Simulation data management for adaptive design of experiments: A litterature review

Gaëtan Blondet, Julien Le Duigou, Nassim Boudaoud, Benoît Eynard

Mechanics & Industry **16**, 611 (2015)

DOI: 10.1051/meca/2015041
www.mechanics-industry.org

Mechanics & Industry

Simulation data management for adaptive design of experiments: A litterature review

Gaëtan Blondet[a], Julien Le Duigou, Nassim Boudaoud and Benoît Eynard

Sorbonne Universités, Université de Technologie de Compiègne, Mechanical Engineering Systems Department, Mechanical Laboratory Roberval UMR UTC/CNRS 7337, CS 60319, 60203 Compiègne Cedex, France

Received 4 December 2014, Accepted 1 June 2015

Abstract – Recent evolutions of computer-aided product development and massive integration of numerical simulations to the design process require new methodologies to manage the continuously increasing flow of data and decrease the computational costs of numerical design of experiments. This paper presents a literature review of Simulation Data Management strategy and adaptive design of experiments methodology to detect possible links between these two fields and identify potential improvements for simulation process shortening. Adaptive design of experiments is based on several methods implying a profusion of different technics. Re-using best practices may help designers to choose relevant methods to reduce computational cost and simulation process duration.

Key words: Simulation lifecycle management / design of experiment / surrogate model / finite element / product lifecycle management

Glossary

SDM	Simulation Data Management
PLM	Product Lifecycle Management
SLM	Simulation Lifecycle Management
CAE	Computer-Aided Engineering
CAD	Computer-Aided Design
DoE	Design of Experiments

1 Introduction

Nowadays, competitiveness and efficiency of companies must be continuously improved to face worldwide competitors. Their processes and products must be continuously optimised with quality, cost and time objectives. As simulation is integrated to the product development process and is used to reach these objectives, the simulation process must reach these objectives too.

By means of recent advances in computing, numerical simulations are required to: (1) understand the product behavior, (2) optimise the product, (3) explore several solutions, and (4) validate the product. Numerical Design of Experiments (DoE) is more and more used to fulfil these four objectives by planning several runs of a numerical model with different parameters [1]. It can also increase product robustness and quality by taking into account product related uncertainties. Nevertheless, the amount of data produced by these simulations is huge, difficult and time-consuming to be extracted, stored and analysed.

This requires an efficient management of product data along its whole lifecycle, which is also known as Product Lifecycle Management (PLM) strategy. It offers to the company the necessary means to control their product along the lifecycle and to improve their processes [2]. In this case, Product data management (PDM) is crucial to reduce times by gathering, classifying and storing data all along the product lifecycle. Simulation Data Management (SDM) may be used to manage data related to Computer-Aided-Engineering (CAE) and computer-aided design (CAD) [3]. SDM, and more generally Engineering Data Management, is defined as a process which aims to organise, structure, store and track produced information, in order to "create a coherent knowledge", from process data and product data [4].

In order to reduce design process cost and time, the simulation process should be shortened. The paper proposes a research survey which focuses on two subjects: (1) re-use of simulation data and (2) reduction of computational costs. First, a global simulation process, focusing on the finite element method used in mechanical

[a] Corresponding author: gaetan.blondet@utc.fr

engineering, is described in Section 2 to identify its main steps and data sources. Second, after a presentation of the different simulation data types, SDM methods and tools are described to analyse the data re-use capabilities of SDM in Section 3. Section 4 presents DoE and methods used to reduce cost and time of simulation process, by re-using data according to different perspectives (e.g. numerical results and best practices), as adaptive DoE and surrogate models. To conclude, a possible coupling between recent DoE methods and SDM strategy is discussed.

2 Simulation process

Numerical simulation is a set of computations representing a physical phenomenon and behaviour. It aims predicting the response of a system subjected to its environment without any physical experiment. As shown in Figure 1, it can be sum up to 3 main steps. This representation can be found in [5, 6]. Details are available in [7].

The first step is the modelling or pre-processing: the physical problem is translated into mathematical equations. This idealisation can be made by different methods. The Finite Element method is the most well-known method and is used for a large range of problems in engineering. The accuracy of the model depends on designer's needs and resources. This step is critical because assumptions are made [8]. Typically, the complete model involves a modelling of a phenomenon (mechanics, thermal, etc.), a system (geometry, materials, parameters, etc.) and its environment (boundary conditions). Then, the problem is discretised with respect to its dimensions (spatial, time and other parameters). Finally, solving algorithm parameters are set. It can imply different methods, depending on the problem, objectives and results expected: iterative methods, DoE, optimisation methods, etc. This step is carried out with CAE softwares, supported by CAD softwares for the geometry definition.

The second step is the solving: equations are formulated from the previously discretised model and solved by the solver program according to chosen algorithms. This step can be improved by optimising the used methods [9]. Finite Element models may be solved in 1 h, 1 day, 1 week or even more. Thus, loops on this step (iterations or DoE) may demand an extremely high computational cost, without certainty about results correctness.

The last step consists in checking results and model validity (post-processing), and storing relevant data. There are five sources of error identified by [8, 10] which must be checked: (1) physical problem interpretation, (2) physical modeling, (3) numerical modeling, (4) solving and (5) results interpretation. Then, the model can be checked and results can be compared to customer's specifications for product validation.

These steps are, most of the time, embedded in CAE systems. Actually, softwares are progressively integrating different tools, for modelling, simulation and validation to cover the whole process and simplify it.

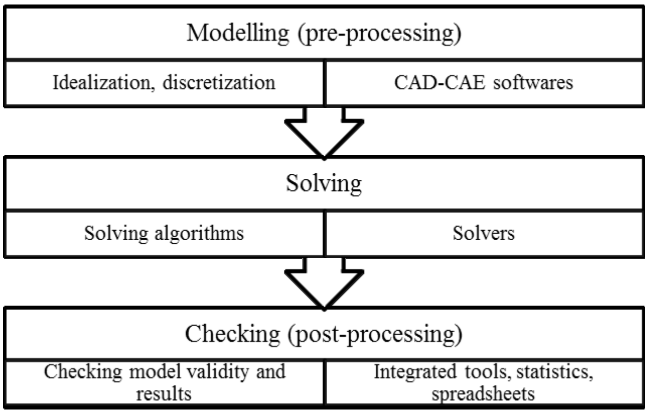

Fig. 1. General simulation process step\methods\tools diagram.

