Title:
Simulation Data Management for Design of Experiments: Concepts and Specifications

Authors:
Gaëtan Blondet, gaetan.blondet@utc.fr, Université de Technologie de Compiègne
Farouk Belkadi, farouk.belkadi@irccyn.ec-nantes.fr, LUNAM Université, IRCCYN - ECN
Julien Le Duigou, julien.le-duigou@utc.fr, Université de Technologie de Compiègne
Alain Bernard, alain.bernard@irrcyn.ec-nantes.fr, LUNAM Université, IRCCYN - ECN
Nassim Boudauud, nassim.boudaoud@utc.fr, Université de Technologie de Compiègne

Keywords:
Adaptive Design of Experiment, Simulation Data Management, Data models, Meta-modelling

DOI: 10.14733/cadconfP.2015.254-258

Introduction:

In a product development process, numerical simulation is a very important step to validate design decisions and to assess product performance along each step of its lifecycle [2]. Designs of Experiment (DoE) are more and more used during the simulation process. To shorten the simulation process, two approaches were identified: (1) the DoE process execution duration reduction and (2) the DoE preparation process shortening by managing and reusing simulation data. The Simulation Data Management for Design of Experiment (SDM4DOE) project aims to develop these approaches to produce an open-source simulation platform dedicated to DoE applications. This platform ought to set and run DoE fast, for complex numerical models, while ensuring traceability of generated data and providing a decision-aid system for DoE preparation. This paper introduces concepts and specification to manage DoE process data. The simulation process and the DoE setting up are firstly described. Then, the selection of methods issue is addressed. The DoE data management system developed to solve this issue is detailed in the third section. To conclude, future works in the SDM4DOE project context are presented.

DoE Process:

The DoE process is based on the simulation process, which consists of three main steps [7]. A parameterized numerical model of the studied system is created. Corresponding outputs are obtained by a specific solver. These outputs are analysed during the post-processing step, for model checking and product validation. All of these steps may generate a large amount of heterogeneous data and could be very expensive and time-consuming. A DoE is a set of experiments defined to assess the numerical model for different configurations of the product. A DoE is defined by its type (distribution of experiments in the design-space), the number and type of its factors (model parameters) and associated levels. A DoE process may have different objectives (exploration, product optimization, sensitivity analysis, etc.) and submitted to different constraints (e.g. computational budget). The computational cost of a DoE is the cost of the numerical model calculation multiplied by the number of experiments. Thus, an optimal strategy is to choose the most efficient DoE and to use a method for reducing the computational cost of each run. An efficient DoE should minimise the number of runs and optimize the space-covering of the runs, according to the objective (exploration, product optimisation…). After the definition and the validation of the numerical model [16], the available factors must be analysed and selected to reduce the DoE cost, by a sensitivity analysis. Only most influent factors are kept, according to a specific output. To analyse accurately these influences, a metamodel (or surrogate model) is created from DoE results and statistical methods, as ANOVA. The metamodel can be reused to replace the numerical model for other studies.

Methods Selection:

Designs of Experiment
A large amount of different types of DoE are available [9] for numerical simulation. The DoE type selection depends on its properties (space-filling, uniformity, etc.), objectives (exploration, optimization, etc.) and constraints (computational budget, output linearity properties, etc.). DoE type selection can be a long and difficult operation. Although some classifications exist [22], they are not exhaustive. There remains a need for classifying DoE. It is important, for this purpose, to identify and capitalize all related data (DoE type, numerical model, objectives and constraints, etc.).

Metamodels
Metamodels are used to replace a costly numerical model by a function (e.g. a polynomial function) faster to be assessed, for a specific output. Metamodeling includes three steps: (1) selection of the type of metamodel, (2) the training step from a DoE and (3) the validation. Types of metamodels are numerous [5], [13]. Each type has specific properties (accuracy, non-linear output approximation, cost, etc.). The training step begins with the execution of a DoE to obtain a set of responses from the initial numerical model. The metamodel internal parameters (e.g. maximal degree a polynomial function) shall be set to fit as closely as possible these results [23]. The validation step is made by statistical methods as cross-validation methods [13]. The time needed to select and tune each method (DoE, metamodel and internal parameters, validation method, etc.) involved can be longer than the time saved for execution [21]. Recommendations and benchmarks exist for the selection of some DoE-metamodel couples [9], [14], [23]. However, these classifications are non-exhaustive. Solutions have been proposed to automatically select the appropriate metamodel [10], but they can increase the computational cost. The data model developed in the following sections will capitalize metamodelling data and the results for each instance of the DoE process. This may lead to reuse data and help the designer to select and tune DoE and metamodel.

