COMPUTER-AIDED DESIGN & APPLICATIONS, 2016
VOL. 13, NO. 6, 872–884
http://dx.doi.org/10.1080/16864360.2016.1168236

Towards a knowledge based framework for numerical design of experiment optimization and management

Gaëtan Blondet[a], Farouk Belkadi[b], Julien Le Duigou[a], Alain Bernard[b] and Nassim Boudaoud[a]

[a]Sorbonne Universités, Université de Technologie de Compiègne, France; [b]LUNAM Université, France

ABSTRACT
Numerical Design of Experiment (DoE) is a powerful tool for product development, used to improve product quality and robustness. However, the simulation process can be highly extended by the DoE process. While methods have been developed to shorten the execution of numerical DoE, the time needed to set up the numerical DoE process is longer and longer. This paper presents a description of the objectives and first results of the SDM4DOE project (Simulation Data Management for Design of Experiment). This project aims to define a set of tools and methods to improve the simulation process involving DoE: data management, data robustness improvement and process shortening. A knowledge-based approach is proposed to solve this main issue, based on a specific knowledge representation.

KEYWORDS
Knowledge management; adaptive design of experiment; simulation data management; knowledge-based system

1. Introduction

Nowadays, numerical simulations are more and more used in a product development process to reach quality, cost and time objectives [5]. By means of recent advances in computing, numerical simulations are required to: (1) understand the product behavior, (2) optimize the product, (3) explore several solutions, and (4) validate the product. Numerical Designs of Experiments (DoE) are used to fulfil these 4 objectives by planning several runs of a numerical model for different configurations [16].

These runs may be numerous and may be based on a costly numerical model. So, the computational cost of a DoE may be very high. The literature survey and the recent experimentation achieved in the SDM4DOE project show that the DoE process is very complex. It consists of multiple sub-processes which must be defined and tuned to prepare the execution [26][41]. It implies lot of iterations and interactions between experts from heterogeneous fields. In addition, designers need to use several methods and tools to support the generation and the exploitation of data and knowledge.

The amount of data produced by the simulation process, and thus by the DoE process, is huge and difficult to be managed [12]. This requires an efficient management of product data along its whole lifecycle. Product Lifecycle Management (PLM) strategy offers the company the necessary means to control their product along the lifecycle and to improve their processes [36][48]. PLM strategy results from the integration of different business processes and from interoperability between several tools [20]. In this context, Product Data Management (PDM) is crucial to reduce times by gathering, classifying and storing data all along the product lifecycle. Simulation Data Management (SDM) is a specific application of PDM for Computer-Aided Engineering applications [17]. SDM, and more generally Engineering Data Management, is defined as a process which aims to organize, structure, store and track produced information, in order to "create a coherent knowledge", from process data and product data [22].

Thus, the DoE process faces up two major issues:

1. The computational cost of the process execution;
2. The lack of a specific data management approach for DoE, which leads to an important loss of time. Data needed to set up the process and data generated need a SDM approach.

SDM4DOE project aims to solve these issues by producing an open-source simulation framework to set up and run numerical DoE. Using this framework, designers would be able to run a DoE fast on complex and costly simulations, and to ensure data traceability and capitalization. The principle of this strategy is to collect the best practices, know-how and expertise in a dedicated knowledge base. Consequently, knowledge could be capitalized

CONTACT Gaëtan Blondet gaetan.blondet@utc.fr

and reused for decision-support in a collaborative product development process.

The aim of this paper is first to address a global survey concerning the main characteristics related to the DoE domain, to pointed out its complexity and the need of innovative ICT-based to cope with these needs. Based on this analysis, an architecture of a decision-support system is proposed, supported by a specific knowledge classification and modeling, to assist designer during the DoE definition. Decision-support systems could be used to help experts and optimize the DoE process in terms of results, time and resources costs. Knowledge-based and data management frameworks might give interesting advantages to the previous systems performance by favouring knowledge capitalization retrieval and reuse [9].

Section 2 presents DoE process and explains the two identified issues. Section 3 details the state of the art of SDM strategy and knowledge management approach. Section 4 introduces developments and proposals made for the SDM4DOE project, as a knowledge-based system and a specific knowledge base for DoE. This section details also the framework and planned validation cases.

2. Design of experiments

Multi-physics simulation gives important information for the selection of the best alternative of technical solutions and to estimate the product performance among its whole lifecycle [6]. DoE supports the implementation of different simulations and their organization according to the variety of physics, parameters and solution alternatives to be analysed.

The DoE process is based on the simulation process which is composed of 3 main steps [14]. A parameterized numerical model of the studied system is created. Then, the solver computes numerical model results, which are analysed during the post-process step. The solving can involve very high computational cost, according the needed level of fidelity. Once the numerical model quality is checked, a DoE can be created.

A DoE is a set of experiments, defined to assess the numerical model for different configurations of the product [29]. DoE can be used to fulfil several objectives, as product optimization, sensitivity analysis, robustness analysis and exploration of different designs. A DoE is defined by its factors and associated levels. Each experiment is defined by a specific set of levels reached by each factor. Then, the cost of a DoE is the cost of a numerical model run, multiplied by the number of experiments. 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 minimize the number of experiments and optimize the design-space-covering, according to the objective (exploration, product optimization . . .).

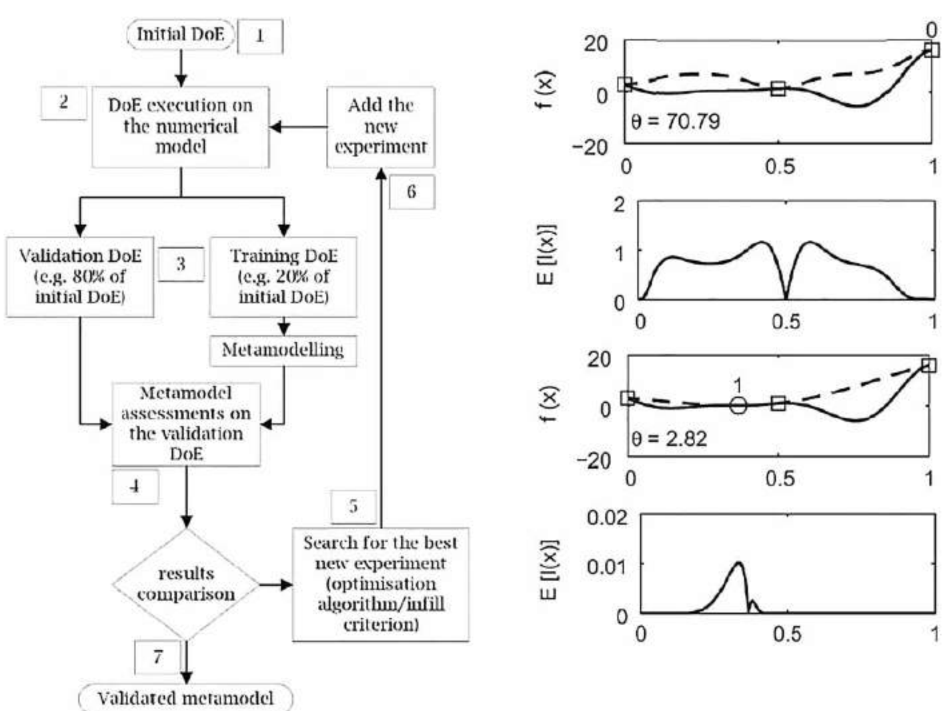

Figure 1. Flowchart of Adaptive DoE strategy and adaptive metamodeling using the expected improvement infill criterion (E[I(x)]) from [25].