Multiple data types are generated from multiple systems at different steps. In addition to that, a simulation may be very expensive (highly nonlinear problem, iterative process, DoE, etc.) and require several softwares, generate a large amount of data and significantly increase computational costs. According to targeted objectives, several improvements can be done on the simulation process, as the use of a SDM system to trace, collect and store information. Regarding to the three steps of simulation process, data can be classified into two groups:

- Long-term data, including user's choices (hypothesis, methods and parameters), valid results, analyses and interpretations. These data have to be stored and archived once the simulation process ends. Collecting them is crucial for linking results with reality, for capitalising best practices and for final validation.
- Short-term data which are used during the solving step, such as intermediate results in computation loops. These data, generated by solving programs, are important to obtain accurate results, but are usually erased once computations are finished.

3 Simulation data management

3.1 Definition and challenges

SDM, covered by Product Data Management, is defined [11] as a technology which uses database solutions to enable users to manage structures of simulation data across the product lifecycle. It can also be defined, as Engineering Data Management defined by [4], as a process which aims to organise, structure, store and track information produced by simulation, in order to "create a coherent knowledge", from process data and product data. Workflow management and administration support may also be added [11–13]. SDM can also be considered as a part of Simulation Lifecycle Management (SLM) [14]. SLM, belonging to PLM, manages the process while SDM manages data. SLM covers collaborative product development, data traceability, decision-support and simulation systems integration for process automation.

Applications of SDM strategy in companies were classified in three maturity levels, defined by [13]:

– Level 1: "Isolated Islands". Local manual processes with low data volume stored in scattered and small databases.
– Level 2: "Archipelago". Many linked databases with general rules and governances.
– Level 3: "centralised/cloud". Automated processes, managing a very huge amount of high-value data stored in centralised database, led by a dedicated Information Technologies department in a collaborative environment.

Benefits of this third level are numerous: fast access to data from anywhere, and for each stakeholder, enhanced decision support, data re-using, automated reporting, improved productivity, centralised and secured database, support for internal policies, etc. [15] assess trends for input data management, used for modeling, by a survey (over 86 companies). Their survey reveals a progressive automation of simulation process, but a complete automation of the simulation process is still too difficult for these companies. Manual interactions with the process are necessary, and a complete automation of the whole process is unnecessary.

Main challenges for SDM were identified by [11,13], as:

– collect, store and retrieve growing volume, velocity and variety of simulation data (Big data problematic);
– manage network;
– improve collaboration and communication;
– ensure data security and integrity;
– integrate all data management system in one.

Thus, SDM aims to shorten the lead time of simulation process. Since simulation process has acquired a strategic role in product development process and may produce a "Data deluge", SDM must support this flow of data in collaborative environment along the product life-cycle [3,16]. So, SDM must involve an efficient data structuration.

3.2 Data models

As previously mentioned, the simulation process is linked to a large amount of different data sources, as input data (assembly, geometry, parameter, hypothesis, etc.), model data (mesh, solving methods, etc.) and output data (results, representations, reports, etc.). Other kinds of data were added to data models to support multi-domain, multi-component and multi-model simulation processes [4,11,17–19], simulation workflow management [20] and simulation loops management [5]. In addition to these multiple data types, SDM has to deal with different data provenance. Four input data sources were identified by [15]: corporate business systems, project specific data, collected data (e.g. measurements) and external reference systems (e.g. laws, standards) [15] described also data properties which should be checked to ensure data quality (e.g. accuracy, timeliness, etc.). Thus, SDM requires methods to deal with heterogeneous data from multiple sources and to ensure good properties.

Data models support data collection, filtration, structuring and control. These functions may enhance data sharing. Data models may also integrate multiple data sources as systems or different designers [21]. To deal with a large variety of data from multiple sources, a solution is the use of a neutral format [4,13]. Hence, standard models have been specified.

One of the most well-known standard data-model is ISO 10303-209. It belongs to the Standard for Exchange of Product data-model family (STEP). The Application Protocol 209 (AP209), created in 2001, improves design analysis and manufacturing productivity of composites and metallic structures. It provides a data model to share CAE data during the whole product lifecycle [22]. This standard involves a neutral file format to avoid interoperability problems between different systems and, thus, to drastically reduce analysis costs. It supports simulation data (models, load cases, results...) and configuration management data (versions) [23]. However, several lacks have been detected by [24]: all of its capabilities are not yet implemented in commercial systems (e.g. complex bonded assemblies support) and representation functions have to be improved.

A new version of AP209 has been released in 2014, with the contribution of CAE vendors, aeronautic and space companies and the LOTAR consortium. The standard ISO 10303 AP 209 edition 2, entitled "Multidisciplinary analysis and design", integrates entities from AP242 (Managed Model Based 3D Engineering) and AP 237 (Fluid dynamics) and has a common core with AP233 (System Engineering) and AP239 (Product Lifecycle Support). This new version provides a data model able to manage multidisciplinary design processes, to ensure links between CAD and CAE models (e.g. by supporting idealised geometries) and involves the integration of SDM systems.

The National Institute of Standards and Technology developed the Core Product Model. It provides an open, independent, generic and expandable product model, able to collect all product data throughout the product's lifecycle [21]. Product is represented as a set of artifacts (components) described by its function, form and behavior [25]. Several extensions were developed, as the Design Analysis Integration Model, which can provide tighter integration of spatial and functional aspects of design [26]. This extension meets the objective of STEP AP209 edition 2 to support the link between CAD and CAE systems.

