 displays a scheme of the linear regression and Student's 95% prediction interval of $\boldsymbol{d_0}$ for a given set of 23 data points.

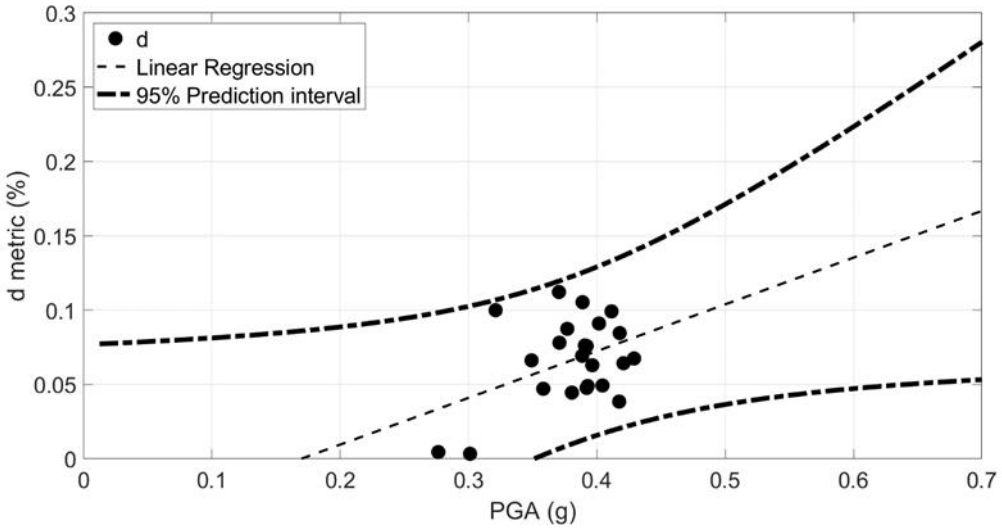


Figure 8: Metric and prediction interval with (Roy and Oberkampf, 2011) method

The black dots are the validation metrics $\boldsymbol{d_0}$ inferred from experimental data and numerical outcomes in the intensity measure domain [0.27 0.43] g, the interstorey-drift with units in percent here. (Roy and Oberkampf, 2011) method is applied. A linear regression (thin dashed lines) and the Student's 95% prediction interval (thick dashed lines) are computed and plotted.

2.5.4. Combination of UNUM, UINP, and UFORM

Let us consider a case where the EDP is the interstorey-drift assessed at a 0.7 g PGA intensity, so outside the validation domain. The propagation of uncertainties is computed to obtain **UINP** at such intensity. The probability-box is constructed. The numerical uncertainties are here taken arbitrarily equal to 0.02% for the sake of the example.

UFORM needs to be evaluated. In this virtual example, if the metric d, i.e., the errors, i.e., the model formulation uncertainties were to be assessed at 0.7 g PGA, engineers conclude that there is 95 % probability that it is greater than 0.05 % and lower than 0.28 % (Figure 8 above). Ultimately, for the final combination of uncertainties at such intensity, **UFORM** will be taken equal to 0.28 at both the left- and right-hand sides of the probability-box formed by **UINP** and **UNUM**.

Figure 9**Erreur ! Source du renvoi introuvable.** presents the total combination of uncertainties.

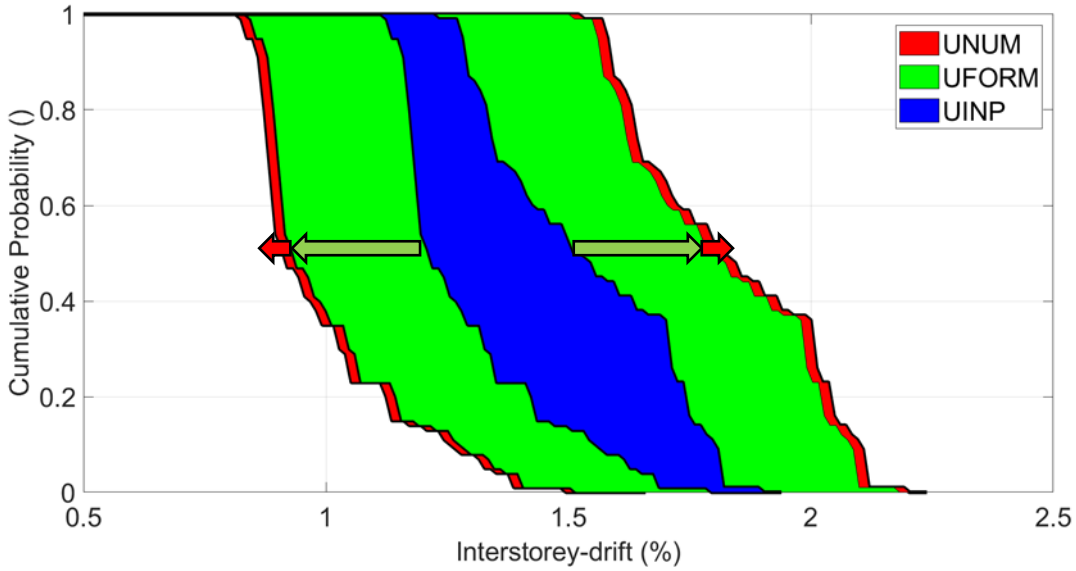


Figure 9: Total combination of uncertainties with (Roy and Oberkampf, 2011) method

For instance, by accounting for the fact that **UFORM** is calculated with a 95 % confidence, the reading of Figure 9 shows that for a 0.7 g, the numerical model will yield with a 95 % confidence the EDP of interest between 0.8 % and 1.4 %. Were the EDP be the interstorey-drift of a given building, this could indicate potentially three different damage states according to (Xiong et al., 2017). A narrower estimation is desirable. It is clearly observable that by applying "blindly" the (Roy and Oberkampf, 2011) in the present framework, the **UFORM** uncertainties are as large as or even larger than the input uncertainties, preventing the engineer from obtaining an accurate estimation.

Next section presents some propositions to improve the accuracy and relevancy of (Roy and Oberkampf, 2011) method.

3. Propositions to improve Roy and Oberkampf method

In the following, the examples given are generic, so the intensities and EDPs units and natures do not matter and are unitless. Three main improvements are proposed in this section and laid out as following:

1) The Overlapping Coefficient (Kwag et al., 2018) is proposed to evaluate the accuracy of a model to an experimental distribution over a validation domain. It can also eventually compare different models, and eventually validate one or several given some criteria. The Overlapping Coefficient is contained between 0 and 1. The closer to 0, the more separated the two distributions are. The closer to 1, the more the two distributions overlap, the more valid the model.
2) The metric used during the validation process can be different from $\boldsymbol{d_0}$. Here are several propositions:
    a. the metric $\boldsymbol{d_0}$ is modified for a more accurate representation of the error. Two metrics are considered, $\boldsymbol{d_m}$ and $\boldsymbol{d_p}$, that respectively represent the areas where the numerical output underestimate and overestimate the expected value;
    b. another metric is proposed: $\boldsymbol{e_{rr}(\alpha_{err})}$. It is the relative difference between the outlines of the probability-box with the experimental value for a given quantile $\boldsymbol{\alpha_{err}}$;
    c. another approach is proposed. The metric is the difference, the error, between the numerical output and the mean or median of the experimental measures. It is not one metric, i.e., the error, calculated for each condition of interest, but it is as many errors as there are propagations samples.

       For example, for each sample of input parameters, the error between the numerical output and the mean or median of the experimental measures is calculated. So, if there are five epistemic and 100 aleatory samples, that totalizes 500 combinations, so 500 errors calculated for each condition of interest over the validation domain. From there, 500 regressions and prediction intervals are computed for each of these combinations.
3) The choice of the regression and prediction methods outside the validation domain:
    a. data points considered as theoretical limit states can be added to obtain more relevant regressions and prediction intervals. For instance, a null value of experimental measure at a null intensity, consequently a null value of numerical EDP, so a null metric;
    b. the prediction intervals might yield unacceptable values or undesired pattern (e.g., a crescent slope rather than an expected decrescent one), especially at small intensities if the commonly used linear regression are automatically and blindly applied of every set of data without previous analysis. So, in this work, it is considered that the regression should be achieved based on the shape of the obtained results (it can be linear, quadratic, exponential, polynomial or others). The choice of the fitting form should vary depending on the application cases.

3.1. Overlapping Coefficient

The overlapping coefficient $\boldsymbol{OC}$ (Kwag et al., 2018) can be used as a global validation metric to objectively quantify the resemblance between two probability density functions PDF $\boldsymbol{pdf_1}$ and $\boldsymbol{pdf_2}$ like shown in Figure 10. $\boldsymbol{OC}$ corresponds to the common area between two curves, is between 0 and

1, has no units. The $\boldsymbol{OC}$ can as well be computed between two normalized curves. A null $\boldsymbol{OC}$ value indicates that the PDFs never overlap, while a value of 1 indicates that the two PDFs perfectly overlap each other. The Overlapping Coefficient will be used to compare the distribution of experimental and simulated values over the validation domain. It is up to the decision-makers to set the value of the critical threshold of ***OC***. This single scalar metric can be used to the discretion of the engineer, like others, evaluate model validation.

$$OC = \int_{-\infty}^{+\infty} \min\left(pdf_1(x), pdf_2(x)\right) dx$$

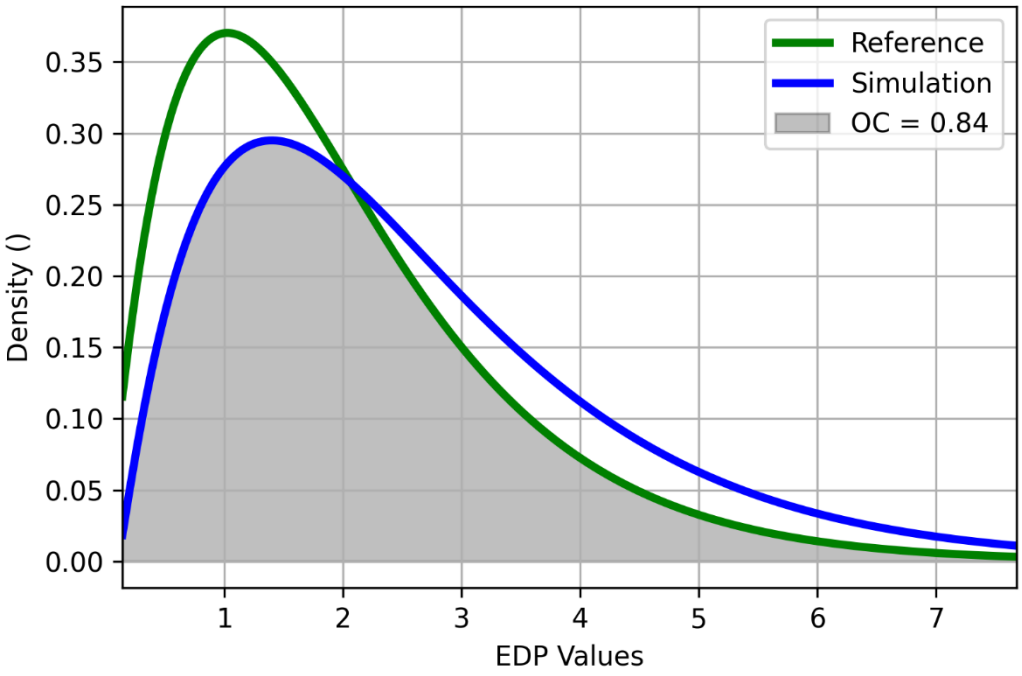


Figure 10: Illustration of the Overlapping Coefficient

3.2. Suggestions of metrics and approaches

Here are some improvements of the area metric used so far in the Validation Process.

3.2.1. From the metric d_0 to the metrics d_m and d_p

One must note that by applying the absolute value operator on the difference of the CPDs, $\boldsymbol{d_0(Num, Exp)}$ is consequently always positive. Therefore, the information regarding whether the numerical model tends to overestimate or underestimate the experimental measurements is lost.

Another issue arises when the experimental and numerical CPDs overlap and yield two areas, one with $\boldsymbol{CPD_{num}(x)}$ greater than $\boldsymbol{CPD_{exp}(x)}$, and vice-versa. Figure 11 below shows such a case, with $\boldsymbol{CPD_{num}(x)}$ represented by the grey area, $\boldsymbol{CPD_{exp}(x)}$ by the red vertical line. In this case, $\boldsymbol{d_0}$ formulation adds in absolute the two green areas (e.g., $\boldsymbol{d_0 = 0.04 + 0.01 = 0.05}$) which is more conservative and less accurate than distinguishing that the model may underestimates the expectations by **0.04** or overestimates them by **0.01**.

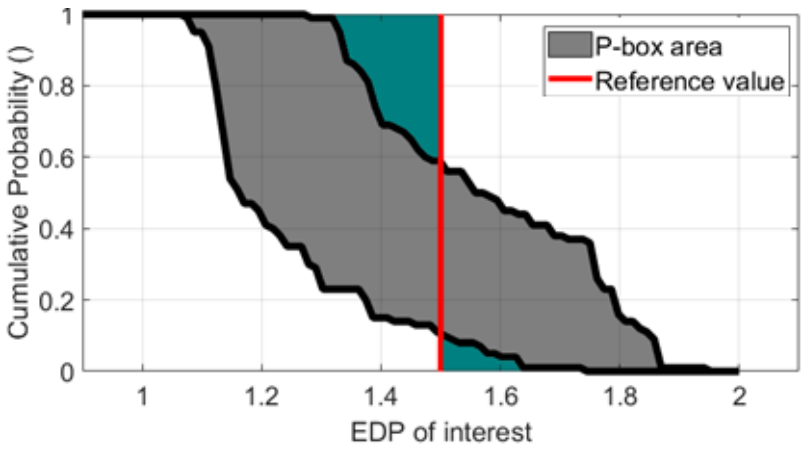


Figure 11: Information loss due to d_0 formulation

Subsequently, it is proposed to separate the areas where the model overestimates and underestimates the foreseen response. Then, for a single model, two metrics are computed, namely $\boldsymbol{d_p}$ and $\boldsymbol{d_m}$, that represents respectively how much the model overestimates and underestimates the experimental measurements. With $\boldsymbol{x}$ corresponding to the EDP inside the domain $[-\infty\,;\,+\infty]$:

$$\boldsymbol{D(x) = CPD_{num}(x) - CPD_{exp}(x)}$$

$$\boldsymbol{D_p(x) = \begin{cases} D(x) > 0 \\ 0\ otherwise \end{cases}}$$

$$\boldsymbol{D_m(x) = \begin{cases} D(x) < 0 \\ 0\ otherwise \end{cases}}$$

$$\boldsymbol{d_p(Num, Exp) = \int_{-\infty}^{+\infty} D_p\ .dx}$$

$$\boldsymbol{d_m(Num, Exp) = \int_{-\infty}^{+\infty} D_m\ .dx}$$

$$\boldsymbol{d_0 = d_p + |d_m|}$$

Their unit is the same as the EDP of interest studied. This proposition shows four different configurations:

The experimental data is larger than the p-box. This means $\boldsymbol{d_p}$ is larger than zero, $\boldsymbol{d_m}$ equals zero, and $\boldsymbol{d_0 = d_p}$. The model tends to underestimate the expected response, as illustrated by Figure 12(a).

The experimental data is smaller than the p-box. This means $\boldsymbol{d_m}$ is smaller than zero, $\boldsymbol{d_p}$ equals zero, and $\boldsymbol{d_0 = |d_m|}$. The model tends to overestimate the expected response, as illustrated by Figure 12(b).

The p-box is both partly greater and partly smaller than the experimental data. This means $\boldsymbol{d_p}$ and $\boldsymbol{d_m}$ are both different than zero, consequently $\boldsymbol{d_0}$ too. There is both a probability to overestimate and to underestimate the expected response, as illustrated by Figure 12(c).

The p-box contains completely the experimental data. This means $\boldsymbol{d_p}$ and $\boldsymbol{d_m}$ both equal zero, consequently $\boldsymbol{d_0}$ too. This is the desired configuration for every test, as illustrated by Figure 12(d).

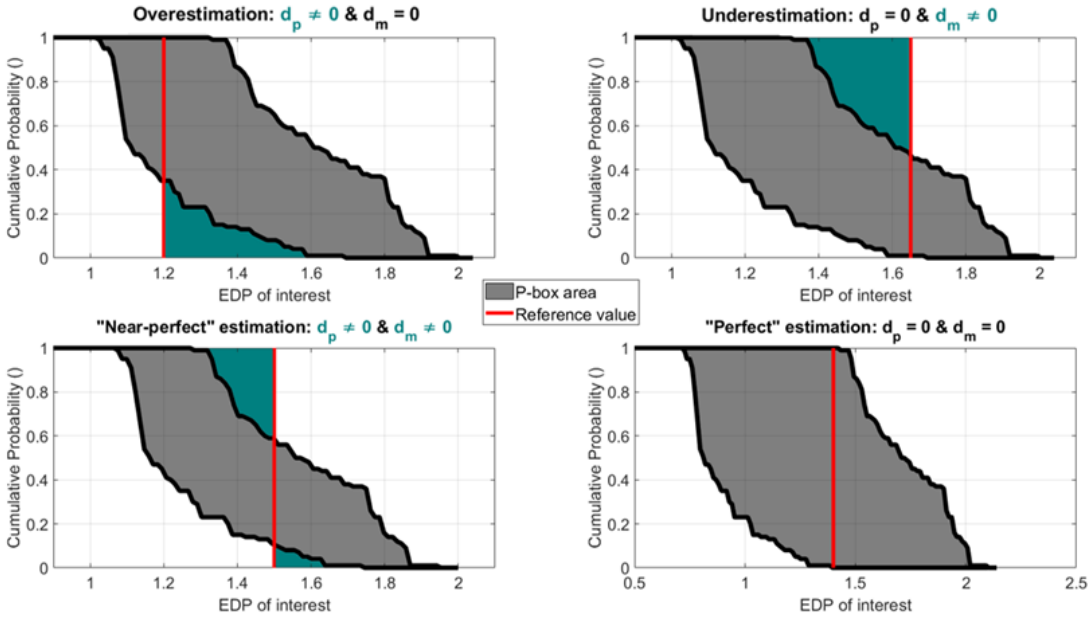


Figure 12: Four different configurations of d_p and d_m

Therefore, it is possible to compute $\boldsymbol{d_p}$ and $\boldsymbol{d_m}$ for each of the data available. By plotting $\boldsymbol{d_p}$ and $\boldsymbol{d_m}$ versus the PGA from the experiments, it is possible to observe an overview of the accuracy of the model, seeing if it either underestimates, or overestimates, the expected response and at which intensities. From there, in case of computations outside the validation domain, it is possible to quantitatively assess **UFORM**. Then, as well as there is now a positive and a negative metric, there are now two positive and negative model formulation uncertainties, respectively $\mathbf{UFORM_p}$ and $\mathbf{UFORM_m}$.

3.2.2. From the area validation metrics d_0, d_m, and d_p to the error e_{rr} for a given quantile α_{err}

Nevertheless, the area validation metrics present one shortcoming: they are not *per se* really EDP values, but areas, i.e., EDP values weighted by probabilities. For example, by looking at Figure 11, one can calculate the $\boldsymbol{d_m}$ and $\boldsymbol{d_p}$ values, which are respectively around 0.04 and 0.01 (no unit of EDP in this example). The experimental value equals 1.5.

For the engineer, these values are intended to represent how far the model underestimates and overestimates what is expected. However, the outlines of the probability-box clearly show that a significant portion of simulations show errors larger in absolute than $\boldsymbol{d_m}$ and $\boldsymbol{d_p}$. Indeed, 36 % of simulations of the right-hand side CPD resulted in EDPs smaller than the experimental measure minus $\boldsymbol{d_m}$, i.e., 1.46. Also, 9 % of simulations of the left-hand side CPD resulted in EDPs larger than the experimental measure plus $\boldsymbol{d_p}$, i.e., 1.51.

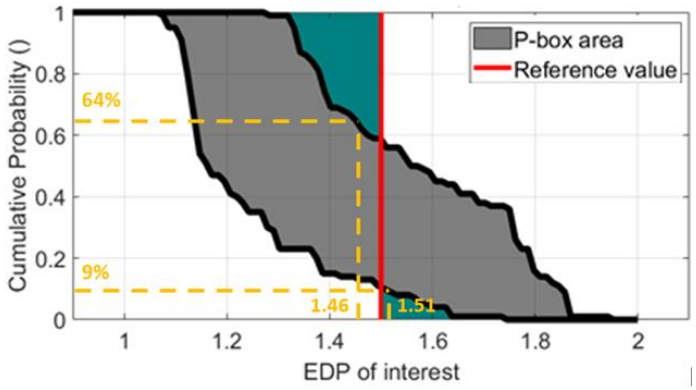


Figure 13: Applicability of d_m and d_p in question

These observations of area validation metrics being significantly smaller than the real differences between experimental and numerical are true throughout many examples. Only when the Probability-box is constituted of vertical CPDs (so, always the same EDP value for all computations) that the area validation metrics equal exactly the differences with the experimental value (because the difference is multiplied by a 100% probability).

A different approach would be to select a fractile α_{err}, for instance 5%, and to read the differences, or errors e_{rr}, between the outlines of the Probability-box and the experimental measure. Figure 14 displays the reading of the differences e_{rr} for a given fractile α_{err} of 5%, namely $e_{rr}(\alpha_{err})$.

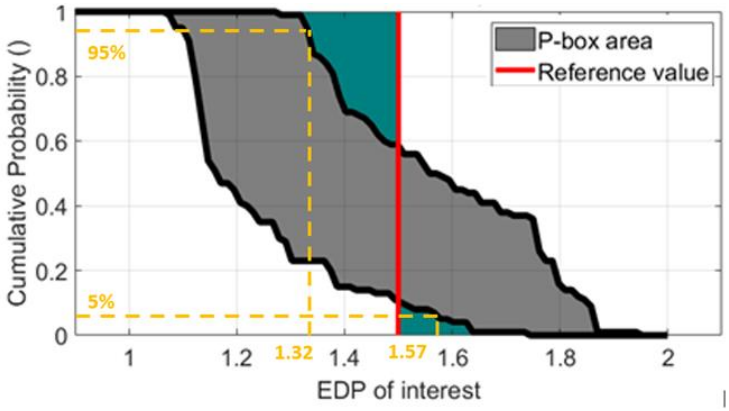


Figure 14: Example of errors e_{rr} reading for a 5% fractile

One can read the EDPs obtained at such quantile, respectively 1.32 and 1.57 for probabilities of exceedance of 95% and 5%. Then, the differences with the experimental value of 1.5 allows the calculation of the errors e_{rr}(95%) and e_{rr}(5%) at theses quantiles, i.e., respectively -0.18 and +0.07. They are notably larger in absolute than d_m and d_p that equal -0.04 and +0.01, whose acceptability is now reconsidered as exists significant probabilities that the numerical/experimental differences might be higher. However, they are "real" values, differences, that are likely to be overpassed with a 10% probability.

This proposition involves setting a quantile α_{err} upon which the reading of the two errors e_{rr}(1-α_{err}) $e_{rr}(\alpha_{err})$ will be done. An error can be either positive, negative, or null depending how the probability-box envelops or not the experimental value. A 10% quantile would have resulted in a negative e_{rr} (90%) (so, again underestimation) but in a null e_{rr} (10%). A 0% quantile would have resulted in the two maximum differences, here respectively -1.9 and +1.3.

3.2.3. Calculating the error for each input parameter sample

Whether it is the $\boldsymbol{d_0}$ metric, or $\boldsymbol{d_m}$ and $\boldsymbol{d_p}$, or $\boldsymbol{e_{rr}(\alpha_{err})}$ and $\boldsymbol{e_{rr}(1-\alpha_{err})}$, the subsequent steps are the same:

- The input parameters are propagated at each experimental condition where data is available. The same distribution of parameters is taken for each condition.
- The metrics are calculated once for each experimental condition (one model output *distribution* being obtained for each EDP value, for example).
- Whatever their nature and level of confidence, a regression and prediction interval of the metric are plotted to estimate the level of correction inside and beyond the validation domain.
- If any study outside the validation domain is intended, the same input parameter propagation is used, and the resulting numerical probability-box is corrected with the values taken from the prediction interval at the condition of interest.

As a reminder, **Figure 8** illustrates the list above.

This subsection presents a different approach. This time, the regression and prediction interval are calculated for the errors of each individual set of input parameters. For instance:

- Let us picture an experimental campaign with three measurements, namely $\boldsymbol{EDP_{expe_j}}$, taken from three different experimental conditions, with $\boldsymbol{j = 1, \dots, 3}$.
- Two epistemic values and 60 aleatory values are sampled, resulting in $\boldsymbol{N = 120}$ different combinations, computed for each experimental condition $\boldsymbol{j}$. $\boldsymbol{EDP_{i,j}}$ are obtained, with $\boldsymbol{i = 1, \dots, N = 120}$ and $\boldsymbol{j = 1, \dots, 3}$.
- For each the three conditions of interest $\boldsymbol{j}$, the errors $\boldsymbol{e_{rr_{i,j}}}$ are calculated, with $\boldsymbol{i = 1, \dots, 120}$ and $\boldsymbol{j = 1, \dots, 3}$.

$$\boldsymbol{e_{rr_{i,j}} = EDP_{i,j} - EDP_{expe_j}}$$

- Figure 14 displays the errors $\boldsymbol{e_{rr_{i,j}}}$ as colored dots for each of the input parameters samples and experimental condition over the validation domain, schemed by the vertical dashed lines.
- Then, for each of the parameter samples $\boldsymbol{i}$, the regression and prediction interval are computed. For a sake of simplicity, only linear regressions (in thin colored lines) are plotted on Figure 14. It represents the behavior of the model response for each parameter input set, and its prediction beyond the validation domain.
- Now, let us picture that the engineer needs an estimation of the EDP at an intensity measure of interest named IMI, outside the validation domain, for instance in seismic analyses a higher ground motion PGA. The same sampling of input parameters is taken to run again $\boldsymbol{i = 1, \dots, N = 120}$ simulations, and obtain 120 new $\boldsymbol{EDP_{i,IMI}}$. An illustration scheme of its histogram and probability-box is displayed in red color in Figure 15.
- For each set $\boldsymbol{i}$ of input parameters, the lecture of the $\boldsymbol{e_{rr_{i,IMI}}}$ is done directly at the abscissa IMI in Figure 14. From there, the corrected values of EDP, namely $\boldsymbol{EDP'_{i,IMI}}$, can be calculated.

$$EDP'_{i,IMI} = EDP_{i,IMI} + e_{rr_{i,IMI}}$$

- Figure 15 displays the new histogram and probability-box of the $EDP'_{i,IMI}$ in green colour. Each individual set of input parameter output having been corrected, it is not abnormal that the shape and width of the new histogram and probability-box in green are different from the original ones in red.