The DoE process [29] begins by a first analysis of relevant factors amongst numerical model design parameters. It is followed by a sensitivity analysis (supported by a first DoE) to select more accurately influent factors according to the studied output. This step must be done to avoid a too high computational cost. During this step a metamodel (or surrogate model), is created, based on the results of this DoE. The metamodel is a function (e.g. a polynomial function) created to reproduce the behavior of the initial model for a specific output. It aims to be faster to be assessed than the initial numerical model. Once the sensitivity analysis is finished, a new DoE can be defined to reach the initial objective (optimization, exploration, etc.). When the metamodel previously created is sufficiently accurate, it can replace the initial numerical model if other runs are needed.

In order to shorten the process, adaptive DoE [7] may be set up. Adaptive DoE, as shown in Fig. 1, are used to create iteratively a dedicated DoE for a specific problem, in order to maximize DoE efficiency in accordance with the objective and constraints of the study. This method consists in:

1. Defining an initial DoE: the type, the number of factors and levels are specified. To ensure process efficiency, the optimal number of initial experiments must be defined. This methods aims to involve as few experiments as possible. DoE types are numerous [26], related to different applications and properties. In a numerical context, space-filling designs are used, as Latin Hypercube Sampling, low-discrepancy sequences and optimal DoE [28]. The choice of a particular kind of DoE is influenced by its space-filling and uniformity properties, the objectives of the study (optimization, exploration, etc.), available resources, constraints, associated numerical model, and by the studied output.
2. Running the experiments on the initial numerical model;
3. Metamodeling: the metamodel type and its submethods are set; internal parameters are computed during a training step based on the initial DoE. Metamodels types are numerous too [49]. Polynomial regressions, support vector regressions, kriging metamodels [11] and artificial neural networks are examples of common metamodels. After the definition, the metamodel is trained on a DoE. The metamodel is fitted with the numerical model behavior. Thus, the selection of the metamodel type depends on the DoE definition [16][49]. Actually, it is common to use a small partition of the DoE, run on the costly numerical model, to train the metamodel. Then, the remaining experiments are used to assess the metamodel predictivity [49]. To keep metamodeling usefulness, the DoE should not be composed of too numerous experiments. If there are too many experiments, the time potentially saved by executing the metamodel instead of the initial numerical model would be lost by running the DoE. This selection depends also on many others parameters: objective of the study, the properties of the studied output (linear or not), the computational cost of the initial numerical model, the computational budget. For instance, a polynomial regression is firstly defined by its degree, and the coefficients are determined during the training step, based on the DoE. The maximal degree and the correctness of computed coefficients depend on the number of available experiments. Too few experiments imply an impossible regression, too many would lead to a very inaccurate metamodel. Furthermore, experiments must be as different as possible (design-space filling and uniformity properties) to ensure a good training. According to the regularity of the function to approximate, an experiment too close from another one would not increase accuracy;
4. Validating the metamodel: the metamodel is assessed on the other partition of the DoE. Results obtained by the metamodel are compared with results previously computed by the initial numerical model. Different statistical methods exists to compute the error [27];
5. Searching for a new experiment: if the metamodel is not validated, a measurement of potential accuracy improvement, the infill criterion [25], is associated with each possibly added experiment. An optimization algorithm will find the experiment which maximizes the infill criterion. The type of infill criterion, the optimization algorithm type and its parameters must be defined. Expertise is needed to set the optimization algorithm. They are classified in two main types: local methods and global methods (metaheuristics). Since the studied model output may be non-linear (unknown in advance), many adaptive methods are based on metaheuristics [25]. A lot of metaheuristics exist [10] and they are hard to be tuned since they are nature-inspired and based on stochastic components. The convergence behavior of these methods cannot be forecast. In addition, local and global methods can be used together to obtain the advantages of both of them [10];
6. Adding the new experiment to the initial DoE and running the initial numerical model on this experiment;
7. Repeating steps 2 to 6 until the metamodel is enough accurate or the maximum of the initial numerical model execution number is reached.

Figure 1 shows the general principle of adaptive DoE. The four diagrams shows an example of results obtainable by this process. On the first iteration, the metamodel (dotted-curve of the top right-hand side diagram) is based on 3 evaluations of the numerical model (solid curve), the infill criterion curve (diagram below) related to the metamodel performance. The experiment corresponding to the highest value of this criterion will be added at the next iteration. Second iteration (bottom diagrams): the metamodel is supported by 4 evaluations and fits better the numerical model than the previous metamodel; the expected improvement criterion is largely decreased.

Thus, the number of potentially costly runs is minimized. But, this method involves new parameters, added to those belonging to DoE and metamodels. The infill criterion and the optimization algorithm must be defined and tuned. The selection of the DoE and metamodel type, and associated parameters and methods, may be a very time-consuming process and requires a good expertise to be set up. The designer would have to choose a particular combination of methods amongst large sets of methods, without any rules in some cases. The time saved by using these methods, aiming to shorten the execution of this process, may be lost because of the time needed to prepare this process. There is a real need to shorten this decision process, to help designers to share their expertise and to classify these methods.

In order to help designers to choose these methods, some classifications are available [26][34][45][47]. However, these classifications are non-exhaustive. They do not cover all types of DoE, metamodels, infill criteria, optimization algorithms, etc. They do not take into account objectives and constraints of the study, the effect of the problem dimensionality or the choice of parameters involved by each method. Several methods were developed to defined automatically some of these parameters [10][30] but they may increase the computational cost and involve new parameters. This confirms the need for complete classifications and comparison of all of these methods to decrease the "analyst time" [43] spent to set up the process. In the next sections, a data and knowledge management approach is proposed to solve this problem. This approach would lead to classify and reuse best practices in a collaborative context. A literature review is presented in the following section about these topics.

