INSTITUT DE FRANCE
Académie des sciences

Comptes Rendus
Mécanique

Francisco Chinesta and Elias Cueto

Conciliating accuracy and efficiency to empower engineering based on performance: a short journey

Volume 351, Special Issue S3 (2023), p. 121-133

Online since: 22 June 2023
Issue date: 26 August 2024

Part of Special Issue: The French "Année de la Mécanique": some views on recent advances in solid and fluid mechanics

Guest editors: Francisco Chinesta (PIMM, UMR CNRS 8006, Arts et Métiers Institute of Technology, Paris, France) and Aziz Hamdouni (LaSIE, UMR CNRS 7356, La Rochelle Université, France)

https://doi.org/10.5802/crmeca.188

The Comptes Rendus. Mécanique *are a member of the Mersenne Center for open scientific publishing*
www.centre-mersenne.org — e-ISSN : 1873-7234

Comptes Rendus
Mécanique
2023, Vol. 351, Special Issue S3, p. 121-133
https://doi.org/10.5802/crmeca.188

Research article / *Article de recherche*

The French "Année de la Mécanique": some views on recent advances in solid and fluid mechanics / *L'année de la mécanique : quelques points de vue sur des avancées récentes en mécanique des solides et des fluides*

Conciliating accuracy and efficiency to empower engineering based on performance: a short journey

Francisco Chinesta [*, a, b] **and Elias Cueto** [c]

[a] PIMM lab, Arts et Metiers Institute of Technology, 151 Boulevard de Hôpital, 75013 Paris, France

[b] CNRS@CREATE LTD, 1 Create Way, 08-01 CREATE Tower, Singapore 138602

[c] Aragon Institute of Engineering Research, Universidad de Zaragoza, Maria de Luna s/n, 50018 Zaragoza, Spain

E-mails: francisco.chinesta@ensam.eu (F. Chinesta), ecueto@unizar.es (E. Cueto)

Abstract. This paper revisits the different arts of engineering. The *art of modeling* for describing the behavior of complex systems from the solution of partial differential equations that are expected to govern their responses. Then, the *art of simulation* concerns the ability of solving these complex mathematical objects expected to describe the physical reality as accurately as possible (accuracy with respect to the exact solution of the models) and as fast as possible. Finally, the *art of decision making* needs to ensure accurate and fast predictions for efficient diagnosis and prognosis. For that purpose physics-informed digital twins (also known as Hybrid Twins) will be employed, allying real-time physics (where complex models are solved by using advanced model order reduction techniques) and physics-informed data-driven models for filling the gap between the reality and the physics-based model predictions. The use of physics-aware data-driven models in tandem with physics-based reduced order models allows us to predict very fast without compromising accuracy. This is compulsory for diagnosis and prognosis purposes.

Keywords. Physics-based modeling, Machine learning, Artificial Intelligence, Data-driven modeling, Model Order Reduction, POD, PGD, Virtual, Digital and Hybrid Twins.

Funding. F. C. is supported by the CREATE-ID ESI-ENSAM research chair. E. C. is supported by ESI Group through the Simulated Reality project, UZ-2019-0060. This work has been partially funded by the Spanish Ministry of Science and Innovation, AEI /10.13039/501100011033, through Grant number PID2020-113463RB-C31 and the Regional Government of Aragon and the European Social Fund, group T24-20R..

Manuscript received 17 August 2022, accepted 28 March 2023.

* Corresponding author

1. The whole picture

At the beginning was science. The so-called modern science, for some historians of science Galileo's heritage, would have to wait a bit more. From the very beginning, the human being created technology to master the natural environment and take profit of it for constructing tools that facilitated the everyday life. Later, was the time to surpass the status quo, to break limits, to make non-natural things—like flying for instance—and engineering became a major protagonist, looking for improving performances. Thus, the dream of Leonardo was flying, the wings were there for making flying possible, or simpler. In many cases the application preceded the scientific understanding.

As Theodore Von Karman said, scientists study the world as it is, engineers create the world that never has been. The first industrial revolution arose from steam engines and transformed industry. Later on, electricity gave rise to the second revolution, the one enabling producing more and faster, and this one not only transformed industry but also the society.

Alessandro Volta was invited two centuries ago at the French academy of sciences in Paris to present his battery. He exhibited the action of applying the electrodes of his battery to a frog. All the attendants were very impressed by the frog jumps. After that exciting session, the legend says that Napoleon asked: Alessandro, do you think that one day, this thing that you call electricity, could serve to something else that making jumps to the frogs? Few generations later, can you imagine one single day of your *everyday* life without electricity?

Later, electronics entered the scene and, with it, automation. Welcome to the third industrial revolution that enabled not only making faster, but making better. Despite all these advances, engineering remained product-based, despite of the fact that the society is looking for performance. The product is a simple way to access to this performance. When we buy an electric drill, we are in fact trying to buy a good quality hole. However, engineering remained product-oriented. Why?

Certainty, due to the fact that even if physics-based models are very rich (in terms of physics), it was difficult to both (i) address the product in its environment, usually too large, too complex, too uncertain, too fluctuating...; and (ii) solve them under stringent real-time constraints, needed for optimal decision in operation, the ultimate goal of engineers and engineering.

2. The three arts of the engineering

The just referred challenges were addressed by empowering three arts of engineering.

2.1. *The art of modeling*

Science begun by empirical observation, something that is often called the first paradigm of science. This is science in the times of Tycho Brahe. But science soon embraced—in fact, with Brahe's disciple, Kepler—a new paradigm: that of scientific theory. The *art of modeling* concerned all the physics throughout all the description scales of materials, processes and structures, a rich and holistic physics-based approach. This *art* allowed us to make accurate predictions concerning materials, manufacturing processes and structural analysis, even in extreme conditions and with the computational facilities available 45 years ago. These were in some cases smaller than the one we have in the smartphone placed in our pocket.