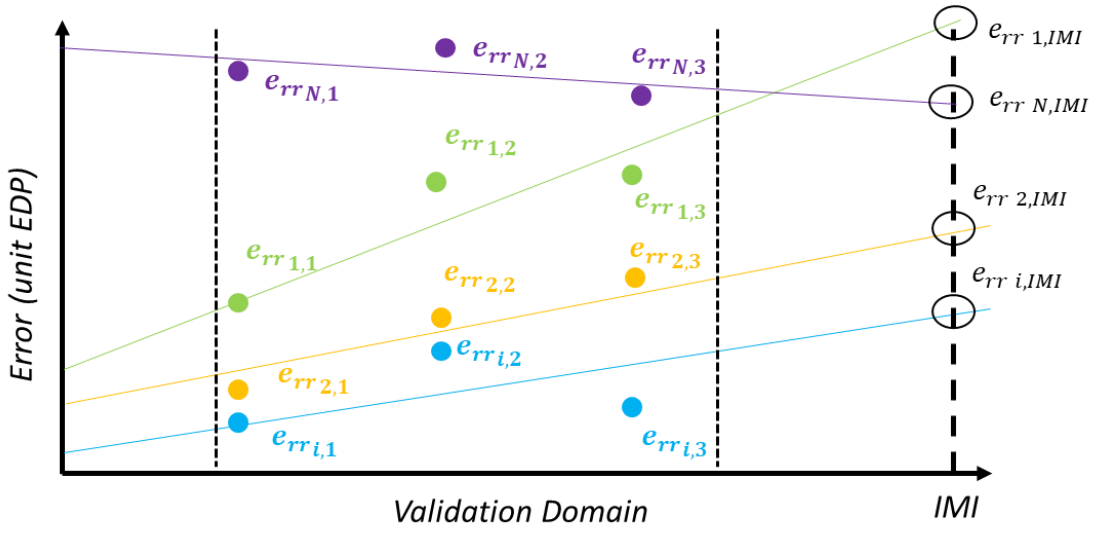


Figure 15: Illustration scheme of regression of the individual errors of each sample of input parameters

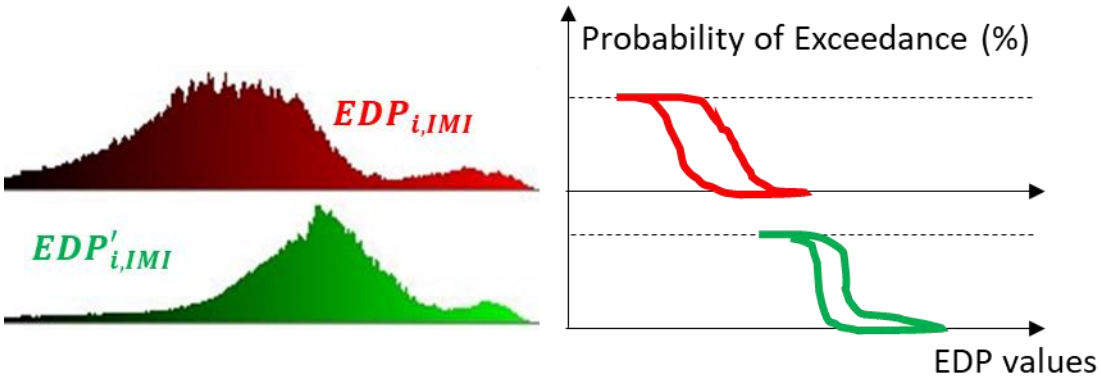


Figure 16: Illustration scheme of the $EDP_{i,IMI}$ histograms and p-box before/after correction

This new approach is more advanced, allows a direct correction on the numerical output and therefore seems richer in information than the previous propositions d_0, d_m, d_p where the measure of the error seems not to be conservative enough. Regarding $e_{rr}(\alpha_{err})$, despite it being of the same nature of the EDP studied, the engineer is bound to set a quantile α_{err} upon which doing the prediction. This puts an additional layer of uncertainty that adds up to the already existing computation of the prediction interval usually taken with a 95% confidence.

3.3. Regression and prediction methods

The present subsections aim simply at highlighting the fact that applying automatically and blindly a linear regression on the data without analyzing first its scatter might lead to inappropriate regression and prediction intervals.

3.3.1. Adding data for more relevant regressions

Adding data

In case where there are strong variabilities of the model accuracy, such method can lead to very wide prediction intervals, and very large, potentially unrealistic, values of **UFORM**. For instance, from null to very low intensities, the model is expected to yield null to very low EDP values, so relatively null or very small uncertainties on the matter. **One additional proposition is to add a zero value at the null intensity** (i.e., d_p=d_m=0 at IM=0) **to provide the regression and prediction method with relevant information.**

Another practical example where adding data is relevant to enrich the regression, whatever its nature, is the frequency loss of a building after seismic solicitation. Indeed, frequency losses are limited EDPs, they cannot be higher than the initial eigenfrequency. Consequently, it can reasonably be assumed that at a certain high intensity to define, the frequency loss reaches an asymptotic threshold. Taking the example of the IDEFIX reinforced concrete beams (METIS report, 2022) (Heitz, 2017), it is known in large intensities that the beams behavior converges towards the steel rebars behavior when the concrete state nears complete damage.

Illustration of adding data and its effects on regression

Adding a null value at the null intensity for a more accurate et realistic regression might increase the slope of the regression, whatever its nature. Figure 17 illustrates three distinct cases of regressions with and without adding a point at null intensity (red dot at origin of axes).

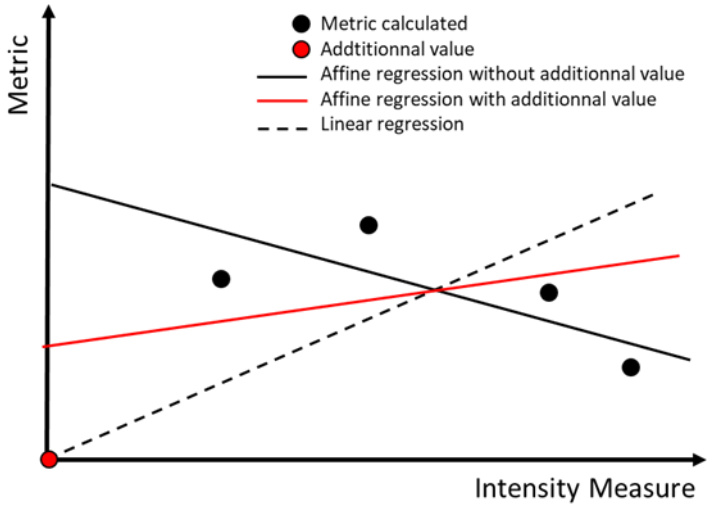


Figure 17: Differences of slope in linear/affine regression with/without null additional data

The decrescent solid black line is an affine regression without accounting for the additional point. Its negative slope predicts smaller values of metric, so a model output more accurate, as the intensity measure grows. The solid red line is an affine regression that accounts for the initial value and has a positive slope in comparison of the solid black line. So, adding an additional null value resulted in

increasing the metric prediction values at higher intensities. Finally, the dashed black line is the linear regression (that naturally accounts for the red dot) and has an even larger slope than the previous one. In the end, conversely, adding a null value seems to increase the metric prediction at large intensities.

In such cases, no regression is satisfying as:

- If the affine solid black or solid red slope is chosen, the low intensities predictions are overconservative.

- If the affine red slope or black dashed is chosen, the decreasing regression pattern is lost, we obtain a positive slope.

N.B.: the number of available experimental measurements in civil engineering frameworks is generally relatively small, i.e., less than 30. This most commonly entails the use of affine/linear regressions. To the opinion of the author, the number and scatter of data taken in the present example, as well as the choice of using 1-D polynomial regression, are representative of many application cases seen so far (Crozet, 2019) (Heitz, 2017) (METIS report, 2022).

3.3.2. Acceptability of regressions and prediction intervals

We saw in the previous subsection the relevancy of adding metric values, but its potential effects on the regression pattern.

First, taking the same illustration plot as before, let's now ignore the affine red line, and plot the 95% prediction intervals as in (Roy and Oberkampf, 2011), as displayed by Figure 18.

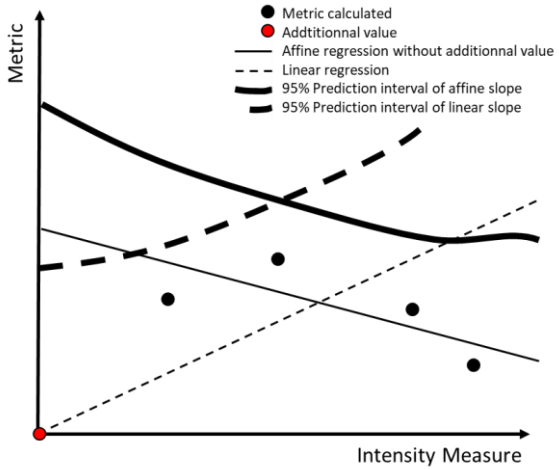


Figure 18: Illustration scheme of prediction intervals from different regressions

As expected, the two upper limit prediction intervals have respectively a negative trend (in solid thick line) and a positive trend (in dashed thick line), as the first curve better represents the model behaviour inside the validation domain and the second curve between the origin and the entrance in the validation domain.

4. Application to SMART2013 experimental campaign

First, the SMART 2013 experimental campaign is presented. Then, the Verification and Validation process is applied, with the presentation of the numerical model used, and the comparison between the original method and the various propositions of improvement.

4.1. Step-by-step summary of the proposed V&V methodology

The application presented in this section will follow the steps below:

1. Analysis of the experiments, measurements, uncertainties, correlations. Choice of the EDP and IM of interest. Definition of the validation domain and the experimental reference.

2. Hypotheses, first numerical modelling, and calibration of input parameters. Set the thresholds of relative error, overlapping coefficient, metrics, probabilities, trend, shape, number of convergence issues upon which the model could be set as "accurate and reliable enough".

3. Resources available for additional studies (parametric and sensitivity analysis, spatial-time discretization, EDP vs. IM correlation, other modelling choices, etc.)? Example: sensitivity analysis's Morris Diagram is displayed. The Young Modulus E has on average more effect on a given EDP than other parameters. The more the knowledge, the faster the simulations, the better.

4. Final choice of the epistemic and/or aleatory variable(s), type and number of samples, IM of interest, and final calibration of the input parameters. Here, 30 Quasi Monte-Carlo correlated samples between Young Modulus and Tensile resistance.

5. For each experiment, so for each IM, run the simulations (i.e., the propagation of uncertainties) and display the output on the validation domain for a first analysis. Here, the 95% grey area interval is plotted with the mean (in black) + 2 times the standard deviation.

6. In a like manner, compute over the validation domain the mean and 95% prediction interval of the Relative Error and the integrated Overlapping Coefficient. Limit the value of the OC to 1.

7. For each IM value, and for each epistemic value, compute the CPD. Then, for each IM, compute the P-Box (in grey area) shaped by the CPDs (thin lines). Plot them with its respective experimental CPD or experimental P-Box (in green area) for comparison. In blue, the area where the experimental P-box is above the numerical P-Box, and in red vice-versa. Here, IM=0.5.

8. For each IM, retrieve the distances between the P-Boxes, sort them in crescent order, and plot the CPD of this distance between the P-Boxes.

9. For each IM value, read the probabilities that the numerical model, by accounting for the epistemic and aleatory variables and their propagation, underestimate/best-estimate, and overestimate the respective measurement. Again, in this example, IM=0.5.

10. For each IM value, report the probabilities on the validation domain.

11. Set a confidence level α_{err}, here of 5%. For each IM value, read the distance, or errors e_{rr}, values at the two quantile values $\frac{\alpha_{err}}{2}$ and $100 - \frac{\alpha_{err}}{2}$.

12. Report the two metrics retrieved, here $\alpha_{err}(97.5\%)$ and $\alpha_{err}(2.5\%)$ respectively in golden downward and upward triangles. If additionnal physical data can be added, e.g., a null metric at IM=0, then add it. Then, compute their respective regression, here affine ones in dashed lines, and the Student's prediction lower and upper intervals. Always check the representativity of the envelop especially where there is not much data and that they present a high scatter pattern.

13. The engineer must decide whether the model is "accurate and reliable enough" over the validation domain and beyond. If not:

a. Obtain more experimental measurements at IM values where there is a lack of information.

b. Improve the numerical modelling strategy and input parameters.

c. Achieve additional studies if resources are available (parametric, sensitivity analysis, spatial-time discretization, EDP vs. IM correlation, other modelling choices, etc.). The more the knowledge, the more accurate the simulations, the faster the simulations, the better.

14. If the model is considered accurate and reliable enough, and if the engineer wants to know the value of the EDP of interest outside the validation domain, then retrieve the metrics $e_{rr}(97.5\%)$ and $e_{rr}(2.5\%)$ on the prediction envelop for the IM of interest.

15. Run the same propagation of uncertainties and input parameters at the IM of interest (outside of the validation domain). Compute the CPDs and the final P-Box. Then translate the two envelops with $-\boldsymbol{e_{rr}(97.5\%)}$ and $-\boldsymbol{e_{rr}(2.5\%)}$.

16. The study outside the validation domain can be achieved for other IM values. The subsequent figure presents the intervals of EDP after correction by the metrics extrapolated outside the validation domain.

4.2. Presentation of SMART2013 campaign

Following the analysis of the SMART2008 experimental campaign (Richard et al., 2008), a new mockup SMART2013 geometrically identical to that of the SMART2008 campaign was built (cf. Figure 19) and tested on the AZALEE vibration table at the TAMARIS facility (CEA) in July 2013. This is a ¼ scale part of auxiliary nuclear building designed according to French guidelines (Richard et al., 2018). The aim is to complete the studies on the ability of buildings to withstand seismic loads and transmit these loads to their equipment. Only the footing of this new model was redefined to improve anchoring conditions on the test fixture compared to SMART2008.

Figure 19: SMART2013 mock-up (DEN, 2014)

The SMART2013 mock-up is subjected to bi-axial horizontal loads. Three sets of seismic were selected for these tests:

1) Signal 1: Design signal (synthetic "SMART2008" signal),

2) Signal 2: Main shock from the Northridge earthquake (recorded by the Tarzana Cedar Hill-1994),

3) Signal 3: Aftershock of the Northridge earthquake (again, recorded by the Tarzana Cedar Hill-1994).

The experimental sequence was carried out as follows:

1) Signal 1: 2 consecutive tests respectively at 50 % and 100 % of the nominal signal level,

2) Signal 2: 5 consecutive trials at 11 %, 22 % (repeated twice due to undesired vertical motion during the first trial), 44 % and 100% respectively of the nominal signal level,

3) Signal 3: 2 consecutive seismic tests at 33 % and 100 % of nominal signal level respectively.

Table 1 presents the successive signals applied on SMART2013 mockup as well as the target and actually achieved PGAs.

	Target PGA	Achieved PGA
Run 07: 50% Design Signal	X axis: 0.10g Y axis: 0.13g	X axis: 0.10g Y axis: 0.14g
Run 09: 100% Design Signal	X axis: 0.20g Y axis: 0.20g	X axis: 0.22g Y axis: 0.23g
Run 11: 11% Northridge Main Shock	X axis: 0.20g Y axis: 0.11g	X axis: 0.21g Y axis: 0.16g
Run 13: 22% Northridge Main Shock	X axis: 0.40g Y axis: 0.21g	X axis: 0.40g Y axis: 0.25g
Run 15: 22% Northridge Main Shock	X axis: 0.40g Y axis: 0.21g	X axis: 0.40g Y axis: 0.24g
Run 17: 44% Northridge Main Shock	X axis: 0.80g Y axis: 0.42g	X axis: 0.60g Y axis: 0.40g
Run 19: 100% Northridge Main Shock	X axis: 1.78g Y axis: 0.99g	X axis: 1.10g Y axis: 1.00g
Run 21: 33% Aftershock Northridge Signal	X axis: 0.12g Y axis: 0.07g	X axis: 0.14g Y axis: 0.14g
Run 23: 100% Aftershock Northridge Signal	X axis: 0.37g Y axis: 0.31g	X axis: 0.70g Y axis: 0.40g

Table 1: Runs applied on SMART2013 mockup

Low-level random excitations were applied between each seismic test to measure the model's frequencies of the model during the tests and monitor their evolution.

As expected, the model showed significant torsional behavior. The initial frequencies measured were 6.28Hz (Bending along X), 7.86Hz (Bending along Y) and 16.50Hz (Torsion along Z). The drop in frequency measured between the start and end of the tests was respectively equal to 60% (Bending along X), 43% (Bending along Y) and 48% (Torsion along Z). Figure 20 displays the numerical SMART2013 and shaking table models and axis.

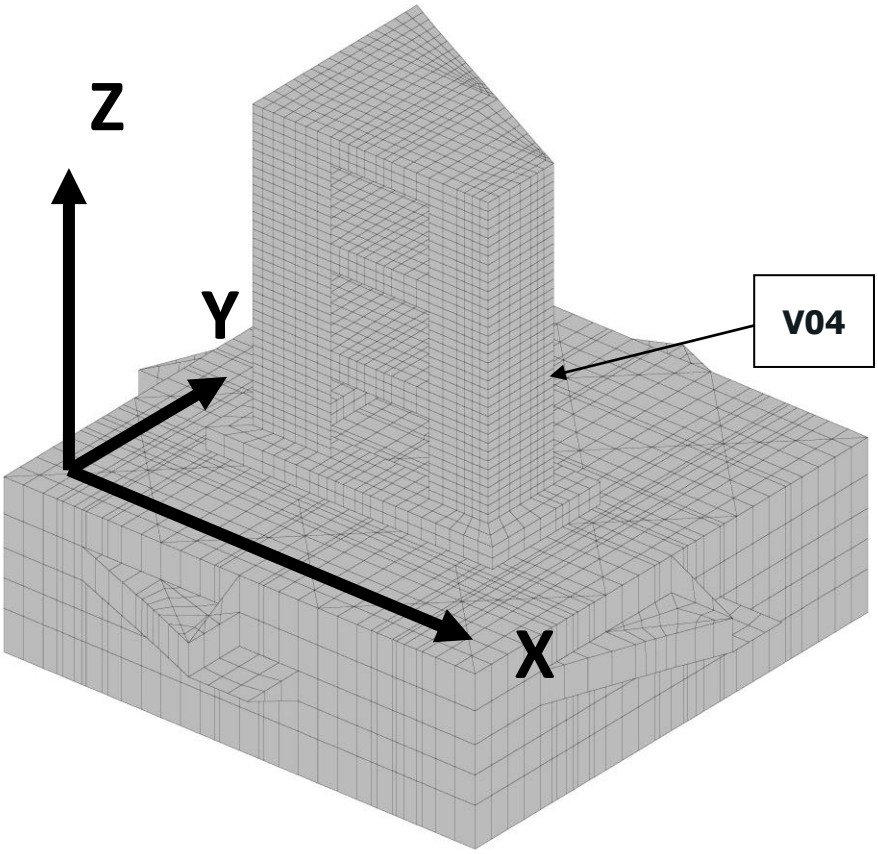


Figure 20: Shaking table and SMART2013 numerical model

Damage occurred mainly at the base of the model and at the top of the less wide wall (labelled 04) at the 1st level. During Run 19 (Signal 2: Main shock of the Northridge earthquake – 100 % nominal level), significant damage occurred at the end and base of wall 04, resulting in a large crack opening (over 10 mm), presumably because of the failure of the steel covering in this area, without jeopardizing the stability of the mock-up (Figure 21). The application of the two aftershock signals (Signal 3) did not generate any additional damage (Figure 22).

Figure 21: Crack at the bottom of wall 04 after Northridge 0.40g (DEN, 2014)

Figure 22: Final state of SMART 2013 (DEN, 2014)

From Run07 to Run23, there were 3 instances where the mock-up significantly showed signs of progressive damage, namely after Run13, Run17, and Run19, as displayed in Table 2.

Run	Mode 1 (Bending - X direction)	Mode 2 (Bending - Y direction)	Mode 3 (Torsion – Z direction)
6	6.3	7.8	16.5
8	6.1	7.7	16.5
10	6.1	7.8	16
12	6	7.9	15.8
14	4.4	6.4	13
16	4.4	6.5	13
18	4	5.7	12.7
20	2.8	4.4	9
22	2.8	4.4	9
24	2.6	4.4	8.6

Table 2: Evolution of eigenfrequencies (DEN, 2014)

4.3. Numerical campaigns for linear and nonlinear analyzes

First, the numerical model of SMART 2013 is presented. Secondly, two numerical models are presented (Number of Runs, Intensity Measure chosen, list of EDP, etc.), the first with only linear analyses (Campaign 1) for all runs, the second this time accounting for nonlinearities (Campaign 2). Then, the validation process is laid out, comparing successively the outcomes of the subsequent processes:

P0: The (Roy and Oberkampf, 2011) method: i.e., the use of the area metric d_0.

P1: The use of d_m and d_p area metrics.

P2: The use of $e_{rr}(\alpha_{err}$=2.5%), the numerical/experimental error e_{rr} at a quantile α_{err}.

P3: The regression/extrapolation is applied on the error of each epistemic-aleatory sampling.

4.3.1. Numerical Model

The numerical model of SMART 2013 displayed by Figure 23 used herein is conceived using the Cast3m software (CEA, 2011) in its 2021 version. It is composed of finite elements of various nature:

A shaking table model may or may not be used. It is composed of elastic linear shell elements, with 18 to 24 degrees of freedom per element.

The foundations are elastic linear cubic elements, with 24 degrees of freedom per element.

The concrete of walls/slabs/beams are shell 5-layers elements with 18 to 24 degrees of freedom per element, with the RICCOQ constitutive law for the nonlinear model (Richard, 2012). It is a damage model with unilateral effect, and asymmetrical softening in compression/tension. The linear model only considers an elastic linear behavior for those elements.

The steel rebars of walls/slabs/beams are shell 4-layers elements with 12 degrees of freedom, with a linear elastic behavior as during the experiments, rebars did not show plastic deformations.

For the same reasons, the column is modeled with elastic linear behavior with multifiber elements.

Totally, with an average 15 centimeters element characteristic length, the full model with the shaking model totalizes 62 000 degrees of freedom. Rayleigh damping matrix is computed between 85% of the first eigenfrequency and 100% of the third eigenfrequency which is in accordance with the recommendations of (ASN, 2017).

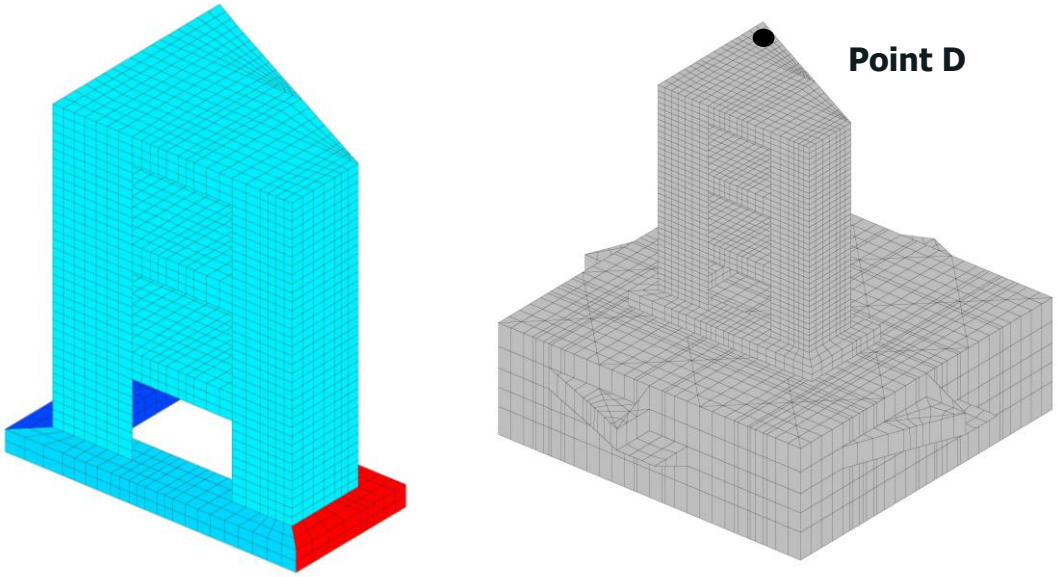


Figure 23: From left to right, the SMART2013 finite element model with and without the shaking table

4.3.2. EDPs

The EDPs studied are:

- the global drifts at the point D at the third floor (where the largest displacements were measured), in directions X and Y;
- the initial three eigenfrequencies;
- the relative drop in frequency of the three eigenmodes;
- the root mean square errors (RMSE) of the pseudo-accelerations over the range 85 % of first eigenfrequency and 100 % of the third one. The closer this mean error to zero, the closer the numerical spectral acceleration is from the experimental one.

4.3.3.Specificities of the modal analysis

Regarding the numerical estimation of the frequency drops on Cast3m software (CEA, 2011), three methods are considered:

The "classic" method of the modal analysis applied at the end of each RUN by accounting for the damage D in the concrete in the effective Young Modulus $E_{eff} = (1 - D) * E$ in the stiffness matrix.

After that SMART2013 has reached its peak displacement and seen D reach its maximal value over time, the input loading is stopped, enabling the structure to freely oscillate until immobility, and assessing both a damping coefficient and a frequency in bending and torsion.

