PRINCIPLES OF VERIFICATION AND VALIDATION

Lesław Kwaśniewski [a], Cezary Bojanowski [b]

[a] Warsaw University of Technology, Faculty of Civil Engineering, Warsaw, Poland
[b] Transportation Research and Analysis Computing Center, Argonne National Laboratory, Argonne, IL, USA

Abstract

This paper discusses the concepts of verification and validation in computational mechanics with special attention to structural fire engineering, by referring to recently published papers and guides on V&V that define some best practices and show directions for future development. The perspective of an analyst, who develops computational models, makes runs, and analyses numerical results mostly using software based on the finite element method, is presented. The considerations emphasize practical problems encountered in the V&V process, potential sources of errors and uncertainties, the importance of sensitivity study, new ideas regarding the relationship between validation and verification, differences between calibration and validation, new aspects of the validation metrics, and guides for designing validation experiments. The discussion is illustrated by computational problem examples.

Keywords: benchmark, calibration, fire, sensitivity study, system response quantity, validation, verification

INTRODUCTION

Wide application of numerical models in structural engineering raises the question about their predictive capability. This question is especially legitimate in the research areas where complex, highly nonlinear structural behaviour is of interest. One of such research areas is the structural fire engineering where interaction of additional effects due to elevated temperatures has to be considered. Among such effects there are: thermal reduction of material properties, generation of additional forces due to constrained thermal deformation, complex thermo, chemical and mechanical effects such as dehydration and vapour pressure leading to premature concrete failure and spalling.

The high, steady interest in computational research for structural fire engineering can be observed based on the simplified statistics presented in Fig. 1 which shows number of related papers recorded in the Google Scholar database (with FIRE in the title and FIRE + “FINITE ELEMENT” anywhere in the article). There can be an expectation for more precise evaluation procedures, specifically dedicated to the considered research area such as structural fire engineering however, in the topic literature definitely dominates opinion that a general procedure applicable to what is called Computational Science and Engineering (CS&E) or Computational Engineering and Physics (CE&P) should be developed, (Oberkampf et al, 2004). The mentioned broad areas encompass many fields of engineering and physics, characterized usually by adjective “computational” such as (computational) fluid dynamics, solid mechanics, and structural dynamics. Even though, it is clear that the expected predictive capability for linear FE static analysis is different than for structural fire engineering, as it is shown schematically in Fig. 2, the same principles of V&V are applicable to all these research fields.

Report (Oden et al, 2006) describes the importance of computer simulation for the development of technical ideas today and predicts a sharp increase in the near future. We are witnessing the continuation of the computer revolution, which, according to Moore's law recognizes (Moore, 1998) a two-fold increase in computing power over the 18 months. In the

80's and 90's it was represented by doubling of the processor clock speed and now represented by an increase of number of transistors that can be packed in a standard chip size. The hardware development is followed by the rapid advancement of numerical programs. For example, based on the finite element method (FEM) commercial program LS-DYNA®, whose source code had 50,000 lines in its early days, in the 70s, now has more than 2.5 million lines in little more than a decade (Kwasniewski, 2009). The improvements in computational capabilities are well illustrated by an example presented in (Belytschko et al, 2000): in the 1970s, a 20 ms crash test simulation using a 300-element vehicle model took about 30 hours of computer time at a cost equivalent to the three-year salary of a university professor. Today's multiprocessor machines allow using a much higher number of finite elements - tens of millions in some FE models. Rapidly increasing number of users or of such programs, with the increasing access to multiprocessor computers with high-performance computing, degrades the computational resource limitations as an excuse for simplified computer simulations. The only limitation left for the use of multiple processors to solve a given problem is scalability of the software for a given problem.