Despite of the encountered computing limitations it was possible for instance, in 1985, to simulate a crash test (involving fast-dynamics, multi-contact, plastic deformation, damage and rupture, the worst among all the imaginable behaviors), as sketched in Figure 1. This was a major accomplishment in computer simulation that can be found in the Computer History Museum at San Jose (California) [1].

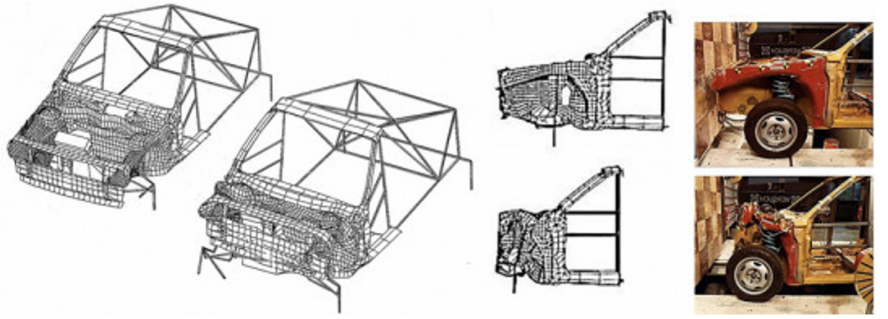

Figure 1. Pioneering crash analyses. Picture courtesy of ESI Group.

However, these accomplishments remained limited by the available computational resources as well as by the characteristic time of calculation and response. Replacing product management by the management of performance needs for faster predictions, for faster and more optimal designs, for faster and better decisions.

2.2. *The art of simulation*

Thus, the *art of modeling* was enriched with a second art, the *art of simulation*—the third paradigm of science—that can be summarized in a single sentence: the ability of applying the best numerical technique to a given problem, ensuring the best accuracy and efficiency. The *art of simulation* is also a long history, more than 50 years old, that began when computers irrupted and replaced the human analogical calculation by its radically new, and much faster, digital counterpart. But the new engineering based on performance, at the beginning of the third millennium, needs to proceed faster, in real-time. Sometimes even faster.

It was at the end of the XX century, that new technologies were proposed and implemented, that without breaking down its physics-based rich descriptions, allowed to speed-up their predictive capacities, i.e., the solution of the complex models that such rich engineering descriptions imply. New capabilities were possible by involving the use of Model Order Reduction (MOR).

To introduce the main idea behind MOR, we consider the picture of a very touristic monument on Paris, the one depicted in Figure 2. When showing this image to our Parisian colleagues, asking them about what monument it is, the response is immediate: it is the *Pantheon*. Even if we are very happy with that response, our next question is, after removing the picture from their eyes: how many columns are placed in its front? There, the responses are quite diverse, 5 for some of the interrogated people, 6 for others, some indicate seven...The surprise arrives when our colleagues are informed on the exact number of columns: just 22. Every response was very far from the exact value.

What does this experiment prove? It proves the fact that we can be sure on something even when we absolutely ignore the details. Thus, MOR could be defined as the mathematical technique able to extract essential features while ignoring accessorial details. Obviously, this definition entails the dependence on the definition of essential and accessorial. For a tourist or a computational algorithm for recognizing monuments from pictures, the essential is recognizing the monument, whereas for the enterprise in charge of the columns maintenance the number of columns seems crucial! Thus, models can and should be reduced with respect to a given goal, and the validity of reduced models depend on the circumstances in which they were learned.

For describing MOR in a more proper way, we consider a field of interest $u(x,t)$, with the quantity $u(\bullet)$ being a scalar, a vector or a tensor field, that depends on the physical space,

Figure 2. A highly visited monument in Paris.

represented by the space coordinates $x \in \Omega \subset \mathbb{R}^{\mathtt{D}}$ ($\mathtt{D} = 3$ when operating in the three-dimensional space) and time $t \in \mathcal{T} \subset \mathbb{R}$.

For computing numerically such a rich field, the first step consists in reducing the number of unknowns—a step called discretization—that approximates the solution from the value of the field at different locations $x_i,\ i = 1, \ldots, \mathtt{N}$, (the so-called nodes in finite elements notation) and at different time instants $t_m,\ m = 1, \ldots, \mathtt{M}$, according to the general form

$$u(x, t_m) \approx \sum_{i=1}^{\mathtt{N}} U_i^m N_i(x), \tag{1}$$

where when considering an interpolative approximation, i.e., $N_i(x_j) = \delta_{ij}$, with δ the Kronecker delta, $U_i^m = u(x_i, t_m)$. The previous expression represents the usual approximation employed in the Finite Elements Method (FEM), where the approximation functions $N_i(x)$ are called *shape functions*.

For computing the $\mathtt{N}$ unknown values of the field at each time instant t_m, an equivalent number of equations is required. For that, physics offers in general the model to which the evolution (in space and time) of the field under consideration is subjected, represented in its most general form as

$$\mathscr{L}(u(x, t); \mu) = F(x, t; \mu), \text{ in } \Omega_\mu, \tag{2}$$

where $\mathscr{L}(\bullet)$ represents a generic linear or nonlinear differential operator, $F(\bullet)$ is the so-called forcing term, and μ a set of parameters. Some of these parameters could affect the domain in which the problem is defined Ω_μ (the so-called geometrical parameters), some of them could affect the physical model itself and others the forcing term. Thus, it is expected that the solution will depend on the choice made for the different parameters grouped in vector μ, i.e., $u(x, t; \mu)$.