Using the same method, but after that the structure is immobilized, loading it with one slight impulse that keeps it in its linear state, and again, measuring its dynamics.

Until now, the first "classic" method is used, as it does not need the structure to oscillate more after stopping the loading, which would result in a significant and undesired additional computation runtime, while the modal analysis method can be applied at any time iteration.

Nonetheless, one should note that the results of the modal analysis are also dependent on the nature of the numerical model itself. Indeed, cracks alternatively open and close depending on the loading state of the structure, i.e., cracks are different bending towards X to the left, or to the right, or when there is no bending. Considering an effective Young's modulus at the end of simulations cannot take the crack closure phenomenon into account, notably at rest as shown by Figure 24. Furthermore, it does not yield a mapping of cracks unlike during experiments. During the numerical campaign, the frequency drops yielded by the modal analysis are much larger than during the experiments.

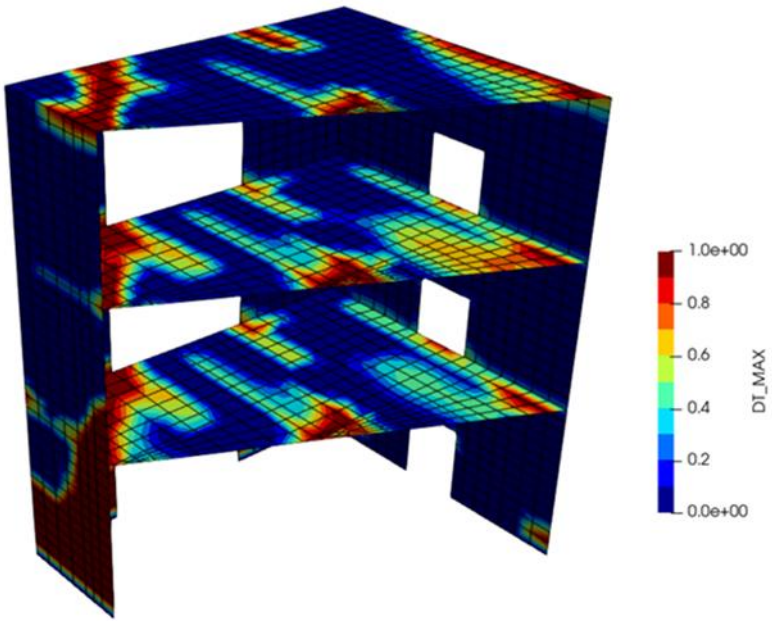


Figure 24: Example of tension damage parameter spread on SMART 2013 (internal variable of RICCOQ constative law)

Consequently, the validation process aims at assessing how accurate a numerical model is to yield expected EDP values, and to predict its behavior over a validity domain. But that does not only depend on the nature of the model, but also on the way the numerical EDPs are calculated. Here, the final eigenfrequencies yielded by the three different bullet points might be very different from each other despite SMART2013 having endured the same loading.

4.3.4. Specificities of the numerical model

To have the initial eigenfrequencies much closer to the experimental ones, the shaking table model is included, despite the additional computational cost. (Richard et al., 2013) states that this choice was made by most of the participants of the benchmark to better match the initial modal characteristics. SMART2013 mass (46t) is heavier than the mass of the aluminium-made shaking table (25t). It entailed a soil-structure interaction during the experiments where they reached the limit capacities of the table. On this topic, (Charbonnel, 2021) has later shown that the characteristic of the hydraulic actuators interfere with the impedances of interaction of the single vibrating table with the model.

Moreover, the foundations are discretized into three zones, as observable in left Figure 23. Each zone sees its Young Modulus decreased to better match the expected values. The foundations dark blue, light blue, and red zones have their Young's modulus respectively degraded to 0.86 GPa, 7.09 GPa, and 18.4 GPa. This is a significant change compared to the median value of 25.7 GPa, but it unexpectedly induced an easier convergence of the numerical simulations.

	Exp. Measures	Without shaking table	With shaking table	With the shaking table & foundation Young modulus decreased
f_1 (Hz)	6.28	8.20	6.50	5.90
f_2 (Hz)	7.80	12.1	9.20	8.20
f_3 (Hz)	16.5	28.5	21.5	17.5

Table 3: Change of initial eigenfrequencies in various model choice

4.3.5. Lists of Runs computed

The runs retained for the numerical campaign are listed here:

Because the intensity measure retained is the PGA, and because Run 09 and 11 have the same PGA, the nonlinear calculations will only include Run 11 to reduce the computational cost.

Because during Run 13 (Northridge 44% nominal level) the vertical mode of SMART 2013 was unexpectedly triggered and stopped with the urgency switch, it was again applied as Run 15. Run 15 being much shorter than Run13, and with no supplement of damage, it will be retained during the numerical planning.

Then, Run 17 and Run 19 are computed with the nonlinear constitutive laws. Runs 21 and 23 are not computed as they did not induce additional damage to SMART 2013.

This totalizes 5 runs when accounting for nonlinearities.

All runs input signals were truncated to gain computational runtimes. Accordingly, it was ensured that the spectral input accelerations along X, Y, and Z axis (on the shaking table table) as well as the floor response spectral accelerations (at point D for horizontal, point F for vertical) are identical before and after the truncation. Figure 25, Figure 26, Figure 27, Figure 28 show the correspondence between the pseudo accelerations between the original (in black) and truncated (in red) runs. A coefficient envelop was added during the first second of the runs to ensure a progressive start of motion of the shaking table numerical model. It is considered that the peak drifts and frequency drops are attained during the selected range.

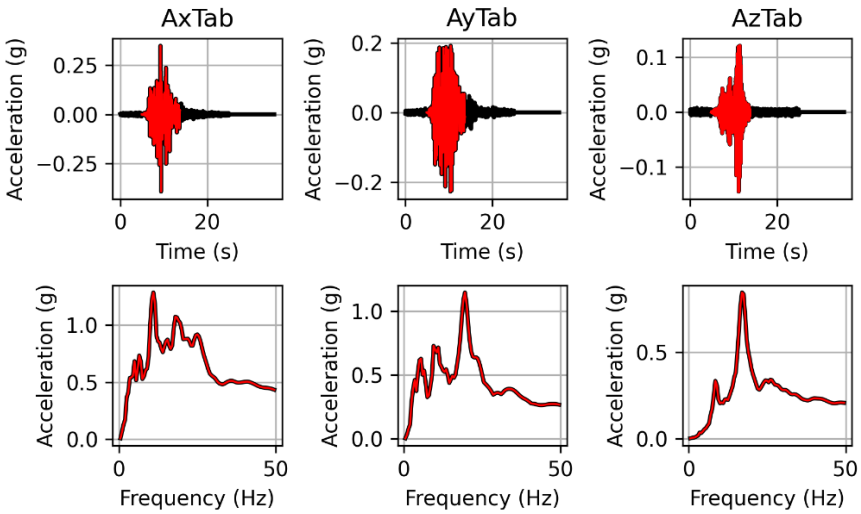


Figure 25: Comparison of original (in black) and truncated (in red) input of Run 15 (damping ratio of 5 %)

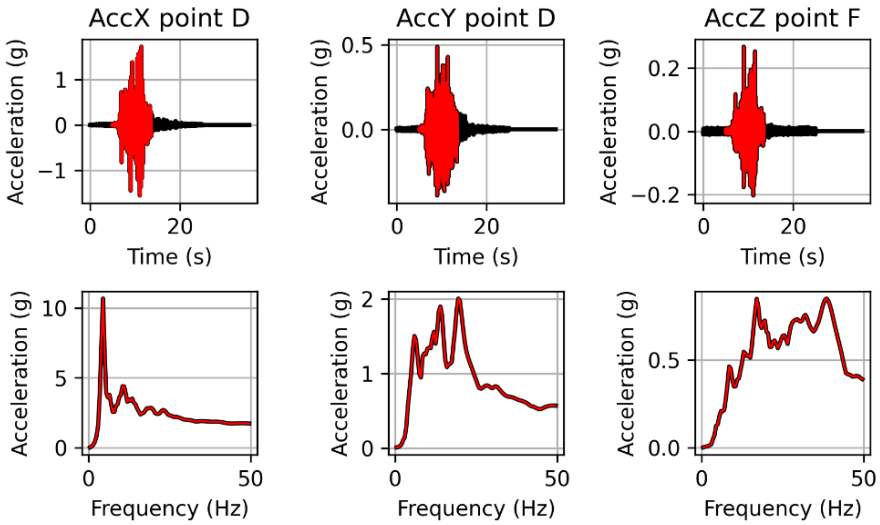


Figure 26: Comparison of original (in black) and truncated (in red) floor spectra of Run 15 (damping ratio of 5 %)

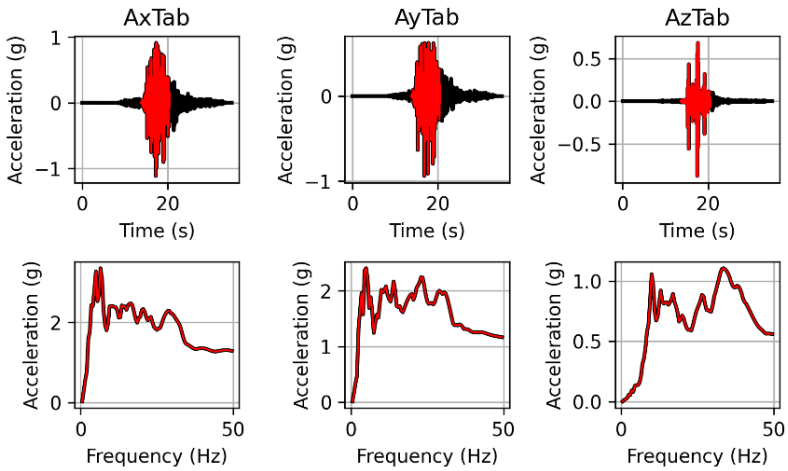


Figure 27: Comparison of original (in black) and truncated (in red) input of Run 19 (damping ratio of 5 %)

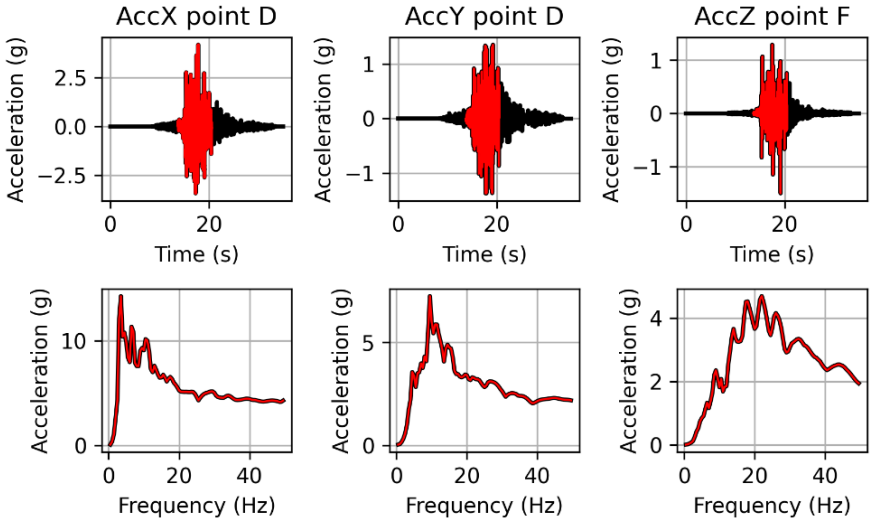


Figure 28: Comparison of original (in black) and truncated (in red) floor spectra of Run 19 (damping ratio of 5 %)

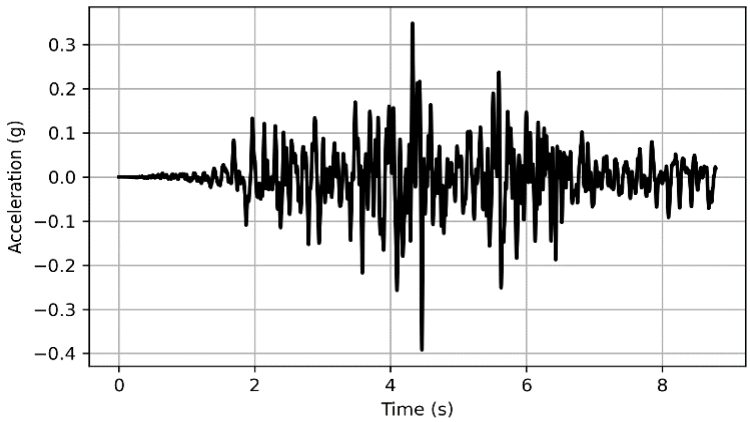


Figure 29: Zoom on truncated Run15

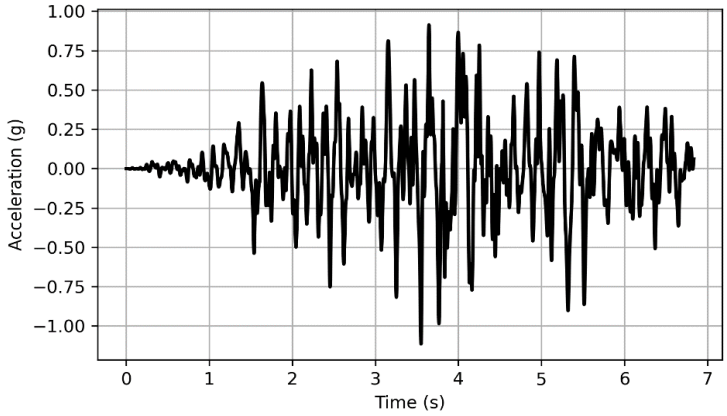


Figure 30: Zoom on truncated Run 19

Throughout the numerical design plan, one important simplification is made. **For each Run, the structure is taken at the beginning in its initial sane state with no damage induced by a precedent Run**. For instance, the triplet 1/30 having SMART initial eigenfrequencies at 6.1 Hz, 8.9 Hz, and 18.0 Hz will have this state at the beginning of each Run, even those where the eigenfrequencies measured experimentally had dropped significantly (after Run 13, Run 17, and Run 19). It is assumed that the frequency drop will not significantly depend on the initial state.

This is a significant simplification and hypothesis done but it significantly helps the present attempt to unravel the Validation process for a structure exhibiting nonlinearities. Future approaches will either (i) account for a global and averaged decrease of the concrete Young Modulus after a damaging loading (ii) or run all the Runs in a row, which are not yet done out of simplicity in the numerical post-treatment and pre-treatment.

Intensity Measures studied: the PGA and the pseudo-accelerations are the IMs used in the present study.

4.3.6. Sensitivity Analysis

Non-exhaustively, additional parametric, sensitivity (Iooss and Saltelli, 2017) (Bouhjiti et al., 2017), EDP versus IM correlation (Buratti, 2012) (Eads et al, 2015) (Langlade, 2021), spatial and time discretization (Subsection 2.3.1) analyses should be performed if resources are available. Indeed, uncertainties propagation is costly, so it is best to perform them with only the most influential parameters and efficient intensity measure. Achieving all these calculations is usually not possible due to a lack of time and resources.

In this subsection nonetheless, a Morris sensitivity analysis (Herman et al. 2013) (Bouhjiti, 2018) is performed onto SMART2013. This method is commonly known as being low in terms of computational cost and allows to compare at a global scale the sensitivity of EDP to the given input parameters distribution inferred from the experimental measurements and data. This first "quick" analysis helps the engineer in justifying the choice between the future variables and deterministic parameters should he not know the nature and behavior of the model and input.

Description of Morris analysis

The method of Morris (1991), in its original form, is based on one-at-a-time (OAT) methods and allows first order comparison of input parameters. It performs a derivation of measures of global sensitivity from a set of local effects, sampled though the parameter space. Each parameter p changes with a Δ_p value, one-at-a-time, creating a trajectory on a grid in the same parameter space. Consequently, one trajectory contains N_p changes, derivations, also called "perturbations". Then, each trajectory yields one estimate of the elementary effect EE_p for each parameter p (i.e., the ratio of the change in model output to the change in that parameter) and follows the following equation:

$$EE_p = \frac{f(x_1, \dots, x_p + \Delta_p, \dots, x_p) - f(x)}{\Delta_p}$$

With f the model, $f(x)$ the previous output of the trajectory. With N_p+1 simulations, one can then estimate the elementary effects of each parameter p with one trajectory and classify them between the less and the most influent on the model responses. However, to ensure the conclusions, the number of trajectories should be multiple, not just unique, because each EE_p depends on the initial point upon which start the trajectory. Literature usually takes up the number of trajectories N_t at 10 as a minimum. The points passed through by the trajectories must now be defined. Each parameter range is discretized n_l times, with n_l commonly known as the "levels", the number of values each parameter can take, with n_l set at a minimum of four by literature. The parameters range and levels can be set deterministically by the engineer or with a probability distribution and Latin Hypercube sampling based upon the experimental measures if any are available (e.g., measures of concrete tensile resistance fitting a log-normal distribution). Consequently, this makes a total of N_t trajectories multiplied by $(N_p + 1)$ simulations, yielding N_t elementary effects EE_p.

From there, the mean μ_p, the mean of absolute values μ_p^* and the standard deviation σ_p of EE_p are calculated for each parameter p. Because the elementary effects could be alternatively positive or negative, the mean μ_p^* is calculated by taking their absolute values as shown by the equation below. A low μ_p and a large μ_p^* would indicate alternative positive and negative EE_p while relatively similar means μ_p and μ_p^* indicates only few changes of signs in the model output.

$$\mu_p = \frac{\sum_{i=1}^{N_t} EE_{p,i}}{\Delta_p}$$

$$\mu_p^* = \frac{\sum_{i=1}^{N_t} |EE_{p,i}|}{\Delta_p}$$

$$\sigma_p = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N_t}\left(EE_{p,i}^2 - \mu_p\right)^2}$$

A visual description of a Morris analysis is the Morris diagram, where each parameter p is plotted with μ_p^* and σ_p as its respective abscissa and ordinate. This allows to compare each input parameter between themselves. For instance, if a parameter reaches the north-east corner of the diagram, the more it has influence on the output of interest, and this influence presents a significant variability as well.

Linear studies: sensitivity of parameters E, ρ, S_d, N_l, [ξ_1,ξ_2] onto X drift, $S_a(f_1)$, S_a(30Hz)

The shaking table is accounted for throughout the present study because known as being key to better capture the modal behavior of SMART2013. A uniform sample is computed for each parameter, the objective being to have a first informative and quick classification of the sensitivity. The input parameters nature and range considered for the linear studies are the following:

- $\boldsymbol{E}$: Young modulus [20 ; 30] GPa
- $\boldsymbol{\rho}$: material density [2000 ; 2500] kg/m^3
- $\boldsymbol{S_d}$: Spatial discretization [0.16 ; 0.24] m
- $\boldsymbol{N_l}$: Number of concrete layers in the multilayer shell elements [3 ; 7]
- $\boldsymbol{[\xi_1,\xi_2]}$: Rayleigh damping coefficients [3 ; 6]%

The outputs of interest are here the longitudinal drift, the floor response spectra at the fundamental first eigenfrequency, and at 30 Hz. In the present document, the number of trajectories N_t is taken at 10 and the parameters ranges are each discretized uniformly with $n_l = 4$ levels, i.e., four values across each range.

It is to be noted that the spatial discretization values considered in this study are all below the critical mesh convergence value as defined by the Richardson's extrapolation presented in 2.3.1 (i.e. $\hat{p} \approx 2$). It does not mean that there is no sensitivity to the mesh size, but rather that this sensitivity tends to decrease in a predicable manner with respect to the mesh size.

Because there is an important variability from the ground motion choice itself (PGA, frequency content, time duration, …), the Morris analysis should account for this point. To do so, $N_r = 19$ different ground motions ranging from 0.05 g to 1.1 g PGA are selected. Because they have strong differences in intensity and frequency content, their respective μ_p^* and σ_p are very different in terms of scale. Consequently, analyzing the sensitivity output of the N_r records plotted in the same diagram would be difficult, or a logarithmic scale should be used. To represent them all on the Morris diagram, for each record r of the total N_r records, each $\mu_{p,r}^*$ and $\sigma_{p,r}$ is normalized by the respective maximum $\mu_{p,r}^*$ and $\sigma_{p,r}$: $\|\mu_{p,r}^*\| = \frac{\mu_{p,r}^*}{\max_{\{i=1,\dots,p\}}(\mu_{i,r}^*)}$

$$\|\sigma_{p,r}\| = \frac{\sigma_{p,r}}{\max_{\{i=1,\dots,p\}}(\sigma_{i,r})}$$

Consequently, all Morris analyses results can be plotted in the same diagram, all points values being between 0 and 1, and enabling an easier relative comparison of the EDP sensitivity.

Figure 31 displays the Morris diagram with the normalized values $\mu^*$and standard deviation $\|\sigma\|$ for each record regarding the longitudinal drift. The more a colored point is close to the (1,1) area, the more effect it has on the drift relatively compared to the other parameters. The Young's modulus E blue points tend indeed to be closer to (1,1) but closely follow by the damping coefficients $[\xi_1,\xi_2]$ and by the spatial discretization S_d. Such a diagram might be totally different considering other EDP.

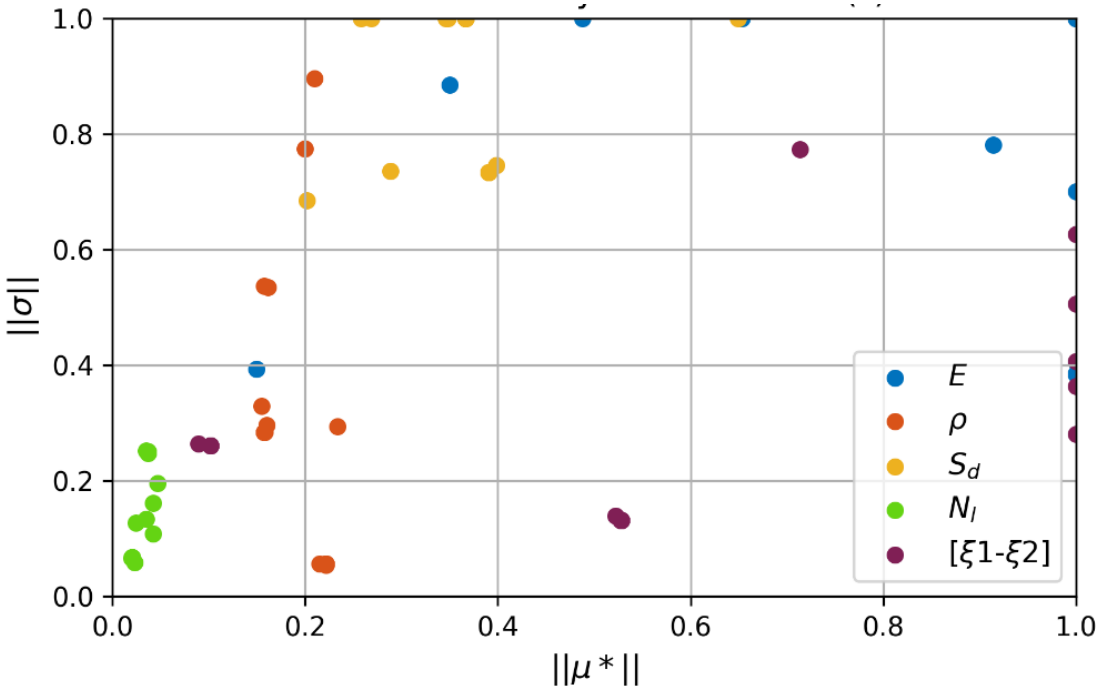


Figure 31: Normalized Morris Diagram for N_r = 19 selected records and five different parameters on the X drift of SMART2013 linear model

Now, to have a more global overview and simplify even more the above Morris diagram, the means μ and standard deviation σ over all the seismic records r of the normalized $\|\mu_{p,r}^*\|$ and $\|\sigma_{p,r}\|$ are computed. This allows to aggregate N_t dots in Figure 31 into one. Now, plot for each input parameter, the means of $\mu_{\|\mu_p^*\|}$ and $\mu_{\|\sigma_p\|}$ as well as the related 95 % prediction interval in the Morris diagram, as shown by Figure 32.

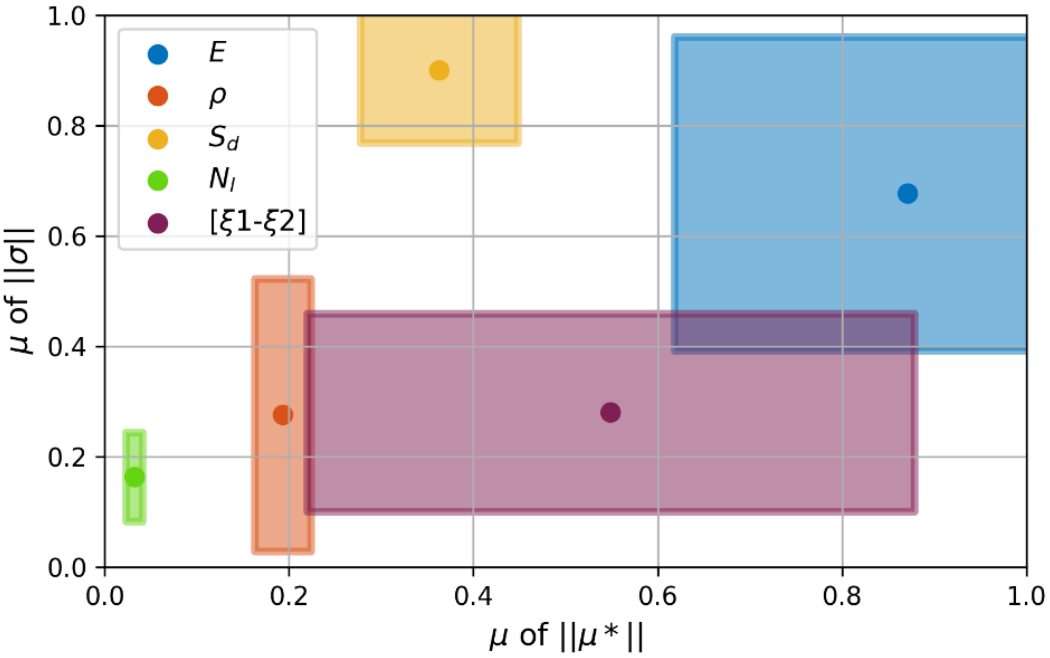


Figure 32: 95% prediction intervals of normalized $\|\mu^*\|$ and $\|\sigma\|$ of Morris Diagram

The material density and number of concrete layers in the model have comparatively little effects on the drift. To keep computational cost as reasonable as possible, S_d is arbitrarily set at 0.2 m. The engineer then decides to take the damping coefficients as epistemic uncertainty and the Young's modulus as the aleatory uncertainty.