Several industrial projects were launched to improve SDM implementation during the design process. SimPDM project is a framework giving recommendation concerning CAD/CAE integration [27]. The proposed SDM system is based on a data model, in order to manage the steps for modelling, solving and checking data, as well as configuration and version metadata. It also provides synchronisation functionality with other Data Management Systems, such as PDM system, after checking operations for data consistency. It is also able to be linked to CAE systems for every simulation steps. Collaborative CAD/CAE Integration project enhanced these recommendations by adding multi-domain and multi-enterprise [28]. CRESCENDO project [29] consisted in the creation of the Behavioral Digital Aircraft. It proposed integration of multi-disciplinary simulations, a collaborative environment with full data traceability and reusability. These methods aim to automate manual time-consuming activities such as modelling steps. It also supports multi-domain and multi-software models. Surrogate modelling methods are also included. These methods are described in the next section. A common language was built on STEP AP233 and AP239 to support such a collaborative process.

3.3 Data re-use for automation of simulation process

Re-using data and best practices from previous analysis is a challenge for industry [11,17,30]. It may shorten simulation processes by taking advantage of accumulated expertise and avoiding long data retrieval operations, rework operations, or other operations with no added value.

As presented in [31], Simulation and Analysis governance strategy, is designed to increase the process efficiency. It recommends re-using models and best practices. It also includes standardised work processes, integration with manufacturing operations, and collaborative engineering across the extended enterprise and over the full product lifecycle. In support of Simulation and Analysis governance, a SDM system should capture all data required for each step of the process. SDM strategy should also support DoE, optimisation and stochastic computations [13,31]. Thus, a SDM system prevents designers and analysts from creating data which already exist.

SDM solutions can be provided by PLM softwares and CAE softwares. PLM softwares offer general solutions to manage projects and related data, as workflows, CAD models, versions, results, etc. CAE softwares may embed DoE and very specific SDM solutions, as automatic report generation, link between numerical models and material databases. Both of them may be used to reuse best practices but it must be enriched to manage the whole DoE process with accuracy.

In addition to data models, ontologies can be used to enrich them. An ontology is a system of fundamental concepts set up to model, represent and describe a specific domain in terms of axiomatic definitions and taxonomic structures [16]. Ontologies are used in data models to specify a common language to efficiently share data between different stakeholders. An ontology covering requirement engineering, mechanical design and numerical simulation was defined [32]. This ontology, coupled with visualisation graph techniques, supports data capitalisation, data re-use and decision-making by representing dynamically relationship between different engineering entities (e.g. link between a specific design and its associated simulations). In manufacturing environment, the ontology ONTO-PDM was proposed [33]. It is a product ontology based on STEP 10303 and International Electrotechnical Commission standards (for manufacturing), to ensure interoperability between different stakeholders. Moreover, it enhances the relevancy, clarity and traceability of information of the product development process. OntoSTEP-NC, an ontology based on STEP-NC (AP238) standard data model was proposed [34]. This standard is designed to ensure bi-directionality exchanges between CAD systems, Computer Aided Manufacturing systems and machines. OntoSTEP-NC enriches this standard and allows a feedback from manufacturing to design and production engineering. This feedback contributes to capitalise and re-use the best practices in Manufacturing Process Management. This feedback also helps the programmer for decision support.

A Universal SDM System was defined [35] to face up to data variety. It enhances data re-use by centralising data storages in one single system. This system aims to be a common structuring organisation among the simulation data, available for every application and every stakeholder. However, its main weakness is that users always need a customised SDM system. According to [36], one static database for everything and everyone is unrealistic. Then, they propose an adaptive database, to ensure a heterogeneous data management from different sources in a collaborative environment.

A method for decision-support related to the modelling step was developed by [8]. It involves re-using previous studies by comparing their similarity to the current study. DoE and a metamodel (detailed further) are used to classify and compare performance of previous studies according to their different assumptions. Then, optimal modelling assumptions are proposed to the designer, regarding to quality, cost and time objectives.

3.4 Synthesis

SLM can be defined as a strategy to manage and automate the simulation process and related data. While a fully automated process is not necessary, this strategy is more and more demanded to manage a continuously growing amount of heterogeneous data with various sources. Data models and ontologies are used to structure these information and process. They are developed to support the entire simulation lifecycle and to ensure interoperability between systems and applications. They are essential for data storage and re-use. They also shorten the simulation process by automating manual and time-consuming tasks. Capitalised designers' choices, validated results and synthesis of analysis may help designers by providing an efficient decision support. However, DoE are not fully supported by the SLM strategy. They are not integrated in data models and ontologies. As shown further, DoE applications increase the need of re-using best practices.

4 Design of experiments for numerical simulation

4.1 Basic methods of DoE

Numerical DoE is a set of numerical experiments defined to assess the numerical model for different configuration, specification or solution. This method consists in exploring a design space, to improve product robustness and quality. It is used for sensitivity analysis, product optimisation or design exploration. A DoE is defined by a number of factors of different types (qualitative, discrete or continuous) and their levels.

As a numerical DoE is applied to a numerical model, each experiment involves an evaluation of the model, and thus computational cost depends on two parameters. First, a run of a Finite Element model may require a huge amount of computational resources to be performed. Second, a DoE involving a large number of runs (i.e. experiments) will increase drastically the needed resources. An efficient DoE should minimise the number of runs, with a distribution adapted to the objective (i.e. significant space-covering for design exploration). Thus, an optimal strategy is to choose the most efficient DoE and to use a method for reducing the computational cost of each run.