3. Data and knowledge management for DoE optimization

Combined with knowledge management methods, Simulation Data Management (SDM) approach may solve this problem and give relevant assistance for the designer from the first stages of the DoE definition. This section presents a literature survey about data and knowledge management tools and methods for simulation support issues.

3.1. Simulation data management

Simulation Lifecycle Management (SLM) strategy [18] is designed to increase the process efficiency. It recommends re-using models and best practices. It also includes standardized work processes, integration with manufacturing operations, and collaborative engineering across the extended enterprise and over the full product lifecycle.

A SDM strategy can support this strategy by capturing all data required for each step of the process. It should also support DoE, optimization and stochastic computations [13][18]. SDM can also manage workflow management and administration support [39][42]. SDM is as a part of SLM [19], which is a specific application of PLM for simulations. SLM manage the process while SDM manage data. SLM covers collaborative product development, data traceability, exchange and reuse, decision-support and simulation systems integration for process automation. An SDM system aims to deal with a growing volume of heterogeneous data in a collaborative context along the product lifecycle [13][39]. Since the DoE process is based on the simulation process and generate a huge volume of heterogeneous data, SDM can be extended to manage it.

For the simulation process, multiple data, as input data (assembly, geometry, parameter, hypothesis, etc.), model data (mesh, solving methods, etc.) and output data (results, representations, reports, etc.) must be managed. These data are now mostly covered by existing data models. Others kinds of data were also taken into account to support multi-domain, multi-component and multi-model simulation processes [3][31][50], simulation loops management [14] and relationships between each product element [35]. The data model specified by the standard "Multidisciplinary analysis and design" ISO 10303–209 edition 2 [33] was specified to manage design data and analysis data in a collaborative context. It covers mechanical parts and assemblies geometric and shape data, associated materials, data related to the product lifecycle, composites structures, mechanical and fluid dynamics analysis and data representations. The links between the initial shape and idealized shapes used for different analysis are supported. The Core Product Model covers functions, forms (geometry and material) and behaviors of product or components [23]. It aims to provide a generic, independent and extensible product model able to capture engineering data. The

Design-analysis Integration data architecture [24] is an extension of CPM which integrates analysis tasks to the product design.

Existing data models and SDM strategy do not cover DoE process. As a part of the simulation process, DoE management can be supported by SDM to structure, capitalize and reuse specific data involved, as DoE definition data, associated simulation model data, collaboration and process data.

3.2. Knowledge-based methods

The definition of knowledge may depend on the context [12]. In a product design context, the knowledge is produced by the "interpretation of information deduced from analysis of data" [12] and can be defined as sharable set of experience and ways of working [44] which can produce additional value [21]. It is important to distinguish data, information and knowledge [2][21]. Data can be reduced to a set of symbol or measures, transformed to information by analysing and organizing data, which produce a real meaning. Knowledge can be classified into several categories. For instance, formal [12] or explicit [21] knowledge is mostly written or expressible while tacit knowledge [1] is entirely owned by each stakeholder. Thus, the first issue is to gather these two kinds of knowledge.

A second challenge concerns the knowledge modeling and representation to reuse and produce new knowledge. Knowledge representations can be classified in multiple categories [13, 49]: pictorial (drawings, etc.), Symbolic (diagrams, ontologies, etc.), linguistics, virtual (numerical models) and algorithm. Knowledge Management consists in capturing, storing, reusing, sharing and creating knowledge to produce added-value [12][44][46]. An interesting approach is the use of ontologies to represent knowledge. 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 [12]. Based on data models, ontologies give sense to data. Concepts are usually defined to build a common taxonomy and improve collaborative works by easily sharing knowledge. [2] proposed an ontology covering requirement engineering, mechanical design and numerical simulation. This ontology supports data capitalization, data re-use and decision-making by representing dynamically relationships between different engineering entities (e.g. link between a specific design and its associated simulations). The ontological approach would be also efficient to represent DoE process knowledge.

In engineering, the variety of viewpoints implies the development of methods and tools promoting knowledge integration upstream the product design process. Experts collaborate among the whole product lifecycle and use different supports to share, exchange and build new knowledge. This aggregation of knowledge highlights the issue of data consistency [38] to prevent conflicts between parameters.

Thus, information and knowledge of a fine granular level such as parameters or constraints should be correctly considered to ensure the success of design process. Indeed, regarding the technical culture of each expertise field, and the variety of software supports, heterogeneous knowledge might be used and represented with different semantics during the same design stage.

DoE is one of representative cases in which knowledge consistency takes great sense. During the DoE working session, experts fulfil different interactions and iterations to identify the best strategy of simulation, taking into consideration all relevant parameters values and points of view of the problem to be solved.

The achievement of a DoE process require the use of a set of tools supporting design, simulation, computing, statistical analysis, computing monitoring, etc. DoE is a collaboration space between actors from different disciplines [3], participating with different roles in DoE session. It concerns also some actors of the engineering activities realized before and after the DoE stage. Furthermore, several tests, choices and decisions are taken during the DoE process throughout lot of iterations and, currently, more than one working session.

Heterogeneous knowledge models are used for this aim. In order to ensure consistency between these knowledge models, meta-models should be proposed within generic semantic and rich representation of concepts and relationships between them [4]. The aim is to support the structure of a common knowledge base along the whole DoE process.

Knowledge-based systems are a specific branch of Artificial Intelligence [37] and are used to solve problems by exploiting capitalized knowledge. An example of knowledge-based system is expert system [37]. Such a system is being developed to manage knowledge between Computer-Aided Design and Computer-Aided Engineering activities [38], better than current PDM and SDM systems can do. A knowledge-based engineering approach was also studied to enrich SDM systems [15]. But none of this works integrated DoE considerations. Next section presents a classification of relevant knowledge produced and shared during a design of experiment process. Then it exposes the proposed knowledge-based system dedicated to DoE optimization and management.

