A new Reliability-Based Data-Driven approach to Simulation-Based Models

Jacobo Ayensa Jiménez*, Mohamed Hamdy Doweidar* and Manuel Doblaré Castellano*
Mechanical Engineering Department, University of Zaragoza, Spain;
Aragón Institute of Engineering Research (I3A), University of Zaragoza, Spain;
Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain.
Campus Río Ebro, Edificio I+D, Mariano Esquillor S. N.
Email:jacoboaj@unizar.es; mohamed@unizar.es; mdoblare@unizar.es

Abstract—Data Science has burst into simulation-based engineering sciences with an impressive impulse. However, data are never uncertainty-free and a suitable approach is needed to face data measurement errors and their intrinsic randomness in problems with well-established physical constraints. We face this problem by hybridizing a standard mathematical modeling approach with a new data-driven solver accounting for the phenomenological part of the problem and able to handle the uncertainty of input data in an intelligent way. The reliability-based data-driven procedure performance is evaluated in a simple but illustrative unidimensional problem showing, in contrast with other data-driven solvers, better convergence, higher accuracy, clearer interpretation and major flexibility.

I. INTRODUCTION

Despite the wide application of Data Science in areas such as marketing and e-commerce [1], social sciences [2] or healthcare [3], there are other fields where very little has been done. An example are the disciplines where physical models and the corresponding mathematical and numerical simulation tools are well established like Computational Physics, Computational Chemistry or Computational Engineering (SBES). A straightforward application of these techniques is dynamic data-driven systems (DDDAS) [4], in which the idea is providing both predictive and learning capabilities to the control system from data acquired from a sufficient set of sensors. This paradigm was settled down by Kalman [5] in the sixties with his groundbreaking filter and is still nowadays a hot topic of research opening up a huge range of possibilities [6].

Unlike these approaches, based on the direct treatment of available data, SBES incorporates, in addition to data, some *a priori* characteristic physical knowledge of the analyzed system. At this point, it is crucial to distinguish between two kinds of knowledge. On one hand, physical general principles, such as conservation and thermodynamic laws that are universally accepted as able to describe the underlying universe structure. On the other hand, phenomenological models, such as macroscopic material constitutive relations. Even if it would be possible to derive the real mechanistic constitutive relations also from first physical principles, the overwhelming number of degrees of freedom involved in the relevant spatial-temporal scales needed for real applications makes this possibility intractable [7]. *Data Analytics* techniques would be very useful in SBES to extrapolate the phenomenological submodel, but now constrained with the mathematical expression of first principles.

One idea in this direction was introduced by T. Sussman [8] for hyperelastic isotropic materials using splines interpolation in what is now known as “What You Prescribe Is What You Get” (WYPWYG) philosophy. Recently, F. Chinesta and coworkers defined a strategy for data-driven Computational Mechanics [9], combining Manifold Learning techniques and a (possibly optimized) directional search strategy inspired in the LaTin method [10]. M. Ortiz and his group presented a material model-free method based on the minimization of the distance between the searched solution and a set of experimental data, using a proper energy norm, while remaining in the equilibrium manifold, or equivalently, a well-posed penalty approach [11]. None of these works take into consideration the inherent inaccuracy of the data. Here, a new family of methods, called reliability-based data-driven solvers are defined. Data-driven solver methodology naturally allows incorporating reliability along the statement of the modeling. With this, simulations become sensitive to measurements precision and incorporate uncertainty considerations. An easy but illustrative one-dimensional problem is used to compare results and to show improvements using this methodology, emphasizing the added value with respect to existing methods.

II. METHODOLOGY

Data-driven solvers may be seen as iterative solvers searching for the intersection of an empirical (data based) manifold and a physical manifold. The first one is in many practical applications experimentally based and has, therefore, a discrete nature. The second is usually established in terms of sound laws particular to the problem in hands, but otherwise, derived from first principles universally accepted as the basis of Physics. For the sake of simplicity, let us consider the elastic three-dimensional problem. In that case, the physical manifold is the set of states that verify global and local equilibrium (i.e. conservation of linear and angular momentum), that in the static case (negligible inertial effects) is written in differential form as $\nabla \cdot \sigma = 0$, where $\sigma$ is the stress tensor. This equation is usually approximated and solved in a discrete form using numerical methods like Finite Elements (FEM). After a

\[
\sigma = \nabla \cdot \sigma = 0
\]
conv

B. Reliability-based data-driven solver

X

cy

constraints only, so they can be written as

Ax

σ

We may, therefore, define a stochastic analogous problem to

before, N

whose dimension, E

the result of experimental tests. We then define a local penalty

be defined in terms of state variables

the uncertainty of each point approximates to zero.

E

tal measurements. The set

=1, \ldots, m, \) of data points, resulting from experimen-

tal points) and Y

interpolation, but in this case, non-smoothness of

oscillations provoking bad convergence. This can be avoided

one is Mahalanobis distance [12], equivalent to choose M = 2(Σ)−1 as metric matrix.