Adaptive DoE
Adaptive DoE are defined to produce the best DoE according to the studied numerical model and the objective of the study. From an initial DoE, only the most efficient experiments are iteratively added. Efficiency is defined by an infill criterion, which defines the potential gain associated to an experiment [12-13]. This process is repeated until a convergence criterion or a maximum number of experiments performed on the original numerical model. The two main types of data involved by adaptive DoE are (1) the infill criterion and (2) the optimization algorithm (evolutionary algorithms, local methods [4-19]) used to find the experiment maximizing the infill criterion. The selection and tuning of these two elements depend on the metamodel used, the objective and constraints of the study [3], [20]. Adaptive DoE can lead to non-negligible time-saving, but it is a complex methods involving long time to be defined. A non-exhaustive infill criteria taxonomy [18] and methods for selection automation [4], which can increase the computational cost, exist. The classification and capitalization of simulation data is still a serious development axis to provide decision support to the designer [6]. This will reduce the preparation time and avoid increasing the computation time. In the context of SDM4DOE project, developments planned to shorten the DoE process (preparation and runs) cover: (1) adaptive metamodelling process, (2) statistical methods for relevant factors selection, (3) data management and capitalization for a decision-aid framework. The first two axis will be implemented in the platform URANIE [15]. The third axis will be developed by the definition of a data model dedicated to DoE process.

DoE Data Management:
Simulation Data Management (SDM) [8] is a part of a larger issue of Product Data Management (PDM) [1]. PDM provides methods and tools to support the structuring, storage in shared repositories, management and sharing of product-related data and processes for its processing. More recently, SDM issues are associated with the new Product Lifecycle Management approaches (PLM) [1]. This covers technical data management systems of the product (PDM), the design support tools (CAD/CAE), manufacturing support tools (CAM) and other ERP applications [11]. Numerical DoE is a concrete

example of the use cases of SDM tools. The realization of a DoE is based on the combined use of a set of design tools, simulation, computation, statistical processing and control of the computing process.

DoE process is a collaborative context in which several experts from different fields with different roles are involved. The effect of DoE results on the global product development process implies many interactions between DoE Process and upstream and downstream activities. Informatics systems and support activities are also concerned.

Moreover, as many choices, tests and decisions are made during the DoE process, several iterations are needed, which are made on several working sessions. Stakeholders working on two different sessions could be changed because of their availability, the objectives and the DoE run. Thus, a DoE run involves a large amount of heterogeneous data. Some of these data needs to be standardized to ensure a good communication between the stakeholders and different tools used.

DoE Data Model

In the context of the SDM4DOE project, a first study was realized to map and classify data involved in a DoE process. The global organisation of main data types are shown in Fig. 3, represented in packages by an UML diagram.

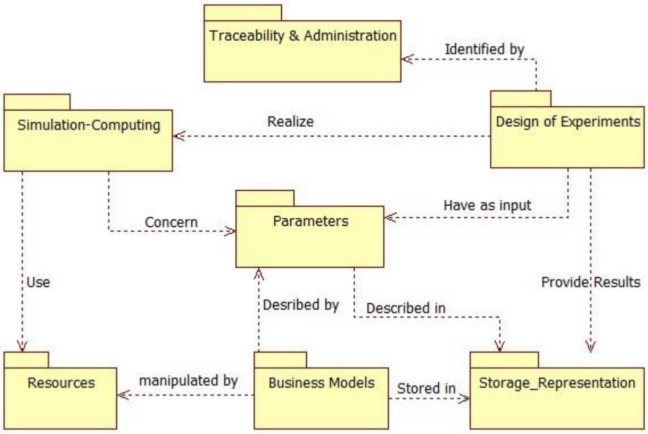

Fig. 3: DoE Data classification.

