An Information-Theoretic Machine Learning for uncertainty analysis of heterogeneous reservoir simulation

Ianyqui F. Costa[1], Liliane de A. Fonseca[2], Ramiro B. Willmersdorf[3], Ézio da R. Araújo[4]

[1,2,4]*Dept. of Civic Engineering, Federal University of Pernambuco*
[1] ianyqui.costa@ufpe.br, [2] liliane.fonseca@ufpe.br, [4] ezio@ufpe.br
[3]*Dept. of Mechanical Engineering, Federal University of Pernambuco*
ramiro.willmersdorf@ufpe.br
Av. Prof. Moraes Rego, nº 1235, 50670-901, Cidade Universitária, Recife-PE, Brasil

Abstract. Deep Neural Networks (DNNs) have shown significant advances in solving time-dependent partial differential equations with smooth coefficients, mainly in high-dimensional and Big-Data problems. However, despite this, they could have poor accuracy. Uncertainty analysis of reservoir simulations offers substantial difficulties because they are low-dimension problems and have heterogeneous reservoir properties and coefficients varying in time. Moreover, they belong to the Small-Data problem class. There is no clear advantage of DNNs over traditional algorithms with those collective properties. Physical Informed NNs (PINN) used to be the main driver of Machine Learning applications of scientific and engineering problems. However, it failed to solve reservoir heterogeneous problems, making it necessary to homogenize subdomains before submitting problems to PINNs. As heterogeneity changes with time, successive homogenization of selected regions should be done, introducing new difficulties. In this paper, we propose to use Information-Theoretic Machine Learning (ITML) to solve the above set of problems, firstly performing a stochastic upscaling that reduces the cardinality and the probabilistic dimension of the original reservoir models and then solving the upscaled model by traditional methods. The upscaling is forced to obey existing geostatistical properties and physical laws, which is more effective than the posterior verification as done by PINNs. Solving the reduced problems by direct methods adds the advantage of more reliable solutions not guaranteed by DNNs solutions in the Small-Data regime. Results that contemplate time-varying coefficients will be described in further publications.

Keywords: Stochastic Reservoir Simulation, DNN, PINN, ITM, upscaling

1 Introduction

Deep Neural Network (DNN) models have shown remarkable results when applied to high dimensional (hundreds of phase states), Weinan [1], and Big-Data (hundred thousand steps) problems, despite their low accuracy relative to traditional models (finite element, finite volume).

Concerning generalization error, the classical statistical learning theory establishes that the *optimal generalization loss* requires selecting a *model capacity* that attains the best balance between underfitting and overfitting, that is, between not having enough capacity to model the training data accurately and having too much, thus prone to adapt to closely to the training data, at the expense of generalization. However, it has long been observed that DNN models can obtain good generalization performance even though measures (typically, the cross-entropy) exhibit characteristics of overfitting, meaning there is a third regime for massively overparameterized models. The transition for this third regime, the *interpolation threshold*, is characterized by a decrease in the generalization error just after its peak. Bornschein [2] shows that large DNN models, over the interpolation threshold, can generalize well even for Small-Data problems.

Due to the operators' complexity and cardinality, real-world stochastic reservoir problems should operate with about two hundred elements of the supervised set. Instead of using a large DNN model to account for generalization, this work develops an Information-Theoretical Machine Learning to upscale the reservoir model and solve it using traditional methods, thus overcoming too the remaining accuracy problem encountered in DNN models.

2 Methodology

Let the reservoir be described by a set of N_R microscale geostatistical realizations of a random field collectively aggregated into an ensemble X, identifying fine scale meshes with N cells. In general, X contains more information than necessary for the accurate solution of a given reservoir simulation problem. The goal is to reduce this ensemble into an ensemble Y over $n < N$ coarse scale meshes and possibly $n_R < N_R$ macroscale realizations. If that reduction is possible, the answer to the specific problem can be encountered with less computational effort and reduced accuracy. Ensemble X can be as detailed as a cellular mesh, containing microscale information that can be compressed into a desired macroscale for the specific objective.

If the properties described by X can be compressed into Y for the solution of a specific problem, what needs to be determined is what information in X must be statistically preserved in such a way that distortion in the desired response statistic can be controlled. At least two measures therefore need to be established. The first is the measure of the fidelity of the representation of X by Y, and the other is the statistical accuracy of the desired response. The physics at the microscale and that of the macroscale need not be the same.