In general, time derivatives involved in the differential operator are approximated by using adequate finite difference schemes, taking profit of the causality that allows computing the present from the past. Thus, the general physics-based model, at time t_m, reads

$$\widetilde{\mathscr{L}}(u(x, t_m); \mu) = \widetilde{F}(x, t_m; \mu), \text{ in } \Omega_\mu,\ t = t_m, \tag{3}$$

that involves the solution at the present time, i.e. $u(x, t_m; \mu)$ that by using the approximation (1) consists of $\mathtt{N}$ unknowns, the $\mathtt{N}$ nodal values of the field at the present time t_m and for the

considered set of parameters μ, i.e. $U_i^m(\mu)$, $i = 1, \ldots, \mathtt{N}$. In the previous equation, $\widetilde{\bullet}$ represents the semi-discretized differential form.

Now, if we define the model residual at present time t_m and for the given choice of the parameters μ, as

$$\mathscr{R}_\mu(u(x,t_m)) = \widetilde{\mathscr{L}}(u(x,t_m);\mu) - \widetilde{F}(x,t_m;\mu), \tag{4}$$

to enforce its nullity, the weighted residual formulation proceeds by enforcing, for any test function $u^*(x,t_m)$ in an appropriate functional space, a vanishing integral

$$\int_{\Omega_\mu} u^*(x,t_m)\, \mathscr{R}_\mu(u(x,t_m))\, dx = 0. \tag{5}$$

Within the Galerkin setting, the test functions are approximated from

$$u^*(x,t_m) \approx \sum_{i=1}^{\mathtt{N}} U_i^* N_i(x), \tag{6}$$

that introduced into the weighted residual formulation Eq. (5) and taking into account the arbitrariness of the $\mathtt{N}$ coefficients U_i^* allows (after some technical manipulations: integration by parts, domain partition into elements, numerical integration, enforcement of essential boundary conditions...) to the algebraic system

$$\mathbb{K}\mathbf{U}^m = \mathbf{F}, \tag{7}$$

where matrix $\mathbb{K}$ could depend on the solution itself $\mathbf{U}^m$, in the case of nonlinear modes, on time, and also on the parameters choice μ. The forcing term $\mathbf{F}$ will depend, in turn, on present and/or past time instants, as well as on the parameters choice. All these dependences are not explicitly indicated for the sake of notational simplicity.

Even if the reduction accomplished is unimaginable, a *double infinity* of points and time instants, condensed into $\mathtt{N}\times\mathtt{M}$ discrete values (nodes and time instants), sometimes the complexity of the involved physics and the associated problem solutions, needs for several millions of nodes where then solution must be computed every mili-second, or even micro-second, as it is the case when analyzing the crash test before referred.

Solving too many algebraic systems, each one containing millions of rows and columns, is feasible with the nowadays computational resources. However, fulfilling the stringent real-time constraints while keeping reasonable the computational resources (if simulation is expected be used in small and medium enterprises, as a consequence of democratizing simulation) is far to be obvious. Despite of the impressive reduction just mentioned, that leaded to the algebraic system (7), its repeated solution continues being, very often, too expensive. MOR could alleviate these difficulties.

Imagine for a while, and without loss of generality, a steady-state linear problem, whose discretized form reads

$$\mathbb{K}\mathbf{U} = \mathbf{F}. \tag{8}$$

This discrete model is able to compute the response $\mathbf{U}$ for any choice of the forcing term $\mathbf{F}$, that is, $\mathbf{U}^1$ results from the forcing $\mathbf{F}^1 = (1,0,\ldots,0)^T$, $\mathbf{U}^2$ from $\mathbf{F}^2 = (0,1,0,\ldots,0)^T$ and so on, until the response $\mathbf{U}^{\mathtt{N}}$ related to $\mathbf{F}^{\mathtt{N}} = (0,0,\ldots,1)^T$. Because of the linearity, superposition applies and the response calculation to any forcing spanned into the canonical basis just considered, is straightforward.

However, in general, only a region of this impressively large domain $\mathbb{R}^{\mathtt{N}}$ of all the possible responses is, and will be, explored by a system during its life. Even the human being, each one of us, and despite the impressive freedom that we have, we explore only a small region of the domain containing all the possibilities open to us.

Thus, one could expect the response $\mathbf{U}$ expressible in a smaller approximation basis, instead of the one of dimension $\mathtt{N}$ just considered. If a reduced approximation basis exists, i.e., all

the expected solutions $\mathbf{U}$ could be approximated by using a reduced basis of size $\mathtt{R}$, i.e., $\mathbf{U} = \mathtt{span}\{\mathbf{R}_1, \ldots, \mathbf{R}_\mathtt{R}\}$, with $\mathtt{R} \ll \mathtt{N}$, then by putting the different vectors composing the reduced basis into the columns of matrix $\mathbb{B}$, i.e. $\mathbb{B} = (\mathbf{R}_1 \ldots \mathbf{R}_\mathtt{R})$, one can write $\mathbf{U} = \mathbb{B}\mathbf{V}$, where $\mathbf{V}$ is a $\mathtt{R} \times 1$ vector, which component V_k represents the contribution of the approximation function $\mathbf{R}_k$ for representing $\mathbf{U}$.

Thus, the algebraic system (9) can be rewritten in its reduced form

$$\mathbb{K}\mathbb{B}\mathbf{V} = \mathbf{F}, \tag{9}$$

that,projected into the reduced basis (Galerkin projection), becomes

$$\mathbb{B}^T\mathbb{K}\mathbb{B}\mathbf{V} = \mathbb{B}^T\mathbf{F}, \tag{10}$$

which defines an algebraic system of size $\mathtt{R} \times \mathtt{R}$.

The only remaining issue is the procedure to be employed for discovering that reduced basis able to span the problem solution. The construction of this basis is based on an appropriate learning procedure. For that purpose, manifold learning can be successfully applied for extracting the so-called slow manifold where the problem solution lives. Principal Component Analysis (PCA) performs well in the case of linear manifolds, whereas nonlinear manifold learning performs very well when the solution manifold exhibits a noticeable nonlinearity [2–4].