DoE were initially used for physical experiments. Classic DoE are presented by [30, 37–40]. It includes very expensive Full Factorial design and Fractional Factorial design, which is a cheaper version as some interactions between several parameters are neglected. It includes also Central composite and Box-Behnken design, limited for 3 or 5 levels, and Doehlert design, involving a more uniform distribution of experiments. All of these designs are based on a pre-selected regression model: they are model-dependent. But, in simulation process, each experiment is made by a deterministic solver. Thus, these DoE, which may involve repeated experiments to take into account experimental uncertainties, are inappropriate. More adapted DoE are used in this context. Latin Hypercube Sampling and orthogonal arrays are largely used as it is simple to build and to use. Furthermore, low-discrepancy-sequences-based-design is based on the minimisation of the discrepancy measure, i.e. the difference between a uniform sampling and the sampling of interest. It includes, amongst others, uniform design, Halton, Faure or Sobol sequences designs. These deterministic-sequences-based-designs are different from low-discrepancy design. Low-discrepancy design is based on stochastic algorithms minimising the discrepancy and is slower computed. These DoE used for numerical simulation are model-independent. Maximal entropy design maximises the amount of information in the distribution of experiments. Model-oriented designs [41, 42] can be used to obtain an optimal design. Several optimal designs exist, depending on the optimised criterion used to define each assessment: D-optimal, A-optimal, I and M-optimal (adapted for Kriging metamodel) [41], etc. If the model is linear in its parameters, these optimal designs are model-independent, otherwise not. However, the choice of the criterion is objective-dependent [1] proposed also to develop criteria for optimal design, as combination of a criterion to "identify the design region in which system performance is optimised" and design criteria "on the prediction error of the true output". Other criteria for this kind of DoE were presented in [43]. The authors give an assessment of presented criteria for sensitivity analysis.

The selection of a DoE method for a specific problem depends on the uniformity of the sampling, and the filling of design space. Moreover, it is also linked on the objectives and the constraints of DoE. DoE types are numerous, related to different applications and properties (see Ref. [1] for additional details on DoE properties). Thus, selection of DoE method may be a very time-consuming process [44] developed a design comparison chart to help designers in their choices, but it is non-exhaustive. Thus, there still exists a need for classification and comparison of DoE methods to support designers' decisions. In order to decrease the computational cost, by reducing the number of runs, several methods have been developed, such as adaptive DoE and surrogate modelling. Moreover, the choice of a specific DoE method will depend on the selected surrogate model.

4.2 Metamodels

Metamodeling, or Surrogate modelling, consists in replacing the costly Finite Element model by a function faster to be evaluated, to approximate a specific response with a lower computational cost. Surrogate models are used in many fields and a large amount of works was found related to this method. Applications for structural mechanics [45], Computed Fluid dynamics [46, 47], electromagnetics [48, 49], discrete event simulation for manufacturing processes [50] or forming process [51, 52] can be mentioned. Furthermore, surrogate models are used to fulfil Finite Element models objectives faster, as model approximation, design space exploration, sensitivity analysis [53] and optimisation [30].

Three steps are required to define a metamodel: (1) the surrogate model type selection, (2) the training and (3) the validation.

With the variety of existing metamodels, the selection of the best one may be difficult. A classification of these methods has been made by [54]. First, statistical learning methods include Response Surface Methods [55], and other polynomial approximations, Kriging [56, 57], Support Machine Vector and Multivariate Adaptive Regression Splines method [58]. Then, machine learning methods cover, amongst others, Artificial Neural Networks and clustering techniques. Finally, instance-based learning methods cover Radial Basis Functions method, which consist of a linear combination of functions approximation to improve polynomial models. This method can be represented as an Artificial Neural Network. Fitness inheritance methods and decision trees are also presented [54]. More details on these methods can be found in [1, 30, 59–61]. Surrogate models are built on some assumptions as function continuities, shape and smoothness [59]. If these assumptions are not valid, (e.g. in non-linear problems), multiple surrogate models can be used together to deal with function discontinuities. Different couples of DoE and metamodel were compared and the strong dependency between both of them is highlighted [1]. Furthermore, they showed a need for a DoE and metamodel classification to help the user regarding the objectives.

Surrogate model selection is problem dependent and a universal method does not yet exist [54, 59, 62]. However, the SUrrogate MOdeling Toolbox platform [63] integrates mathematical methods to select automatically the best metamodel. This solution replaces time lost in metamodel selection and tuning by increasing computational cost. An "automatic surrogate model type selection framework" exists using the Evolutionary Model Selection algorithm [64]. This algorithm dynamically selects the best surrogate model type and parameters. The same kind of algorithm is developed in [62]. However, as it is detailed further, it involves evolutionary algorithms drawbacks. Thus, as the selection is based on stochastic variables, the algorithm convergence is not guaranteed.

The training step is managed by assessing the Finite Element model with a DoE (as efficient as possible, see Sect. 4.1) to determine surrogate model coefficients. Each metamodel is used with its fitting method, as, for instance, the least-square methods, which link the model and results obtained from Finite Element model assessment [40]. The training step strongly depends on the number of assessments: not enough implies a low accuracy, but too much may lead to an overfitted model (learning by heart). This phenomenon means an inability to predict the behaviour beyond these first assessments. The analyst time spent to tune metamodels parameters (and also optimisation algorithms) is important to be took into account [65]. This time may not be negligible for some of the considered metamodels, and must be taken into account with metamodeling time (DoE selection and training step) to obtain a more accurate computational cost.

The validation is done by using another DoE to measure its predictive performance. The mostly used method consists in defining a DoE, using a partition (e.g. 20%) for the training and using the remaining partition (e.g. 80%) for validation.

Metamodel usefulness for optimisation problems was also discussed by [65]. They compared several metamodel-based optimisation and optimisation process without metamodels. They concluded that metamodeling does not always improve the optimisation efficiency. Metamodelling performance decreases with the complexity of the approximated function and depends also on allocated computational budget. However, the functions used during this test were perfectly known (analytical). It will not be the case for a real case study. Thus, the metamodel choice may be more difficult since its performance could be unpredictable.

The choice of the right metamodel is strongly linked to the function to approximate and to the available computational budget. It also depends on the DoE chosen. Recommendations about selection of some DoE and metamodels exists [40], but are not complete. Also, a metamodel considering multiple parameters can be hard and long to be tuned. However, some algorithms were developed to automatically select and tune the metamodel. As metamodeling is not always the most efficient strategy, the choice of using or not metamodels is also important. In this way, this method may be enhanced by re-using long-term data, such as capitalised best practices, for fast selection and tuning.