4.3.7. Epistemic and aleatory uncertainties

Same epistemic uncertainty for linear and nonlinear analyzes (Campaigns 1 and 2)

No sensitivity analysis could be achieved during the study time to evaluate the EDPs sensitivity towards the input parameters and the knowledge of their range of potential values.

The choice of the Rayleigh damping coefficients is commonly prone to be debated, and the values can differ from 2 % to 9 % in the case of SMART2013 depending on the modeling choices, constitutive laws, loading intensity (Richard et al., 2016), and norms and guidelines. Consequently, The Rayleigh damping is set as the epistemic uncertainty. Both damping coefficients ξ_1 and ξ_2 respectively applied at the frequencies f_{R1} and f_{R2} of the Rayleigh damping matrix are sampled and propagated together. To follow (ASN, 2017) recommendations, f_{R1} and f_{R2} are chosen as 85 % of initial f_1 and 100 % of initial f_3, both frequencies enveloping most of the modal behavior of SMART2013.

(DEN, 2014) relates that the damping coefficients of the first bending mode f_1 and third torsion mode f_3 equal respectively 2.6 % and 5.5 %, which allows to establish a global range of uncertainty. A conservative investigation would lead to have four cases where the pair ξ_1 and ξ_2 take the values 2 % and 6%, which totalizes four configurations. However, in the peak intensities of Run17 and Run19, strong numerical instabilities occurred, that could have originated from either time discretization choice or small damping values, or both. Considering the large computation cost of nonlinear simulations, it is decided as a first application to increase the limit ranges to 3 % and 6 %.

Also, the four damping evolutions (depending on the combination of values of ξ_1 and ξ_2) are plotted and displayed by Figure 33. It shows that, over the f_{R1} and f_{R2} range, the two limit cases [3 %; 3 %] and [6 %; 6 %] are envelops and contain the two curves [3 %; 6 %] and [6 %; 3 %]. Consequently, it is decided to reduce the numerical campaign from four epistemic cases to two epistemic cases of damping coefficient pairs. This directly divides by two the number of computations, and so the time duration of the numerical campaign.

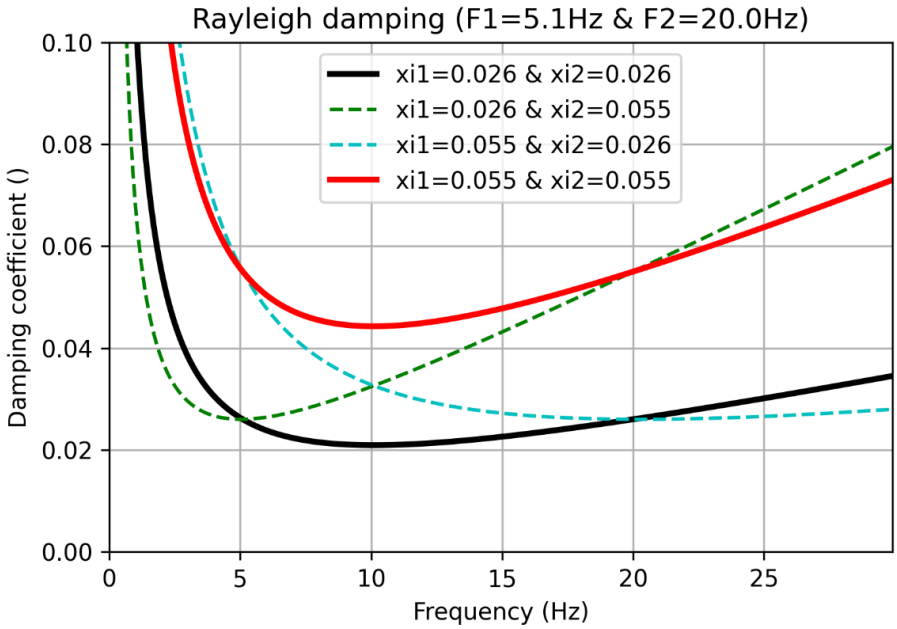


Figure 33: Rayleigh damping evolution over frequency: envelop cases (solid lines)

Aleatory uncertainty for linear analyses (Campaign 1)

The Young's modulus $\boldsymbol{E}$ is chosen as the main aleatory uncertainty. Again, no sensitivity analysis was made to confirm this point, but the author sees its direct impact on global rigidity through the Young's modulus and on the damage progress through the structure.

According to Wilks formulation (Porter, 2019), a minimum amount of 59 samples from a given distribution are necessary to ensure with a 95 % level of confidence that the minimum and maximum values of the 59 values are respectively smaller and larger than the 5% and 95% quantiles of the distribution. Consequently, whatever the nature of the aleatory variable, at least 59 samples are necessary, and 60 will be ultimately taken for the linear analyzes.

Aleatory uncertainty for nonlinear analyses (Campaign 2)

The concrete material properties are chosen as the main aleatory uncertainty, namely the Young's modulus $\boldsymbol{E}$, the tensile resistance $\boldsymbol{f_t}$, and the tensile fracture modulus $\boldsymbol{G_f}$. Again, no sensitivity analyzes was made to confirm this point, but the author sees their direct impact on global rigidity through the Young Modulus and on the damage progress through the structure. These three parameters are selected as one set, i.e., one sampling. However, the triplets are not drawn independently but drawn by accounting for the correlation between $\boldsymbol{E}$, $\boldsymbol{f_t}$, and $\boldsymbol{G_f}$ inferred from the measurements. A multivariate lognormal Quasi-Monte-Carlo sampling is achieved to do so.

However, considering, (i) the three to five days span time duration for each simulation, (ii) the two epistemic cases, (iii) the three to four runs to compute (11, 15, 17, and 19), (iv) the calculations resources available, (v) the objective mainly pedagogical of the present document, it was decided by the author to reduce the number of samples to 30. For the 30 samples, there are no correlated parameter values that fall outside the 1 % quantile for their respective distributions. Figure 34 displays respectively the correlated Quasi Monte-Carlo draws of $\boldsymbol{E}$ and $\boldsymbol{f_t}$. It can be compared to the empirical relationship provided by Eurocode 2 and deduced from the following equations:

- $E_{cm} = 22\left(\frac{f_{cm}}{10}\right)^{0.3}$ with E_{cm} the Young's modulus, f_{cm} the mean compressive strength;
- $f_{ctm} = 0.30 f_{ck}^{2/3}$ with f_{ctm} the tensile strength and f_{ck} the characteristic compressive strength;

- and $f_{ck} = f_{cm} - 8\,\mathrm{MPa}$. However, it has been shown experimentally by (Larrard, 2010) that a better approximation of f_{ck} is $f_{ck} = f_{cm} - 20\,\%$.

Finally, it can be deduced that:

$$E_{cm} = 22\left[\left(1.2\left(\frac{f_{ctm}}{0.3}\right)^{\frac{3}{2}}\right)/10\right]^{0.3}$$

It can be seen on Figure 34 that the cloud of points is out of the empirical prediction from Eurocode 2. The values of Young's modulus are lower than the expected ones for a standard concrete. However, it is consistent with the values indicated in the experimental reports (Richard et al., 2016) i.e. between 24.4 GPa and 29.5 GPa depending on the location in the structure.

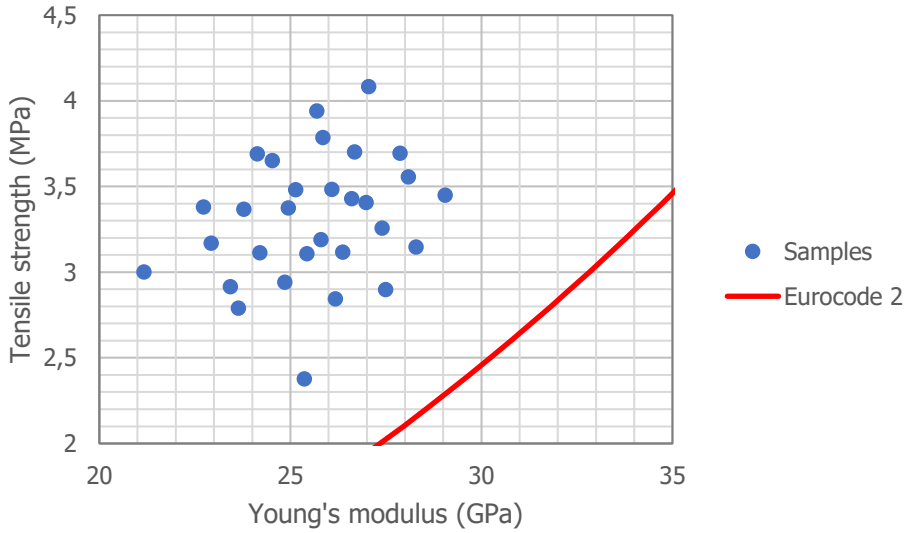


Figure 34: 30 Quasi Monte-Carlo correlated draws of E and f_t for nonlinear analyzes and empirical relationship provided by Eurocode 2 (red curve)

One can note in figure 34 that Figure 35**: Mean and standard deviation** displays the evolution of the means and standard deviation of the linear campaign 60 samples (in blue) and nonlinear campaign 30 samples (in golden). Ultimately, the mean (resp. standard-deviation) of the 30 samples is less than 1% (resp. 5%) different than the mean (resp. standard-deviation) of the 60 samples. The author considered it to be a small "enough" difference to justify that 30 samples should be enough to consider the two campaigns comparable.

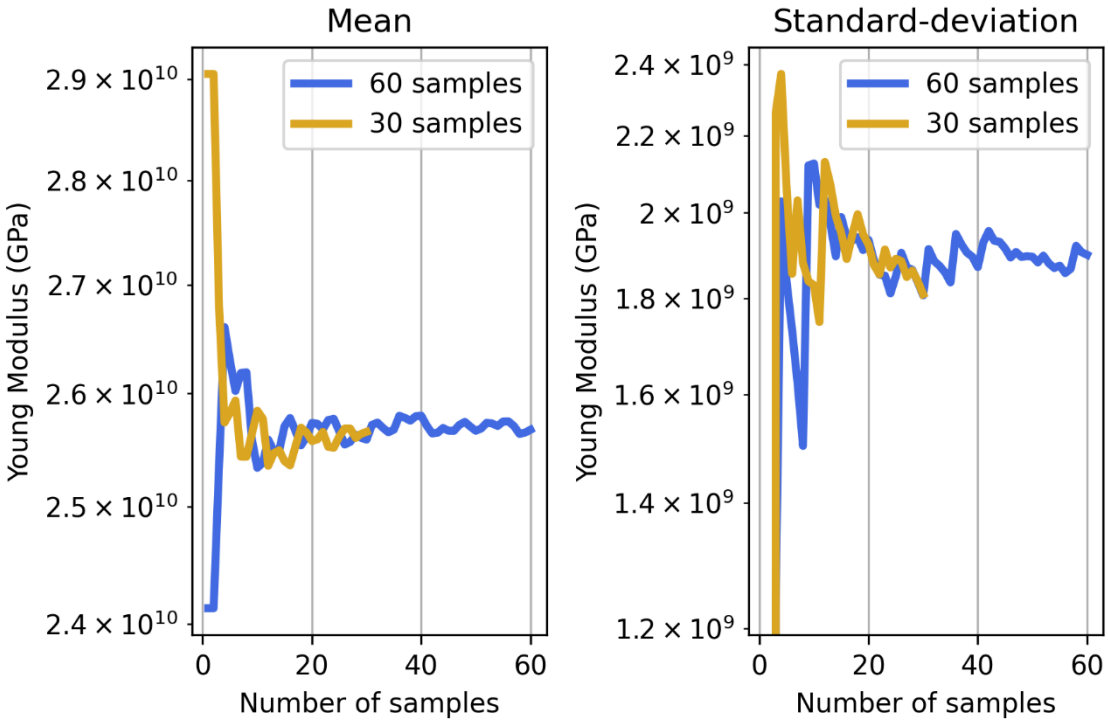


Figure 35: Mean and standard deviation of Young's modulus for 30 and 60 samples

4.3.8. Computation outside the validation domain

Because the Run 19 should have reached a 1.78 g PGA as desired (see Table 1), but failed due to experimental limitations, the 1.78 g PGA response to such intensity is still unknown, potentially undoable experimentally. Let us now use the previous numerical campaign to predict the response.

First, a 1.78 g PGA must be chosen. In the present example, it is the original Northridge signal recorded at Tarzana Cedar Hill station in 1994 that is used (Ancheta et al., 2013) and displayed on Figure 36. It is different from Run 19 because the jacks of AZALEE table in EMSI laboratory did not allow the acceleration at the base of the structure to properly reproduce the input signal. Figure 37 highlights these differences between the main run along X axis, both in terms of time histories, peaks, and spectra trends.

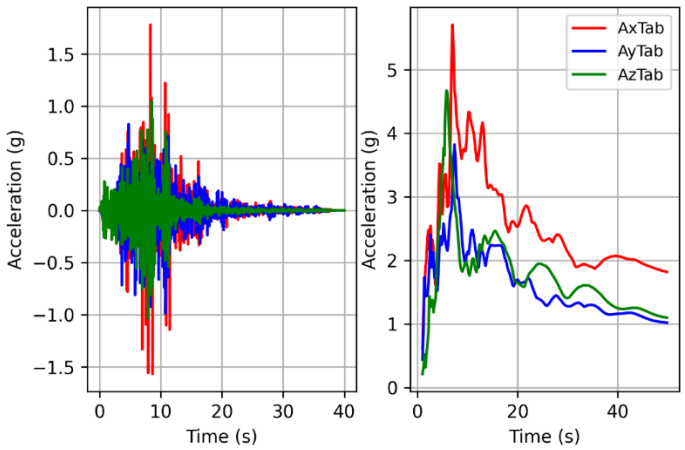


Figure 36: Original Northridge 1.78 g main PGA for Campaign n° 1

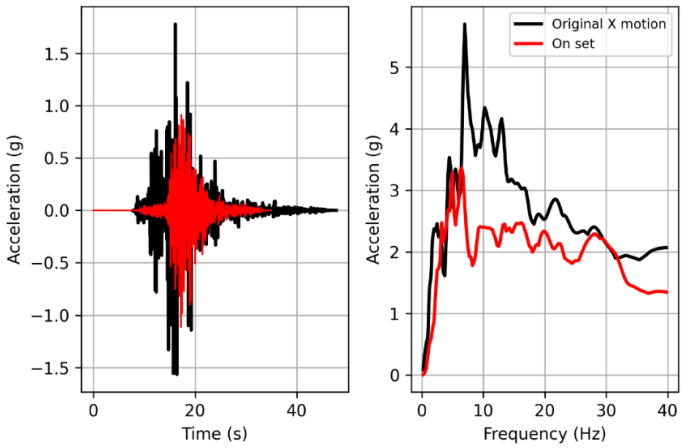


Figure 37: Differences between the original and achieved Northridge motion of the shaking table

However, because numerical convergence difficulties are likely to happen as such intensities causing time-consuming numerical campaigns, another smaller-intensity record is considered. It is the same Northridge record but scaled at 0.83g PGA. This intensity was chosen to be specifically in the middle of the PGAs of Run17 and Run19, respectively 0.56 g and 1.11 g to be the farther away from these experimental runs where data is missing and of interest. Furthermore, the simulations are expected to converge in much less iterations than at PGA 1.78 g.

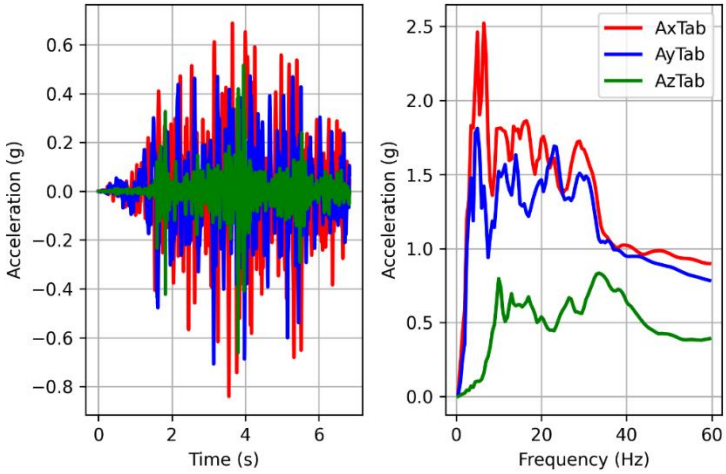


Figure 38: Northridge 0.83 g PGA scaled record

4.4. Validation Process based on Campaign n°1

In the present subsection, the Validation Process is unraveled. Runs 07, 11, 15, 17, and 19 are applied to SMART under linear elastic assumptions for demonstrativeness purposes. Run 09 was removed because it has the same PGA as Run 11. The global structural drift of the mock-up along the longitudinal X axis is studied first, then the initial eigenfrequencies, and the floor response spectral longitudinal acceleration at the 3rd floor (point D) at a 30 Hz frequency, namely Sax(30 Hz). The area validation metrics $\boldsymbol{d_0}$ (section 2.5), $\boldsymbol{d_m}$ and $\boldsymbol{d_p}$ (section 3.2.1), the errors $\boldsymbol{e_{rr}}$ at a given quantile $\boldsymbol{\alpha_{err}}$ of 2.5% (section 3.2.2), are compared, as well as the different approach (section 3.2.3) of calculating the difference between the simulations with the experimental measure for each input parameter sample. The longitudinal maximum drift is the EDP of interest.

Remark: for the sake of simplicity, only the X-direction motion is considered for the validation process in this report. In an industrial context, both X and Y directions should be considered for an unsymmetrical structure such as SMART2013.

4.4.1. X drift: Experimental vs. Numerical comparison over the validation domain

First, the numerical output 95 % interval (shaded area) and experimental measurements (in green dots) are displayed by the following Figure 39. Run 07 and 11 capture the experimental measurements. The numerical model and parameters distribution underestimate the experimental longitudinal drifts beyond 0.2 g where nonlinearities start to appear.

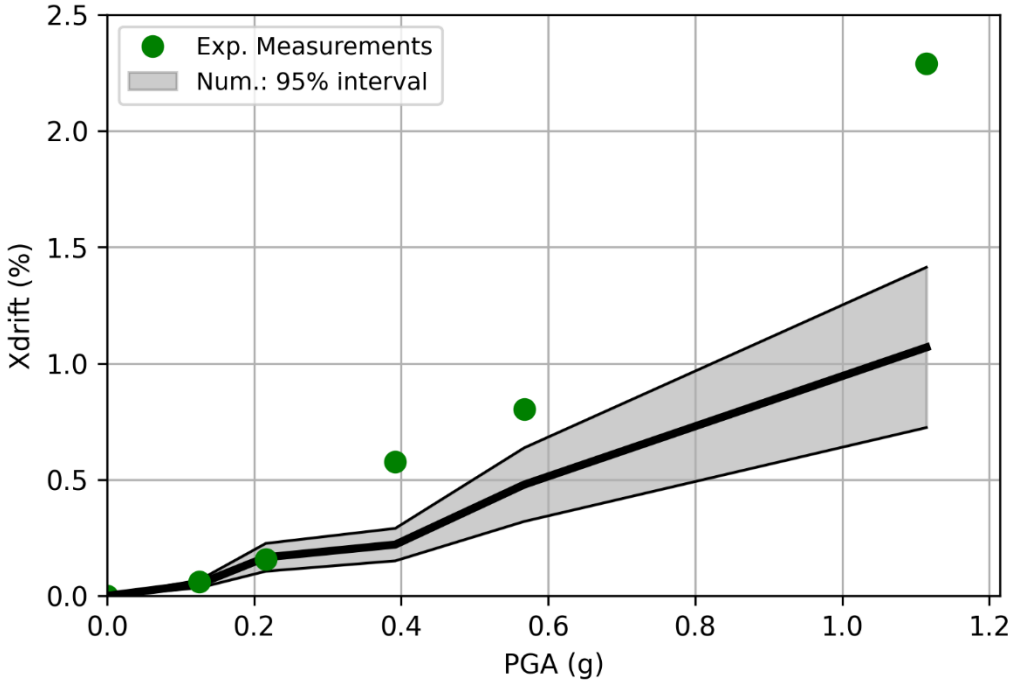


Figure 39: Numerical and Experimental comparison of X drifts values at point D (linear elastic model)

Following Figure 40 summarizes the numerical probability-box tendencies to underestimate (in blue), overestimate (in red), and to contain within its envelop (in black) the experimental measurements under the studies assumptions, i.e., epistemic and aleatory variables and distributions. In the linear elastic domain, from 0 to 0.2 g, the probability-boxes of Run 07 and 11 envelops with a 60 % probability the expectations. Then, it decreases down to 0 % because of constant underestimations because nonlinearities, that tend to increase displacements, are not accounted for.

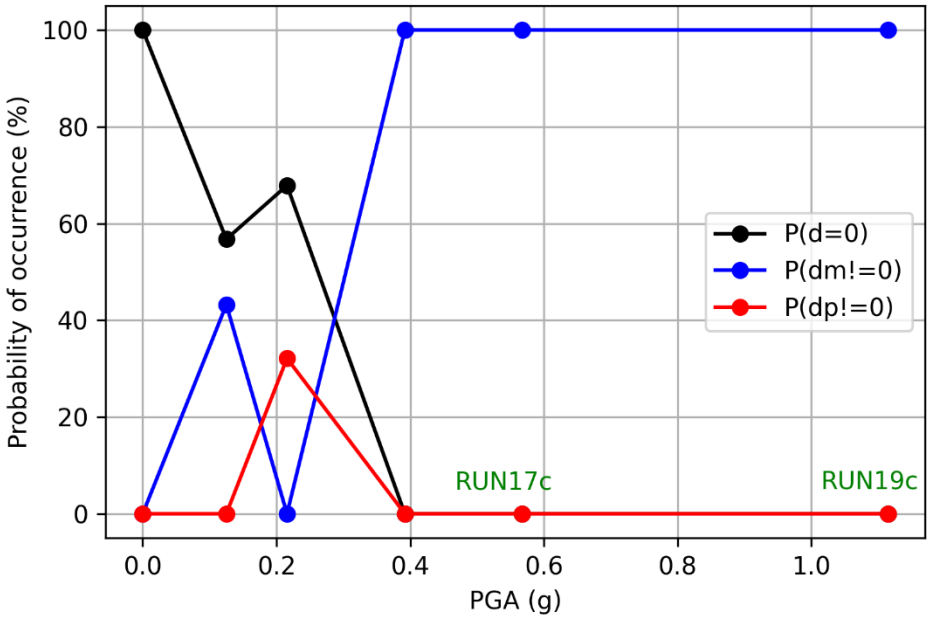


Figure 40: Model best/under/over estimations of X drift at point D in the validation domain (linear elastic model)

Now, to better quantify the model accuracy over the validation domain, Figure 41 below presents the median and 95 % prediction interval of interpolated relative errors (on the left hand-side figure), and the evolution of the integrated overlapping coefficient (on the right hand-side figure) from 0 g to 1.11 g for the X drifts values at point D.

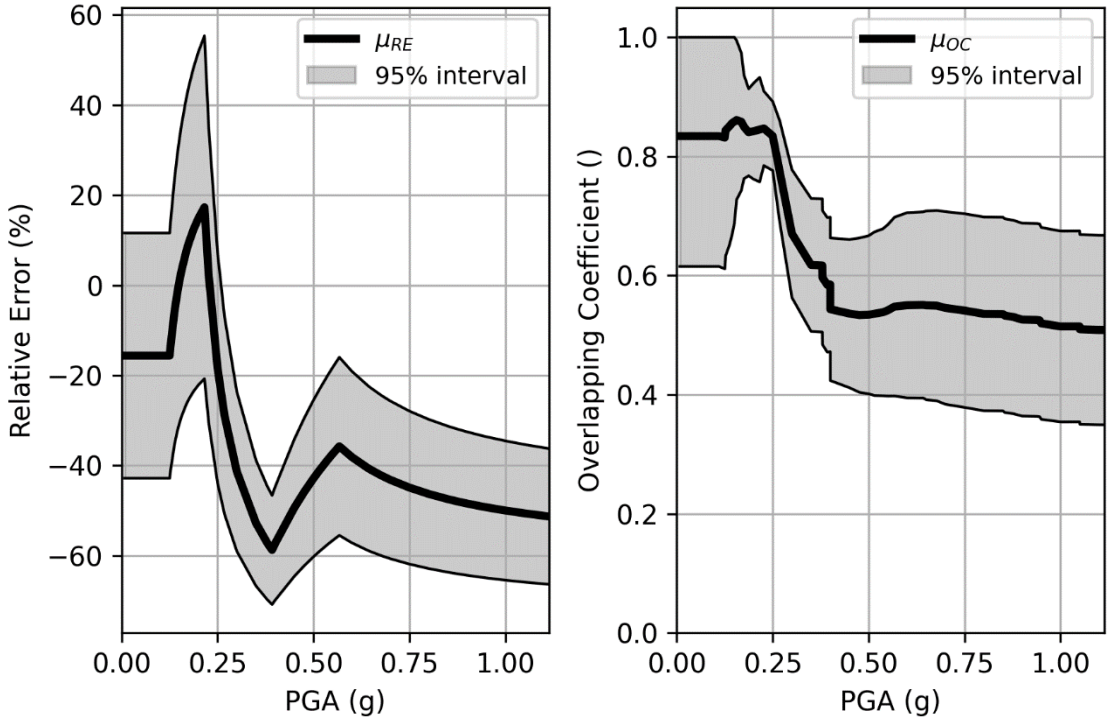


Figure 41: Relative error and overlapping coefficient evolution for X drifts at point D (computations vs experiments, linear elastic model)

Both figures allow to quickly spot the mains areas of discrepancies. As expected, the linear elastic model shows better validity capabilities in the elastic range up to 0.22 g (run07 to run11) than when damage clearly started to appear (run13 to run19). The more the intensity increases, the less the linear model is valid, accurate as the medians of relative errors and overlapping coefficient decrease

respectively from -20 % to -50 % and from 0.8 to 0.5. The chief-engineer can then decide whether the model needs improvement, or for instance at which seismic intensity experimental tests should be performed to obtain more measurements. Considering nonlinearities (Section 4.5) should allow for a reduction of the relative error and an increase of the overlapping coefficient in the large intensities, to the cost of a higher computation time.