The temporal scale will not be explicitly addressed in this text, without losing generality. This will be done in a future text when the methodology is applied to history matching as a component of stochastic control of reservoir management and development.

2.1 Mutual information

Replacing the X field with a smaller Y field implies the possibility of assigning the same value to large regions of the domain. A mutual information indicator measures the average information that knowledge of Y can provide about X and vice versa.

The hypothesis that the large regions of the heterogeneous domain of interest can be assigned the same values makes it possible to describe the random field X by the smaller field Y. A measure of the ability of Y to approximate X can be given by the mutual information indicator given by,

$$I(X,Y) = \iint p_{XY}(x,y) log\ \frac{p_{XY}(x,y)}{p_X(x)p_y(y)} dxdy, \tag{1}$$

where $p_{XY}(x,y)$ is the joint density of (X,Y), and $p_X(x)$ and $p_Y(y)$ are the respective marginal distributions. Its minimum value is zero when the two fields are independent, that is, $p(X,Y) = p_X(x)p_Y(y)$. Its maximum value is known as the entropy of X, which occurs when the two fields are identical $X \equiv Y$. Defining joint entropy as the indicator itself, and developing,

$$H(X,Y) = -\int p_X(x) log\ p_X(x) dx\ + \iint p_{XY}(x,y) log\ \frac{p_{XY}(x,y)}{p_X(x)p_y(y)} dxdy\ . \tag{2}$$

The first term in the second member of eq. (2) is the entropy of X,

$$H(X) = -\int p_X(x) log\ p_X(x) dx, \tag{3}$$

while the second is the conditional entropy.

In information theory, $I(X,Y)$ is also known as the information transmission rate R of a continuous channel.

2.2 Interscale response fidelity

The measure of fidelity is known as the measure of distortion of the response provided by the two fields. This measure can be scalar, vector, or matrix. Without loss of generality, it will be assumed that the measure is a scalar, since the other measures must always be transformed into one or a set of scalar measures. For example, let the distance be given by,

$$d(X,Y) = (r(X) - r(Y))^2, \tag{4}$$

where $r(X)$ is the response due to the microscale and $r(Y)$ to the macroscale, usually functions implicitly defined by a numerical simulator. For each given stance defined as in eq. (4), there will be a need for its gradient relative to Y.There are no statistics in eq. (5), that is, the random functions $r(X)$: $R^N \to R$, e $r(Y)$: $R^N \to R$ must be understood as an ensemble, or set, of samples, or realizations, $r(X_i)$, $i = 1, \dots, n_R$. In this sense, $d(X,Y)$: $R^N \times R^N \to R^+$ is a collection of $N_R \times n_R$ samples in one dimension. Furthermore, it will be admitted, without loss of generality, that the microscale is represented only by a random variable Y. This means that the macromesh has a single element, cell or block, with a single property that, essentially, is the best stochastic representation. The distance, in this case, will be in the application of $d(X,Y)$: $R^N \times R \to R^+$.

It is important to note, at this stage, that no reference is made to the physics that prevails in the geometric domains. It is possible, for example, to evaluate $r(X)$ in complex carbonate physics, using Darcy-Stokes or Stokes-Brinkman with Darcyan fractures, or even in compositional models, and $r(Y)$ with Black-Oil physics.

The objective is twofold. First, find a compression scheme from X to Y that leads to a minimum distortion of the desired response, and then determine its value. Determining the optimal transformation involves evaluating the posterior distribution of $Y, p_Y(y)$. For the stochastic scheme to be complete, it is necessary to evaluate the distortion in a probabilistic way. One possibility is that the distortion, D, be the distance's mathematical expectation, but more meaningful functions for certain applications, such as percentiles and risk measures on the density's tails, can be chosen,

$$D = E[d(X,Y)]. \tag{5}$$

By hypothesis, the field describing the microscale is given by the ensemble of N_R geostatistical realizations. Thus, evaluation of the distortion, its expectation, or any other convenient stochastic measure, must be done with the Monte Carlo technique, or some simplifying variant. Since X is already sampled it is possible to write, using the relation between the joint density and the conditional density,

$$D = E[d(X,Y)] = \sum_{X,Y} p_{XY}(X,Y)d(X,Y) = \sum_{X,Y} p_X(X)p_{XY}(Y|X)d(X,Y), \tag{6}$$

where $p_{XY}(Y|X)$ is the conditional density of Y. Its minimum, with respect to $p_{XY}(Y|X)$, will be a Dirac's delta function that assigns to each x of X the Y that minimizes $d(X,Y)$.