4.3 Adaptive design of experiments

This method can be found in the literature with several names: Adaptive DoE, metamodelling adaptive-recursive approach [65], sequential design, Variable Fidelity Modelling [66] or active learning [63]. Adaptive DoE is used to create iteratively a dedicated DoE for a specific problem, in order to maximise DoE efficiency. This method may fulfill several objectives as metamodel fitting, optimisation or design-space exploration. This method is based on two main steps: (1) searching for new experiments from an initial DoE and (2) selecting the best experiment to add to the initial DoE. Definition of best experiment depends on the chosen infill criterion used for selection, which is linked to the study objective. Then, these steps are repeated until a convergence criterion or a maximum number of experiments is reached.

The main issue concerning the development of dynamically adaptive DoE is the choice of an infill criterion [59]. Many developments were made for optimisation problems. A typical framework for Surrogate Based Optimisation is described in [46]. Here, the infill criterion is chosen to increase intensification, in order to find faster the global optimum of the objective function. The authors used a combination of an adaptive updating method and a real-time updating performed by an evolutionary algorithm, to refine the DoE around optima. This method aims both to optimise the surrogate model for: (1) fitting well with the objective function and (2) to obtain the optimum of the objective function. The Surrogate Based Optimisation framework is detailed and discussed in [59]. In the same approach, a particular Surrogate Based Optimisation framework, the Efficient Global Optimisation, using Kriging metamodel and Genetic Algorithm is used in [67]. Expected Improvement criterion involves a measure of possible improvement. This criterion is largely used [68–73]. The Particle Swarm Optimisation Intelligent Sampling method, which combines a Particle Swarm Optimisation method, used to add new experiments, and adaptive response surface methods metamodel are used [52]. An algorithm combining Kriging metamodel and Particle Swarm Optimisation algorithm for optimisation is also presented by [74]. Another application can be found in [75], with a very specific metamodel and an elitist Genetic Algorithm for forming applications. A taxonomy was presented in [69] to select the infill criteria related to the metamodel used, but only for polynomial metamodels.

These adaptive methods are based on metaheuristics, in order to search for a new experiment. Metaheuristics are algorithms used to solve complex optimisation problems. A recent survey presents these methods [76]. Their main properties are nature-inspired and based on stochastic components. These methods are largely used to create an adaptive DoE algorithm [46, 52, 67, 74, 75]. These methods were divided into two groups [76]: single-solution based metaheuristics (e.g. simulated annealing, Variable Neighborhood Search, Tabu Search, etc.) and population-based metaheuristics. Population-based methods can be split up into two other sub-classes: evolutionary computation methods (e.g. Genetic Algorithm, cultural and coevolutionary methods, etc.) and Swarm intelligence methods (e.g. Ant colony, Particle Swarm Optimisation, artificial immune systems, etc.) [76–78]. Another assessment of several metaheuristics was made by [79]. As for DoE and surrogate models, selection and tuning are difficult. Methods were developed, as Adaptive metaheuristics and hyper-heuristics to select and tune automatically metaheuristics [76].

For metamodel fitting, a simple criterion consists in selecting the experiment related to the estimation which maximizes an error measurement between the Finite Element model and the metamodel, such as variance of estimation provided by a Kriging metamodel [41, 42]. A class of infill criteria is related to contour approximation, which is close to metamodel fitting [68] have listed several criteria of this class, based on the uncertainty of each experiment (margin uncertainty for a given trust-region or confidence intervals [80]). A margin indicator function can be defined to set a trust-region around the function (contour) to approximate. The goal is to select the closest experiment from the function. There exist also the margin probability function, and the expected feasibility function. A sub-class, called One-step-look-ahead criteria, includes the Weighted Integrated-Mean-Square-Error criterion [41]. The Expected Improvement-based criterion for contour estimation was also developed [81]. Reference [38] developed the adaptive Wootton, Sergent, Phan-Tan-Luu's algorithm to build a space-filling design able to deal with high-dimensional problems (number of parameters > 20).

To be successful, all of these methods used to search for and select infill experiments have to do as much design-space exploration as exploitation (intensification) [59]. The goal is to give accurate results without missing any optimum or falling into a local optimum. Each algorithm has a particular manner to achieve this equilibrium.

4.4 Synthesis

Adaptive DoE methods may involve metamodelling to reduce the computational cost and metaheuristics to search for new experiments. Many infill criteria exist to select the most efficient experiment and to sequentially add it to the DoE. This section has not covered all of these criteria, since they are very numerous. Although the Expected Improvement criterion seems to be the most used, a clear assessment would be valuable to choose the most efficient criterion. The efficiency of adaptive DoE method strongly depends on surrogate model, metaheuristics and criteria used. While hyper-heuristics and adaptive metaheuristics are being developed to shorten the simulation process, none of these methods are based on best-practices re-using. It could shorten the process by supporting decision by already known results, instead of run new computations. There exists a real need to develop methods able to compare, classify and select the right methods according to a specific problem. Thus, a capitalisation of successful combination of these elements could help designers to shorten especially pre-processing step of numerical simulation.

5 Discussion

In this paper, two opposite ways to shorten the simulation process are presented. Firstly, the amount of data and data heterogeneity are constantly increasing with computers and networks capabilities. Thus, a SDM strategy has to be set up to shorten pre-processing and post-processing steps, by re-using best-practices. Moreover, it may lead to an automated process by gathering, storing, classifying, retrieving and re-using data. Nevertheless, some manual actions are still mandatory, at least to check the process. Secondly, mathematical methods can be used to shorten the solving step, especially for DoE, as metamodelling and adaptive DoE. These methods re-used short-term data (e.g. intermediate results).

A challenge for SDM, and SLM, is to efficiently re-use long-term data (e.g. best practices) to shorten simulation process. This involves a specific data structure to classify data, and a strategy to retrieve the most relevant data. A lot of works were done to manage the big flow of data. Recent developments support different formats, versions and sources to guarantee traceability and fast access to the highest-quality data available. To structure data and enhance collaborative simulation process, standard data models, as STEP AP 209 ed2, were developed. These data models can manage complex simulation process, involving multi-domain and multi-software project in a collaborative environment. However, management of DoE just begins to be developed, but not for adaptive DoE and technics described in this paper. In the context of DoE applications, several methods were developed to reduce:

- the number of runs, by re-using results from previous runs from a DoE to adapt it dynamically;
- the computational cost of each run, by creating a surrogate model.