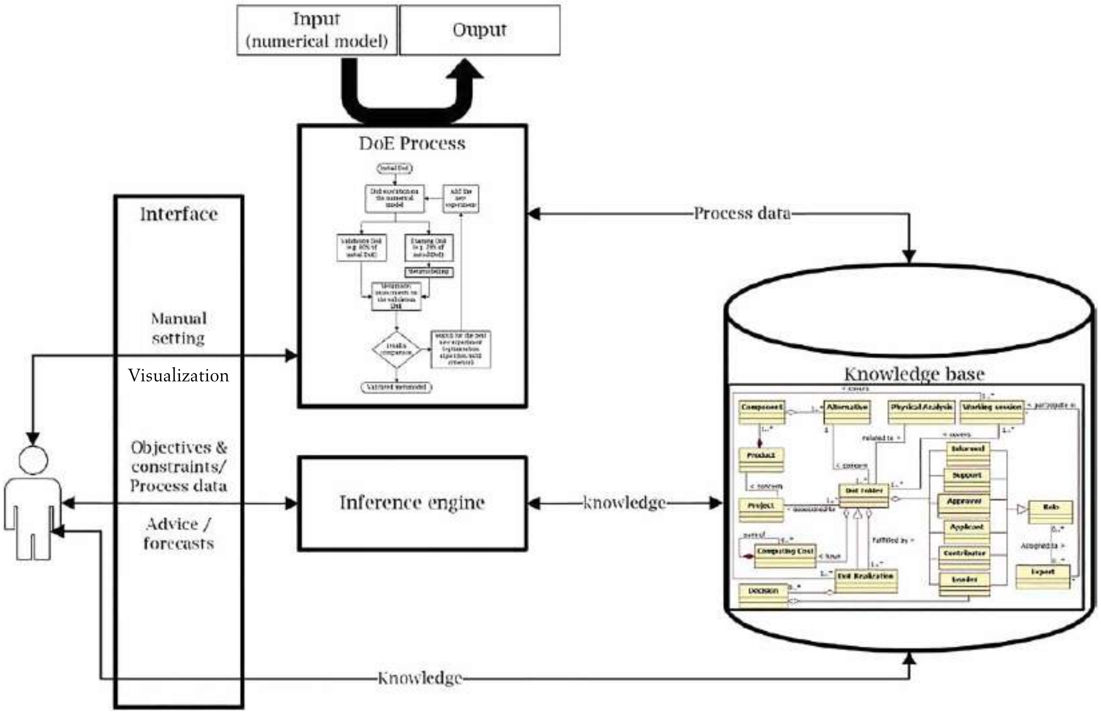

Figure 2. Knowledge-based expert system principles for DoE process.

4. Proposition of Decision-aid system for DoE optimization

As a large volume of heterogeneous data are created during a simulation process in a collaborative context, a SDM approach leads to capitalize and reuse data. But, these data were created and defined by expert rules and know-how of different stakeholders. To go further in data reuse, a simulation knowledge management approach may be set up.

A knowledge-based expert system consists of a knowledge base, an inference engine and a user interface [32][37]. According to its architecture, such a system would be able to deduce a result from input (forward chaining) or to find inputs corresponding to a given results (backward chaining), or even both of them. The inference engine can be based on a set of logical rules, a semantic approach or an ontology, a stochastic approach or a fuzzy logic engine. Thus, the system reuses knowledge to explain results or to give advice.

The DoE process analysis detailed in previous section will be used to build the knowledge base. This base will consist of theoretic and explicit knowledge. Then, new knowledge will be captured by the help of the data model specified in Section 4.

The inference engine would be able to give advice and to forecast the process results. Since many methods involved in the DoE process have a random behavior or are difficult to be tuned up because of their complexity, the expert system would be based on a mix of logical rules and stochastic or fuzzy logic approaches.

Thus, the system would be able to propose several solutions: impossible, possible and certainly feasible solutions which might be executed and probably more efficient. Proposed solutions would be chosen according to objectives and constraints of the study, and resources available.

As shown on Fig. 2, the designer could use this system for several tasks: (1) to obtain advice for a specific problem. The system would propose a successful set of methods and parameters according to previously specified objectives, constraints and resources; (2) to diagnose the efficiency of a user-specified DoE process; (3) to obtain knowledge from the knowledge base, or to add directly knowledge into it. Thus, the definition of parameters and methods used for a specific instance for DoE process would be semi-automatic. However, the possibility to accept, modify or deny the advice or the diagnostic is kept. In this case, the user-interface would allow the designer to define parameters and methods manually.

4.1. *Knowledge base definition*

To implement the proposed knowledge-based solution supporting Design of Experiments optimization, open-source software architecture is under development in the SDM4DOE project to be proposed as a SaaS

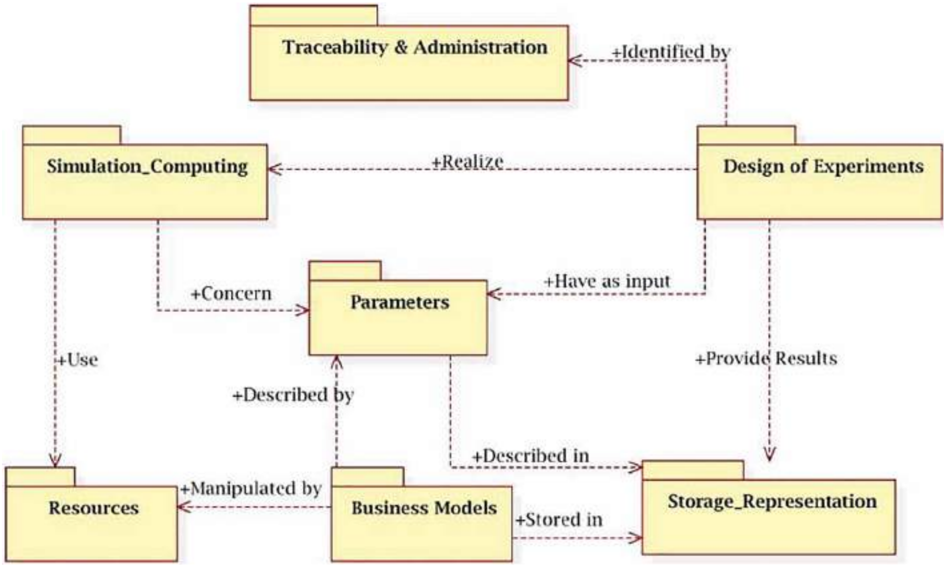

Figure 3. DoE knowledge classification.

framework. In this framework, all knowledge models will be implemented in the knowledge base as a combination of data base, documents and specific files headers with specific parsers to support knowledge retrieval. The SDM4DoE interface includes different functions to support knowledge capitalization, visualization and extraction from the knowledge base.

The framework is based on several components:

1. A user interface integrated in a web browser and supported by the library JQuery UI, which communicates with the web server by https protocol;
2. An Apache web-server to receive tasks from the user;
3. An application-server, Django, receives information from the web-server and determines actions in order to satisfy user's tasks.
4. Softwares used to execute these actions: Alfresco for data classification and retrieval, URANIE and ROOT for statistical computations, Gmsh for finite-element mesh visualization, Code_Aster for finite-element model solving and Slurm to allocate computational resources.