When the reduced basis is extracted by using PCA, the projected algebraic system, Eq. (10) represents the so-called Proper Orthogonal Decomposition (POD) based model order reduction, widely employed in many domains of science and engineering [5]. The main issues remain the ones related to the reduced basis generality and adaptation, the error quantification, the sampling strategy for extracting the reduced basis (some improvements were proposed in the context of the so-called Reduced Basis method [5–9]), addressing multi-parametric models and the efficient treatment of nonlinear models that needs advanced techniques (hyper-reduction [10, 11], Empirical Interpolation Methods (EIM) [12] and its discrete counterpart [13], cross approximations [14], among many other alternatives).

Until now, and despite of the fact of grouping all these techniques under the denomination of *model order reduction*, the subjacent physical models were never reduced. All them come from the considered physics, with all their richness encapsulated into the matrix $\mathbb{K}$, resulting from a standard discretization. The only thing that was reduced was the basis in which the response $\mathbf{U}$ is expressed, as a consequence of the reduced nature of the forcing that remains constrained into a certain manifold ω (with intrinsic dimension $\mathtt{R}$ in the discussed case) of the whole space $\mathbb{R}^\mathtt{N}$.

An alternative procedure, more aligned with the model order reduction rationale, consists of calculating action/reaction pairs, i.e., $\left(\mathbf{U}^k, \mathbf{F}^k\right)$, $k = 1, \ldots, \mathtt{K}$, and looking for the lowest rank matrix $\mathbb{K}^\mathtt{LR}$, i.e., the most reduced model, minimizing, by using a certain p-norm, the cost $\mathscr{C}(\mathbb{K}^\mathtt{LR})$ [15]:

$$\mathscr{C}\left(\mathbb{K}^\mathtt{LR}\right) = \sum_{k=1}^{\mathtt{K}} \left\| \mathbb{K}^\mathtt{LR}\mathbf{U}^k - \mathbf{F}^k \right\|_p, \tag{11}$$

as employed for discovering dynamical systems within the so-called Dynamic Mode Decomposition (DMD) [16, 17].

In [15] the just reduced model extractor was extended for addressing nonlinear models, in [18] for operating within a reduced basis and also for extracting the reduced transfer function that approximates the inverse of model. Moreover, in [19], the procedure was adapted for constructing stable time integrators. We will come back later to the construction of models from data by using machine learning (ML) techniques. Before that, we will address the parametric dimension.

In the parametric case, POD-based techniques continues to perform well, as soon as the reduced basis is extracted from a training that covers as much as possible the parametric domain.

Then, the solution can be projected into the reduced basis and model parameters inferred online by using performant data assimilation techniques, for instance Bayesian or Kalman filters.

Proper Orthogonal Decomposition with Interpolation, PODI, [20] has been largely employed. When using the PODI, the coefficients affecting each of the vectors composing the reduced basis, are assumed to depend on the parameters, whose quantification is performed by using standard regressions. As a main limitation of this technique we can cite the number of parameters that complexities the construction of the regression.

For improving efficiency, several reduced bases are calculated at different locations of the parametric space, instead of a unique reduced basis in the whole parametric space. When different reduced bases are computed for different choices of the parameters, they must be interpolated for covering the whole parametric space before using it for approximating the solution and proceeding from the weighted residual form. Interpolation is a tricky issue and some efficient procedures were proposed for performing that interpolation, as for example the employ of Grassmannian manifolds [21].

Another way of computing parametric solutions of parametrized partial differential equations consists in applying the so-called Proper Generalized Decomposition, PGD, based on the separated representation of the solution. This was originally proposed for defining non-incremental transient solutions [22]. In the parametric setting, the PGD proceeds by assuming a fully separated representation of the problem solution, where parameters are assumed extra-coordinates [23]. Thus, if for the sake of simplicity we assume a model depending on a single parameter μ, the parametric solution approximation $u^{\mathtt{Q}}(x,t,\mu)$ reads

$$u^{\mathtt{Q}}(x,t,\mu) = \sum_{i=1}^{\mathtt{Q}} X_i(x)\,T_i(t)\,M_i(\mu), \tag{12}$$

with $x \in \Omega$, $t \in \mathscr{T}$ and $\mu \in \mathscr{I}$.

To compute the different unknown functions involved in the separated representation, Eq. (12), the usual weighted residual form is extended according to

$$\int_{\Omega\times\mathscr{T}\times\mathscr{I}} u^*(x,t,\mu)\mathscr{R}(u(x,t,\mu))\;dx\;dt\;d\mu = 0. \tag{13}$$

Within a Galerkin framework, when looking to the q^{th} functional product, the trial and test functions to be employed within the integral form, Eq. (13), read respectively

$$u^q(x,t,\mu) = \sum_{i=1}^{q-1} X_i(x)\,T_i(t)\,M_i(\mu) + X_q(x)\,T_q(t)\,M_q(\mu) \equiv u^{q-1}(x,t,\mu) + X_q(x)\,T_q(t)\,M_q(\mu), \tag{14}$$

and

$$u^*(x,t,\mu) = X^*(x)\,T_q(t)\,M_q(\mu) + X_q(x)\,T^*(t)\,M_q(\mu) + X_q(x)\,T_q(t)\,M^*(\mu). \tag{15}$$

The separated representation constructor, deeply described in [24], proceeds by using an alternate directions, fixed point algorithm that computes the unknown function at the enrichment iteration q: $X_q(x)$ from $T_q(t)$ and $M_q(\mu)$ taken at the previous iteration, $T_q(t)$ from $X_q(x)$ and $M_q(\mu)$, and finally $M_q(\mu)$ from $X_q(x)$ and $T_q(t)$. The iteration continues until reaching the fixed point, and then the next function product, $q+1$ is considered.

The main difficulties in applying the just described procedure are the necessity of performing an affine decomposition of the problem residual involved in Eq. (13) for ensuring the efficiency of the sequential calculation of each function that involves the corresponding problem coordinate. Such affine decomposition is not direct, mainly in the case of nonlinear models [14, 23]. The other difficulty is related to the procedure intrusiveness that makes difficult its use in tandem with usual commercial software.