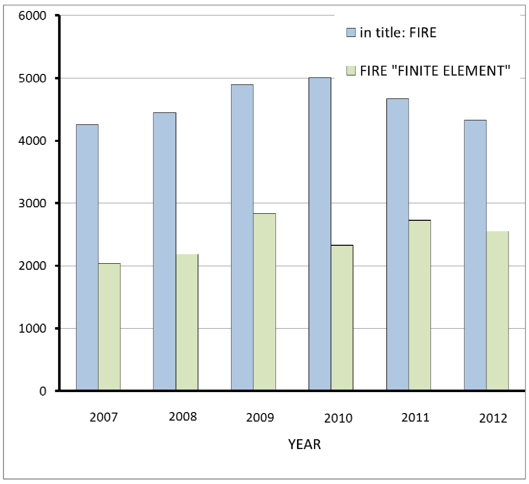

Fig. 1 Number of articles according to Google Scholar

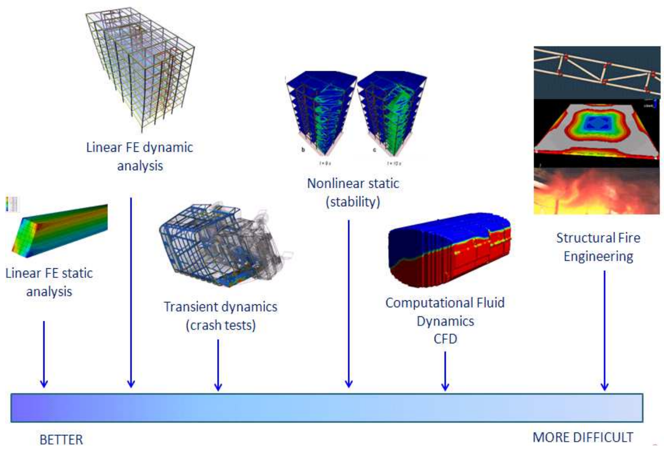

Fig. 2 Predictive capabilities of computer simulations

Despite the rapid hardware and software development there are many contradictory opinions about the reliability of computer predictions (Babuska & Oden, 2004), best expressed by a famous statement: "Essentially, all models are wrong, but some are useful" (Box & Draper, 1987). It is almost impossible to model all the aspects of a complex event, yet valuable conclusions from a series of simulations can be concluded if proper tools and statistical measures are used through the V&V procedures. Early in the development of the finite element method, the Journal of Applied Mechanics rejected FE papers for being insufficiently scientific (Belytschko et al, 2000). Today's general attitude is definitely evolving towards more acceptance of computer predictions, and the numerical results obtained using the dominant FE method are present in numerous technical and scientific papers from many different research areas.

1 MODELLING, VERIFICATION AND VALIDATION

Today verification and validation (V&V) is recognized as the primary method for evaluating the confidence of computer simulations (Oberkampf et al, 2004). The relationships between activities involved in the development of mathematical and computational models and in their verification and validation, are often schematically presented using diagrams such as the one shown in Fig. 3 (Kwasniewski, 2009). In Fig. 3, the boxes represent four main concepts: reality of interest, mathematical model, computer models, and validation experiments. Reality of interest relates itself to two aspects: to the physical system containing objects as well as to the processes intended for analysis. Reality of interest can apply to existing objects or to new solutions (prototyping) but always refers to somehow defined physical objects, for example to a structural element subjected to furnace test or to a whole structure subjected to full scale fire test.

The mathematical (or conceptual) model comprises all assumptions and definitions characterizing the mathematical representation of the reality of interest formulated generally as a system of partial differential equations (PDEs) complemented by boundary and initial conditions (Oberkampf et al, 2004). The transition from reality of interest to mathematical model depends on the objective of the analyses, understanding of physics, the analyst's experience, and resources. Formulation of the mathematical model is the first step in the model development and the first source of errors.

Usually, physical problems of a practical nature, represented by such mathematical models, cannot be solved analytically due to the complexity of, for example, their geometry. To find the solution, a mathematical model is replaced by an approximate computer (computational) model using the process of numerical discretization, which replaces PDEs with sets of algebraic (matrix) equations more suitable for computers. The discretization of space and time can be done using procedures such as the finite element, finite difference, finite volume, and boundary element methods. In solid mechanics and structural dynamics, space discretization is dominantly done with the finite element (FE) method. Time domain of transient events is discretized with finite difference method. In practice FE model development, especially when a commercial code is used, requires many decisions on selection among numerous options.

The last box in Fig. 3 contains a set of validation experiments designed using the validation hierarchy (Oberkampf et al, 2004). The objective of these tests is to increase the accuracy and predictive capability of computer models. Specially designed additional experiments are supposed to provide answers for the questions raised during model development and to quantify the model's uncertainties.

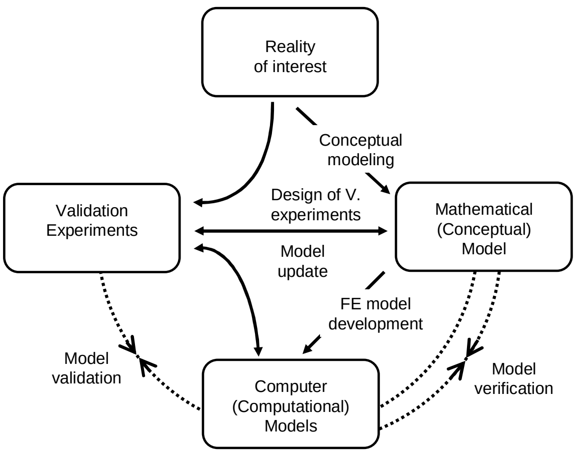

Fig. 3 Relations between modelling, verification and validation (Kwasniewski, 2009)