4.2. DoE knowledge classification

Based on the analysis of DoE working session and experts' interview realized during the SDM4DOE project, a first work of clarification of main concepts used by experts is done. The classification of this knowledge is then achieved according to the role of each kind of knowledge in the DoE process. By means of UML package diagram [8], figure 3 gives an overview of knowledge and data used during DoE. Seven categories are identified.

- "Design of Experiments" concept: this first category of data describes the global the DoE and its main properties. It concerns, for example: the goal, the DoE type (full, fractional, etc.) and the concerned physical analysis in the DoE (electro-mechanic, fluid dynamic, etc.)
- "Traceability and administrative data": links the DoE folder to the global environment in which this DoE is created. This environment is defined by: the project reference, the related product and/or parts, real working sessions during the DoE, involved actors and their roles, traceability of choices and decisions, etc.
- "Parameters" package: is the central node of the global DoE data model. This concept handles the classification of all DoE involved parameters regarding their business nature (geometry, mechanic constraints, etc.), position for each related DoE steps (Input/output or support), nominal and interval of admissible variation of value, etc. This concept proposes great advantages for the definition of standard semantic as a common codification for all business fields involved in the DoE process.
- "Business Models" package: Based on the concept of product model, this package of data can improve a consistent management of all business models (CAD model, FEM, other simulations, ANOVA, etc.). The standard parameters are linked to the unified business models.

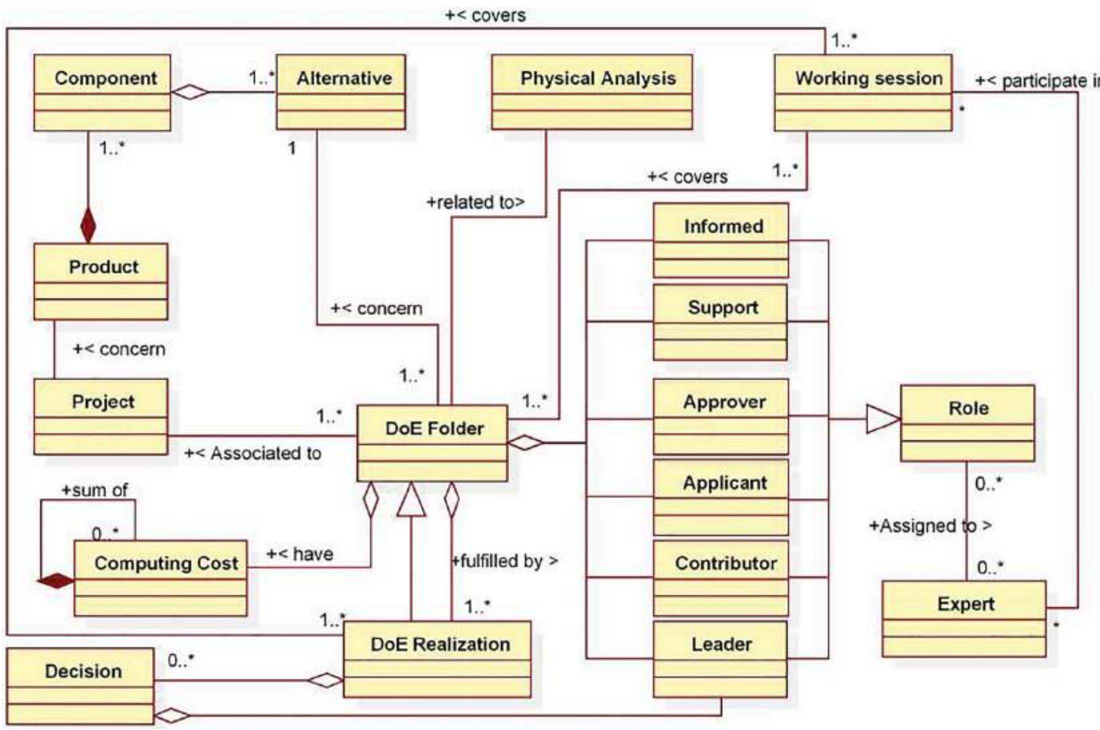

Figure 4. The DoE folder.

- "Simulation and computing" package: includes all analysis methods and simulation-computing scripts available for the current DoE. This may be useful to help actors to find the best method or to use the robust script according their type of problem and the nature of physical analysis. For more traceability, this category of data is linked to all concerned parameters.
- "Resources" package: as a complement of the precedent data package, this category aims to classify the different computing software and clusters for each kind of computing. It will include additional properties such as treatment frequency, resources consumption, cost, etc. the implementation of such model in decision support for DoE can help actors the rapid definition of the DoE strategy face to several computing alternatives.
- "Storage and representation" package: This category of data supports the classification of representation format of parameters. The data of the different DoE used are currently described on different formats and stored in different files. According to the business tool used for the creation of this data, its position may be different. Then, the implementation of such category of data help computing clusters to rapidly find data sources and to use the appropriate parsing method to improve computing efficiency.

4.3. Example: the DoE Folder

Figure 4 illustrates the main concepts required for the representation of all administrative data for traceability issues. The concept of "DoE folder" is the classifier of all DoE realizations according to one alternative solution of the product (or component) and one physical analysis. Each DoE realization consists on the concrete selection of factors and their levels, but also the execution of all simulations required by the selected physical analysis.

The DoE folder is associated to a specific project and it is the space of interaction of several experts. Each expert takes one or more roles in the DoE process. These roles can be changed during the real physical meetings. These meetings are represented by the class "working session", in which one or more DoE folders and/or realization might be occurred.

For traceability and future reuse perspective, the different decisions taken during concrete realization are stored in the related classes. The total computing costs of different realizations is described in the DoE folder as an indicator for future choices of simulation methods and scripts in similar situations.