The main interest of computing such parametric solutions $u(x,t,\mu)$ is the ability of deriving real-time engineering, by computing any solution related to any parameters choice in almost real-time, fact that makes possible simulation, optimization, inverse analysis, simulation-based control and uncertainty propagation, all them under the stringent real-time constraints.

To overcome the difficulties related to the complex nonlinear problems and the procedure intrusiveness, one option consists in constructing metamodels (also known as surrogates, response surfaces, virtual charts or vademecums). The construction of these metamodels can be performed by considering a simple workflow: (i) defining a sampling of the parametric space; (ii) computing a high fidelity solution for each parameter choice; (iii) using an adequate regression for extending the solution known at the points in the sampling everywhere in the parametric space.

There are many alternatives for performing the first and third tasks, trying to conciliate: (i) active learning; making use of a sparse and very reduced data set, considering the best goal-oriented sampling, aimed at roughly linearly scaling with the dimension of the parameters space; (ii) the high fidelity solutions post-compression by extracting first and using then (by projection) reduced bases; (iii) rich enough approximation bases while avoiding overfitting, based on the use of sparse regularizations for enforcing parsimony (elastic-net, ridge, lasso, ...) [25]; (iv) using orthogonal basis for evaluating sensibilities in a direct manner; and (v) efficiently addressing the high-dimensional spaces induced by the multi-parametric models, where the use of separated representations are specially suitable [26–29].

Other alternatives for performing regressions consist of using state-of-the-art techniques, like Support Vector Regression, SVR, [30], Decision Trees [31] or their Random Forest counterpart [32], Deep Neural Networks [33, 34]. These last are very efficient when a large enough amount of data is available.

The use of those parametric solutions (or their parametric transfer functions counterpart) makes it possible to link system components and connect components with their environment, to constitute the so-called *Augmented Virtual Prototype*, AVP, a nominal virtual replica of a real system. AVP allowed to address rich descriptions of physical systems in almost real-time. But are those descriptions and the associated predictions in agreement with observations and measurements on the physical system?

In some cases, it is. In many others, however, this is not the case and significant deviations appear, limiting predictive capabilities. An epistemic *ignorance* seems to persist in our conceptualization and subsequent description of the physical reality. Models are models, but sometimes reality seems to contain something else, the so-called *ignorance* (the part of the reality that our models ignore), and that, as discussed below, constitutes an immense source of opportunities. As Stephen Hawking once said: *The greatest obstacle is not ignorance—it is the illusion of knowledge*.

2.3. *Towards the art of decision making in complex systems of systems*

Advanced procedures are urgently needed, not only for performing faster, but for performing well. The *art of decision making* irrupted almost everywhere, facilitated by our nowadays global interconnected word, where the Internet of Things, IoT, is fully deployed. This is at the heart of the fourth industrial revolution, with *data* as one of its protagonists, but not the only one. Smart cities and nations, with the human in the loop, are attracting interest of scientists, engineers, politicians and decision makers. This fourth industrial revolution has its counterpart in science: we speak of the fourth paradigm of science, where large scientific infrastructures produce huge volumes of data that are readily incorporated into the scientific pipeline.

Complex systems of systems, containing many interconnected components, are concerned by uncertainty and variability that limit the performances of physics-based models. On the other hand, data, even when available, rarely cover the system extension (and scales) in space and time, and then, diagnosis and prognosis based exclusively on data encounter their own limitations. The hybrid paradigm, combining both, the existing knowledge and the collected data representing the considered reality, seems a valuable route for conciliating accuracy and efficiency.

As commented, data becomes a major protagonist. However, the word *data*, simple at a first view, encompass a rich essence. First, there is a vast typology: list, images, time series, fields, graphs… Second, extracting the intrinsic data dimensionality is also a tricky issue, needing for efficient linear and nonlinear dimensionality reduction procedures [2], also known as manifold learning techniques, able to discriminate useless and useful. In recent times, some authors speak even of the fifth paradigm of science, in which data are produced by simulation in order to feed scientific discovery and theories.

But, how to realize that a model is not accurate enough? The simplest way: by making measurements (collecting data) and comparing these measurements with the predictions based on the existing knowledge, provided by the state-of-the-art models, that can be very efficiently solved as discussed above.

When these measurements, even adequately assimilated into the models (model calibration), does not result in a calibrated model able to describe with high fidelity the observed reality, something seems missing in the reality representation. This should be modeled to improve designs and decisions.

One possibility is offered by Artificial Intelligence, AI. Machine Learning, a major protagonist of AI, is able to create predictive models from available or collected data, with an additional added value, the fact of providing that prediction in almost real-time.

However, creating models from scratch, simply because the existing models based on physics were not accurate enough, is not the best choice. Creating a model based on data from scratch needs a lot of data, and in engineering and technology, data is synonym of cost. Sometimes data collection also implies to consider an ethical dimension, to fulfill existing regulations, or must address technical difficulties, related for example to the sensors placement, data transfer…

In our works we advocated by the alliance between both (i) the analogical world of knowledge and physics-based models, and (ii) the more recent digital world of data, manipulated by more and more powerful (accurate, frugal and explainable—certifiable—) techniques of Artificial Intelligence.

This new alliance is materialized in the so-called Digital or Hybrid Twins, in which physics-based models are enriched to decrease their intrinsic ignorance, in a pragmatic way, from data representing the deviation between predictions and measurements. The hybrid paradigm not only allows reducing the amount of needed data (now it is only expected to describe the gap between the reality and the physics-based predictions), but also the ability of explaining the part of the model based on the existing physics or knowledge, and then facilitating the design or decision certification.