Nonetheless, these methods used to shorten the solving step need a long time to be selected and tuned. Thus, they actually need long-term data re-use and expertise, and so, a SDM strategy capitalising best practices. Furthermore, methods re-using short term data are being developed to automatically choose best methods and parameters, as optimisation methods (Evolutionary Model Selection, Efficient Global Optimisation) and hyper-heuristics. These methods may increase the computational cost to shorten the process. So, there is still a need for adaptive strategies suggestion to take advantages from capitalisation and -classification of long-term data. Finally, these two ways converge to a SLM strategy need.

This paper aimed also to show lacks in SDM, as shown in (Fig. 2). Data management is available for Finite

Fig. 2. State-of-the-art of SDM applied to DoE and adaptive techniques.

	Data models	metadata management	Re-use data	
			Short-term	Long-term
Finite Element model	X	X	X	X
DoE			X	
Adaptive DoE			X	

Element models, as data models (e.g. STEP AP203) exist to deal with its data and metadata (i.e. version, format, etc.). However, DoE is not included in these models, neither for Adaptive DoE and all methods involved, as surrogate models and optimisation methods. Short-term data re-use is available for each category (e.g. for iterative computations, model convergence analysis, adaptive DoE, memory based-metaheuristics, automatic method selection). Best practices re-use is possible for Finite Element models, since all tools needed exist. But, without any structure to capitalise data, long-term data re-use, for DoE and related methods, is almost impossible.

Thus, the development of a specific data model, or ontology, for adaptive DoE will lead to re-use data enhancements. It would help designers to choose the best fitted set of method to a specific study, based on previous studies and academic knowledge by avoiding tasks with no added value, like manual data retrieval and comparison, useless runs and tuning operations. It should also include metadata management, like version, related configuration. Such improvements will also lead the simulation process to be more automated and optimised for specific applications.

6 Conclusion

The importance of the simulation process is increasing in the product design process. As companies may have to manage product lifecycle in a collaborative context, the Simulation Lifecycle Management strategy was developed. With the evolution of computers and networks, the amount of simulation data is continuously increasing. Design of experiment is a method involving a production of a huge amount of data. Thus, the possible contributions of Simulation Data Management on DoE were analysed.

SDM involves a structuration of simulation data. This structuration is provided by data models. Although recent data models support multi-domain complex simulations integrated with multiple softwares, DoE are not yet clearly implemented. Moreover, Adaptive DoE and related technics are not supported too. DoE may involve a very high computational cost if a large number of experiments on a Finite Element model are demanded. Many different DoE exist and it implies some difficulties to select the best DoE. To reduce the computational cost, many methods were developed, as Adaptive DoE and surrogate modelling. But, these methods may increase the modelling step time. It may be very difficult to select and tune each technic used to obtain an adaptive DoE.

A SDM strategy should be set to capitalise best practices related to DoE. It may avoid useless operations by using a centralised database. Such a strategy may enhance decision support for designers.

Acknowledgements. This work is done in the French FUI project SDM4DOE. We also thank all consortium partners for their contribution during the development of ideas and concepts proposed in this paper.

References

[1] V.C.P. Chen, K.-L. Tsui, R.R. Barton, M. Meckesheimer, A review on design, modeling and applications of computer experiments, IIE Transactions 38 (2006) 273–291

[2] J. Stark, Product lifecycle management, 21st century paradigm for product realisation, 1st edition, Springer, 2004

[3] CIMDATA, Beyond Simulation Data Management, 2012

[4] G. Feng, D. Cui, C. Wang, J. Yu, Integrated data management in complex product collaborative design, Computers in Industry 60 (2009) 48–63

[5] S. Charles, Gestion intégrée de données CAO et EF. Contribution à la liaison entre conception mécanique et calcul de structures, Thèse, Université de Technologie de Troyes, 2005

[6] E. Patelli, et al., General purpose software for efficient uncertainty management of large finite element models, Finite Elements in Analysis and Design 51 (2012) 31–48

[7] V. Adams, A Designer's Guide to Simulation with Finite Element Analysis, NAFEMS, 2008

[8] Y. Benhafid, N. Troussier, N. Boudaoud, Z. Cherfi, Méthode d 'aide à l'idéalisation de modèles issus de la CAO pour le calcul de structures, Mechanics & Industry 6 (2005) 289–295

[9] X.-S. Yang, S. Koziel, L. Leifsson, Computational Optimization, Modelling and Simulation: Recent Trends and Challenges, in: Proceedings of International Conference on Computational Science, ICCS 2013, Procedia Computer Science, Elsevier B.V., Barcelona, Spain, 2013, pp. 855–860

[10] F. Pourroy, Apport des éléments finis à la conception mécanique, in: Méthodes et Outils Pour La Conception, Les Techniques de L'Ingénieur, Paris, 1999, pp. 475–483

[11] M. Norris, Business Value from Simulation Data Management-a Decade of Production Experience, NAFEMS, 2012

[12] ProSTEP, Automotive CAE Integration, Annex: Simulation Data Management, 2010

[13] S. Chari, Addressing Engineering Simulation Data Management (SDM) Challenges, White Paper, Cabot Partners, 2013

[14] P. Lalor, Simulation Lifecycle Management, NAFEMS – Benchmark Journal, 2007, pp. 16–19

[15] A. Skoogh, T. Perera, B. Johansson, Input data management in simulation – Industrial practices and future trends, Simul. Modell. Practice Theory 29 (2012) 181–192

[16] S.K. Chandrasegaran, K. Ramani, R.D. Sriram, et al., The evolution, challenges, and future of knowledge representation in product design systems, Computer-Aided Design 45 (2013) 204–228

[17] P. Graignic, T. Vosgien, M. Jankovic, V. Tuloup, J. Berquet, N. Troussier, Complex System Simulation: Proposition of a MBSE Framework for Design-Analysis Integration, in: Proceedings of Systems Engineering Research (CSER'13), Procedia Computer Science, Elsevier B.V., Atlanta, Georgy, 2013, pp. 59–68