This is an optimization problem that requires more computational effort than desired. Instead, we try to solve the augmented Lagrangian with the restriction that the indicator $I(X,Y)$ is less than or equal to a pre-specified value $I(X,Y)$. The constrained minimization problem is used to find using the Deterministic Simulating Annealing algorithm in a parallel form (see Costa [3,4] for details). The algorithm carries a high computational cost, but it finds the $p_{XY}(Y|X)$ in a nonparametric form, does not require any assumption about its functional form. The distribution $p_{XY}(Y|X)$ has the form of a Gibbs distribution.

The ensemble Y, and the probability of each one of the n_R coarse-scale realizations can solve any uncertainty reservoir simulation problem in inexpensive, accurate, and correct predictions. The macro ensemble model achieves an exceptional generalization for Neumann boundary conditions (different number and disposition of wells and their rates).

3 Applications

The above methodology is applied to a two-dimensional reservoir with a two-phase flow of oil and water and depletion by water injection.

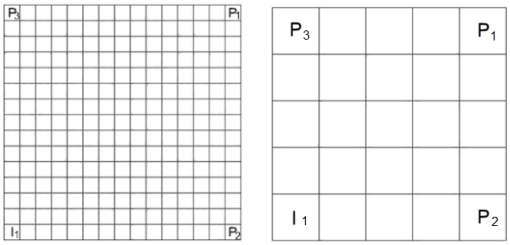

Figure 1 - Microscale (15x15) and microscale mesh (5x5)

The reservoir has a 15×15 cells upscaled for 5×5 cells as shown in Figure 1. The permeability realizations in the X, Y, and XY directions were estimated from a $Lognormal(245.5896, 304.3432)$, and the correlation matrix is given by:

$$\sigma := \begin{bmatrix} 1.50 & -0.595 & 0.85 \\ -0.5 & 2.00 & 0.65 \\ 0.5 & -0.80 & 2.00 \end{bmatrix}$$

The porosity is constant and equal to 0.25, the viscosity of water is $1\ cP$ and that of oil is $10\ cP$, the gravity of water is $1000\ kg/m^3$ and that of oil is $700\ kg/m^3$. The initial pressure in the reservoir is $1\ bar$ and the fluid is incompressible. There is one injection well, and four production wells, one in each corner. The injection well operates with a rate of $0.5\ m^3/day$. The wells operate at bottom-hole pressure of $1\ bar$. The production period is $600\ days$, with measurements taken every $120\ days$.

A set of $N_R = 1000$ micro realizations are given. The procedure starts with simulating 1000 realizations of the fine grid. These are the only simulations of the fine grid. 25 upscaling operations were performed, one for each macrocell, with Dirichlet boundary conditions. The final number of microscale realizations used was $n_R = 10$. The computational cost was significantly reduced.

The problem is determining the new field of absolute permeabilities for the coarse grid given in Figure 1. Each set of 3×3 cells are upscaled to one microcell.

The reservoir simulations used in this example were performed with MRST-2017 and the upscaling was done with the values taken from the simulations in a single-phase (water) problem when the velocities were in a steady-state regime.

Figure 3 shows the cumulative oil and water of the original given models and the upscaled models. To compare the impact of the proposed methodology, the flow of water and oil from the wells was analyzed at both scales.

It should be observed that the wells are located at the center of the corner cells in the two grids. This means there are substantial errors in the location of the four wells in the macro models. This induces some disagreements in the production over time. However, the cumulative production of the reservoir agrees well.

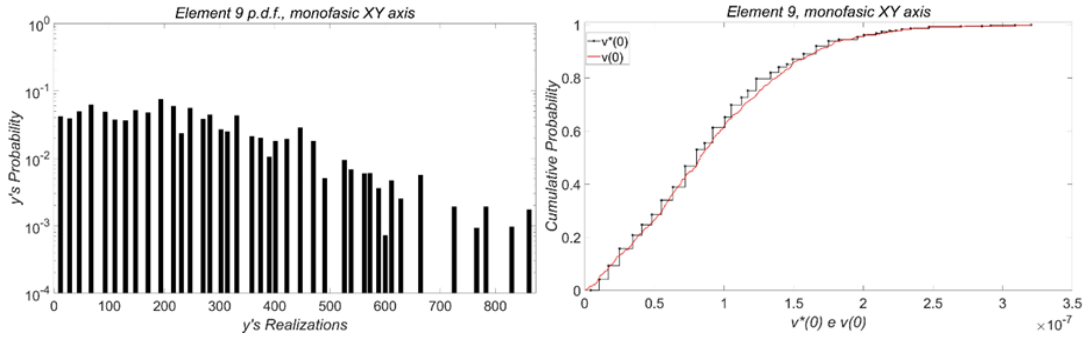

Figure 2: Quantized probability density and corresponding cumulative distribution of one macrocell.