In order to reduce the amount of data to model the observable phenomena (when models do not exist or are too inaccurate) or for enriching the existing physics-based models within the hybrid paradigm previous introduced, physics-aware (also known as physics-informed) Artificial Intelligence seems to be the most appealing route.

The so-called Physics Informed Neural Networks, PINN, [35, 36] consider the approximation of the unknown function $u(x,t)$ as a regression problem defined on an adapted neural network. Then, as soon as the physics is assumed fully known and adequately described by a partial differential equation, the derivatives involved in the differential operator can be applied on the NN and the residual nullity is enforced from the NN loss function.

Sometimes, not everything concerning the physics is fully known. If we consider a hyperelastic material, the best option consists in learning the free energy by constructing a regression linked to the state variables, in such a way that its derivatives leads to the constitutive equation. Then, the free energy is learned to be consistent with the collected data on the structural component, under the equilibrium constraints.

Similar procedures apply in the so-called Structure Preserving NN (also known as Thermodynamic Informed NN) where the free energy and the dissipation potential are computed in such a way that energy balance and entropy production are ensured [37–42].

In those works, inspired from the GENERIC framework [43], the state $\mathbf{Z}$ evolution, $\dot{\mathbf{Z}}$, reads

$$\dot{\mathbf{Z}} = \mathbb{L}\nabla_Z \mathscr{H} + \mathbb{M}\nabla_Z \mathscr{S}, \tag{16}$$

where the first term of the right-hand member represents the reversible evolution (Hamiltonian contribution) whereas the second one represents the dissipative contribution, with $\mathscr{H}$ and $\mathscr{S}$ the energy and the entropy respectively.

Learning matrices $\mathbb{L}$ (known to be skew-symmetric) and $\mathbb{M}$, symmetric and positive semi-definite, as well as both potentials $\mathscr{H}$ and $\mathscr{S}$ (subjected to some constraints: the Jacobi identities as well as the consistency conditions $\mathbb{L}\nabla_Z \mathscr{S} = \mathbf{0}$ and $\mathbf{M}\nabla_Z \mathscr{H} = \mathbf{0}$) is performed from the existing data concerning the state time evolution $\mathbf{Z}(t_i)$, $i = 1, 2, \ldots$ The learnt model has very interesting properties, as are the ones related to energy conservation and positive dissipation, giving rise to stable and accurate time integrators.

The main issue found when learning such thermodynamic-aware models is not the regression implementation, but the choice of the variables in the state vector, an issue discussed in [44]. In very small systems the state variables are easily identified, however in large (continuous) systems such a choice is far of being trivial. In that case, many options exist. One among them consists of performing a dimensionality reduction. The use of most of manifold learning dimensionality reduction has as main drawback the unavailability of performing the inverse mapping for coming back from the reduced space to the departure one.

In the general nonlinear case, an appealing alternative consists in the use of NN-based (sparse) autoencoders [45], where encoding and decoding is learned, while models operate (and are learnt) in a transparent way in the internal layer of reduced dimension where active reduced coordinates (the so-called latent variables or latent space) act. Autoencoders allow to learn the best model representation, that in many cases, other than reducing the dimension of the model, makes possible reducing the model complexity (and therefore its nonlinearity), enabling the use of simpler and cheaper regression techniques operating at the level of the latent space, where sometimes linear regressions suffice.

In the same way that data allows enriching the existing knowledge (i.e., models based on physics), the existing knowledge and physics-based models allow to replace the big data by a smart data paradigm, by simply responding three main questions: (i) what data to collect, (ii) when and (iii) where. For instance, for having an idea on the temperature in Paris we can use two thermometers, one placed inside our apartment, the other in the balcony, and take the temperatures at mid-night, early in the morning and at noon. Two thermometers and 3 measures, to have a reasonable approximation of the temperature evolution each day in Paris. However, for taking that decision about the measurements we used our knowledge. We thus avoid using one million thermometers distributed all along Paris and collecting the temperature at each one, each millisecond. This is expensive and useless. This example illustrates the difference between big-data and smart-data. Smarter seems definitively better!

Reconciliation and alliance between physics-based and data-driven models, leads to the so-called Hybrid Twin, HT [46]. It constitutes a digital replica of a material, process, structure, component, systems or systems of systems, able to replace or substitute the real system for antic-

ipating future responses, to access to its intimate state, while retaining the following features, qualities and functionalities: (i) accuracy guaranteed by the hybrid approach; (ii) frugal, being based on the smart-data paradigm; (iii) holistic throughout all the involved physics and description scales; (iv) providing real-time responses by invoking advanced model order reduction techniques; (v) explicable and certifiable; (vi) adaptable; (vii) reliable; (viii) resilient; (ix) informed -IoT; (x) systemic to address the system within its environment; and (xi) usable.

The HT of a component, of a system, or of a system of systems can and must interact with the human being. The last continues to be, even today, the main contributor in many domains where imagination, intuition, complex abstraction and interpretability are needed, as well as qualitative reasoning, including the emotional dimension.

It is well known that people remember very differently what they learnt after two weeks: 10% to 20% of what they heard or read, 90% of what they did (acted upon) in a real or virtual experience. Thus, we must experience physics, we must be protagonists and not only observers.

Data produced by the HT must be contextualized to be profitable, interpretable and exploitable by humans. New human-centric facilities irrupted and were combined with the hybrid paradigm. Among them the immersive virtual reality, VR, augmented reality, AR, [47] in multiphysics evolving settings, and hybrid reality, HR, in which the human interacts with the virtual world (a digital replica of the reality) in the form of holograms that can be experienced visually, from the touch, and even by integrating other sensorial sensations: smell or taste. Recording emotions could lead to better adapt a product to each specific final user, e.g. adapt the car to the driver, adapt the text-book to the reader, adapt a cosmetic cream to the customer...