The main stage of DoE folder definition consists on the selection of factors from the business model (FEM file). Then, the concrete realization of the DoE will start by the definition of variables inputs as instances of a sub

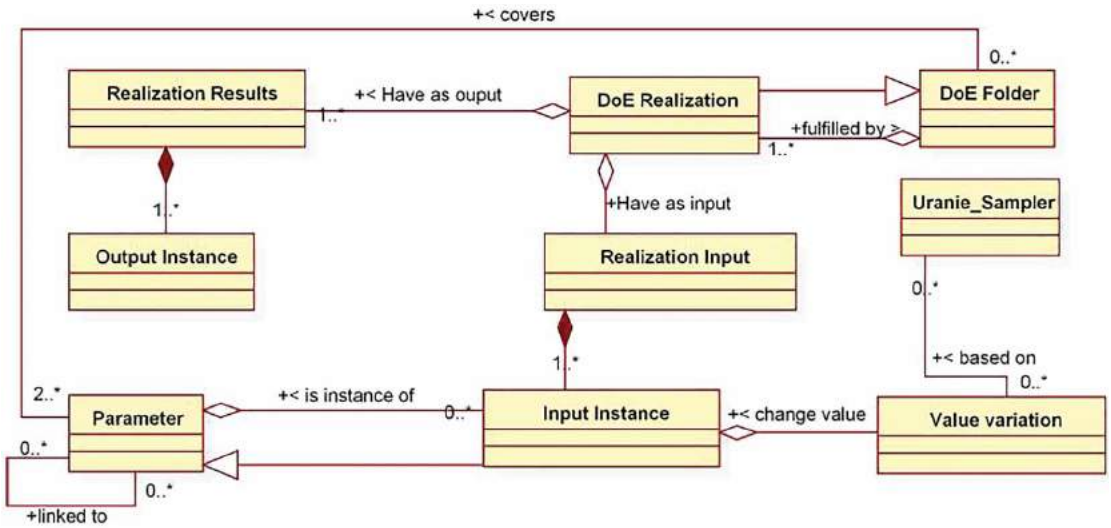

Figure 5. The DoE Realization.

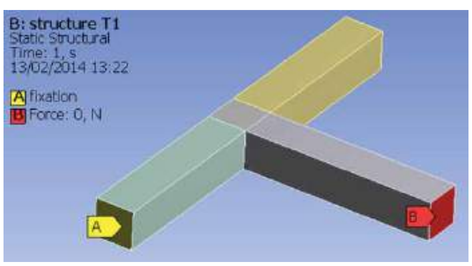

Figure 6. T-shape beam finite element model.

set of the selected parameter (Fig. 5). The value variation of these instances could be obtained through fixed number, a mathematical formula or other random sampling generated by the Sampler module of Uranie software. The result of a realization is obtained as a set of output instances after a set of simulations and mathematical treatments (optimization, ANOVA . . .).

4.4. Application

As an example, a first test was done with a very simple numerical model (Fig. 6). This model is a T-shape beam of an elastic isotropic material. It models the linear elastic behavior of the structure, submitted to a tri-axial constant load (B) and with two clamped tips (A). Strain and stresses are computed by finite-element method, with a validated mesh. Three factors, the Young's Modulus and 2 components of the load vector, were selected after a first sensitivity analysis, in order to run a DoE and compute a metamodel on the strain.

The designer has to define the DoE type and levels for selected factors, the metamodel type, associated parameters and sub-methods. He has to choose the validation strategy to evaluate the predictivity of the metamodel. For example, the designer may choose a DoE based on low-discrepancy sequence (Halton design) with 100 experiments, a kriging metamodel with an exponential variogram type, trained on 30 experiments and validated on remaining experiments (70). This definition may be efficient regarding to the objective. Each experiment has a low computational cost, the metamodel is fast to be trained and compute an uncertainty associated to the result. But other strategies would be more efficient, and several manual attempts would be done before finding an optimal strategy. The knowledge-based system will capitalize these attempts to be used for decision-aid.

To achieve this process, the user fulfils a set of interactions with the knowledge-based framework to design the best DoE process regarding his problem. Fig. 7: illustrates the global scenario of the expected DoE process. At the beginning, by creating new DoE project in the framework interface, the user introduces all relevant data describing his problem to be studied. It concerns all administrative aspects, parameters, and objectives of the project. A research mechanism is then launched in the knowledge base to find similar case studies regarding the inputs of the user. The similar cases studies are analysed by the inference engine and a set of acceptable methods, properties and parameters are displayed to the user as a decision support. The user can select the best combination of methods and properties but also defines the factors and levels. He can also decide to consider other alternative methods types or properties and in this case, the inference engine will evaluates this new combination.

After the validation of all methods properties, the user executes the DoE by selecting the available computing clusters. The interface run the related programs and