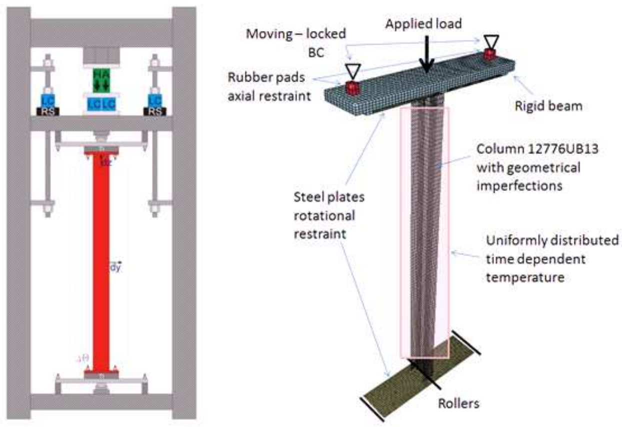

Fig. 4 Example parametric study of furnace test on rotationally restrained steel columns in fire

The solid lines indicate the activities of general model development, including conceptual modelling, computer (FE) model development (i.e., software implementation), and design of validation experiments. Some of these activities are marked with arrows on both sides to show their interactive character, and especially good cooperation between analysts and experimentalists is recommended (Kwasniewski, 2009). The assessment activities, marked with dashed lines, involve verification and validation. Verification and validation should be performed with respect to assumed objectives defining the quantities of interest. The difference between verification and validation is probably most accurately expressed by

Roache's informal statement: "Verification deals with mathematics; validation deals with physics" (Roache, 1998). Verification uses comparison of computational solutions with highly accurate (analytical or numerical) benchmark solutions and among themselves, whereas validation compares the numerical solution with the experimental data. Verification comprises of model and code verification stages. The analyst is usually provided with a software and the code verification stage is usually performed by for software developer. The objective of V&V practices, which is generally to corroborate (mathematical and computational) model for its intended use, can be practically split into three tasks: to detect and separate the model's significant discrepancies, to remove and reduce removable and unavoidable errors, and to evaluate uncertainties in the results. A very important aspect of the V&V process is the proper determination of sources for all significant errors. The dashed line representing validation connects experiments with both the computer and mathematical models. Although validation involves direct comparison of computational results and experimental data, the differences encountered have their sources in both models. It has been pointed out (ASME, 2006) that verification should precede validation, but even the most extensive verification cannot remove all errors (e.g., due to discretization) so validation evaluates the whole modelling process, and some of the errors that originated in different modelling phases cannot be completely separated, compare (Schwer, 2006).

2 DIFFICULTIES WITH EXPERIMENTAL VALIDATION IN STRUCTURAL FIRE ENGINEERING

Experimental validation in the structural fire engineering through comparison between numerical results and experimental data obtained using furnace tests is difficult and has many limitations which are not only economical but also are due to inevitable uncertainties characterising the specimen behaviour (Gillie, 2009). Practically, always limited number of measurements during such tests cannot provide entire information about the space and time distribution of temperatures, evolution of boundary conditions, or generation of additional forces due to constrained thermal and mechanical deformation. The limitations of experimental validation increase the importance of verification which is supposed to deliver evidence that mathematical models are properly implemented and that the numerical solution is correct with respect to the mathematical model.

The problems with experimental validation of computer simulations of structures subjected to fires can be illustrated using the following parametric study (Kwasniewski et al, 2013) where a furnace test (Ali F & O'Connor, 2001) on restrained steel columns was replicated using a coupled structural-thermal numerical calculations, see Fig. 4. The objective of the study was to identify and quantify all possible modelling parameters which can affect the numerical results. The study was focused on improving prediction capabilities for the purpose of virtual testing.

Common model calibration was replaced by experimental validation and extensive parametric study. The calibration is understood here as a posteriori procedure where through repeated calculations with modified input parameters we try to find an "optimal" set of input data which can provide the model's response closest to the actual experimental data. It can happen that due to superimposing of errors we can get good correlation between experimental and numerical results for a wrong model, defined by incorrect input parameters. Often, such a situation can be detected when the model is used for a different case with changed input conditions. Also, a complex model with only some of the input parameters "correctly" calibrated should give a response different from the experimental data due to the indeterminacy of other parameters. This is why validation based on more than one experiment is considered as more reliable (Oberkampf et al, 2004).