This human-centric HT constitutes the major protagonist of an incipient Augmented Intelligence, that does not aim at replacing the human, but enriching him to face the ultimate art, the *art of decision-making*. The one that is not based on a simple reptilian pattern recognition: I am hungry, I eat... necessary, but primitive. Here we are speaking about the art of decision making, that has another more elaborated dimension making use of the two sides of the brain, the rational and the creative.

3. Final remarks

We recently experienced the impact of COVID-19 pandemic, everywhere, and adapted to it in an active and constructive manner. The world is definitively changing, evolving, and things will be definitively different even when pandemic will definitively disappear. During this pandemic, many people worked at home, some activities seems being nowadays resilient enough, but other remain much less, as for example production machines and chains.

Without any doubt, a resilient, human centric world, allying real and virtual, cold digital data and warmer emotions, all them harmoniously entangled, will constitute the new revolution. It will be not only an industrial one. This time, more than ever, it will have a huge human and societal dimension.

Engineering is succeeding in this new framework, sculpting a new world, with major protagonists: HT (combining real-time physics and real-time physics-informed AI), blockchain, quantum computing, multi-sensorial and emotional human-centric... empowering human creativity and imagination, enabling moving further, better and faster... It is not a far dream, it is becoming a reality, that results from an intimate combination of the three just revisited arts of engineering: the art of modeling, the art of simulation and the art of decision making.

Conflicts of interest

The authors declare no competing financial interest.

Dedication

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Acknowledgments

Authors are supported by their respective ESI Group research contracts, whose support are gratefully acknowledged.

This research is part of the programme DesCartes and is supported by the National Research Foundation, Prime Minister Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

[1] ESI Groups, "The History of Crash Simulation", 2019, https://www.youtube.com/watch?v=NU3cB9ArSZg.

[2] J. A. Lee, M. Verleysen, *Nonlinear dimensionality reduction*, Information Science and Statistics, Springer, 2007.

[3] S. T. Roweis, L. K. Saul, "Nonlinear dimensionality reduction by Locally Linear Embedding", *Science* **290** (2000), no. 5500, p. 2323-2326.

[4] L. van der Maaten, G. E. Hinton, "Visualizing data using t-SNE", *J. Mach. Learn. Res.* **9** (2008), p. 2579-2605.

[5] F. Chinesta, A. Huerta, G. Rozza, K. Willcox, "Model Order Reduction", in *Encyclopedia of Computational Mechanics Second Edition* (E. Stein, R. de Borst, T. Hughes, eds.), John Wiley & Sons, 2017, p. 1-36.

[6] K. Veroy, A. T. Patera, "Certified real-time solution of the parametrized steady incompressible Navier–Stokes equations: Rigorous reduced-basis *a posteriori* error bounds", *Int. J. Numer. Methods Fluids* **47** (2005), no. 8-9, p. 773-788.

[7] K. Veroy, C. Prud'homme, D. Rovas, A. T. Patera, "*A Posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations*", in *Proceedings of the 16th AIAA Computational Fluid Dynamics Conference*, American Institute of Aeronautics and Astronautics, Inc., 2003.

[8] Y. Maday, A. T. Patera, G. Turinici, "*A Priori* convergence theory for reduced-basis approximations of single-parameter elliptic partial differential equations", *J. Sci. Comput.* **17** (2002), no. 1-4, p. 437-446.

[9] Y. Maday, A. T. Patera, G. Turinici, "Global *a priori* convergence theory for reduced-basis approximation of single-parameter symmetric coercive elliptic partial differential equations", *C. R. Math. Acad. Sci. Paris* **335** (2002), no. 3, p. 289-294.

[10] D. Ryckelynck, "A priori hyperreduction method: an adaptive approach", *J. Comput. Phys.* **202** (2005), no. 1, p. 346-366.

[11] T. Chapman, P. Avery, P. Collins, C. Farhat, "Accelerated mesh sampling for the hyper reduction of nonlinear computational models", *Int. J. Num. Meth. Engrg.* **109** (2017), no. 12, p. 1623-1654.

[12] M. Barrault, Y. Maday, N. C. Nguyen, A. T. Patera, "An "empirical interpolation" method: Application to efficient reduced-basis discretization of partial differential equations", *C. R. Math. Acad. Sci. Paris* **339** (2004), no. 9, p. 667-672.

[13] S. Chaturantabut, D. C. Sorensen, "Nonlinear model order reduction via discrete empirical interpolation", *SIAM J. Sci. Comput.* **32** (2010), no. 5, p. 2737-2764.

[14] J. V. Aguado, D. Borzacchiello, K. S. Kollepara, F. Chinesta, A. Huerta, "Tensor representation of on-linear models using cross approximations", *J. Sci. Comput.* **81** (2019), p. 22-47.

[15] A. Reille, N. Hascoet, C. Ghnatios, A. Ammar, E. Cueto, J.-L. Duval, F. Chinesta, R. Keunings, "Incremental dynamic mode decomposition: A reduced-model learner operating at the low-data limit", *C. R. Méc. Acad. Sci. Paris* **347** (2019), no. 11, p. 780-792.

[16] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data", *J. Fluid Mech.* **656** (2010), p. 5-28.

[17] M. O. Williams, I. G. Kevrekidis, C. W. Rowley, "A data-driven approximation of the Koopman operator: extending dynamic mode decomposition", *J. Nonlinear Sci.* **25** (2015), no. 6, p. 1307-1346.

[18] A. Reille, V. Champaney, F. Daim, Y. Tourbier, N. Hascoet, D. González, E. Cueto, J.-L. Duval, F. Chinesta, "Learning data-driven reduced elastic and inelastic models of spot-welded patches", *Mechanics & Industry,* **22** (2021), article no. 32 (17 pages).

[19] A. Sancarlos, M. Cameron, J.-M. Le Peuvedic, J. Groulier, J.-L. Duval, E. Cueton, F. Chinesta, "Learning stable reduced-order models for hybrid twins. Data Centric Engineering", *Data-Centric Engineering* **2** (2021), article no. E10.