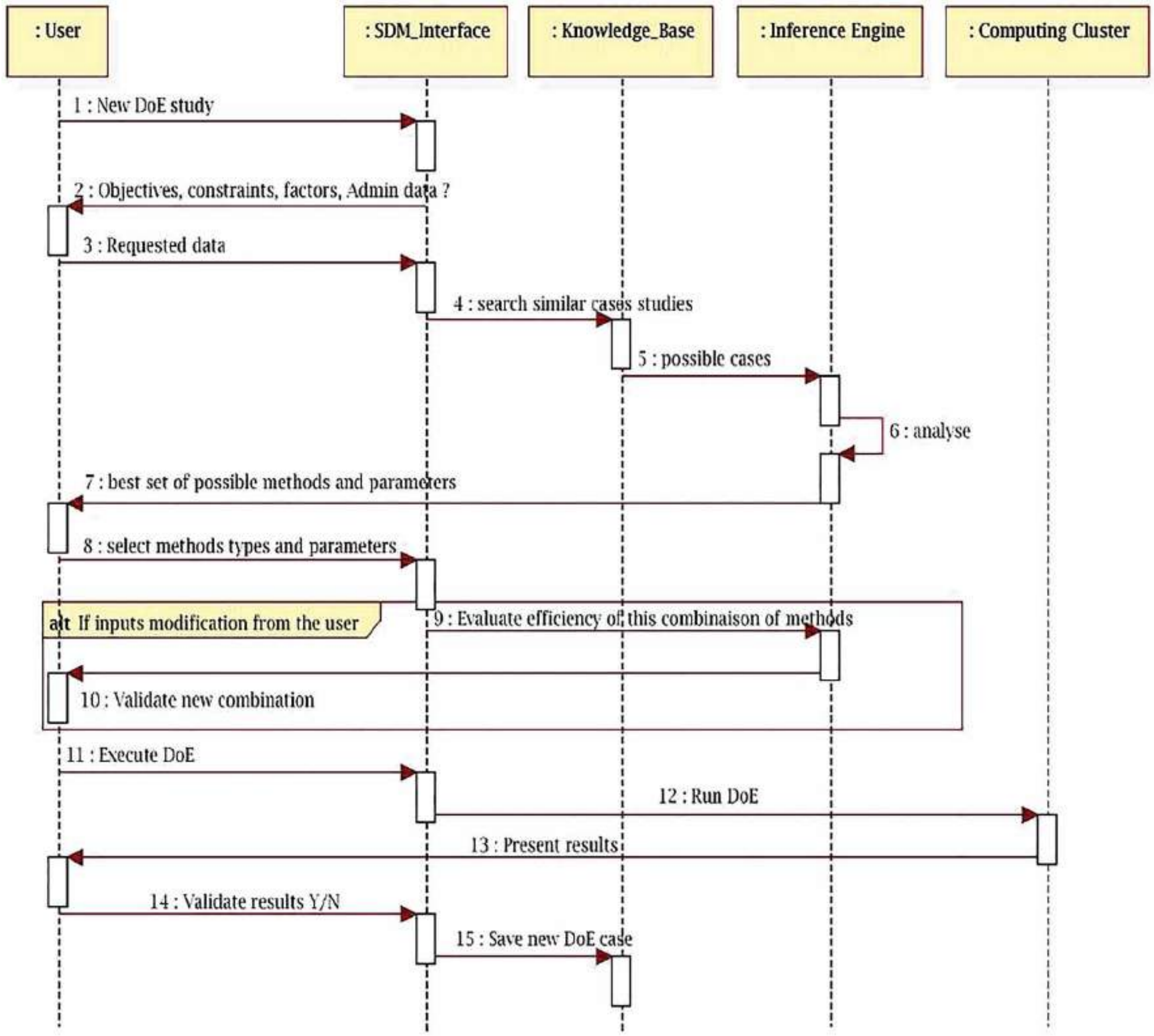

Figure 7. Scenario of knowledge reuse for DoE.

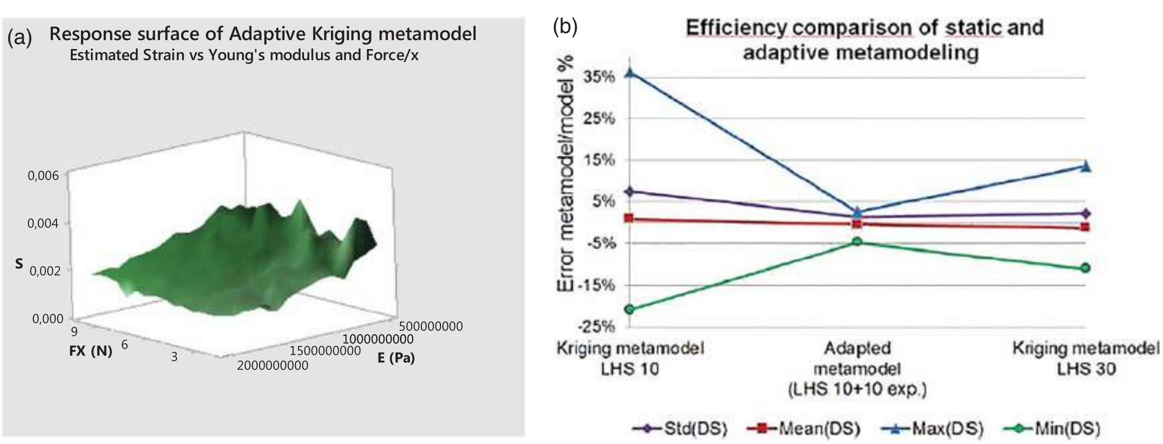

Figure 8. Final metamodel obtained: (a) corresponding response surface, (b) computational cost saved by adaptive strategy.

manages its execution in different clusters. At the end, the results are displayed and stored in the knowledge base with the user report including his evaluation of the results and the decisions regarding it.

The main result of the DoE process is a metamodel that can be reused by designers for future studies (Fig. 8). This metamodel was produced rapidly by an adaptive strategy. Beginning from 10 experiments, it was sufficient to add 10 experiments iteratively to reach better accuracy than a metamodel trained from 30 experiments. The preparation time of this process was also minimized by using the knowledge -based system. The knowledge base was enriched by this application and further application will be defined faster.

5. Conclusion

Although the product quality can be considerably increased thanks to the application of a DoE process, the cost and time involved may slow the product development process. The state of the art and the experience shared by experts in the project shows a complexity and diversity of issues, on computation methods, statistical processing and also data management. Despite the central role of DoE in the overall design process, methodological developments and softwares are limited compared to the complexity of DoE process. The potential gains of metamodeling and adaptive DoE methods are reduced by preparation time. It becomes necessary to support the use of such methods by a data model, leading to a decision support system. The knowledge-based system proposed in this paper would alleviate this main drawback by giving advice or forecasts to the designers. The capitalization of knowledge is ensured by a specific knowledge representation. By the mean of such a system, knowledge management would be guaranteed all along the product life-cycle, the DoE process would be shorten and new methodologies might be explored. This framework will be validated with two industrial cases: a nuclear power plant subjected to an earthquake, and a dynamic analysis of an engine support. These two cases will involve complex analysis with high dimensional problems, in order to validate the efficiency of this strategy to shorten the DoE process.

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

ORCID

Gaëtan Blondet http://orcid.org/0000-0003-3225-0277
Farouk Belkadi http://orcid.org/0000-0003-2783-4147
Julien Le Duigou http://orcid.org/0000-0001-8723-2506

References

[1] Arduin, P.-E.; Le Duigou, J.; Abel, M.-H.; Eynard, B.: Knowledge Sharing in Design Based on Product Lifecycle Management System, ICoRD'15–Research into Design Across Boundaries Volume 2, Bangalore, India, 2015, 507–17.

[2] Assouroko, I.; Ducellier, G.; Boutinaud, P.; Eynard, B.: Knowledge management and reuse in collaborative product development – a semantic relationship management-based approach, International Journal of Product Lifecycle Management, 7(1), 2014, 54–74. http://dx.doi.org/10.1504/IJPLM.2014.065460

[3] Assouroko, I.; Ducellier, G.; Eynard, B.; Boutinaud, P.: Semantic-based approach for the integration of product design and numerical simulation, International Conference on Product Lifecycle Management, Eindhoven, NL, Inderscience Enterprises Ltd., 2011.

[4] Belkadi, F.; Dremont, N.; Notin, A.; Troussier, N.; Messaadia, M.: A meta-modeling framework for knowledge consistency in collaborative design, Annual Reviews in Control, 36(2), 2012, 346–58. http://dx.doi.org/10.1016/j.arcontrol.2012.09.016

[5] Benhafid, Y.; Troussier, N.; Boudaoud, N.; Cherfi, Z.: Méthode d'aide à l'idéalisation de modèles issus de la CAO pour le calcul de structures, Mechanics & Industry, 6(3), 2005, 289–95. http://dx.doi.org/10.1051/meca:2005032

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

[7] 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.

[8] Booch, G.; Rumbaugh, J.; Jacobson, I.: The Unified Modeling Language user guide, Addison Wesley Longman Publishing Co., Inc. Redwood City, CA, USA, 1999.