In the considered parametric study the comparison of the numerical results and the experimental data was presented for the relationships between column's average temperature and axial force, axial displacement, and lateral displacement in the middle section. Three critical modelling characteristics were determined: material behaviour, geometrical imperfections, and longitudinal variation of the column temperature. It was found that the postponed buckling occurring at higher furnace temperatures is due to a non-uniform temperature distribution along the column, caused by heat transfer at the partially insulated furnace openings. The study shows how the modelling factors, initially ignored, may affect the numerical results without calibrating the FE model. In the authors' opinion it is not possible to correlate better numerical results with the existing experimental data without reducing model uncertainties (e.g. imperfection magnitudes and loading variation) through additional experiments and measurements. It seems that due to many uncertainties characterizing the fire experiments, with often their wide variation, it is not justifiable to show the comparison between numerical and experimental results in a traditional deterministic way, where only two numbers or curves (i.e. experimental and numerical) are presented.

3 BENCHMARK PROBLEMS AND VERIFICATION

Verification is supposed to deliver evidence that mathematical models are properly implemented and that the numerical solution is correct with respect to the mathematical model. Due to the high complexity of mostly nonlinear problems that are practically important, such verification can be conducted only empirically using "a posteriori" approach where the reasoning is based on the experience coming from repeated calculations. A standard example is the posteriori error estimation based on numerical results for different mesh resolutions. According to (AIAA, 1998) verification can be conducted through tests of agreement between a computational solution and four types of benchmark solutions: analytical, highly accurate numerical solutions of an ODE or PDE problem, and manufactured solutions. In contrast to numerical solutions used in the validation stage, the numerical solutions applied for verification can represent mathematical models with little physical importance.

The importance and usefulness of benchmark studies for specific areas of CS&E such as structural fire engineering is postulated in many papers and conference proceedings. A benchmark example should satisfy the following requirements. The problem considered should be relatively simple, easy to understand. The considered case can show little of practical meaning. It is supposed to be used for verification of computational models not to solve an engineering problem. The complete input data must be provided in an easy to follow way. All assumptions regarding material properties, boundary conditions, temperature distribution, loading conditions, large/small deformations and displacements should be identified. If a numerical solution is considered as a benchmark problem the mesh density study should also be considered and it should be shown that provided results are within the range of asymptotic convergence. One should also consider as a part of verification to use alternative numerical models e.g. different codes or solid vs. shell finite elements (if possible). Publishing a benchmark study we claim that this is a reliable solution. Hopefully, this assumption will be verified by other users. Benchmark problems can serve for code developers but are probably the most helpful for code users who can verify their modelling assumptions, as most of the errors are due to the analyst's mistakes.

4 COMPARISON BETWEEN EXPERIMENTAL AND NUMERICAL RESULTS

The soundness of an experiment as a source of data for validation depends also on the relationship between the application and the validation domains (Oberkampf et al, 2004). The application domain defines the intended boundaries for the use and predictive capability of the computational model. The validation domain characterizes the representation capabilities of the experiment. When a complex system is modelled, there is a need for many validation experiments capturing different physical aspects of the system (e.g., different loading scenarios, boundary and initial conditions) on different level of complexity of the model. Unfortunately, due to high cost of furnace tests, the experiments are rarely repeated and the probability distribution of the test results is undefined. This distribution can be dramatically different, depending on the selection of the so called system response quantity (SRQ). Some of the researchers acknowledge large discrepancies between the experiment and the computation especially for concrete structural elements subject to elevated temperatures when the important role of moisture transport on the spalling mechanism is not sufficiently captured in the computational model (Heijden & Bijnen, 2007). The need for multiple experiments and computational probabilistic analysis can be best described by Fig. 5 presenting the difference that can be measured between a single experiment and a single simulation and the actual means of a given measure. Sensitivity of this measure to a given parameter is represented by the shape (width) of the distribution function. When comparing simulated results to just one experimental result the analyst has no confidence about representativeness of the experiment result. In the process of calibrating the computational model to just one experiment actually more errors can be introduced in the model and its predictive capability can be negatively affected for a different set of initial parameters