[18] W. Wang, R. Mo, Y.T. Jin, Y. Zhang, Research on Multi-dimensional Simulation Data Management Model for Complicated Product, in: Proceedings of 2010 International Conference on Electrical and Control Engineering, IEEE, Wuhan, China, 2010, pp. 2948–2951

[19] I. Assouroko, G. Ducellier, B. Eynard, P. Boutinaud, Semantic-based approach for the integration of product design and numerical simulation, in: Proceedings of International Conference on Product Lifecycle Management, 2011

[20] H. Yin, Y. Gao, H. Yan, J. Wang, Simulation Data and Process Management System in the Development of Virtual Prototype, in: Proceedings of 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, IEEE, Toronto, Ontario, 2010, pp. 152–155

[21] S.J. Fenves, S. Foufou, C. Bock, R.D. Sriram, CPM: A Core Model for Product Data, 2006, pp. 1–14

[22] S. Charles, B. Eynard, P. Bartholomew, C. Paleczny, Standardization of the Finite Element Analysis Data-Exchange in Aeronautics Concurrent Engineering, J. Comput. Inform. Sci. Eng. 5 (2005) 63–66

[23] B.-K. Park, J.J. Kim, A sharable format for multidisciplinary finite element analysis data, Computer-Aided Design 44 (2012) 626–636

[24] K.a. Hunten, A. Barnard Feeney, V. Srinivasan, Recent advances in sharing standardized STEP composite structure design and manufacturing information, Computer-Aided Design 45 (2013) 1215–1221

[25] J.S. Gero, M.A. Rosenman, A conceptual framework for knowledge-based design research at Sydney University's design computing unit, Artif. Intell. Eng. 5 (1990) 65–77

[26] S.J. Fenves, R.D. Sriram, Y. Choi, B. Gurumoorthy, G. Mocko, Master Product Model for the Support of Tighter Integration of Spatial and Functional Design, 2003

[27] ProSTEP/IViP, Simulation Data Management. Recommendation. Integration of Simulation and Computation in a PDM Environment (SimPDM), PSI 4, Version 2.0, 2008

[28] G. Syal, CAE-Process and Networks: A Methodology for Continuous Product Validation, Ph.O. Thesis, Karlsruhe Institute of Technology, 2013

[29] CRESCENDO Consortium, CRESCENDO Forum Handbook, CRESCENDO Forum, Toulouse, 2013, p. 160

[30] G.G. Wang, S. Shan, Review of Metamodeling Techniques in Support of Engineering Design Optimization, J. Mech. Design 2006 (2006) 415–426

[31] CIMDATA, Simulation & Analysis Governance, 2014

[32] I. Assouroko, G. Ducellier, P. Boutinaud, B. Eynard, Knowledge management and reuse in collaborative product development – a semantic relationship management-based approach, Int. J. Product Lifecycle Manage. 7 (2014) 54–74

[33] H. Panetto, M. Dassisti, A. Tursi, ONTO-PDM: Product-driven ontology for Product Data Management interoperability within manufacturing process environment, Adv. Eng. Inform. 26 (2012) 334–348

[34] C. Danjou, J. Le Duigou, B. Eynard, Closed-Loop Manufacturing process based on STEP-NC, Int. J. Interactive Design Manufact. (2015) http://dx.doi.org/10.1007/s12008-015-0268-1

[35] Y. Xue-rong, L. Jia-hong, L. Ying, L. Feng, F. Xiang-jun, W. Yuan-lan, Study of Universal Simulation Data Management System, in: 2009 International Conference on Information Technology and Computer Science, IEEE, 2009, pp. 333–338

[36] L. Yu, W. Gao, Q. An, J. Zhao, D. Liang, D. Gao, Data resource management according to customer requirements, Math. Comput. Modell. 54 (2011) 895–901

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

[38] A. Beal, M. Claeys-Bruno, M. Sergent, Constructing space-filling designs using an adaptive WSP algorithm for spaces with constraints, Chemometrics and Intelligent Laboratory Systems 133 (2013) 84–91

[39] J. Gauchi, Plans d'expériences optimaux: un exposé didactique, La Revue de Modulad 33 (2005)

[40] T.W. Simpson, J. Peplinski, P.N. Koch, Metamodels for computer-based engineering design: survey and recommendations, Eng. Comput. 17 (2001) 129–150

[41] V. Picheny, D. Ginsbourger, O. Roustant, R.T. Haftka, N.-H. Kim, Adaptive Designs of Experiments for Accurate Approximation of a Target Region, J. Mech. Design 132 (2010) 81–89

[42] S. Gazut, J.-M. Martinez, G. Dreyfus, Y. Oussar, Towards the optimal design of numerical experiments, IEEE Trans. Neural Networks 19 (2008) 874–882

[43] E. Janouchová, A. Kuèerová, Competitive comparison of optimal designs of experiments for sampling-based sensitivity analysis, Comput. Struct. 124 (2013) 47–60

[44] S.M. Sanchez, H. Wan, Work smarter, not harder: a tutorial on designing and conducting simulation experiments, in: Winter Simulation Conference, 2012, pp. 1929–1943

[45] M.J. Pais, F.A.C. Viana, N.H. Kim, Enabling high-order integration of fatigue crack growth with surrogate modeling, Int. J. Fatigue 43 (2012) 150–159

[46] E. Iuliano, D. Quagliarella, Proper Orthogonal Decomposition, surrogate modelling and evolutionary optimization in aerodynamic design, Comput. Fluids 84 (2013) 327–350

[47] T. Braconnier, M. Ferrier, J.-C. Jouhaud, M. Montagnac, P. Sagaut, Towards an adaptive POD/SVD surrogate model for aeronautic design, Comput. Fluids 40 (2011) 195–209

[48] S. Koziel, S. Ogurtsov, L. Leifsson, Knowledge-Based Response Correction and Adaptive Design Specifications for Microwave Design Optimization, in: Proceedings of International Conference on Computational Science, ICCS 2012, Procedia Computer Science, Elsevier B.V., Omaha, Nebraska, 2012, pp. 764–773