[9] Bosch-Mauchand, M.; Belkadi, F.; Bricogne, M.; Eynard, B.: Knowledge-based assessment of manufacturing process performance: integration of product lifecycle management and value-chain simulation approaches, International Journal of Computer, Integrated Manufacturing, 26(5), 2013, 453–73. http://dx.doi.org/10.1080/0951192X.2012.731611.

[10] 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.

[11] 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.

[12] 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.

[13] Chari, S.: Addressing Engineering Simulation Data Management (SDM) Challenges, 2013, 1–14.

[14] 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.

[15] 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, 1719–1730.
[16] 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.
[17] CIMDATA: Beyond Simulation Data Management, 2012.
[18] CIMDATA: Simulation & Analysis Governance, 2014.
[19] CIMDATA: Simulation Lifecycle Management, 2011.
[20] Dhuieb, M.A.; Belkadi, F.; Laroche, F.; Bernard, A.: Interoperability Framework for Supporting Information-Based Assistance in the Factory, IFIP 11t, International Conference on Product Lifecycle Management PLM'2014, Japan, 2014.
[21] Ermine, J.: Management et ingénierie des connaissances. Modèles et méthodes, 2014.
[22] Feng, G.; Cui, D.; Wang, C.; Yu, J.: Integrated data management in complex product collaborative design, Computers in Industry, 60(1), 2009, 48–63. http://dx.doi.org/10.1016/j.compind.2008.09.006.
[23] Fenves, S.J.; Foufou, S.; Bock, C.; Sriram, R.D.: CPM: A Core Model for Product Data, National, Institute of Standards and Technology (NIST), 2006.
[24] Fenves, S.J.; Sriram, R.D.; Choi, Y.; Gurumoorthy, B.; Mocko, G.: Master Product Model for the Support of Tighter, Integration of Spatial and Functional Design, NIST, 2003, 31.
[25] 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.
[26] 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.
[27] Gazut, S.: Conception et mise en œuvre de nouvelles méthodes d'élaboration de plans d'expériences pour l'apprentissage de modèles non linéaires., Ph.D. Thesis, Université Paris-Sud XI, 2007.
[28] Gazut, S.: Conception et mise en œuvre de nouvelles méthodes d'élaboration de plans d'expériences pour l'apprentissage de modèles non linéaires., Ph.D. Thesis, Université Paris-Sud XI, 2007.
[29] 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.
[30] Gorissen, D.; Dhaene, T.; Demeester, P.; Couckuyt, I.; Crombecq, K.: A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based Design, Journal of Machine Learning Research, 11, 2010, 2051–5.
[31] Graignic, P.; Vosgien, T.; Jankovic, M.; Tuloup, V.; Berquet, J.; Troussier, N.: Complex System Simulation: Proposition of a MBSE Framework for Design-Analysis, Integration, Systems Engineering Research (CSER'13), vol. 16, Atlanta, USA, Elsevier B.V., 2013, 59–68. http://dx.doi.org/10.1016/j.procs.2013.01.007.
[32] Ikram, A.; Qamar, U.: Developing an expert system based on association rules and predicate logic for earthquake prediction, Knowledge-Based Systems, 75, 2015, 87–103. http://dx.doi.org/10.1016/j.knosys.2014.11.024.
[33] ISO: 10303-209:2014, Industrial Automation Systems and Integration – Product Data Representation and Exchange – Part 209: Application Protocol: Multidisciplinary Analysis and Design, 2014.
[34] Jones D.R.: A Taxonomy of Global Optimization Methods Based on Response Surfaces, Journal of Global Optimization, 21(4), 2001, 39. http://dx.doi.org/10.1023/A:1012771025575.
[35] Kibamba, Y.: Spécification et développement d'un environnement collaboratif du cycle de vie des données de simulation numérique, Ph.D. Thesis, Université de Technologie de Compiègne, 2011.
[36] Kibamba, Y.; Eynard, B.; Troussier, N.; Derigent, W.: PLM – Digital Engineering and Numerical Simulation for Collaborative Product Development, International Conference on Research into Design, ICoRD'09, Bangalore, India, 2009, 216–223.
[37] Kiritsis, D.: A review of knowledge-based expert systems for process planning. Methods and problems, The International Journal of Advanced Manufacturing Technology, 10, 1995, 240–262. http://dx.doi.org/10.1007/BF01186876.
[38] Monticolo, D.; Badin, J.; Gomes, S.; Bonjour, E.; Chamoret, D.: A meta-model for knowledge configuration management to support collaborative engineering, Computers in Industry, 66, 2014, 11–20. http://dx.doi.org/10.1016/j.compind.2014.08.001.
[39] Norris, M.: Business Value from Simulation Data Management-a Decade of Production Experience, NAFEMS, 2012.
[40] Owen, R.; Horváth, I.: Towards Product-Related Knowledge Asset Warehousing in Enterprises, the Fourth International Symposium on Tools and Methods of Competitive Engineering, 2002, 155–170.
[41] 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.
[42] ProSTEP: Automotive CAE Integration. Annex: Simulation Data Management, 2010.
[43] Razavi, S.; Tolson, B. a.; Burn, D.H.: Numerical assessment of metamodeling strategies in computationally intensive optimization, Environmental Modelling & Software, 34, 2012, 67–86. http://dx.doi.org/10.1016/j.envsoft.2011.09.010.
[44] Sainter, P.; Oldham, K.; Larkin, A.; Murton, A.; Brimble, R.: Product Knowledge Management within Knowledge-Based Engineering Systems, Design Engineering Technical Conference, Baltimore, Maryland, 2000.
[45] Sanchez, S.M.; Wan, H.: Work smarter, not harder: a tutorial on designing and conducting simulation Experiments, Winter Simulation Conference, Berlin, Germany, IEEE, 2012, 1929–1943.
[46] Serban, A.M.; Luan, J.: Overview of Knowledge Management, New Directions for Institutional Research, 113, 2004, 5–16. http://dx.doi.org/10.1002/ir.34.
[47] 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, 129–150. http://dx.doi.org/10.1007/PL00007198

[48] Stark, J.: Product lifecycle management, 21st century paradigm for product realisation, Springer, 2004.

[49] Wang, G.G.; Shan, S.: Review of Metamodeling Techniques in Support of Engineering Design Optimization, Journal of Mechanical Design, 2006, 415–26. http://dx.doi.org/10.1115/DETC2006-99412.

[50] Wang, W.; Mo, R.; Jin, Y.T.; Zhang, Y.: Research on Multi-dimensional Simulation Data Management Model for Complicated Product, International Conference on Electrical and Control Engineering, Wuhan, China, IEEE, 2010, 2948–51. http://dx.doi.org/10.1109/iCECE.2010.719.