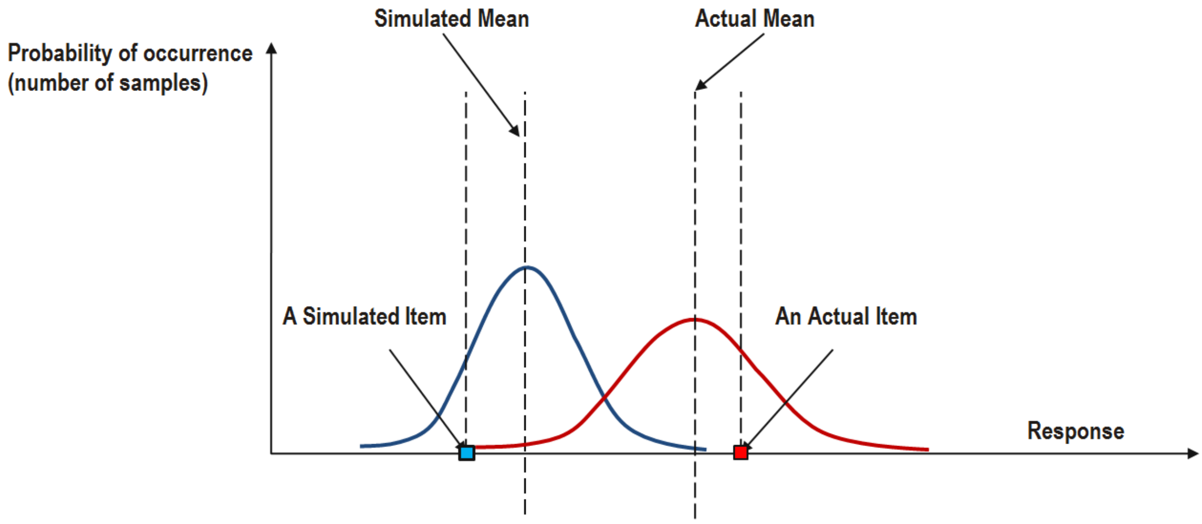

Fig. 5 Difference between simulated and experimental values for a single and item and a population of results

4.1 Validation domains

The ideal situation, possible only for simple systems, is when the validation domain completely overlaps the application domain. This means that the available set of the validation experiments covers all possible parameters defining the computational model within its intended application. When complex systems are analysed, it is sometimes infeasible or even impossible to conduct all necessary experiments to verify all features of the computational model. An example of such a situation is the global analysis of structures in the fire (Foster, 2007). There have been only a few full-scale experimental fire tests (i.e., the Cardington tests) conducted so far, but there are numerical capabilities for such complex analysis. The extreme, theoretical situation is when all possible or available experiments are too far from the application of interest and there is no overlap between the validation domain and the application domain. The credibility of such a computational model, validated only through extrapolation, is obviously much smaller. To improve the predictive capability of computation in such cases, hierarchical validation is introduced where closer correlation of the domains is possible for lower-level experiments and then the gained confidence is extrapolated to the global model.

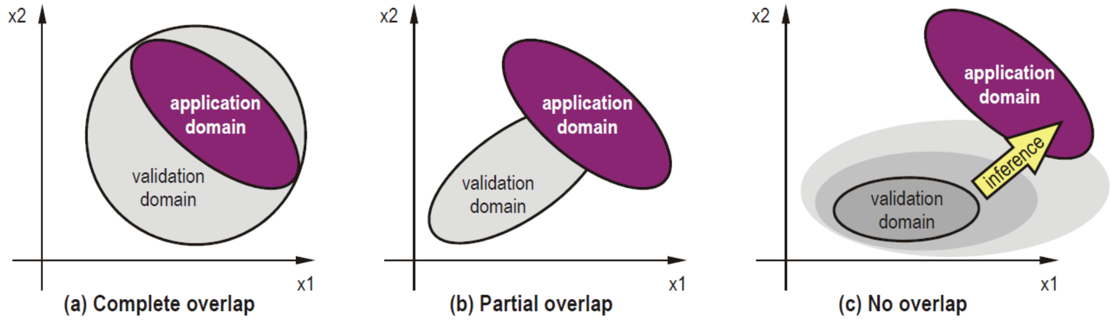

Fig. 6 Possible relations between validation and application domains (Thacker et al, 2004)

4.2 Validation metrics

Another important issue affecting the outcome of the comparison between the experimental and numerical results is which parameter (SRQ) we select for the comparison and how it is represented, deterministically or in a probabilistic manner. In (Oberkampf et al, 2004), the authors distinguish six levels of validation comparisons, see Fig. 6. In the first level approach, the simplest and the most common in today's practice, a strictly qualitative comparison is done using plots over the domain, for example, showing the deformation of a structure. The second level represents a more quantitative but still fully deterministic comparison of the numerical and corresponding experimental, single value input-response pairs, using tables or plots. The second and third levels are most common in papers and reports dealing with computational analyses. In the next, higher levels of comparison, the nondeterministic nature of experimental data with both errors and uncertainties is taken into account. Instead of single values for the input and the corresponding result, there are value bars with the centre point representing the mean value and the length equal to two standard deviations (Oberkampf et al, 2004). The value bars provide information on the probability distribution estimated based on the multiple experiments and can be applied to both uncertainties in the input and in the results.