4.4.2. Example: Zoom on Run 07

Figure 42 presents, among the five possible Runs, the X drift probability-box of Run 07 simulations with a 0.12 g PGA.

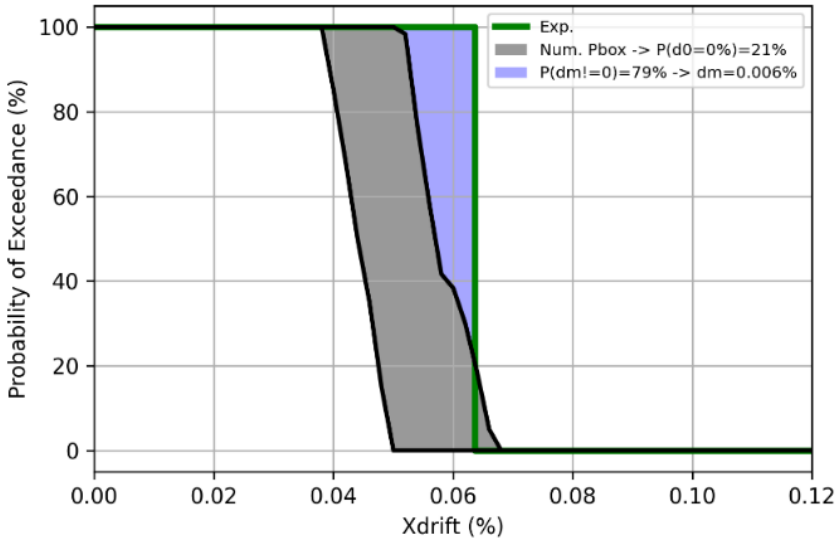


Figure 42: 0.12g PGA Run 07: Experimental Measure and Probability-box of X drift at point D (linear elastic model)

Outputs (in black lines between 0.038 and 0.067) are mainly smaller than the experimental measurements (0.063 in green). The whole set of propagated input and resulting output, shaping the probability-box globally tends to underestimate the experimental measures in 21 % of cases. Over all the runs, regarding X and Y drifts, the numerical model and input parameters assumptions tend to underestimate what is expected. Here, $\boldsymbol{d_0}$ = $\boldsymbol{d_m}$ = 0.006 %, and $\boldsymbol{d_p}$=0 % because the p-box here never completely overpasses the experimental measures.

Next Figure 43 presents the CPD of the population of errors $\boldsymbol{e_{rr}}$ from the external CPDs of the probability-box, still for the Run 07, for a given quantile $\boldsymbol{\alpha_{err}}$ here set at 10 %. Then, two metrics are retrieved, $\boldsymbol{e_{rr}}$(2.5 %) and $\boldsymbol{e_{rr}}$(97.5 %).

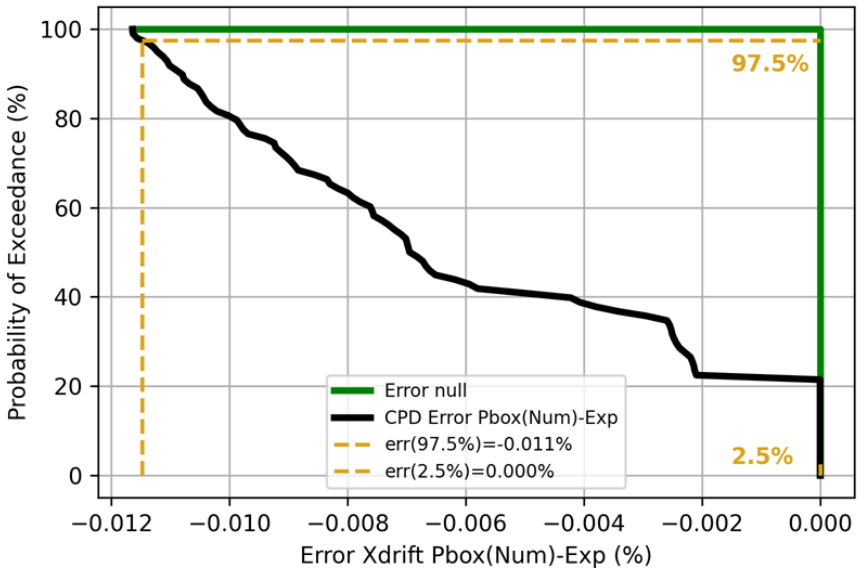


Figure 43: Run 07: Error e_{rr} of the X drift p-box for a given quantile $\alpha_{err} = 2.5\ \%$ (linear elastic model)

The CPD of the p-box error is represented with a higher discretization than the CPDs of Figure 42. As expected from the previous Figure, in 21 % of cases, the error is null.

Here, we then have e_{rr}(2.5 %) equaling 0 % as highlighted in golden in the lower right part of the plot, i.e., the probability-box never overestimates with a 97.5 % confidence. On the contrary, α_{err}(97.5 %) equals -0.011 %, i.e., the numerical p-box are in absolute smaller than 0.011 % with a 97.5 % confidence. One noticeable difference with the area validation metrics is the value itself. d_0 and d_m are indeed averagely twice smaller than e_{rr}(97.5 %). By reading Figure 43 again, there is only a 43 % probability that the CPDs values are smaller in absolute than 0.006 % (d_0 and d_m), consequently a 57 % probability that they are larger in absolute which is consequent.

4.4.3. X drift regressions and prediction intervals

Figure 44 below displays on the left-hand side the metrics d_0 (in black dots), d_m and d_p (respectively in blue and red dots), and e_{rr}(97.5 %) and e_{rr}(2.5 %) (respectively in golden downward and upward triangles) calculated for each Run. On the right-hand side figure, for each Run, for each epistemic uncertainty value (in blue for ξ_1=ξ_2=3 %, in orange for ξ_1=ξ_2=6 %) the 60 errors between the numerical X drift and its related experimental measure. Overall, the area validation metrics and errors have a linear trend, and that will be ultimately this type of regression that will be used.

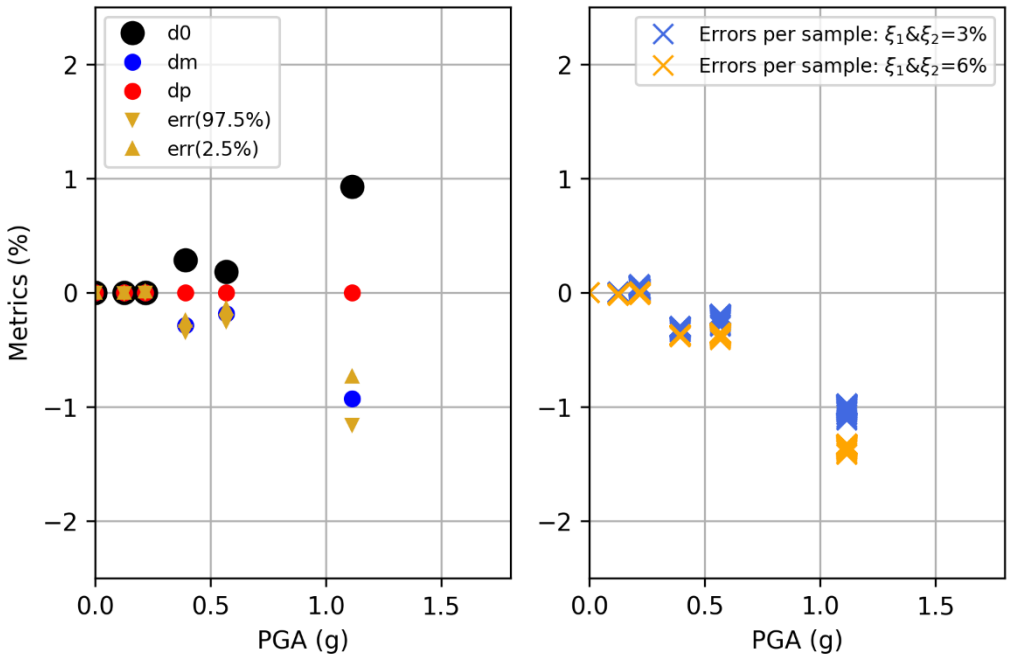


Figure 44: X drift: List of metrics Versus Intensity Measure (linear elastic model)

On the left hand-side Figure, $\boldsymbol{d_0}$ is positive which is expected by its definition. We obtain as expected $\boldsymbol{d_0} = |\boldsymbol{d_m}| + \boldsymbol{d_p}$, which can be reduced to $\boldsymbol{d_0}=|\boldsymbol{d_m}|$ since $\boldsymbol{d_p}$ is always null, the probability-box never overpassing the experimental data. It is clearly visible here that $\boldsymbol{d_0}$ is not suitable to understand the model trends on a given EDP over the validation domain.

Looking at $\boldsymbol{e_{rr}}$(97.5 %) and $\boldsymbol{e_{rr}}$(2.5 %), the golden triangles are relatively close from each other and from $\boldsymbol{d_m}$ over the validation domain. It gives the extra knowledge that there are no significant variations between $\boldsymbol{e_{rr}}$(97.5 %) and $\boldsymbol{e_{rr}}$(2.5 %), so EDPs are not significantly dependent over the input epistemic and aleatory parameters.

Studying now the right hand-side Figure, the plots of the individual errors calculated for each simulation again confirm the numerical model tendencies to underestimate the experimental data.

Next Figure 45 shows the same data as in Figure 44 (the dots in smaller size for more clarity) as well as the linear regressions over all the Runs (i) on the left-hand side Figure for each metrics, and (ii) on the right hand-side Figure, for each simulation error.

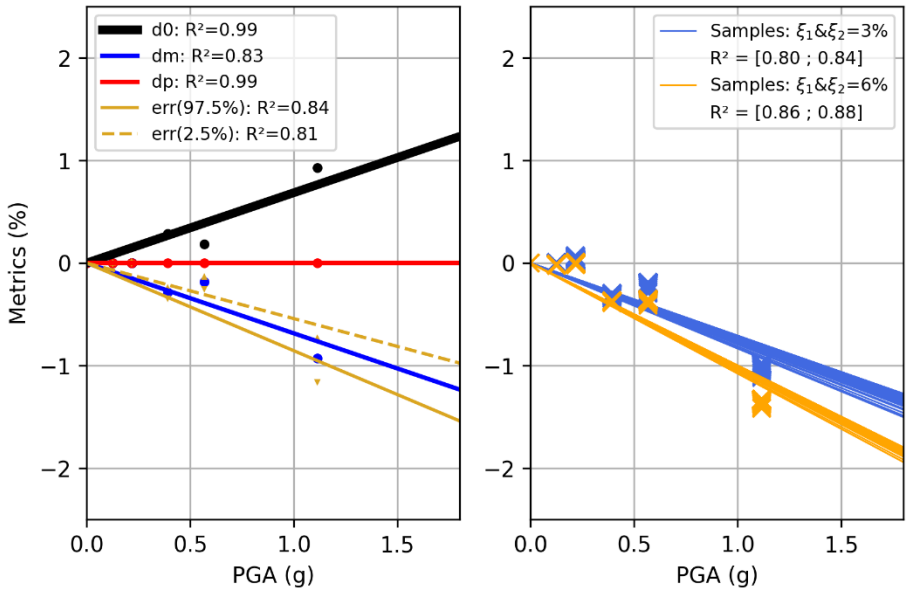


Figure 45: Linear regressions of metrics of X drift at point D (linear elastic model)

The following Figure 46 displays the lower and upper 95 % prediction intervals of the same abovementioned metrics and errors. For instance, the lower 95 % prediction interval of $\boldsymbol{d_m}$ (in blue) and the upper 95 % prediction interval of $\boldsymbol{d_p}$ (in red) draw the global envelop of the metrics $\boldsymbol{d_m}$ and $\boldsymbol{d_p}$.

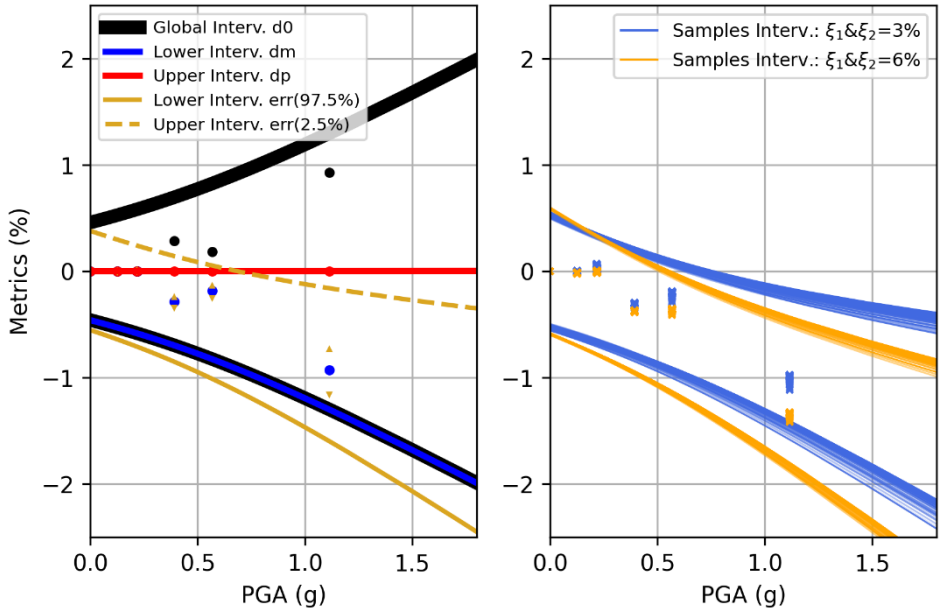


Figure 46: 95% Student Prediction Intervals of metrics of X drift at point D (linear elastic model)

The differences between the various approaches are visible on the last Figure. First, $\boldsymbol{d_0}$ does not account for any signs, but only absolute values. The approach by (Roy and Oberkampf, 2011) then spreads the upper prediction interval symmetrically over and below zero, which results in an overconservative output when correction is needed outside the validation domain.

Regarding $\boldsymbol{d_m}$ and $\boldsymbol{d_p}$ that each embodies the numerical model tendency to respectively under and overestimate the measurements, the global envelop shaped by the two blue and red curves are more informative and less conservative.

Regarding $\boldsymbol{e_{rr}}$(97.5 %) and $\boldsymbol{e_{rr}}$(2.5 %), their respective lower and upper intervals have the same decrescent slope and relatively close values, which will ultimately bring more accuracy than with the $\boldsymbol{d_m}$ and $\boldsymbol{d_p}$ envelop.

Finally, on the right hand-side figure, for each epistemic case of damping coefficients combination, the 60 lower and upper 95% prediction intervals are plotted.

For all cases, the larger absolute value of metrics and errors will be retained if an EDP correction value is asked outside the validation domain.

4.4.4. Response 1st peak frequency evolution results

From now on, the model considered in the nonlinear one. To use SMART2013 feedbacks and conclusions in other potential projects, the eigenfrequencies evolution due to damage are not plotted in absolute but rather relatively to their respective initial eigenfrequency, hence the unit in percentages. In Figure 47, the experimental measurements show a steady decrease over the validation domain in green dots, starting obviously at 100% (initial sane state), the first noticeable decrease arriving at 0.4 g (run 13) and falling at 45 % (Run 19 at 1.1 g).

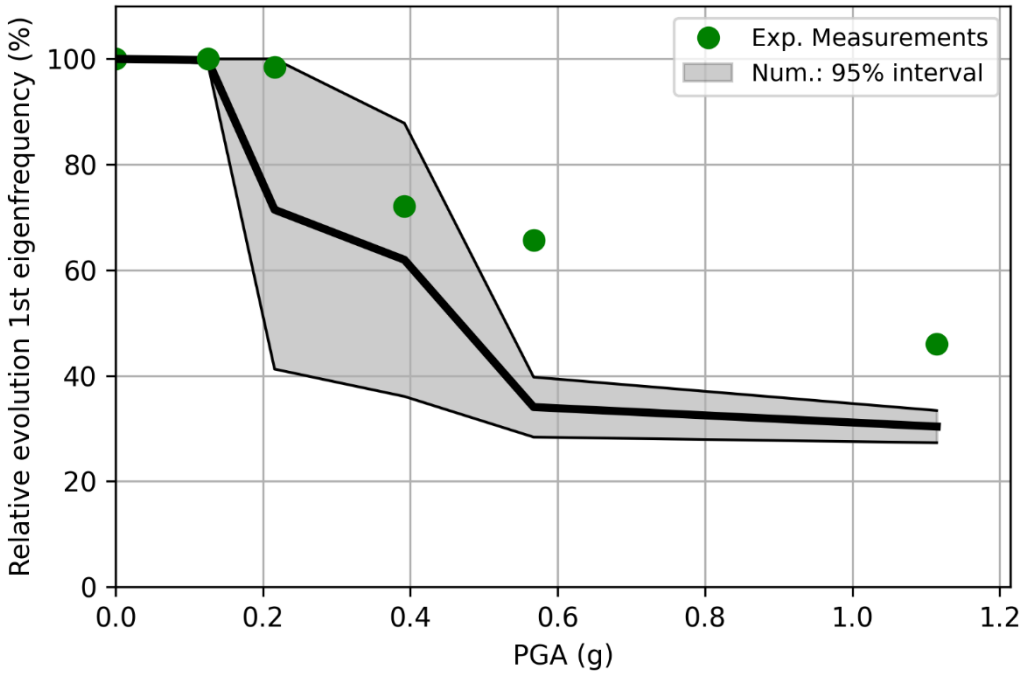


Figure 47: Relative evolution response 1st peak frequency

The median curve in thick black line of the numerical simulations show that SMART2013 model endures more damage than desired, the frequency loss starting sooner than expected and ending at a lower state 20 % below the final measured state. This might be because the mock-up sees its cracks close after each test, regaining some stiffness, while the numerical model does not account for this phenomenon. Same observations can be made on the second and third eigenfrequencies. One should notice here that the 95 % prediction interval narrows down as the intensity grows larger than 0.5 g.

Figure 48 presents the relative error and the overlapping evolution over the validation domain of the relative evolution of the 1st eigenfrequency. As expected, both respectively have values equal to 0 and 1 at the beginning of the seismic runs as no damage appears. In case where several different models

and/or behavior laws are candidate for a future simulation, the plots of these two coefficients over the validation domain can help the decision-makers to choose and select the most appropriate one.

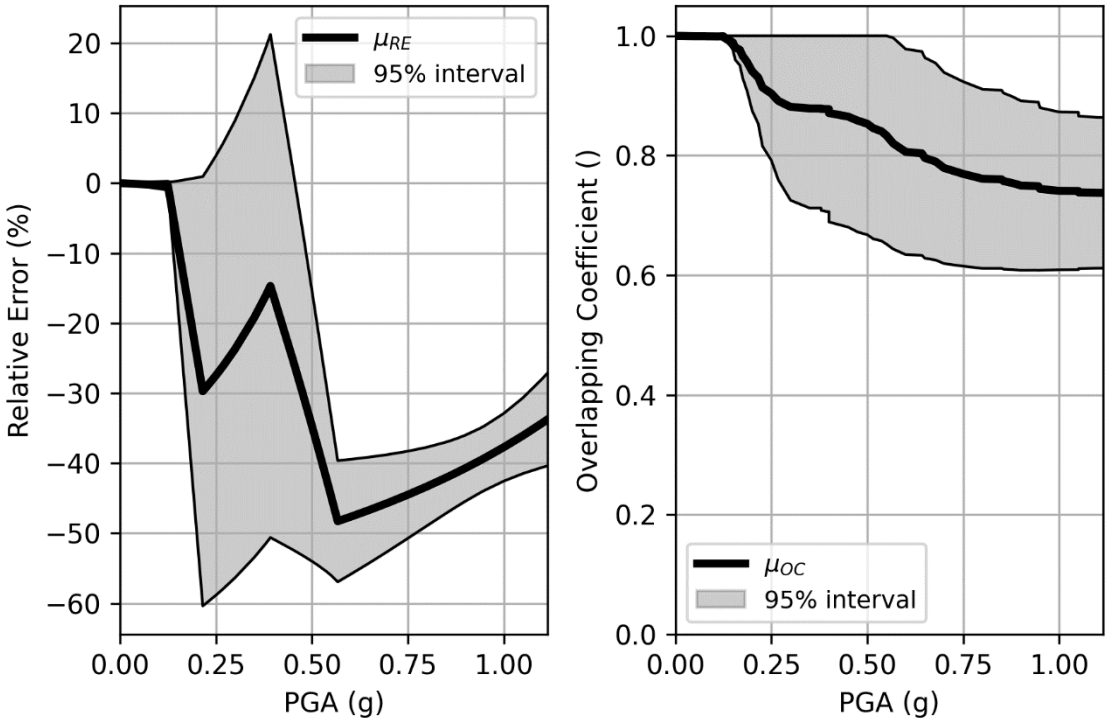


Figure 48: Relative Error and Overlapping coefficient over the validation domain of the relative evolution of response 1st peak frequency

Figure 49 displays the 95 % prediction intervals of the metrics and approaches introduced so far. In this case, because the metrics dots have a bell-shaped trend (first increasing, then decreasing, in absolute), a second order polynomial regression is chosen before computing the prediction interval. By looking at Figure 49 and to the opinion of the authors, this choice seems relevant over the validation domain and up to 1.8g. This choice can be challenged at higher intensities as the slope and interval would increase then higher and higher. Thus, the decreasing trend of the metrics in the large intensities can be represented, which would not have been possible with a linear or affine regression.

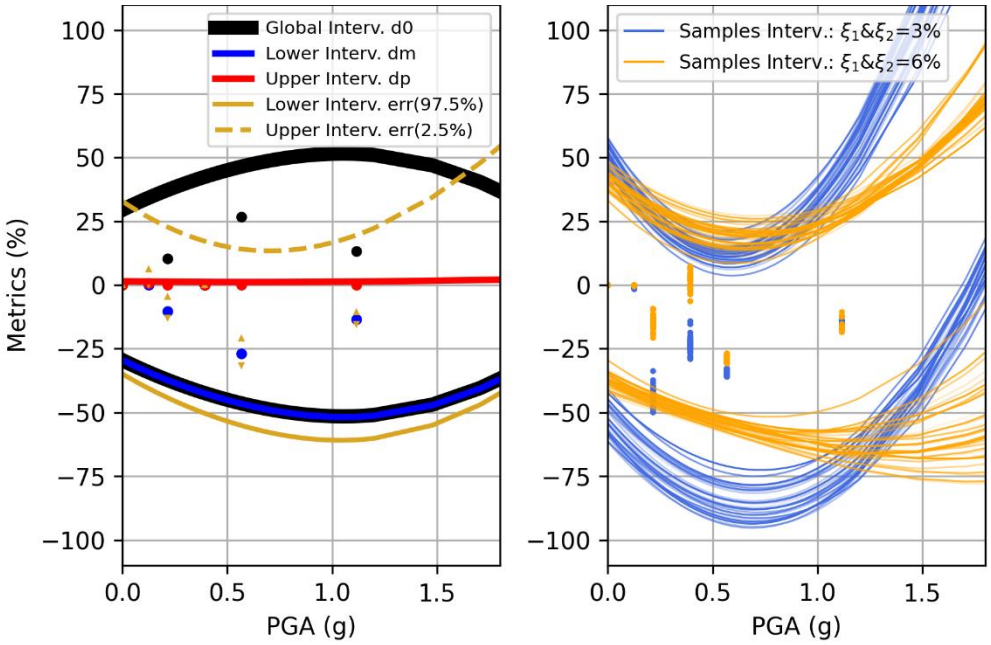


Figure 49: 95% prediction intervals of relative evolution of 1st eigenfrequency metrics

4.5. Comparison between Campaigns n°1 (linear analyses) and n°2 (nonlinear analyses)

Figure 50 displays the comparison between the medians μ (solid lines) and 95 % prediction intervals (shaded areas) of the drift for the linear and nonlinear studies (respectively in royal blue and golden) confronted to the experimental values.

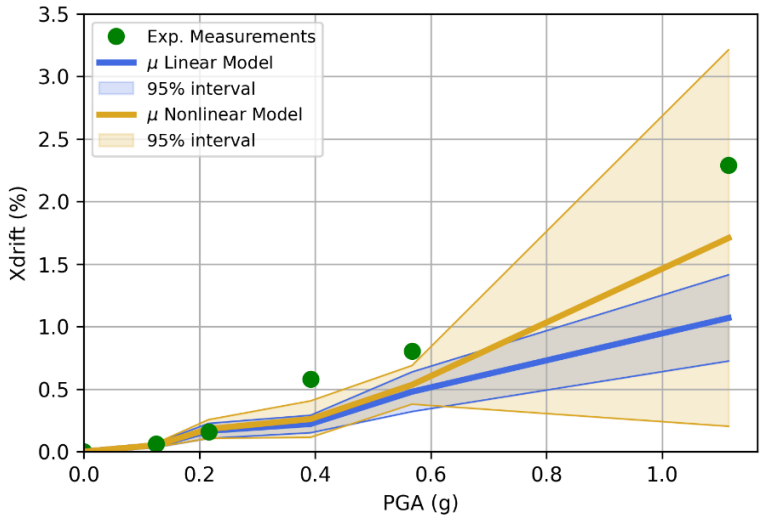


Figure 50: Comparison between Linear and Nonlinear studies

Accounting for nonlinearities leads to increases in the drift, i.e., the median curve gets closer to the experimental measurements as the record intensities grows. Conversely, the scatter increases strongly as well, leading the 95 % prediction envelop to widen significantly. One must recall here that, due to computational runtimes and this report due date, the nonlinear campaign was achieved with the material properties of the concrete sampled 30 times, unlike the 60 times of the linear studies. The standard-deviation will be expected to have converged at 30 samples.