- The package "Design of Experiments": this first category of data aims to describe the main properties of DoE: objectives, the type, the nature of the studied phenomenon, etc.
- The package "Traceability and Administration" is used to link the DoE folder with its administrative and operational environment. This connection is done by several concepts such as: the reference of the project, the product and / or component associated with the DoE, the stakeholders involved in the working sessions, the timing of these working sessions with related decisions, etc.
- The "Parameters" package is the central node of the DoE data model. Through this concept the different types of parameters (factors) are classified according to their nature, their input/output status in the different steps of the DoE process and the possible intervals of variation of their value. The concept of "parameter" is used to define other specific properties but also to define a standard for a common semantic for codification of parameter names between enterprise business platforms and SDM platform.
- The package "Business Models" is used to manage in a uniform way all types and versions of business models involved in a DoE process to facilitate their sharing. It also links parameters to business models. It mainly involves CAD models, FEM and metamodel data.
- The package "Storage-Representation" provides important information to SDM for data location identification for extraction and exploitation. Because of their heterogeneity and diversity of their sources, data are stored in different places and in different formats (databases, header files, etc.) according to their nature.

- The package "Simulation-Computing" lists all types of simulation scripts and possible methods of analysis in a DoE process context and according to a given study. It can also link these different types of processing with concerned parameters.
- The package «Resources» completes the definition of processing by proposing a classification of all softwares and methodological resources supporting computations. This will allow SDM users to quickly define the various steps of the DoE. The concept of "computational cost" allows a better characterization of the various alternatives of computations to facilitate the selection of the DoE execution mode.

Principle of the SDM framework

To implement a DoE data management system, open-source software architecture is being developed in the SDM4DOE project. This architecture is developed as a SaaS web service, based on several open-source solutions. In a further development, the platform will be composed generic components to integrate other simulation softwares, including commercial softwares.

The SDM platform related to DoE will provide users a set of features for DoE processing and management of administrative and technical data during a work session, via a web-based interface. It also allows you to view and navigate the previous PEN traceability data. To achieve processing features, the URANIE open-source software integration is planned.

The fulfilment of a DoE is based on the link between the SDM system and the company's business platform, in which the computation chain is implemented, supervised by post-processing software Salomé. These services include the simulation software Code_Aster, in which acceleration methods of computations will be implemented. Finally, in order to support the data handled by the DoE, the various models described in the previous section will be implemented, according to the study and the nature of the data in a database or in the headers of documents associated with different business files.

Conclusion:

In this paper, a first inventory of issues related to numerical design of experiments was made to shorten the process. The first results of the SDM4DOE project show that existing SDM tools are not adapted to the heterogeneous nature of the DoE process data. Indeed, the optimization of a DoE process is based on the effective management of technical data, administrative data and also data related to the traceability of decisions. At this moment, a set of concepts has been proposed and is being refined and validate. The next step consists in integrating these concepts to develop a first version of the SDM platform. On the other hand, the mastery of the DoE complexity will be based on a standard codification proposal, on both business files and data types, to ensure the integration of SDM4DOE solution for all types of business computing platforms. This work will be subjected to a validation step by using this solution on complex systems.

Acknowledgement:

This work is funded by the French FUI project SDM4DOE and labelised by French competitiveness clusters SYSTEMATIC and ID4CAR. We also thank all consortium partners for their contribution during the development of ideas and concepts proposed in this paper. http://www.id4car.org/SDM4DOE.html

References:

[1] Assouroko, I.; Ducellier, G.; Belkadi, F.; Eynard, B.; Boutinaud, P.: Improvement of engineering design and numerical simulation data exchange based on requirements deployment: a conceptual framework, 7th International Conference on Product Lifecycle Management, Bremen, Germany, 2010, 487–97.

[2] Bernard, A.; Chenouard, R.: Multi-physics simulation for product-service performance assessment, 6th CIRP Conference on Industrial Product-Service Systems, 16, Windsor, Ontario, Canada, 2014, 21–5. http://dx.doi.org/10.1016/j.procir.2014.03.002.

[3] Bingham, D.; Ranjan, P.; Welch, W.J.: Design of Computer Experiments for Optimization, Estimation of Function Contours, and Related Objectives, Statistics in Action: A Canadian Outlook, 2014, 109.

[4] Boussaïd, I.; Lepagnot, J.; Siarry, P.: A survey on optimization metaheuristics, Information Sciences, 237, 2013, 82–117. http://dx.doi.org/10.1016/j.ins.2013.02.041.

[5] Castric, S.; Denis-Vidal, L.; Cherfi, Z.; Blanchard, G.J.; Boudaoud, N.: Modeling Pollutant Emissions of Diesel Engine based on Kriging Models: a Comparison between Geostatistic and Gaussian Process Approach, 14th IFAC Symposium on Information Control Problems in Manufacturing, INCOM'12, vol. 14, Bucharest: 2012, 1708–15. http://dx.doi.org/10.3182/20120523-3-RO-2023.00038.