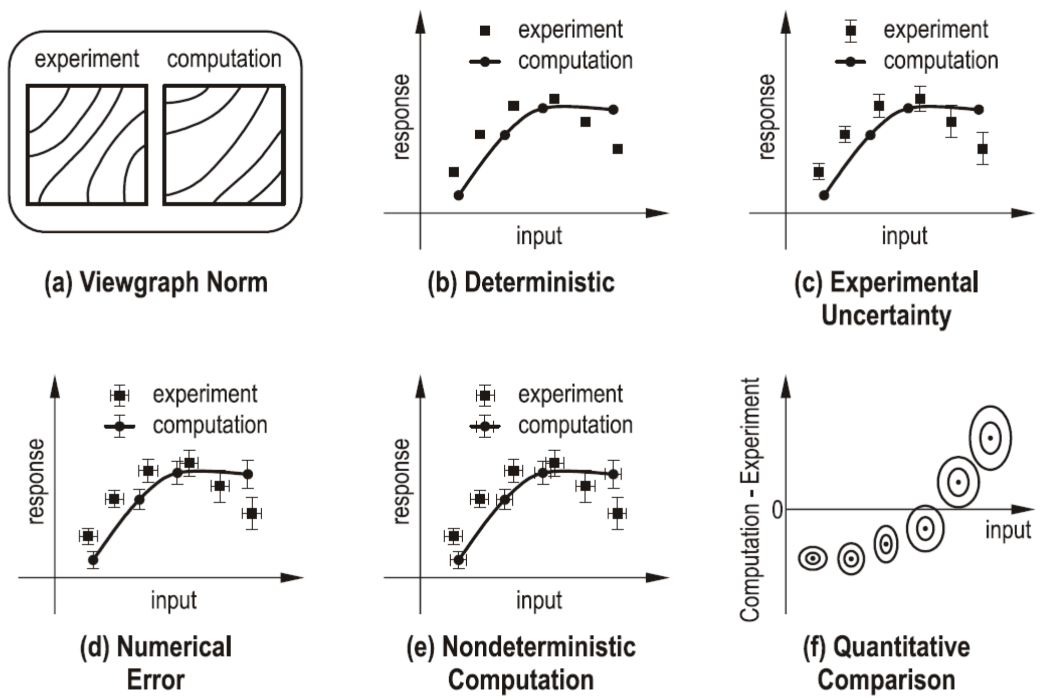

Fig. 7 Quality levels of validation metrics (Oberkampf et al, 2004)

The highest levels are represented by comparison methods where additionally the computational results are treated as nondeterministic, with their own input and output uncertainties. The necessary information is provided by repeated computations for the input variation determined by the experimentally estimated probability distributions (Szabó, 2008). The ideal approach according to (Oberkampf et al, 2004) would show the difference between the computational and experimental probability distributions over the whole possible range of the input quantity, but that would require an enormous effort for any real life application. It should be remembered that for a nonlinear system the relationship between the input and the output can be very complex, and for example, the mean values of the response do not have to be equal to the response for the mean values of the input parameters, compare Fig. 5.

4.3 System response quantity

As already mentioned, the validation procedure is based on the comparison between computational results and experimental data. Generally, an experiment can provide much less information than the calculation. The measurements for a quasi-static experiment on an engineering structure usually give us loadings, displacements, and strains. A dynamic experiment provides time histories of loads, strains, displacements, and accelerations. In thermal analysis, the spatial distribution of temperatures is measured. The measurements are done for a limited number of selected locations. Not all the experimental output data is equally representative and has the same importance for comparison with the computation. For example, strain in a uniform beam subject to bending is a local quantity related to internal forces in the considered cross-section. However, the maximum deflection of the same beam is a more representative quantity as it is the result of all deformations along the beam and depends on the whole distribution of internal forces and on boundary conditions. The correlation between the experimental and computational displacements for such a case is more important for the purpose of validation than comparison of local strains. The comparison of stresses instead of strains is more common in engineering practice but requires recalculation of experimentally directly measured strains using material properties affected by their uncertainties. In structural dynamics, we get a better correlation between smoother time histories of displacements than between their second time derivatives—time histories of accelerations that are rougher. The selection of the system response quantity (SRQ) is often limited by the experiment output, e.g., for the earthquake or crash analyses time histories of accelerations are the basic output information.