The following Figure 51 displays the comparison between the linear and nonlinear campaigns for the relative errors and the integrated overlapping coefficients over the validation domain.

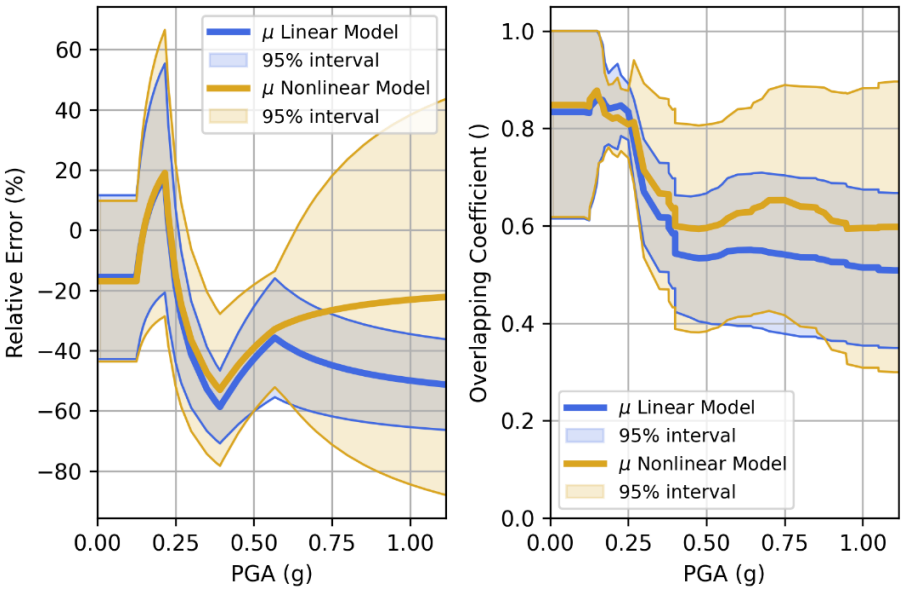


Figure 51: Comparison of relative errors and overlapping coefficients between linear and nonlinear analyses in terms of X drift values at point D

These two graphics allow for comparing both campaigns and for deciding which model should be used depending on the range of seismic intensity, or the targeted level of confidence, and should account for the computing runtime. Their patterns change notably when PGA reaches 0.25 g and beyond. Below that, nonlinearities are present but have small effects on the output. By looking specifically at the medians, the nonlinear model should be chosen for any intensity because the golden line is mostly over the blue one, i.e., the relative error is closer to zero, and the overlapping coefficient is closer to 1. However, the nonlinear golden intervals are much wider, leading in the range [0.8 ; 1.1] g to potential larger relative errors [-80 ; -90] %, and smaller overlapping coefficients [0.30 ; 0.35].

Figure 52**: Linear and Nonlinear campaigns comparison of 95 % prediction intervals of metrics d_0 (left hand-side) and d_m & d_p (right hand-side)** below compares the 95 % prediction intervals of the area validation metrics d_0 (left hand-side figure) and d_m & d_p (right hand-side figure) of the linear (blue areas) and nonlinear (golden areas) campaigns.

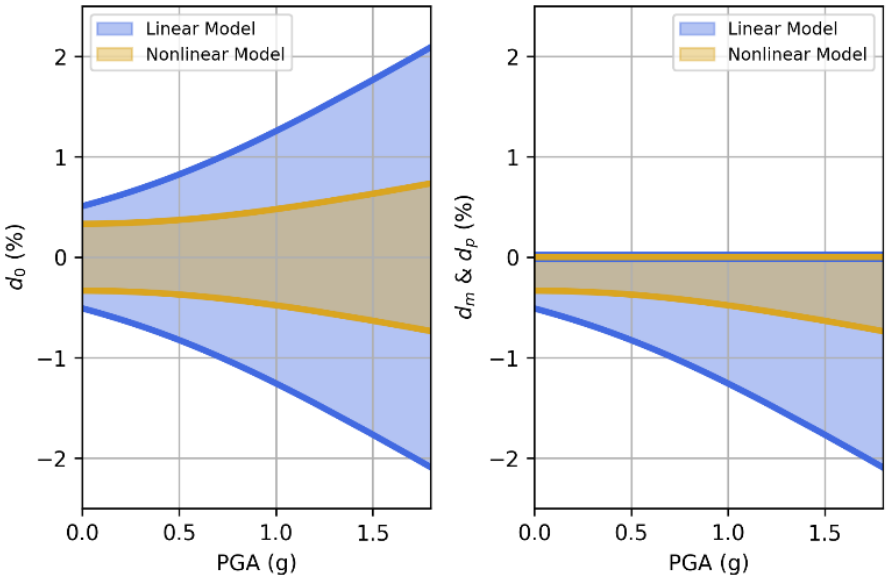


Figure 52: Linear and Nonlinear campaigns comparison of 95 % prediction intervals of metrics d_0 (left hand-side) and d_m & d_p (right hand-side)

The nonlinear model yields smaller area validation metrics over the domain compared to the linear model. Nonetheless, as abovementioned, area validation metrics are less informative as error metrics. Figure 53**: Linear and Nonlinear campaigns comparison of 95 % prediction intervals of metrics $e_{rr}(\alpha_{err} = 2.5\ \%)$ (left hand-side) and e_{rr} per combination (right hand-side)** below compares the 95 % prediction intervals of the error metrics $e_{rr}(\alpha_{err}$=2.5 %) (left hand-side figure) and e_{rr} per combination (right hand-side figure) of the linear (blue areas) and nonlinear (golden areas) campaigns.

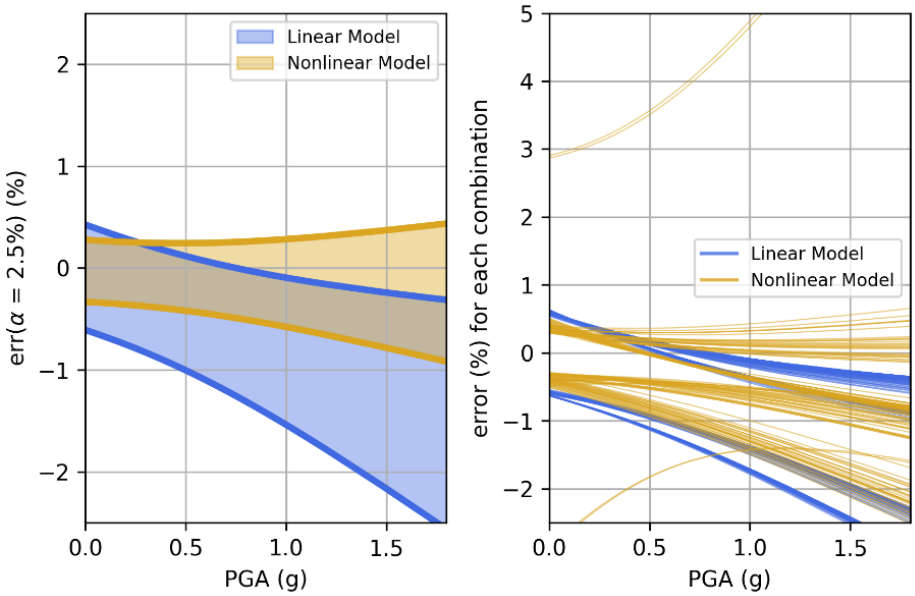


Figure 53: Linear and Nonlinear campaigns comparison of 95 % prediction intervals of metrics $e_{rr}(\alpha_{err} = 2.5\ \%)$ (left hand-side) and e_{rr} per combination (right hand-side)

Here for the $e_{rr}(\alpha_{err}$=2.5 %) metric in the left hand-side figure, nonlinear simulations yielded a 95 % prediction interval of smaller width and more centered around zero than the linear model. Regarding the direct error e_{rr} metrics calculated for each combination in the right hand-side figure, the conclusion is similar, most of the golden intervals are closer to the zero abscissa and with a narrower width. One notices nonetheless, among the 60 golden intervals, two of them are outliers, with significantly wide intervals caused by the nonlinearities, and could be discarded.

4.6. 1.78 g PGA study case outside the validation domain: simulations and corrections by the metrics

For the analysis to be consistent, the same input epistemic and aleatory parameters than precedingly are used. This totalizes 120 outputs, divided in two epistemic cases, and displayed at PGA 1.78 g on Figure 54. Only the linear campaign could be fully achieved at 1.78 g PGA, the nonlinear simulation campaign being stopped because of time-consuming simulations and numerical convergence difficulties at such seismic intensity. EDPs extend from 0.85 % to 1.50 %, which are smaller than drifts calculated at a smaller 1.1 g PGA. The resulting numerical CPDs are displayed in green on Figure 55. However, there is no experimental reference at 1.78 g to compare to. Then, an extrapolation of the metrics computed at lower PGA has to be done.

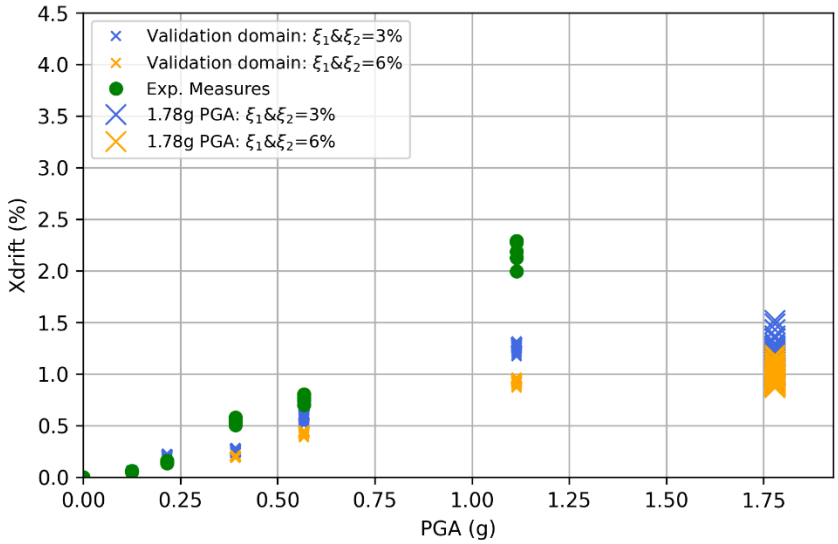


Figure 54: X drifts (coloured crosses) at 1.78 g PGA (linear model)

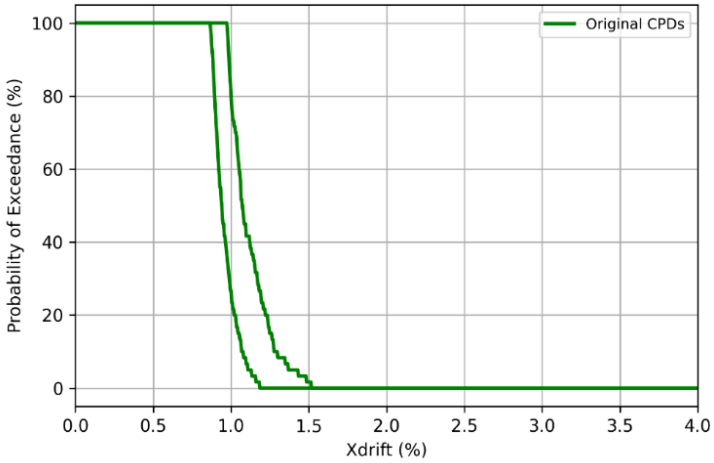


Figure 55: Numerical CPDs of X drift 1.78 g PGA Run before correction (linear model)

Now will successively come the corrections by the metrics extrapolated at the PGA of interest. The reading of each correction can be done using Figure 46. First starting with $\boldsymbol{d_0}$ metric.

At abscissa 1.78 g, the lower and upper black intervals of $\boldsymbol{d_0}$ yield respectively a correction of -1.85 % and +1.85 %. Consequently, in the method by (Roy and Oberkampf, 2011), the left and right hand-sides CPDs are respectively offset by -2.07 % and +2.07 %, while also being limited by the 0 threshold. The resulting probability-box in Figure 56 in grey area extends from 0 to 3.58 % which is the threshold to be overpassed with a 95 % confidence.

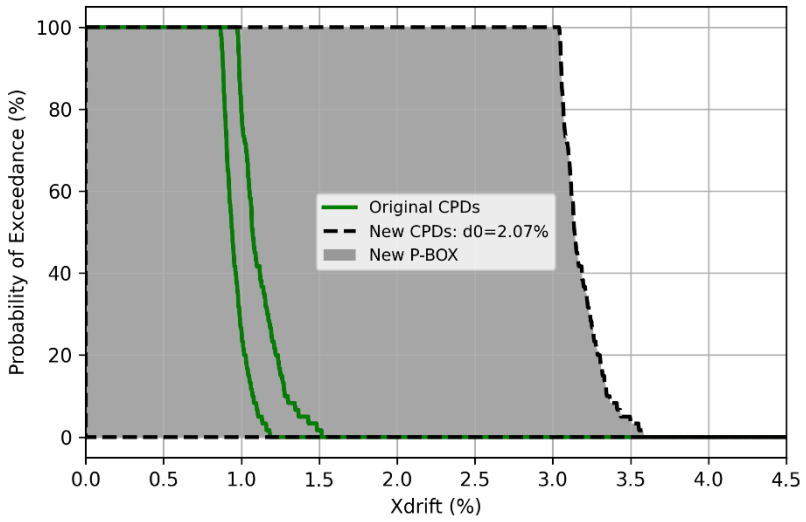


Figure 56: Correction of CPD X drifts with d_0 metric on both sides (linear model)

Now looking at the area validation metrics $\boldsymbol{d_m}$ and $\boldsymbol{d_p}$, their respective blue and red envelops in Figure 46 yield -2.07 % and 0.0 % at PGA 1.78 g. The underestimation is now clearly visible unlike with $\boldsymbol{d_0}$. The right-hand side CPD is then translated to the right by a 2.07 % value. The resulting probability-box extends now from 0.85 % to 3.58 % with a 95 % confidence on the upper threshold, as displayed on Figure 57. One could argue that, since numerical model shows no sign of overestimating the measurements, then the left hand-side CPD could as well be shifted to the right also by 2.07 % which would result to be as conservative and acceptable as before, but more accurate. Such a case would result in a probability-box extending from 2.87 % to 3.58 %. However, until now, the method is strictly applied without considering specificities to reduce intervals furthermore.

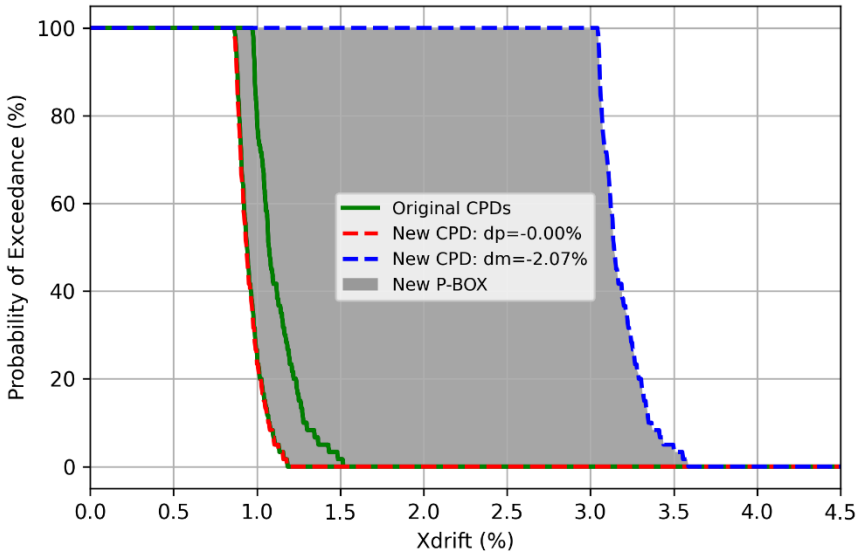


Figure 57: Correction of CPD X drifts with d_m and d_p metrics (linear model)

Now looking at the $\boldsymbol{e_{rr}}$ with a $\mathbf{2.5}$ % quantile, the golden solid $\boldsymbol{e_{rr}}(\mathbf{97.5}\,\%)$ and dashed $\boldsymbol{e_{rr}}(\mathbf{10}\,\%)$ prediction intervals of Figure 46 respectively yield corrections values of -0.31 % and -2.54 %. Then, the left and right hand-sides original CPDs are offset respectively by +0.31 % and +2.54 %. Now, the final probability-box displayed in Figure 58 extends from 1.15% to 4.07%. The global interval is wider.

But one must remember here that, not only the prediction intervals are calculated with a 95 % confidence, but also that the errors themselves are calculated with a 5 % quantile, i.e., there is a 2.5 % probability that errors $\boldsymbol{e_{rr}}$ are larger than $\boldsymbol{e_{rr}}(\mathbf{97.5}\,\%)$ in absolute, which is non neglectable.

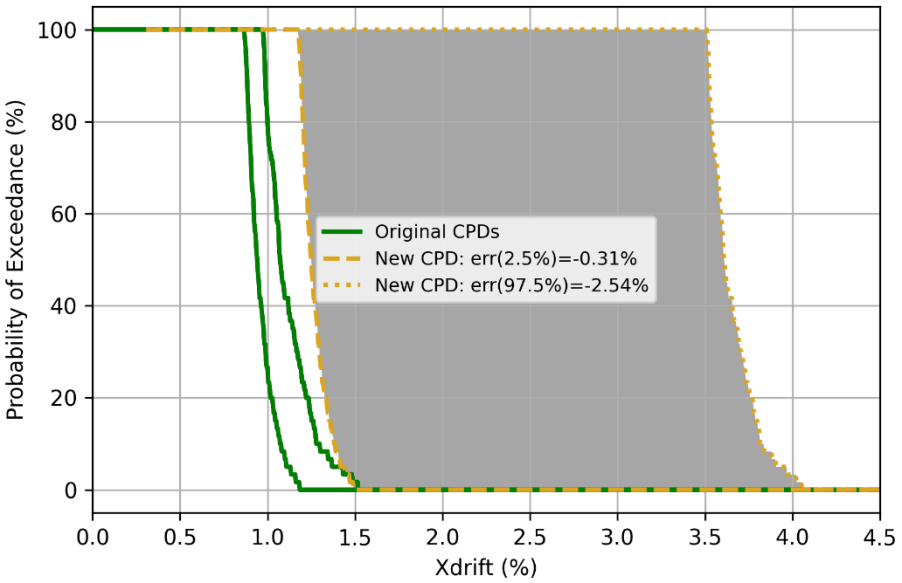


Figure 58: Correction of CPD X drifts with e_{rr}(α_{err}=2.5 %) (linear model)

Finally, now is presented the approach where all sample cases are corrected individually by their respective 95 % prediction interval. Figure 46 presents all of them, classified by the damping coefficient values. For each sample case, the drift correction extrapolated at 1.78 g is taken and added to the drift output of the same sample case computed at 1.78 g PGA. A new set of drift values is obtained and plotted again as two new CPDs in royal blue in Figure 59 below. The new probability-box extends now from 3.24 % to 4.10 %.

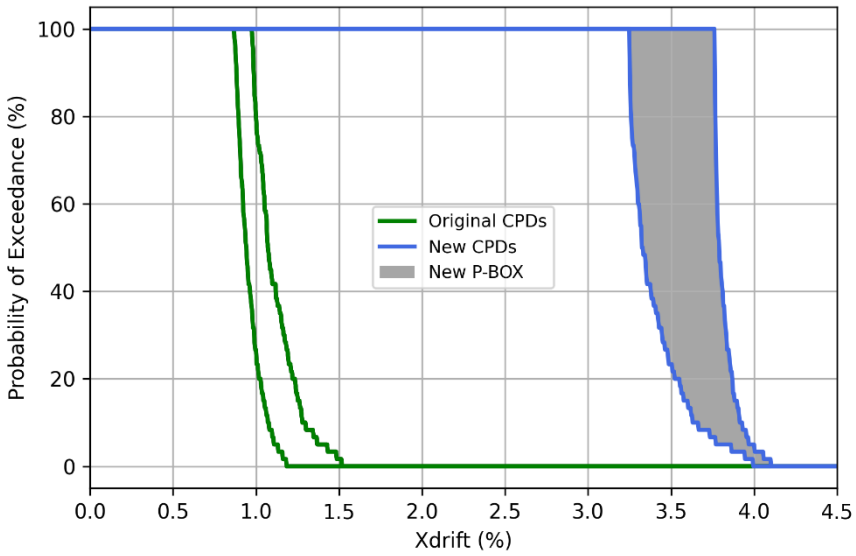


Figure 59: Correction of CPD X drifts for each input sample (linear model)

Last Figure 60 and Figure 61 allow for comparing the different probability-boxes and seeing the differences in width and value between them. Eventually, the application of each method (from the previous three) should depend on the level of precision one aims at performing a validation process through a given metric. Indeed, in our case, the precision increases with the complexity of the post processing of the selected metrics. A graded approach, as described in (AIEA, 2014) based on the importance of the outcome at sight seems the right way to address the problem.

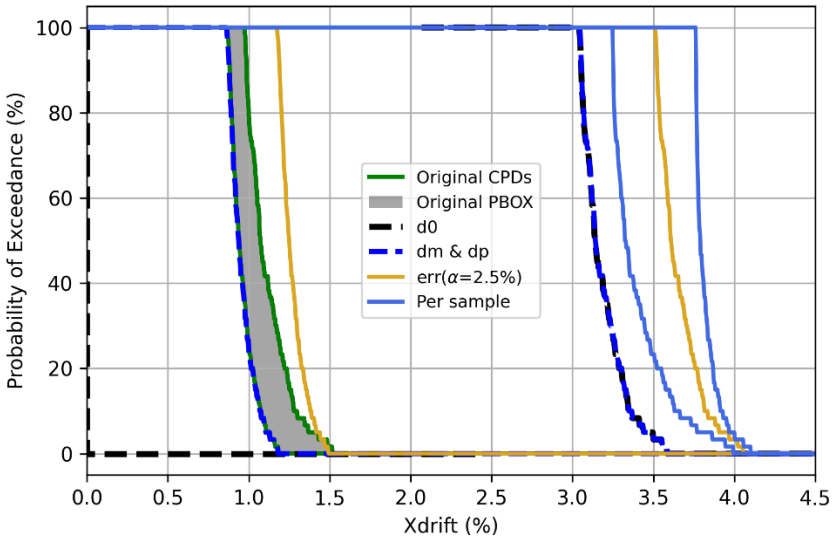


Figure 60: Comparison of corrected X drifts P-boxes (linear model)

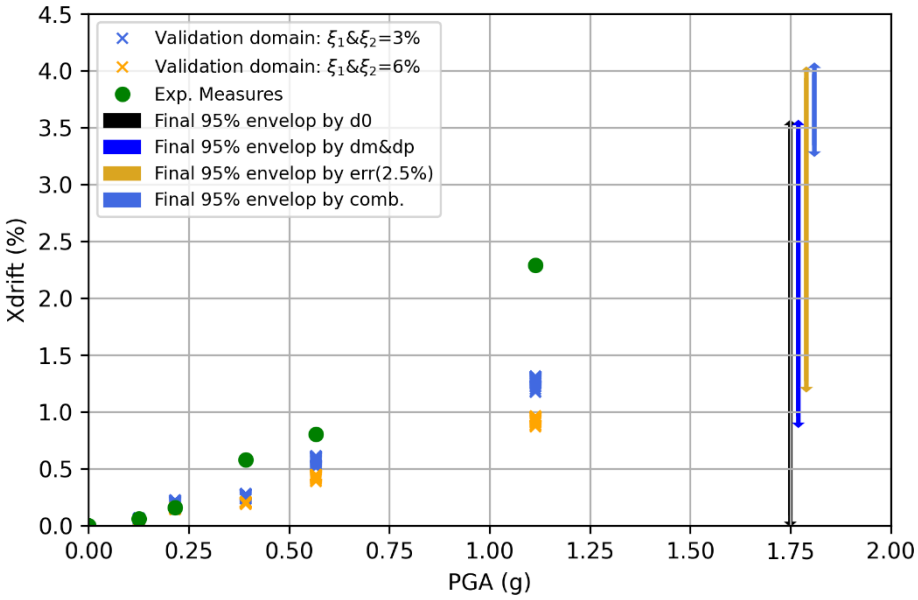


Figure 61: 95% X drift envelops at 1.78 g PGA after correction by the different metrics (linear model)

4.7. 0.83g PGA study case: simulations and corrections by the metrics

The precedent subsection 4.6 considered the original recorded 1.78 g PGA Northridge motion applied on SMART linear model, nonlinearities could not be accounted for due to convergence difficulties and expensive computational runtimes. The numerical output outside the validation domain were corrected by the various metrics of the four different methods presented throughout Section 3. The present subsection aims at doing the same corrections but on the output of a smaller intensity record of Northridge 0.83 g PGA where both linear and nonlinear campaigns could be achieved.

From Figure 62 and Figure 67, the original CPD is displayed, then corrected respectively by the area metrics $\boldsymbol{d_0}$, $\boldsymbol{d_m}$ & $\boldsymbol{d_p}$, and errors metrics $\boldsymbol{e_{rr}}(\boldsymbol{\alpha_{err}}$=2.5 %) and error $\boldsymbol{e_{rr}}$ for each combination, before being altogether displayed. Again, in a like manner as in subsection 4.6, the error metrics and

approaches yield more acceptable intervals, larger upper limits, with the latter method also with a narrower interval. The engineer can then read the upper 95 % limit expected at 0.83 g PGA.

