Physics-based & Data-driven models: the winning team

Francisco Chinesta[1] and Elías Cueto[2],
[1] ESI Group Chair @ Arts et Métiers Institute of Technology, París, France
[2] ESI Group Chair @ I3A, University of Zaragoza, Zaragoza, Spain

In the previous industrial revolution, virtual twins emulating a physical system were considered as the major protagonists of simulation-based engineering. This type of twin was usually based on numerical, yet static, models that were used, often separately and independently, in designing, manufacturing and testing complex systems and their components. They were, however, not expected to accommodate or assimilate data. The reason is that the characteristic time of standard simulations was, and still is even today, not compatible with the real-time responses needed when prediction are required in-operation, as nowadays needed in the context of connected systems.

The subsequent generation of twins, called digital twins, allows real-time decision making by using powerful data analytics, machine learning and artificial intelligence on the abundant collected data. Thus, predictive and operative maintenance, and data-based control can be possible. However, creating a data-based model from scratch is expensive and sometimes requires too much data that can be difficult or even impossible to collect. Moreover, in science and engineering decisions and designs must be certified, and for that, the employed rationale explained, a real issue for nowadays artificial intelligence based procedures.

Hence, a hybrid paradigm seems more pertinent. Hybrid twins also include predictions from physics-based models. However, persistent and biased deviations from the physical measurements are now interpreted as ignorance about some hidden physics taking place, and thus provide an opportunity to learn on-the-fly. Hybrid Twins embrace physics-based models (that should accommodate real-time queries) and "deviation data-driven models", the last intended to fill the gaps on inherent epistemic ignorance in the physics-based model.

Thus, the "Hybrid" framework allows combining data and models, mathematical physics and artificial intelligence. On the other hand, it allows operating with a more consequent and reasonable amount of data (within the so-called smart-data paradigm), because it serves to model the gap more than creating the models from scratch. Finally, as soon as engineering systems are certified from the physics-based model, the data-driven model of the deviation (ignorance) serves for improving prediction, and can be viewed as a bonus or surplus.

Data was present in industry from the very beginning. However, it has served traditionally for calibrating models and for validating designs. These data contributed to generate knowledge enabling the training of experts. Today, new technologies facilitate massive data acquisition that in most cases remain without analysis. This is often due to the lack of appropriate tools for treating that data at the required rates, or simply to the inadequacy of these techniques for extracting the hidden knowledge behind data.

Today the industrial reality is large. For example, testing machines used for calibrating material models produce vast amounts of data (X-tomography, Microscopy, Laser velocimetry, ...) in particular in form of sequence of images or time-series. On the other hand, production machines provide, in general, much less data, because in most of cases its acquisition is very expensive, in some cases technologically challenging, and often simply impossible.

Thus, the main three factors associated to data collected in industry concern (i) quantity, (ii) characteristic time of the response, and (iii) data quality (noise or bias).

Our framework concerns the hybrid paradigm in which data (within an engineered artificial intelligence framework) will enrich models exhibiting limited accuracy, and models should help data to become smarter, by informing what data, at what scale, where and when it should be collected.

Model enrichment is based on the use of the gap between measurements and model predictions. The interest of using a model is twofold; first, it allows moving faster on a solid foundation, and second, the better are the models, the smaller are the deviations, implying an almost linear or slightly nonlinear behavior, both making possible their approximation from few data, instead of the vast amount of data needed for creating a model from the scratch.

The hybrid paradigm can be expressed as: *Reality = First order physics-based model, manipulated from the tools of applied mathematics and computer science PLUS a data-driven correction learned on-the-fly, based on data manipulated from adequate (engineered) artificial intelligence techniques.*

This framework makes it possible operating in both, the big-data limit as well as in the scarce-data limit, and both must accommodate with the characteristic time of the industrial process. When data is abundant and we proceed offline, the use of deep-learning is an excellent candidate. However, when data is very scarce and the data-driven model must adapt in time very fast, other techniques able to operate in those circumstances could and should be considered, analyzed, developed, improved and tested on real use cases.

In the industrial context, AI has six major axes: (i) visualizing multidimensional data; (ii) classifying data; (iii) modeling the input / output relationship enabling quantitative predictions – the art of modeling –; (iv) certifying those predictions; (v) explaining them, that is, extracting knowledge from the available data, and (vi) apply them in production (taking advantage of the online adaptation –hybrid twin paradigm–) and training final users and students on the use of AI techniques for adding value to industrial technology.

---

[1] F. Chinesta, E. Cueto, E. Abisset, J.L. Duval and K. El Khaldi, *Virtual, Digital and Hybrid Twins: A New Paradigm in Data-Based Engineering and Engineered Data*, Archives of Comput. Meth. in Engrg., **27**, 105-134 (2020).