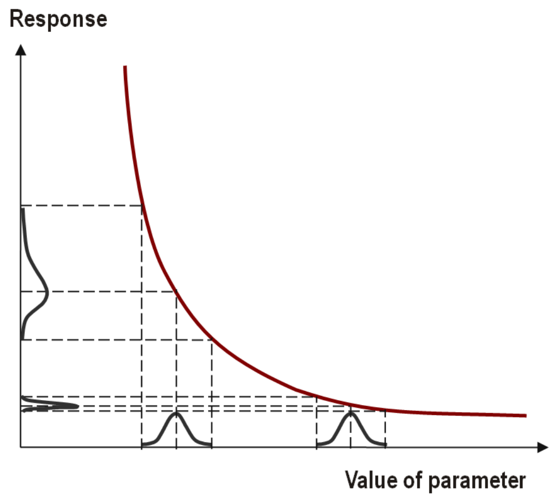

Fig. 8 Different sensitivity of a response on a parameter variation

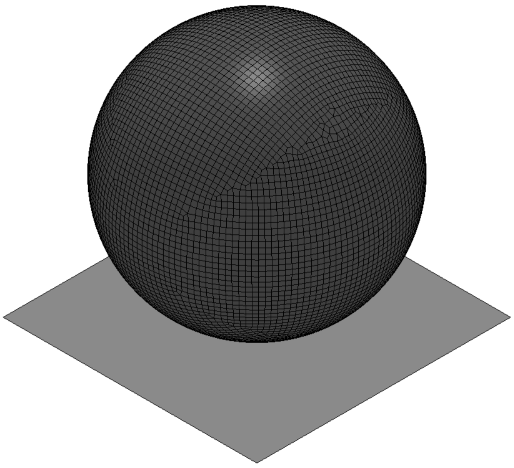

Fig. 9 Model of a hollow glass ball subject to gravity acceleration

Also the range in which we test given response has a great impact on accuracy of the results. If in a given range the response is not sensitive to variation of the input parameter the accuracy of prediction can be high (See Fig. 8). In some other range the response may be highly sensitive to the variation of the input parameter and the accuracy of the prediction may be significantly lower.

Let’s introduce the general idea of predictive capabilities of numerical calculations (i.e. computer simulations) through a simplified example problem of structural mechanics. In this problem we consider a test where a hollow glass ball with external radius of 25mm and the wall 1mm thick is falling under gravity from a prescribed height (2.0 m) and hits a rigid surface. The schematic of the test is shown in Fig. 9. The question is how precisely we can predict the considered process using available nowadays software. From an experimental point of view this test can be performed multiple times and probabilistic values for input parameters characterizing glass as well as response can be measured. Geometrical imperfections on macroscopic level can be also measured for the ball. Although experiments are controlled by many parameters we don’t have to measure all of them to be able to perform the tests. For example, not knowing failure parameters in the glass we can drop a hundred balls and measure a radius occupied by all the shattered pieces of the ball. To perform an equivalent simulation many more parameters has to be measured or provided to the analyst. Performed here simulation in LS-DYNA software required detailed material properties of the glass including failure and erosion criteria to allow for material separation. Such process, although non-physical, is often used in simulations that pertain to material separation. If any of the SRQ is related to the failure it may be predicted with large error or uncertainty. Mesh size and mesh pattern are directly influencing the patterns of cracks that can develop only through the eroded elements. On the other hand for example the maximum force that is exerted by the ball on the ground is less affected by the mesh pattern and primarily depends on the mass and the drop height of the ball.