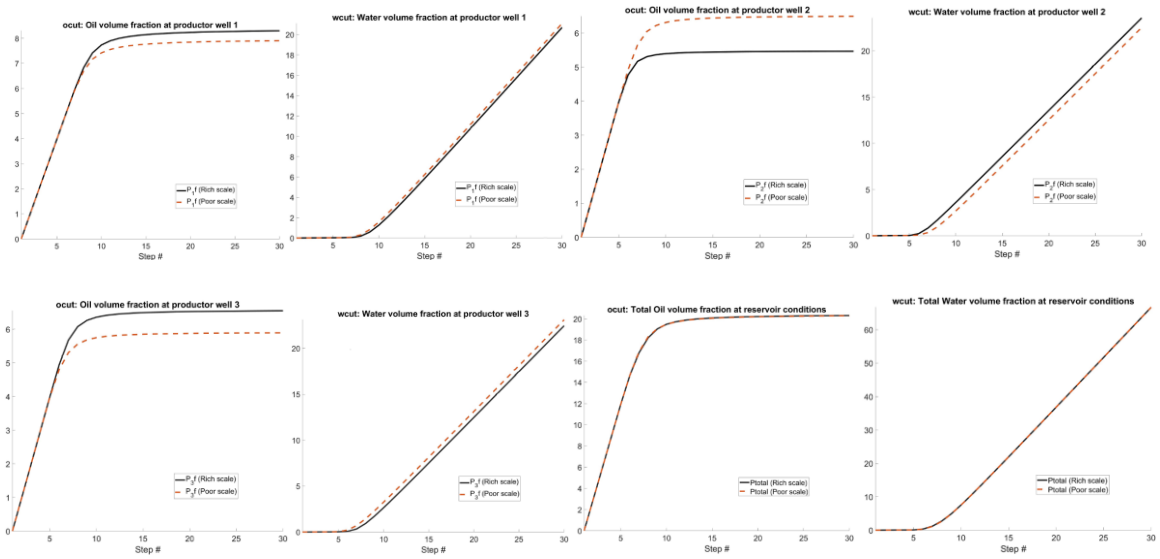

Figure 3 – Cumulative oil and water of the production wells.

4 Conclusions

The stochastic upscaling works well as shown in the two-dimensional application above and in similar small examples not shown in this paper. It remains to solve the problem of downscale to capture fluid velocities, saturations, and pressures in the microcells.

Acknowledgements. The Authors acknowledge the Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP) and Financiadora de Estudos e Projetos (FINEP), by the financial support provided by the Programa de Recursos Humanos da ANP – PRH-47.1 supported by investments by Petroleum firms qualified in Cláusula de P, D&I of Resolução ANP nº 50/2015, with support to scholarships and complementary funds for realizations of research and professional formation in Geology and Civil Engineering for the Petroleum and Gas Industries.

Authorship statement. The authors thereby confirm that they are the sole liable persons responsible for the authorship of this work and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors or has the permission of the owners to be included here.

References

[1] J. Chen, S. X. Xi, Weinan E., Z. Young. "BridgingTraditional and Machine Learning-based Algorithm for Solving PDEs: The Randon Feature Method", Journal of Machine Learning, 2022.
[2] j. Bornschein, F, Visin, S. Osindero. "Small Data, Big Decisions: Model Selection in the Small-Data Regime", ArXiv, 2020.
[3] I. F. Costa, L. A. Fonseca, E. R, Araújo. "Heterogeneous Stochastic Multiscale," CILAMCE2017, XXXVIII Ibero-Latin American Congresso on Computational Methods in Engineering held in Florianópolis, Santa Catarina, Brasil, from November 5th to 8th, 2017.
[4] I. F. Costa. "Um Método para Mudanças de Escalas Heterogêneas e Estocásticas., 2018.