[49] S. Koziel, S. Ogurtsov, L. Leifsson, Physics-based Surrogates for Low-cost Modeling of Microwave Structures, in: Proceedings of International Conference on Computational Science, ICCS 2013, Procedia Computer Science, Elsevier B.V., Barcelona, Spain, 2013, pp. 869–878

[50] A. Negahban, J.S. Smith, Simulation for manufacturing system design and operation: Literature review and analysis, J. Manufact. Syst. 33 (2014) 241–261

[51] U.A. Dabade, R.C. Bhedasgaonkar, Casting Defect Analysis using Design of Experiments (DoE) and Computer Aided Casting Simulation Technique, in: Proceedings of Forty Sixth CIRP Conference on Manufacturing Systems 2013, Elsevier B.V., Setubal, Portugal, 2013, pp. 616–621

[52] W. Hu, L.G. Yao, Z.Z. Hua, Optimization of sheet metal forming processes by adaptive response surface based on intelligent sampling method, J. Mater. Process. Technol. 197 (2008) 77–88

[53] J. Luo, W. Lu, Sobol' sensitivity analysis of NAPL-contaminated aquifer remediation process based on multiple surrogates, Comput. Geosci. 67 (2014) 110–116

[54] L. Shi, K. Rasheed, A survey of Fitness Approximation Methods Applied in Evolutionnary algorithms, in: Computational Intelligence in Expensive Optimization Problems, Springer, 2010, pp. 3–28

[55] G.E.P. Box, K.B. Wilson, On the experimental attainment of optimum conditions, J. Royal Statist. Soc. 13 (1951) 1–45

[56] D.G. Krige, A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand, J. Chem. Metall. Mining Soc. South Africa 52 (1951) 119–139

[57] G. Matheron, Principles of geostatistics, Economic Geology 58 (1963) 1246–1266

[58] J.H. Friedman, Multivariate Adaptive Regression Splines, The Annals of Statistics 19 (1991) 1–141

[59] A.I.J. Forrester, A.J. Keane, Recent advances in surrogate-based optimization, Progress Aerospace Sci. 45 (2009) 50–79

[60] V. Dubourg, B. Sudret, F. Deheeger, Metamodel-based importance sampling for structural reliability analysis, Probab. Eng. Mech. 33 (2013) 47–57

[61] S. Castric, L. Denis-Vidal, Z. Cherfi, G.J. Blanchard, N. Boudaoud, Modeling Pollutant Emissions of Diesel Engine based on Kriging Models: a Comparison between Geostatistic and Gaussian Process Approach, in: Proceedings of 14th IFAC Symposium on Information Control Problems in Manufacturing, INCOM'12, Bucharest, Romania, 2012, pp. 1708–1715

[62] D. Lim, Y. Jin, Y. Ong, B. Sendhoff, Generalizing Surrogate-Assisted Evolutionary Computation, IEEE Trans. Evol. Comput. 14 (2010) 329–355

[63] D. Gorissen, T. Dhaene, P. Demeester, I. Couckuyt, K. Crombecq, A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based Design, J. Machine Learning Res. 11 (2010) 2051–2055

[64] D. Gorissen, T. Dhaene, F. De Turck, Evolutionary Model Type Selection for Global Surrogate Modeling, J. Machine Learning Res. 10 (2009) 2039–2078

[65] S. Razavi, B.A. Tolson, D.H. Burn, Numerical assessment of metamodelling strategies in computationally intensive optimization, Environ. Modell. Software 34 (2012) 67–86

[66] Z.-H. Han, S. Görtz, R. Zimmermann, Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function, Aerospace Sci. Technol. 25 (2013) 177–189

[67] I. Couckuyt, F. De Turck, T. Dhaene, D. Gorissen, Automatic surrogate model type selection during the optimization of expensive black-box problems, in: S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu (Ed.) Proceedings of Winter Simulation Conference, IEEE, Phoenix, Arizona, 2011, pp. 4274–4284

[68] V. Dubourg, Méta-modèles adaptatifs pour l'analyse de fiabilité et l'optimisation sous contrainte fiabiliste, Thèse, Université Blaise Pascal-Clermont II, 2011

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

[70] D. Huang, T.T. Allen, W.I. Notz, N. Zheng, Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models, J. Global Optim. 34 (2006) 441–466

[71] C. Chevalier, Fast uncertainty reduction strategies relying on Gaussian process models, Thèse, University of Bern, 2013

[72] F. Jurecka, M. Ganser, K.-U. Bletzinger, Update scheme for sequential spatial correlation approximations in robust design optimisation, Comput. Struct. 85 (2007) 606–614

[73] N. Courrier, P.-A. Boucard, B. Soulier, The use of partially converged simulations in building surrogate models, Adv. Eng. Software 67 (2014) 186–197

[74] H. Liu, S. Maghsoodloo, Simulation optimization based on Taylor Kriging and evolutionary algorithm, Appl. Soft Comput. 11 (2011) 3451–3462

[75] M. Ejday, L. Fourment, Optimisation multi-objectifs à base de méta-modèle pour les problèmes de mise en forme, Mechanics & Industry 11 (2010) 223–233

[76] I. Boussaïd, J. Lepagnot, P. Siarry, A survey on optimization metaheuristics, Inform. Sci. 237 (2013) 82–117

[77] J. Rada-Vilela, M. Johnston, M. Zhang, Population statistics for particle swarm optimization: Resampling methods in noisy optimization problems, Swarm and Evol. Comput. 17 (2014) 1–23

[78] B. Yannou, F. Cluzel, M. Dihlmann, Evolutionary and interactive sketching tool for innovative car shape design, Mechanics & Industry 14 (2013) 1–22

[79] L. Gacôgne, Comparaison entre PSO et autre heuristiques d'optimisation avec opérateurs implicites, in: Séminaire Optimisation Par Essaim Particulaire, Paris, France, 2003, pp. 1–12

[80] E. Jack Chen, M. Li, Design of experiments for interpolation-based metamodels, Simul. Modell. Practice Theory 44 (2014) 14–25

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