Under normality conditions, D₂

follows a non-central chi-
squared distribution with n = 2N degrees of freedom and non-centrality parameter λ (u − μ)TΣ−1(u − μ). In

particular, expected value writes μ(D₂) = 2N + d₂ and variance σ²(D₂) = 2(2N + 2d₂).

III. Results

We evaluate the performance of different data-driven

solvers, including the reliability-based one proposed herein. As

it could be predicted, the main problem of the linearization

approach appears when dealing with irregular (non-smooth)

empirical manifolds. This is typical in Physics when working

with models that have discontinuities, like in many mechanical

problems such as plasticity, damage, fracture and contact prob-

cles. A very basic unidimensional trivial problem exemplifies

well their main pathologies.

For a simple uniaxial loaded rod, with F = 100 kN, A = 200 cm² and L = 10 m, this problem may be easily solved

through traditional model-based techniques. The solution is

based on the combination of three equations. Equilibrium

equation, σA = F, compatibility equation, ε = \frac{u}{L}. For this

problem to be mathematically closed, we need a mathematical

relation, i.e. a model, relating the internal (state) variable

stress, σ, and the measurable variable strain, ε, what is known

as a constitutive relation of the material σ = f(ε).

Let us consider that the constitutive relation is not known

and the material behavior could be linear, smoothly nonlinear

or non-smoothly nonlinear. In any case, what we have to

describe the material behavior is a considerable amount of

experimental pair values \((ε, σ)\), \(E = \{(ε^j; σ^j)\}_{j=1, \ldots, m}\). For testing data-driven solvers based on linearization, let us com-

pare the computed results when considering a non-smoothly

nonlinear behavior and using the well-known iterative tangent

Newton-Raphson method, with the analytical results obtained

through the exact linear model.

Figure 2 shows solution points for the DD and RBDD

solvers for a material where the uncertainty associated with

this stochastic approach of the problem. A very recommended

one is Mahalanobis distance [12], equivalent to choose M = 2(Σ)−1 as metric matrix.

}
the actual material behavior is not homogeneous: in the elastic zone, where the material is very well characterized, uncertainty is low, but it increases when strains are higher. RBDD solver is sensitive to this variation, while DD solver is not. For complete information, Figure 2 is complemented by the statistical properties summarized in Table I.

![Fig. 1. Performance of different tested solvers.](image)

![Fig. 2. Performance of DD and RBDD solvers for heterogeneous uncertainty.](image)

**TABLE I**

<table>
<thead>
<tr>
<th>Method</th>
<th>Squared distance</th>
<th>Expected value</th>
<th>Variance</th>
<th>95%-Confident bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD</td>
<td>3.90 x 10^-2</td>
<td>3.33</td>
<td>11.32</td>
<td>10.06</td>
</tr>
<tr>
<td>RBDD</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**IV. CONCLUSION**

A new reliability-based data-driven (RBDD) solver has been formulated for data-driven simulation-based engineering problems, allowing uncertainty considerations in the input data that are, therefore, not considered as uncertainty-free, but of random nature. The data-driven simulation problem is defined as a constrained stochastic optimization problem.

It has been shown that selecting a proper uncertainty dependent distance, the Mahalanobis distance, results in very good statistical properties as well as easily interpretable optimal distances. RBDD solvers present a meeting point between theoretical sciences, through epistemological constraints, and experimental sciences, through uncertain real world data. The elegance of the mathematical formulation enables many analysis and theoretical considerations for the whole spectrum of Continuum Physics. The ease of combining the presented concepts with all trendy Data Science and Deep Learning tools opens up huge possibilities for facing the most challenging problems.

**ACKNOWLEDGMENT**

The authors gratefully acknowledge the financial support from the Spanish Ministry of Economy and Competitiveness (MINECO MAT2016-76039-C4-4-R, AEI/FEDER, UE), the Government of Aragon (DGA) and the Biomedical Research Networking Center in Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN). CIBER-BBN is financed by the Instituto de Salud Carlos III with assistance from the European Regional Development Fund.

This work is part of the paper *A new reliability-based data-driven approach for noisy experimental data with physical constraints*, in publication process.

**REFERENCES**


Jacobo Ayensa Jiménez is a Ph.D. student at the Mechanical Engineering Department, University of Zaragoza, Spain. He received his B.Sc. in mathematics and M.Sc. in Civil Engineering at the CFIS center, in the Polytechnic University of Catalonia, Spain, and his M.Sc. in Advanced Design in Mechanical Engineering in the University of Seville, Spain. He began his professional career in the research department of Abengoa working on the uncertainty incorporation in mechanical systems design, such as CSPP troughs and heliostats. Then he moved to the academic world for starting his Ph.D at the Aragón Institute of Engineering Research, where he is focused in mathematical modeling of tumour microenvironments.