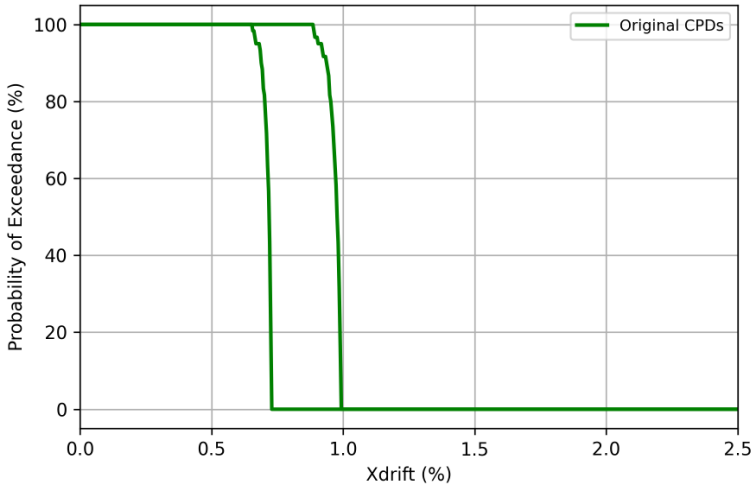


Figure 62: CPDs of X drift 0.83 g PGA Run before correction

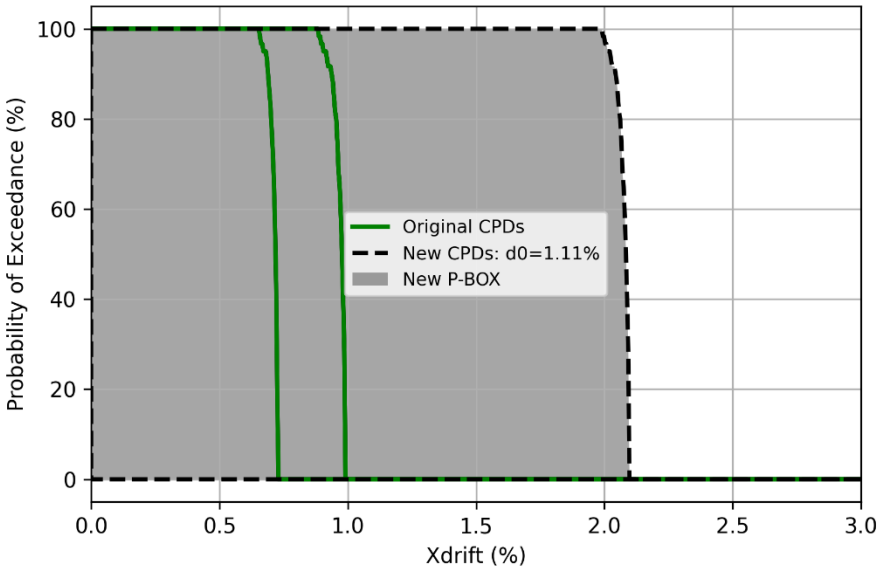


Figure 63: Correction of CPD X drifts with d_0 metric on both sides

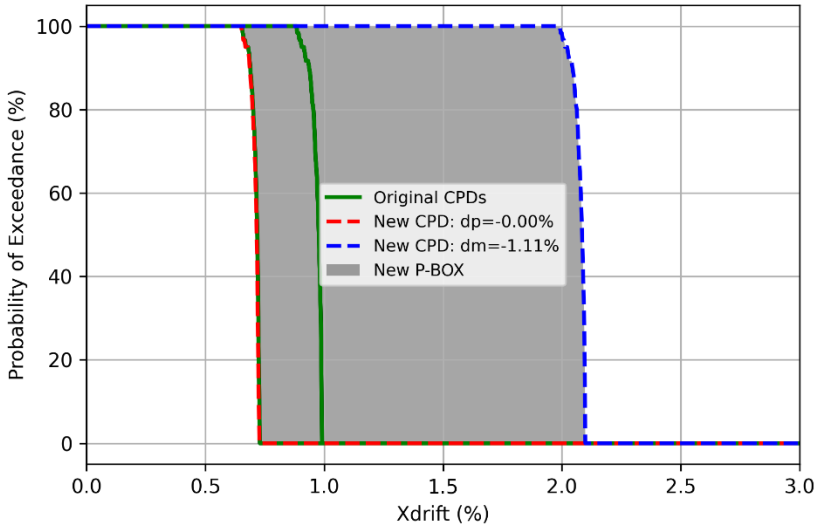


Figure 64: Correction of CPD X drifts with d_m and d_p metrics

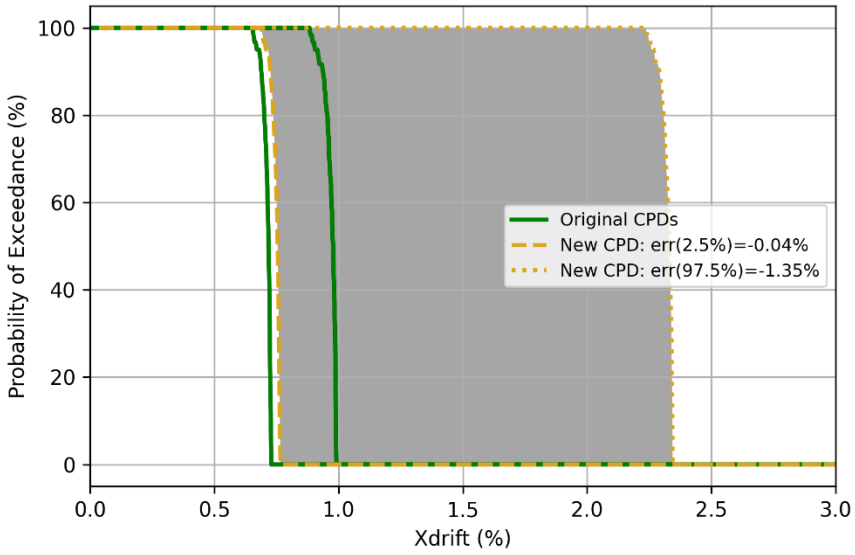


Figure 65: Correction of CPD X drifts with $e_{rr}(\alpha_{err}$=2.5%)

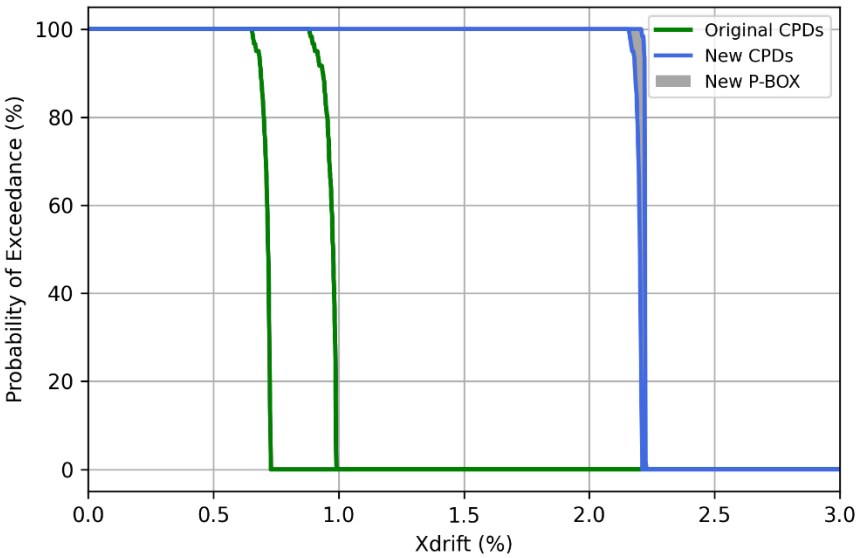


Figure 66: Correction of CPD X drifts for each input sample

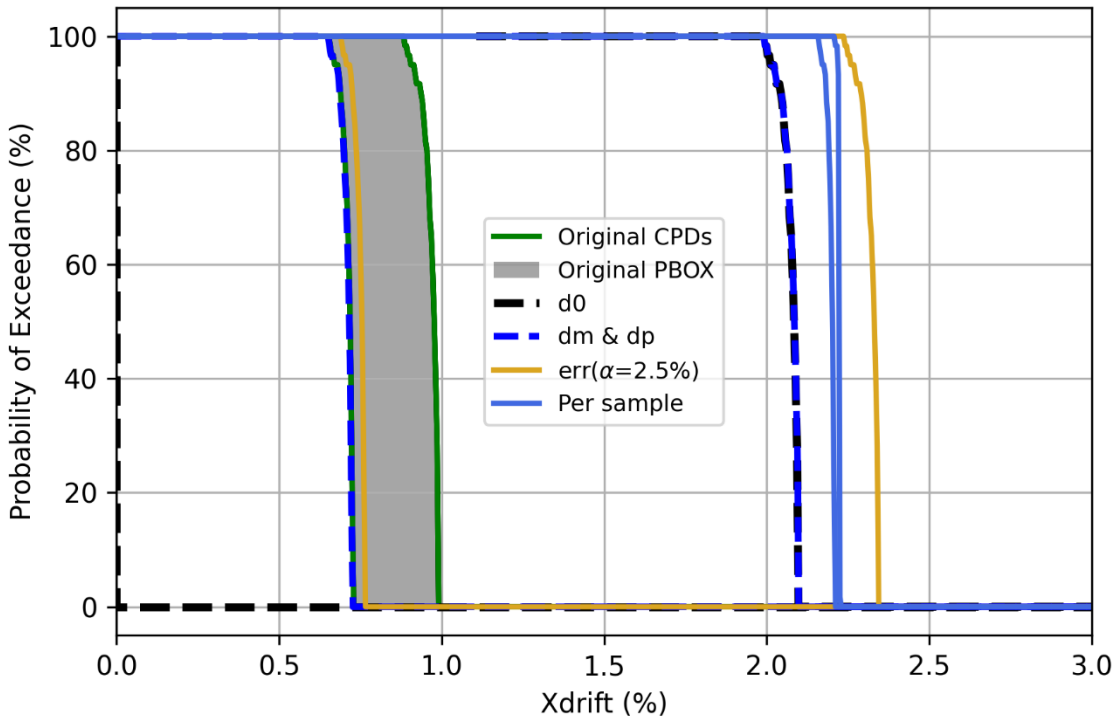


Figure 67: 95% X drift envelops at 0.83 g PGA after correction by the different metrics

5. Conclusions and recommendations

In the present report, the following points have been successively presented:

- The definitions of verification and validation as well as the state of literature on the matter.
- The validation flowchart step by step with the article by (Roy and Oberkampf, 2011) and its approach used for illustration. An introduction to the sources of uncertainties is made.
- Several propositions of improvement of (Roy and Oberkampf, 2011) approach: for instance, they include modifying the area validation metric d_0, or using rather the error e_{rr} at a given quantile in order to provide richer information regarding the overestimation or underestimation tendency of the model.
- SMART2013 experimental campaign, including its dimensions, modal behavior, the various seismic tests retained.
- The validation process applied onto SMART2013 with the various approaches abovementioned. Linear and nonlinear numerical models are considered. Seismic records outside the validation domain are also studied.

Overall, the following conclusions and recommendations are made by the author:

- Global remarks:
  - The variety of application cases in the civil engineering field is vast, and no unique method exists to validate all existing studies.
  - Accounting for uncertainties is necessary to obtain an informed and rigorous outcome. However, assessing uncertainties highly depends on the number of available experimental measurements.
  - The chief-engineers willing to evaluate the capabilities of a numerical model should know beforehand the indicator and the value of the associated critical threshold upon which a model will be considered valid 'enough' or not.
  - The author points out that a model can be invalid for a particular EDP of interest, but valid for another EDP. The validation process is EDP dependent.
  - If a numerical model shows poor accuracy but high repeatability and small scatter, the output may be corrected by the metrics regression and confidence interval computation at a given condition of interest.
  - There is no unique regression and confidence interval computation method. It highly depends on the number of measurements available, and on the engineer knowledge of the physical phenomenon at stake and capabilities of the numerical model.
  - The Validation process can also be used for the comparison analysis of various numerical models.
- Recommendations:
  - This report covers 4 validation metrics (d_0, d_m & d_p , $e_{rr}(\alpha_{err})$, e_{rr} per input combination), the selection of each metric depends on the level of precision one aims at. A graded approach based on the importance of the outcome and the precision at sight seems the right way to perform such selection.
  - Depending on the application cases, numerical resources, and time availability, the number of propagated samples will change and must be justified. Sensitivity analyses and EDP-IM correlation should be performed, if possible, to justify rigorously the choice of the variables to propagate, eliminate the less influent, and reduce ultimately the computational cost.
  - **Adding relevant physical data based on existing feedbacks or common-sense analyses is recommended** to improve the computations of the regression and prediction interval. For instance, setting a null drift at a null seismic intensity, or a threshold for eigenfrequencies drops where it is known that it cannot be lower than zero.
  - **Adding relevant physical relations or properties** between inputs, or EDPs based on known physical phenomena.
  - Considering a systematic linear regression might lead to non-negligible errors and mismatches. One should achieve a regression following the shape of the metrics distribution over the validation domain.

- If possible, experimental uncertainties have to be accounted for by considering experimental p-boxes in a similar fashion as model output p-boxes.

6. Bibliography

AIAA. (2002). *Guide for the verification and validation of computational fluid dynamics simulations*.

AIEA (2014). *Use of a Graded Approach in the Application of the Management System Requirements for Facilities and Activities*. IAEA-TECDOC-1740 ¦ 978-92-0-105114-1, 2014

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S., ... & Donahue, J. L. (2013). *PEER NGA-west2 database* (Vol. 172). Berkeley, CA: Pacific Earthquake Engineering Research Center.

ANS. (1987). *Guidelines for the verification and validation of scientific and engineering computer programs for the nuclear industry* [Norme]. American Nuclear Society.

ASME. (2007). Guide for Verification & Validation in Computational Solid Mechanics—PTC 60/V&V 10. *ASME*.

ASN. (2017). *Qualification des outils de calcul scientifique utilisés dans la démonstration de sûreté nucléaire—1ère barrière*. ASN.

Bouhjiti D., (2018). *Analyse probabiliste de la fissuration et du confinement des grands ouvrages en béton armé et précontraint* (Doctoral dissertation, Université Grenoble Alpes).

Bouhjiti D., Julien Baroth, Frédéric Dufour, Benoit Masson (2017). Analyse de sensibilité de la fissuration des grands ouvrages en béton armé et précontraint aux effets d'échelles - Cas de la maquette VERCORS*. CFM 2017 - 23ème Congrès Français de Mécanique*.

Buratti, N. (2012, September). A comparison of the performances of various ground–motion intensity measures. In Proceedings of the 15th world conference on earthquake engineering, Lisbon, Portugal (pp. 24-28).

Capdevielle, S. (2017). Élément poutre multifibre avec gauchissement pour la modélisation non linéaire des structures en béton armé. *Academic Journal of Civil Engineering*, *35*(1), 496-503.

Charbonnel, P.-E. (2021). Fuzzy-driven strategy for fully automated modal analysis: Application to the SMART2013 shaking-table test campaign. *Mechanical Systems and Signal Processing*, *152*, 107388.

Commissariat à l'énergie atomique et aux énergies alternatives (2011). Assurance Qualité Cast3m, Cast3m : Guide de Validation http://www-cast3m.cea.fr/index.php?xml=maj2011#validation

Crozet, V., Politopoulos, I., Yang, M., Martinez, J. M., & Erlicher, S. (2018). Sensitivity analysis of pounding between adjacent structures. *Earthquake Engineering & Structural Dynamics*, *47*(1), 219-235.

Da Veiga, S., Gamboa, F., Iooss, B., & Prieur, C. (2021). *Basics and trends in sensitivity analysis: Theory and practice in R*. Society for Industrial and Applied Mathematics.

De Biasio, M. (2014). *Ground motion intensity measures for seismic probabilistic risk analysis* (Doctoral dissertation, Université de Grenoble).

Direction de l'Energie Nucléaire (2014). SMART 2013 Test report.

Devore, J. L. (2015). *Probability and Statistics for Engineering and the Sciences*. Cengage Learning.

Dimov, I., & Georgieva, R. (2010). Monte Carlo algorithms for evaluating Sobol'sensitivity indices. *Mathematics and computers in simulation*, *81*(3), 506-514.

Eads, L., Miranda, E., & Lignos, D. G. (2015). Average spectral acceleration as an intensity measure for collapse risk assessment. Earthquake Engineering & Structural Dynamics, 44(12), 2057-2073.

Görtler, et al., (2019). A Visual Exploration of Gaussian Processes, Distill.

Heitz, T. (2017). *Comportements non-linéaires locaux et modélisation numérique de l'amortissement dans les structures de génie civil en dynamique* [Thèse de doctorat].

IAEA. (1999). *Verification and validation of software related to nuclear power plant instrumentation and control*. IAEA.

IEEE. (1984). IEEE Standard Dictionary of Electrical and Electronics Terms. *IEEE Transactions on Microwave Theory and Techniques*, *32*(11), 1521-1521. https://doi.org/10.1109/TMTT.1984.1132885

Iooss, B., & Saltelli, A. (2017). Introduction to sensitivity analysis. Handbook of uncertainty quantification, 1103-1122.

Kat, C.-J., & Els, P. S. (2012). Validation metric based on relative error. *Mathematical and Computer Modelling of Dynamical Systems*, *18*(5), 487-520. https://doi.org/10.1080/13873954.2012.663392

Kwag, S., Gupta, A. & Dinh N. (2018). Risk-informed framework for verification and valildation. *Reliability Engineering and System Safety, 169,* 380-393.

Herman, J. D., Kollat, J. B., Reed, P. M., & Wagener, T. (2013). Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models. *Hydrology and Earth System Sciences*, *17*(7), 2893-2903.

Langlade, T. (2021). Vulnerability and risk analyzes of structures subjected to earthquake-induced building pounding with a non-smooth contact dynamics method (Doctoral dissertation, Université de Lyon).

Larrard, T. D. (2010). Variabilité des propriétés du béton : caractérisation expérimentale et modélisation probabiliste de la lixiviation. Ph.D. thesis, *École Normale Supérieure de Cachan*.

Oberkampf, W. L., & Barone, M. F. (2006). Measures of agreement between computation and experiment : Validation metrics. *Journal of Computational Physics*, *217*(1), 5-36. https://doi.org/10.1016/j.jcp.2006.03.037

Mara, T. A. (2009). Extension of the RBD-FAST method to the computation of global sensitivity indices. *Reliability Engineering & System Safety*, *94*(8), 1274-1281.

MathWorks – *Editeur de MATLAB et Simulink*. https://fr.mathworks.com/

IRSN (2022) Technical Report: IRSN-2022-00707 –Verification and validation of nonlinear mechanical model based on IDEFIX experimental campaign.

NPARC Alliance (2022): *Examining Spatial (Grid) Convergence*. (s. d.). Adresse https://www.grc.nasa.gov/www/wind/valid/tutorial/spatconv.html

Pedregosa et al., 2011. Scikit-learn: Machine Learning in Python, *JMLR* 12, pp. 2825-2830.

Porter, N. W. (2019). Wilks' formula applied to computational tools: A practical discussion and verification. *Annals of Nuclear Energy*, *133*, 129-137.

Rasmussen, C. E., Williams C. K. I., 2006. Gaussian Processes for Machine Learning*, the MIT Press, ISBN 026218253X*. c 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

Richard, B., Voldoire, F., Fontan, M., Mazars, J., Chaudat, T., Abouri, S. & Bonfils, N. (2018). SMART 2013: Lessons learned from the international benchmark about the seismic margin assessment of nuclear RC buildings. *Engineering Structures*, *161*, 207-222.

Richard, B., Cherubini, S., Voldoire, F., Charbonnel, P. E., Chaudat, T., Abouri, S., & Bonfils, N. (2016). SMART 2013: Experimental and numerical assessment of the dynamic behavior by shaking table tests

of an asymmetrical reinforced concrete structure subjected to high intensity ground motions. *Engineering Structures*, *109*, 99-116.

Richard, B., Martinelli, P., Voldoire, F., Corus, M., Chaudat, T., Abouri, S., & Bonfils, N. (2015). SMART 2008: Shaking table tests on an asymmetrical reinforced concrete structure and seismic margins assessment. *Engineering Structures*, 105, 48-61.

Richard, B. (2012). SERIES/ENISTAT Project. Preliminary numerical time history analysis. Internal CEA Technical report RT-12-013/A.

Richard, B., Lebon, G., & Ragueneau, F. (2012). 3D MODELLING OF CONCRETE FOR EARTHQUAKE ANALYSIS: DAMAGE MECHANICS AND PLASTICITY COUPLING. *Blucher Mechanical Engineering Proceedings*, *1*(1), 3269-3277.

Richard B. & Ragueneau F. (2012). Continuum damage mechanics-based model for quasi-brittle materials subjected to cyclic loadings: formulation, numerical implementation and applications. Engineering Fracture Mechanics.

Roache, P. J. (1994). Perspective: A Method for Uniform Reporting of Grid Refinement Studies. *Journal of Fluids Engineering*, *116*(3), 405-413. https://doi.org/10.1115/1.2910291

Roy, C. J. (2005). Review of code and solution verification procedures for computational simulation. *Journal of Computational Physics*, *205*(1), 131-156. https://doi.org/10.1016/j.jcp.2004.10.036

Roy, C. J. (2010). *Review of discretization error estimators in scientific computing*. 29.

Roy, C. J., & Oberkampf, W. L. (2011). A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing. *Computer Methods in Applied Mechanics and Engineering*, *200*(25-28), 2131-2144. https://doi.org/10.1016/j.cma.2011.03.016

Thacker, B. H., Doebling, S. W., Hemez, F. M., Anderson, M. C., Pepin, J. E., & Rodriguez, E. A. (2004). *Concepts of Model Verification and Validation* (p. 41). Los Alamos National Laboratory. http://inis.iaea.org/search/search.aspx?orig_q=RN:36030870

Trucano, T. G., Easterling, R. G., Dowding, K. J., Paez, T. L., Urbina, A., Romero, V. J., Rutherford, B. M., & Hills, R. G. (2001). *Description of the Sandia validation metrics project*.

Trucano, T. G., Pilch, M., & Oberkampf, W. L. (2002). *General Concepts for Experimental Validation of ASCI Code Applications*. Sandia National Labs., Albuquerque, NM (US); Sandia National Labs., Livermore, CA (US). https://doi.org/10.2172/800777

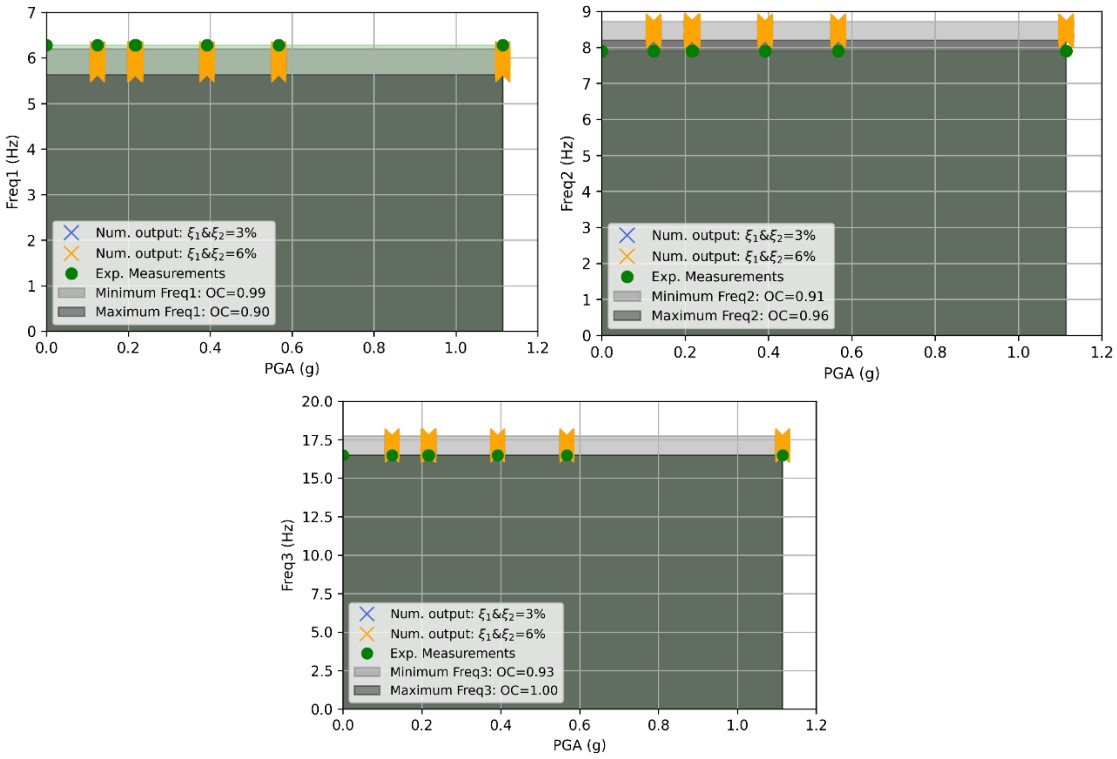
Wikipedia contributors. (2022, August 22). Gaussian process. In Wikipedia, The Free Encyclopedia. Retrieved 13:12, October 18, 2022, from https://en.wikipedia.org/w/index.php?title=Gaussian_process&oldid=1105870192

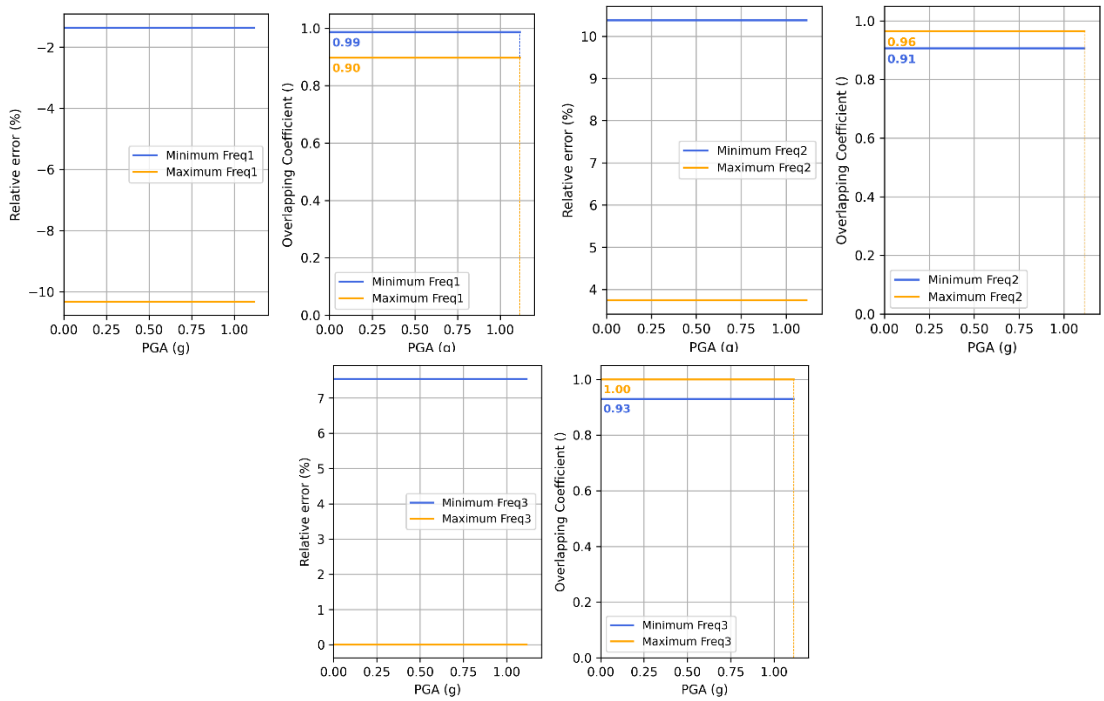
Xiong, C., Lu, X., Lin, X., Xu, Z., & Ye, L. (2017). Parameter determination and damage assessment for THA-based regional seismic damage prediction of multi-story buildings. *Journal of Earthquake Engineering*, *21*(3), 461-485.