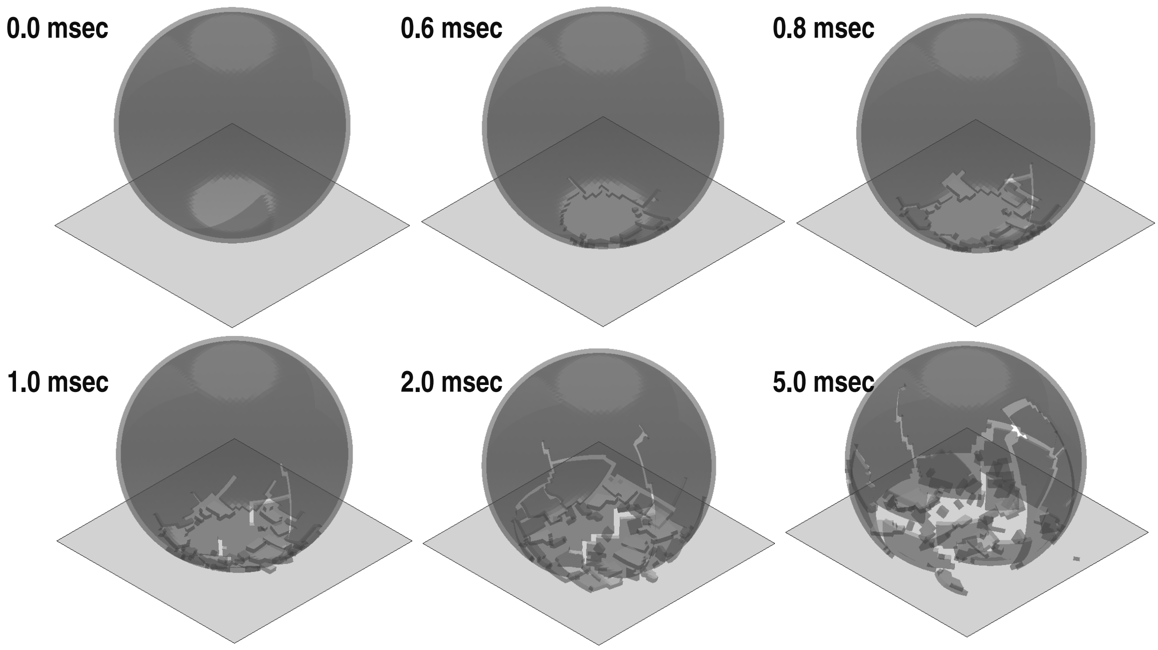

Fig 10. Evolution of glass ball failure upon impact on rigid surface

5 SUMMARY

Recently in the literature, some guidelines for improving validation procedures have been formulated. Validation hierarchy, placed in opposition to the model calibration common today, and new validation metrics are examples of such improvement concepts. The examples presented in this paper show how much the result of validation can be affected by the selection of the system response quantity and that there is no universal metrics. The importance of the comparisons using view graphs, often considered in the literature as lower-level practice, is also emphasized as an efficient method for checking the physical validity of mathematical models. What should be recommended, especially for complex problems with practical meaning, is the design of simple experimental tests placed on different levels of hierarchical validation. Such simple and less expensive tests can provide more valuable material for comparison than costly experiments on entire structure.

REFERENCES

AIAA. Guide for the Verification and Validation of Computational Fluid Dynamics Simulations, American Institute of Aeronautics and Astronautics, AIAA-G-077-1998, Reston, VA, 1998.

Ali F., O'Connor D. 2001. Structural performance of rotationally restrained steel columns in fire, Fire Safety Journal 36.

ASME Guide for Verification and Validation in Computational Solid Mechanics, The American Society of Mechanical Engineers, ISBN #: 079183042X, 2006.

Babuska I., Oden J. T., Verification and validation in computational engineering and science: basic concepts, Computer Methods in Applied Mechanics and Engineering, Elsevier, 193 (36-38) 2004, 4057-4066.

Belytschko T., Liu W. K., Moran B., Nonlinear finite elements for continua and structures, Wiley New York, 2000.

Box G. E. P., Draper N. R., Empirical model-building and response surfaces, John Wiley & Sons., pp. 669, 1987.

Foster S., Chladna M., Hsieh C., Burgess I., Plank R., Thermal and structural behaviour of a full-scale composite building subject to a severe compartment fire, Fire Safety Journal, 42- 3, 183-199, 2007.

Gillie M., Author's Analysis of heated structures: Nature and modelling benchmarks, Fire Safety Journal 44, pp. 673–680, 2009.

Kwasniewski L., Ali F., Balcerzak M., Coupled structural-thermal calculations for restrained steel columns in fire, Journal of Structural Fire Engineering vol. 4, no. 1, March 2013, pp. 59-70, 2013.

Kwasniewski L., On practical problems with verification and validation of computational models, Archives of Civil Engineering, 55 (3), 323-346, 2009.

Moore G.E., Cramming more components onto integrated circuits, Proceedings of the IEEE, Institute of Electrical and Electronics Engineers, 86 (1), 1998, 82-85.

Oberkampf W.L., Trucano T.G., Hirsch C., Verification, validation, and predictive capability in computational engineering and physics, Appl. Mech. Rev. 57 (5), 345–384, 2004.

Oden J. T., Belytschko T., Fish J., Hughes T. J., Johnson C., Keyes D., Laub A., Petzold L., Srolovitz D., Yip S., Revolutionizing engineering science through simulation, National Science Foundation Blue Ribbon Panel Report, (65), 2006.

Roache P.J., Verification and Validation in Computational Science and Engineering, Hermosa Publishers, Albuquerque, NM 1998.

Schwer L.E., An Overview of the ASME Guide for Verification and Validation in Computational Solid Mechanics. LS-DYNA Anwenderforum, Ulm 2006.

Szabó B.A., Actis R.L., On the role of hierarchic spaces and models in verification and validation, Comput. Methods Appl. Mech. Engrg. doi:10.1016/j.cma.2008.04.025, 2008.

Thacker B. H., Hemez F. M., Anderson M. C., Pepin J. E., Rodriguez E. A., Concepts of Model Verification and Validation., Los Alamos National Laboratory, 2004.

Van der Heijden G. H. A., Van Bijnen R. M. W., Pel L., Huinink H. P., Moisture transport in heated concrete, as studied by NMR, and its consequences for fire spalling, Cement and Concrete Research, Volume 37, Issue 6, 894-901, June 2007.