[20] H. V. Ly, H. T. Tran, "Modeling and control of physical processes using Proper Orthogonal Decomposition", *Mathematical and Computer Modelling* **33** (2001), no. 1-3, p. 223-236.

[21] D. Amsallem, C. Farhat, "Interpolation method for adapting reduced-order models and application to aeroelasticity", *AIAA J.* **46** (2008), no. 7, p. 1803-1813.
[22] P. Ladevèze, *Nonlinear Computational Structural Mechanics. New Approaches and Non-Incremental Methods of Calculation*, Mechanical Engineering Series, Springer, 1999.
[23] F. Chinesta, A. Leygue, F. Bordeu, F. Bordeu, J. V. Aguado, E. Cueto, D. González, I. Alfaro, A. Ammar, A. Huerta, "PGD-Based *Computational Vademecum* for Efficient Design, Optimization and Control", *Arch. Comput. Methods Eng.* **20** (2013), no. 1, p. 31-59.
[24] F. Chinesta, R. Keunings, A. Leygue, *The Proper Generalized Decomposition for Advanced Numerical Simulations. A primer*, SpringerBriefs in Applied Sciences and Technology, Springer, 2014.
[25] S. L. Brunton, J. L. Proctor, N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems", *Proc. Natl. Acad. Sci. USA* **113** (2016), no. 15, p. 3932-3937.
[26] R. Tibshirani, "Regression shrinkage and selection via the lasso", *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **58** (1996), no. 1, p. 267-288.
[27] D. Borzacchiello, J. V. Aguado, F. Chinesta, "Non-intrusive *Sparse Subspace Learning* for parametrized problems", *Arch. Comput. Methods Eng.* **26** (2019), no. 2, p. 303-326.
[28] R. Ibáñez, E. Abisset-Chavanne, A. Ammar, D. González, E. Cueto, A. Huerta, J.-L. Duval, F. Chinesta, "A multi-dimensional data-driven sparse identification technique: the sparse Proper Generalized Decomposition", *Complexity* **2018** (2018), article no. 5608286.
[29] A. Sancarlos, V. Champaney, J.-L. Duval, E. Cueto, F. Chinesta, "PGD-based advanced nonlinear multiparametric regressions for constructing metamodels at the scarce-data limit", *Adv. Model. and Simul. in Eng. Sci.* **10** (2023), no. 4.
[30] M. Awad, R. Khanna, "Support Vector Regression", in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Apress, 2015, p. 67-80.
[31] C. W. Kirkwood, "Decision Tree primer", 2002, https://www.public.asu.edu/~kirkwood/DAStuff/refs/decisiontrees/index.html.
[32] L. Breiman, "Random Forests", *Mach. Learn.* **45** (2001), p. 5-32.
[33] J. Schmidhuber, "Deep learning in neural networks: An overview", *Neural Netw.* **61** (2015), p. 85-117.
[34] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*, Adaptive Computation and Machine Learning, MIT Press, 2016.
[35] M. Raissi, P. Perdikaris, G. E. Karniadakis, "Physics informed deep learning (part I): data-driven solutions of nonlinear partial differential equations", https://arxiv.org/abs/1711.10561, 2017.
[36] M. Raissi, P. Perdikaris, G. E. Karniadakis, "Physics informed deep learning (part II): data-driven discovery of nonlinear partial differential equations", https://arxiv.org/abs/1711.10566, 2017.
[37] D. González, F. Chinesta, E. Cueto, "Thermodynamically consistent data-driven computational mechanics", *Comput. Mech.* **31** (2019), p. 239-253.
[38] D. González, F. Chinesta, E. Cueto, "Learning corrections for hyper-elastic models from data", *Front. Mater.* **6** (2019).
[39] Q. Hernandez, A. Badias, D. González, F. Chinesta, E. Cueto, "Deep learning of thermodynamics-aware reduced-order models from data", *Comput. Methods Appl. Mech. Eng.* **379** (2021), article no. 113763.
[40] Q. Hernandez, D. González, F. Chinesta, E. Cueto, "Learning non-Markovian physics from data", *Journal of Computational Physics* **428** (2021), article no. 109982.
[41] B. Moya, D. González, I. Alfaro, F. Chinesta, E. Cueto, "Learning slosh dynamics by means of data", *Comput. Mech.* **64** (2019), p. 511-523.
[42] B. Moya, A. Badías, I. Alfaro, F. Chinesta, E. Cueto, "Digital twins that learn and correct themselves", *Int. J. Numer. Methods Eng.* **123** (2020), no. 13, p. 3034-3044.
[43] M. Grmela, H. C. Öttinger, "Dynamics and thermodynamics of complex fluids. I. Development of a general formalism", *Phys. Rev. E* **56** (1997), no. 6, p. 6620-6632.
[44] R. Ibanez, P. Gilormini, E. Cueto, F. Chinesta, "Numerical experiments on unsupervised manifold learning applied to mechanical modeling of materials and structures", *Comptes Rendus. Mécanique* **348** (2020), no. 10-11, p. 937-958.
[45] G. E. Hinton, R. Zemel, "Autoencoders, minimum description length and Helmholtz free energy", in *Advances in Neural Information Processing Systems*, vol. 6, Morgan-Kaufmann, 1993, p. 3-10.
[46] F. Chinesta, E. Cueto, E. Abisset-Chavanne, J.-L. Duval, F. El Khaldi, "Virtual, Digital and Hybrid Twins: A New Paradigm in Data-Based Engineering and Engineered Data", *Arch. Comput. Methods Eng.* **27** (2020), p. 105-134.
[47] A. Badias, S. Curtit, D. González, I. Alfaro, F. Chinesta, E. Cueto, "An Augmented Reality platform for interactive aerodynamic design and analysis", *Int. J. Numer. Methods Eng.* **138** (2019), p. 125-138.