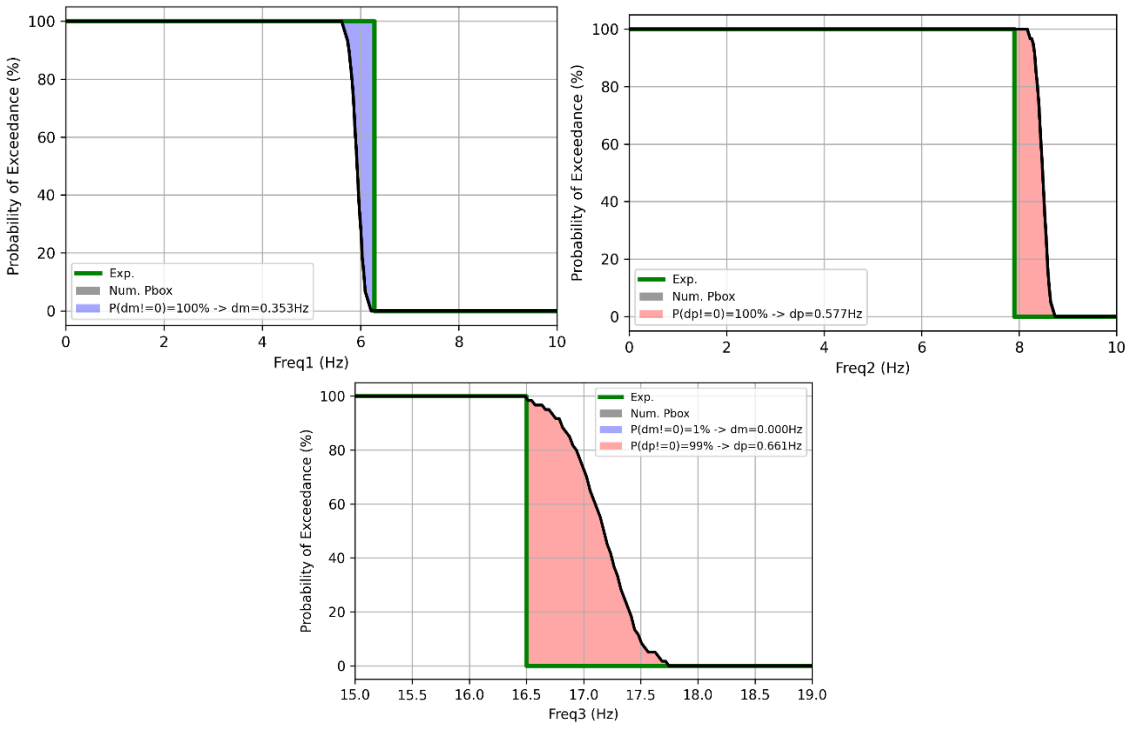
7. Annexes

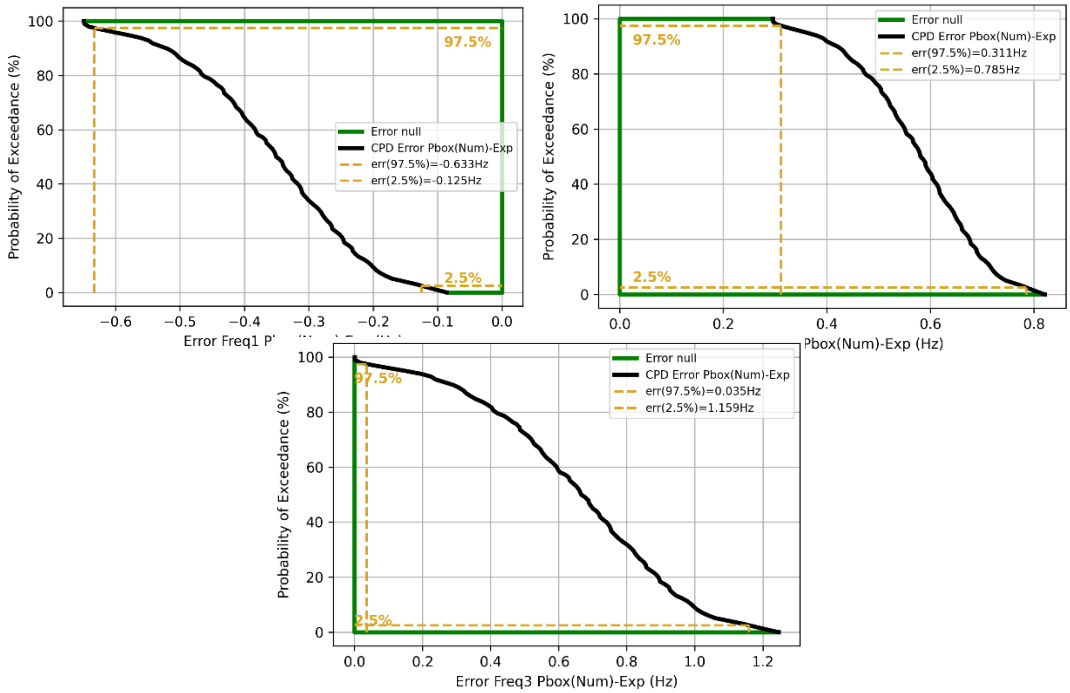
7.1. Eigenfrequencies and Floor response spectra of Campaign n°1

7.1.1. Eigenfrequencies




Relative error (%)
PGA (g)
Minimum Freq1
Maximum Freq1
Overlapping Coefficient ()
0.99
0.90
Minimum Freq1
Maximum Freq1
Relative error (%)
Minimum Freq2
Maximum Freq2
Overlapping Coefficient ()
0.96
0.91
Minimum Freq2
Maximum Freq2
Relative error (%)
Minimum Freq3
Maximum Freq3
Overlapping Coefficient ()
1.00
0.93
Minimum Freq3
Maximum Freq3


Probability of Exceedance (%)
Exp.
Num. Pbox
P(dm!=0)=100% -> dm=0.353Hz
Freq1 (Hz)
Probability of Exceedance (%)
Exp.
Num. Pbox
P(dp!=0)=100% -> dp=0.577Hz
Freq2 (Hz)
Probability of Exceedance (%)
Exp.
Num. Pbox
P(dm!=0)=1% -> dm=0.000Hz
P(dp!=0)=99% -> dp=0.661Hz
Freq3 (Hz)



7.1.2. Floor response spectra Sax(30Hz)

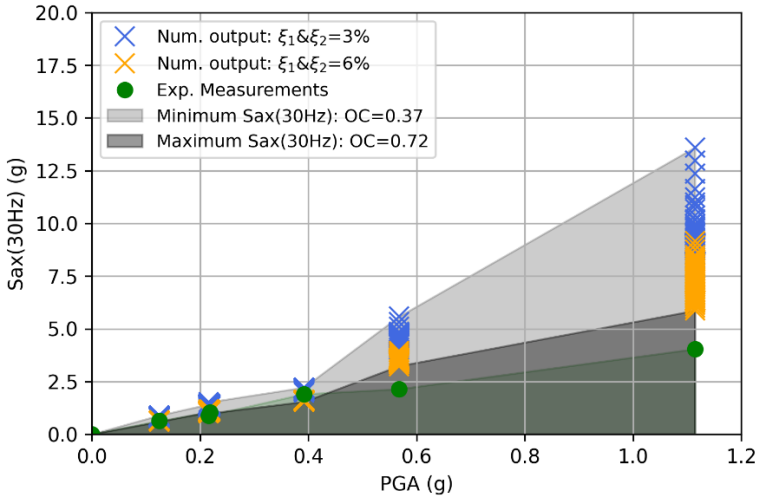


Figure 68: Numerical and Experimental comparison of Sax(30Hz) values

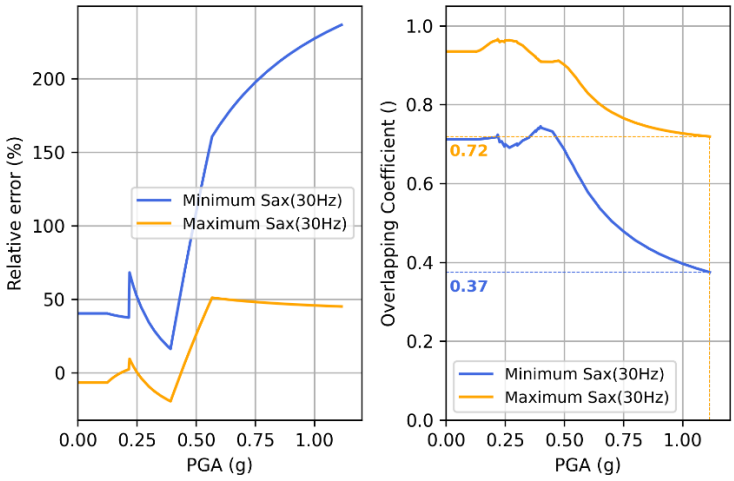


Figure 69: Evolution of relative error and overlapping coefficient

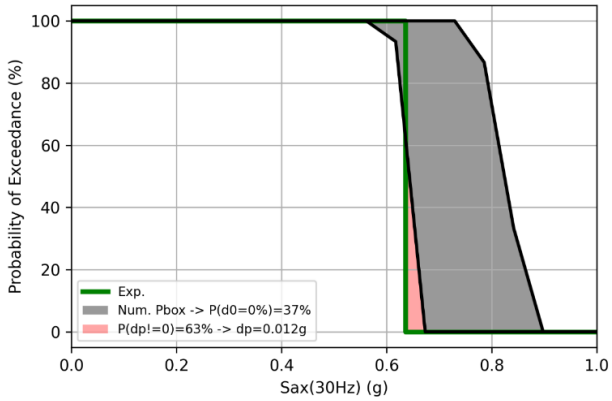


Figure 70: 0.12g PGA Run 07: Experimental Measure and Probability-box of Sax(30Hz)

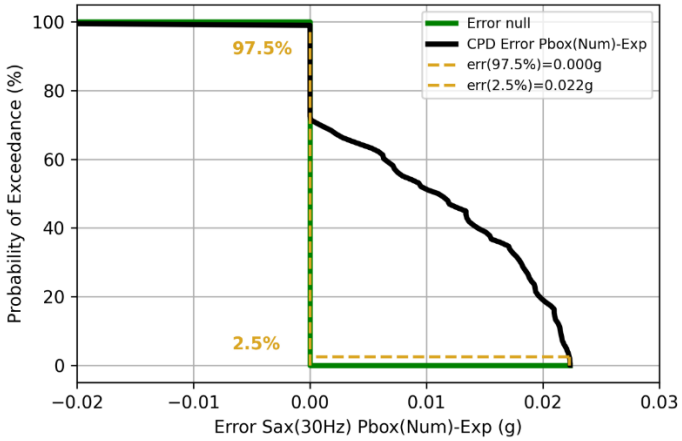


Figure 71: Run 07: Error e_{rr} of Sax(30Hz) p-box for a given quantile $\alpha_{err} = 2.5\%$

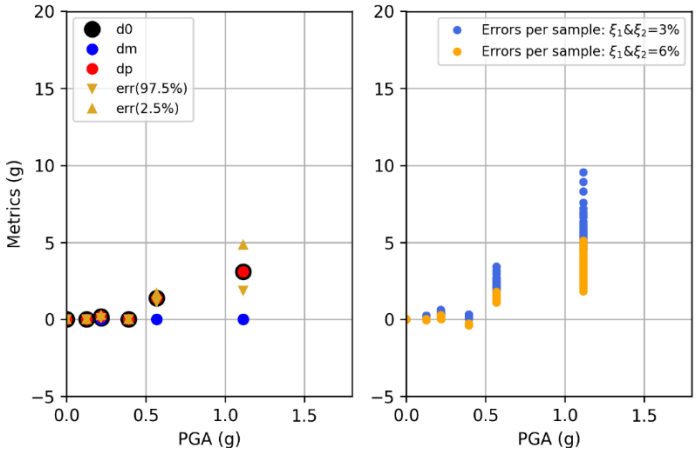


Figure 72: Sax(30Hz): List of metrics Versus Intensity Measure

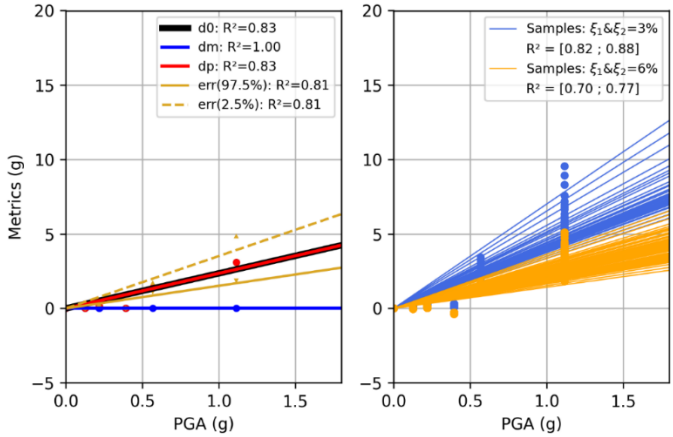


Figure 73: Linear regressions of metrics of Sax(30Hz)

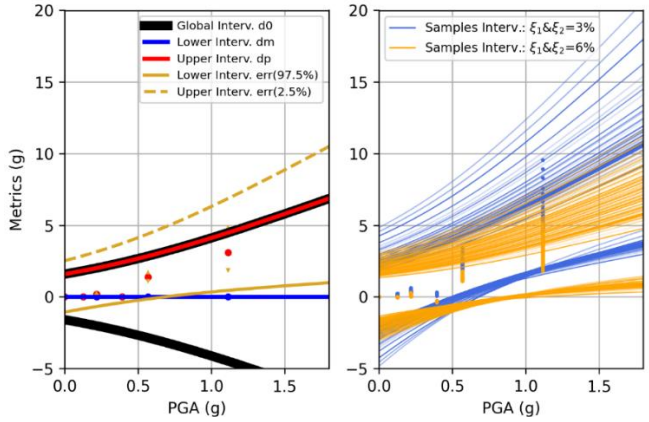


Figure 74: 95% Student Prediction Intervals of metrics of Sax(30Hz)

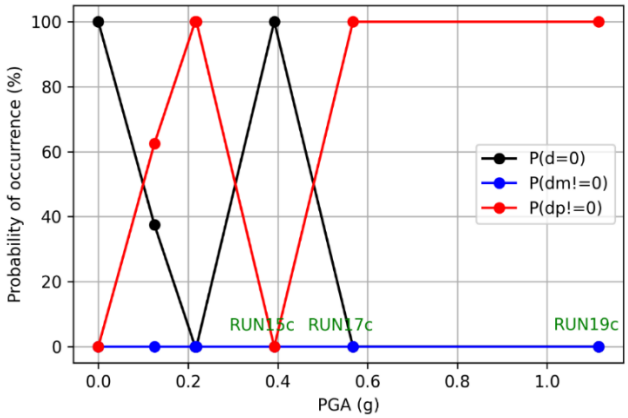
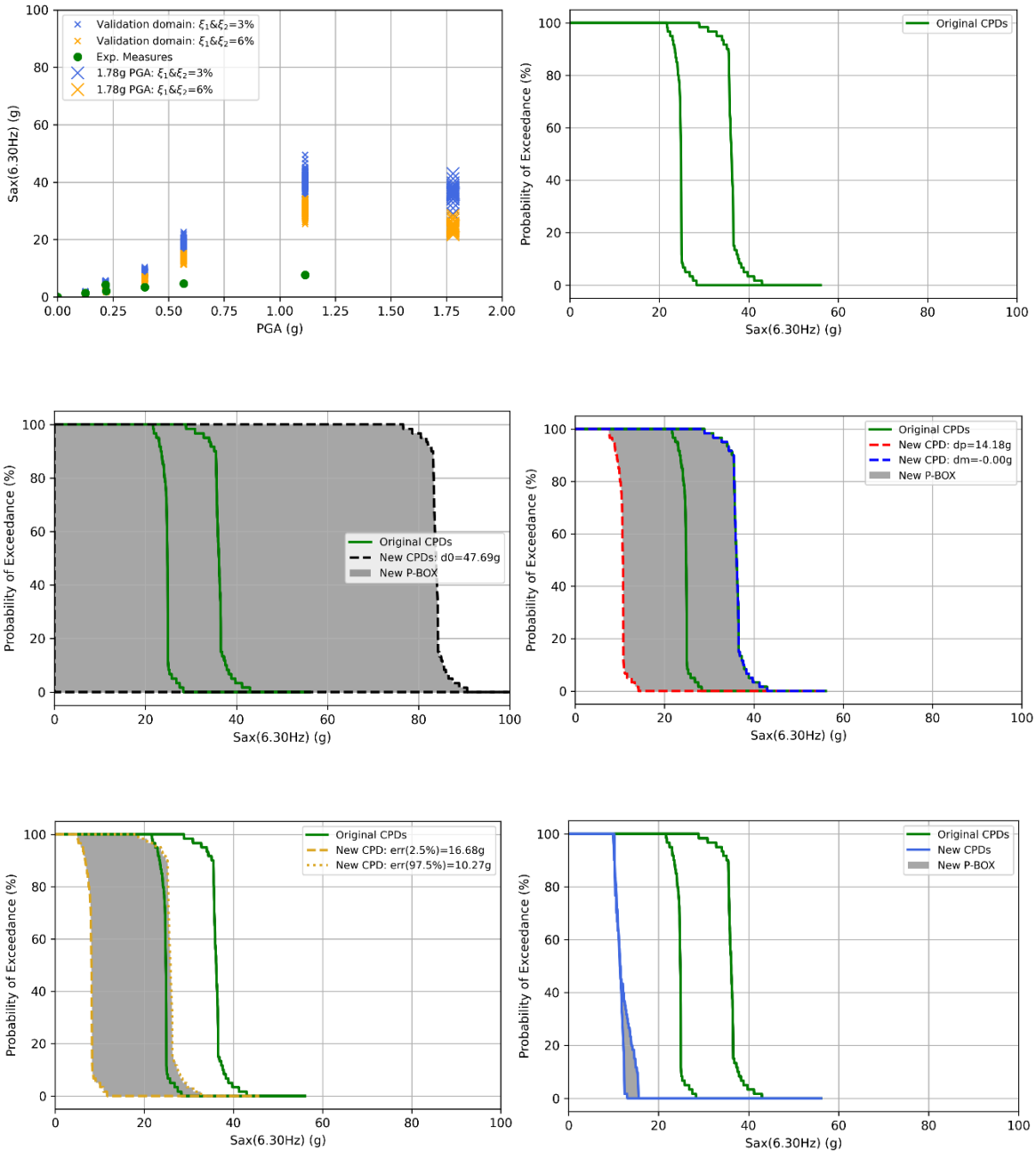
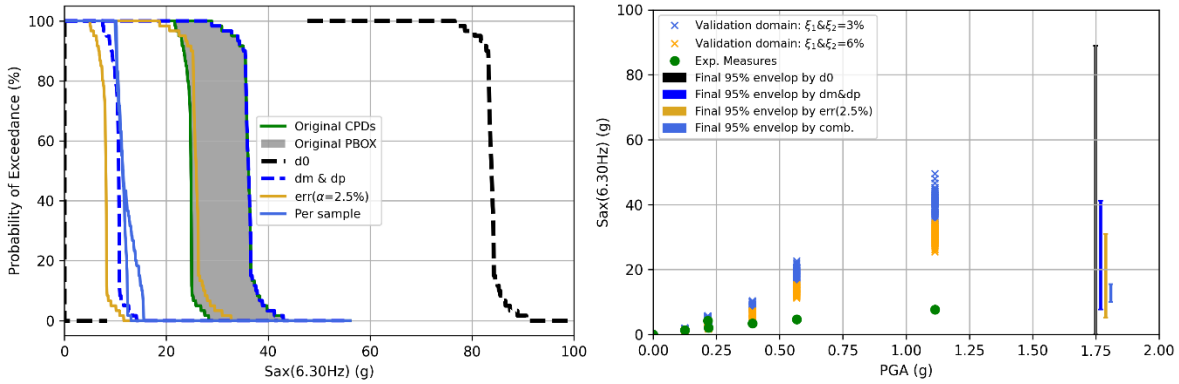


Figure 75: Model best/under/over estimations of Sax(30Hz) in the validation domain





7.2. Attempts to estimate frequency drops with Continuous Wavelet Transform

Figure 76 below is a "*fbsp*" continuous wavelet transform performed on the longitudinal acceleration at D point at the third floor of SMART2013 for a numerical simulation run with Run 15.

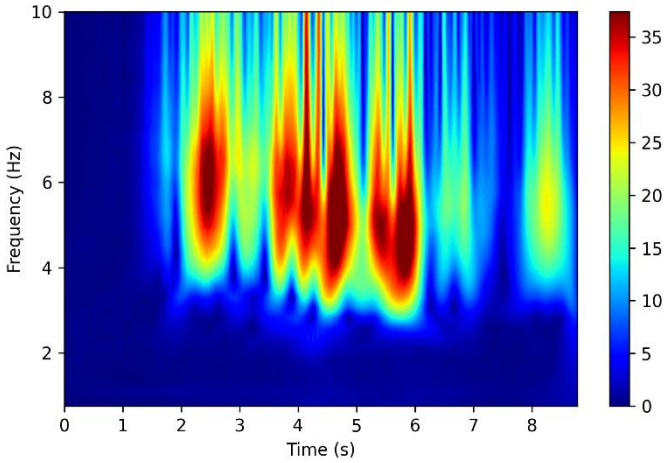


Figure 76: Continuous Wavelet Transform example: X axis

A decrease of frequency is visible from 2s to 6s. However, the spread of intensity over the frequency axis prevents a reliable reading of the exact evolution. The participation of several modes along the longitudinal axis may prevent the intensity peaks to narrow. Figure 77 below shows the maximum values of intensity of the same above signal over a 0.3s window translated from left to right.

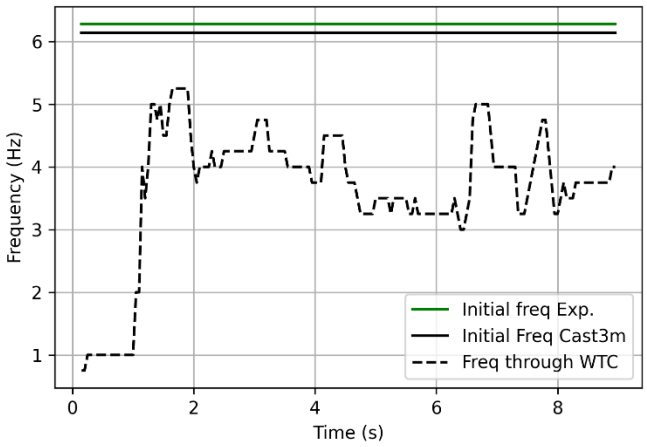


Figure 77: Maximum values of intensity along the Run 15 example simulation

One should notice that the initial eigenfrequencies is not reproduced here at the beginning, and that the variations of intensity are important towards the end of the run, ranging from 3Hz to 5Hz after the 6th second. Regarding the transversal axis and the study of the frequency drop of the 2nd mode, as shown by Figure 78 below, the reading is even more difficult because again blurred by the participation of other modes, notably the third torsion mode.

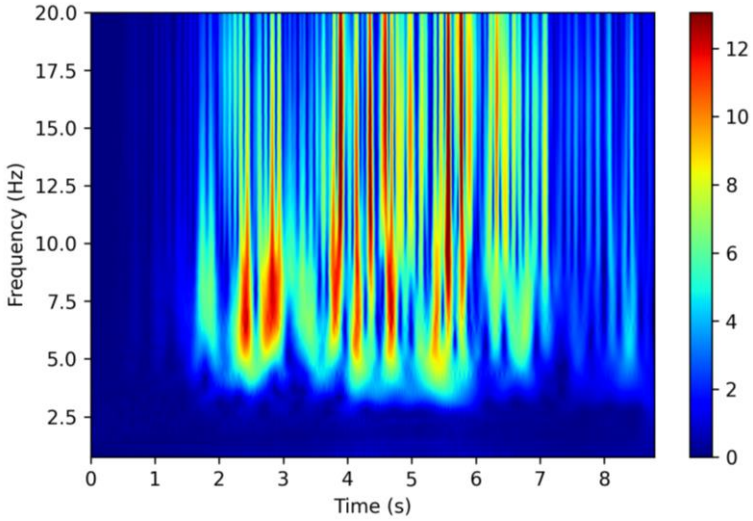


Figure 78: Continuous Wavelet Transform example: Y axis