[6] Chandrasegaran, S.K.; Ramani, K.; Sriram, R.D.; Horváth, I.; Bernard, A.; Harik, R.F.; Gao, W.: The evolution, challenges, and future of knowledge representation in product design systems, Computer-Aided Design, 45(2), 2013, 204–28. http://dx.doi.org/10.1016/j.cad.2012.08.006.

[7] Charles, S.: Gestion intégrée de données CAO et EF. Contribution à la liaison entre conception mécanique et calcul de structures, Ph.D. Thesis, Université de Technologie de Troyes, 2005.

[8] Charles, S.; Ducellier, G.; Eynard, B.: CAD and FEA integration in a simulation Data Management environment based on a knowledge-based system, TMCE, Ljubljana: 2006, 1719–30.

[9] Chen, V.C.P.; Tsui, K.-L.; Barton, R.R.; Meckesheimer, M.: A review on design, modeling and applications of computer experiments, IIE Transactions, 38(4), 2006, 273–91. http://dx.doi.org/10.1080/07408170500232495.

[10] Couckuyt, I.; De Turck, F.; Dhaene, T.; Gorissen, D.: Automatic surrogate model type selection during the optimization of expensive black-box problems, Winter Simulation Conference, Phoenix, Arizona, 2011, 4274–84. http://dx.doi.org/10.1109/WSC.2011.6148114

[11] Dhuieb, M.A.; Belkadi, F.; Laroche, F.; Bernard, A.: Interoperability Framework for Supporting Information-Based Assistance in the Factory, IFIP 11th International Conference on Product Lifecycle Management PLM'2014, Japan, 2014.

[12] Dubourg, V.: Méta-modèles adaptatifs pour l'analyse de fiabilité et l' optimisation sous contrainte fiabiliste, Ph.D. Thesis, Université Blaise Pascal-Clermont II, 2011.

[13] Forrester, A.I.J.; Keane, A.J.: Recent advances in surrogate-based optimization, Progress in Aerospace Sciences, 45, 2009, 50–79. http://dx.doi.org/10.1016/j.paerosci.2008.11.001.

[14] Franco, J.: Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes, Ph.D. Thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne (ENSM-SE), 2008.

[15] Gaudier, F.: URANIE: The CEA/DEN Uncertainty and Sensitivity platform, Sixth International Conference on Sensitivity Analysis of Model Output, 2, 2010, 7660–1. http://dx.doi.org/10.1016/j.sbspro.2010.05.166.

[16] Giles Jr., H.F.; R. Wagner Jr., J.; Eldridge M., M.I.: Design of Experiments, Extrusion: The Definitive Processing Guide and Handbook Plastics Design Library, 2014, 291–308. http://dx.doi.org/10.1016/B978-1-4377-3481-2.00025-9.

[17] Lghali, A.: Surrogate Based Optimization Using Kriging Based Approximation, 2012.

[18] Lim, D.; Jin, Y.; Ong, Y.; Sendhoff, B.: Generalizing Surrogate-Assisted Evolutionary Computation, IEEE Transactions on Evolutionnary Computation, 14(3), 2010, 329–55. http://dx.doi.org/10.1109/TEVC.2009.2027359

[19] Picheny, V.; Ginsbourger, D.; Roustant, O.; Haftka, R.T.; Kim, N.-H.: Adaptive Designs of Experiments for Accurate Approximation of a Target Region, Journal of Mechanical Design, 132(7), 2010, 81–9. http://dx.doi.org/10.1115/1.4001873.

[20] Razavi, S.; Tolson, B. a.; Burn, D.H.: Numerical assessment of metamodelling strategies in computationally intensive optimization, Environmental Modelling & Software, 34, 2012, 67–86. http://dx.doi.org/10.1016/j.envsoft.2011.09.010.

[21] Sanchez, S.M.; Wan, H.: Work smarter, not harder: a tutorial on designing and conducting simulation experiments, Winter Simulation Conference, Berlin, Germany, IEEE, 2012, 1929–43. http://dx.doi.org/10.1109/WSC.2012.6465307

[22] Simpson, T.W.; Peplinski, J.D.; Koch, P.N.; Allen, J.K.: Metamodels for computer-based engineering design: survey and recommendations, Engineering with Computers, 17(2), 2001, 129–50. http://dx.doi.org/10.1007/PL00007198