XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGENEERING

Non-Intrusive Data-driven Surrogate Models for Predicting Turbidity Currents Deposition from 3D Simulations

Roberto M. Velho, Adriano M. A. Cortes, Renato N. Elias, Alvaro L. G. A. Coutinho
High Performance Computing Center - NACAD, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, roberto.velho@nacad.ufrj.br, adriano@nacad.ufrj.br, renato@nacad.ufrj.br, alvaro@nacad.ufrj.br

José J. Camata
Computer Science Department, Federal University of Juiz de Fora, Juiz de Fora, Minas Gerais, Brazil, jose.camata@ufjf.br

Fernando A. Rochinha
Department of Mechanical Engineering, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, faro@mecanica.coppe.ufrj.br

Gabriel M. Guerra
Department of Mechanical Engineering, Fluminense Federal University Rio de Janeiro, Niteroi, Brazil, gguerra@id.uff.br

Tiago H. F. Jesus, Thais C. A. Empinotti, Paulo L. B. Paraizo
Petrobras, Brazil, tiago.jesus@petrobras.com.br, thaisalmeida@petrobras.com.br, paraizo@petrobras.com.br

Abstract. Non-intrusive data-driven methodologies offer a powerful approach for constructing surrogate models to address complex parametric problems. POD-DL integrates deep neural networks and follows a multi-step dimensionality reduction process, beginning with a linear reduction via Proper Orthogonal Decomposition (POD) and followed by a nonlinear reduction using a deep autoencoder. In a previous work [1], we applied this scheme to study gravity currents under the 2D lock-exchange configuration [2]for multiple angles of the initial configuration. The angle of the channel served as a parameter, i.e., different angles generated different dynamics that the surrogate model learned. The regression neural network could then predict the dynamics for unseen angles. In the current work, we extend the methodology to more realistic scenarios, a 3D channel-basin configuration for turbidity currents, adapted from [3], analyzing variations in inlet velocities and sediment concentrations. Synthetic data were generated through large-scale parallel finite element simulations, via the EdgeCFD library [4], allowing the surrogate model to predict the quantities of interest for unseen parameter values.

Keywords: Model order reduction, data-driven models, machine learning, coupled systems, gravity currents, residual-based variational multi-scale (RBVMS).

1 Introduction

Turbidity currents, also known as gravity currents or density currents, are fluid flows driven by small differences in local density [5]. These density variations create pressure gradients that initiate and sustain the flow. Such differences may arise from changes in salinity, temperature, or the presence of suspended particles. When particles contribute to the density contrast, the resulting flows are referred to as particle-laden or particle-driven currents. Particle-driven flows are a crucial subclass of buoyancy-driven fluid motions, characterized by horizontal propagation as a denser fluid displaces a lighter one. A classic example is the release of a heavier fluid behind a lock gate: once the gate is lifted, the denser fluid sinks and spreads horizontally along the bottom, forming a gravity current. We name such an example as the *lock-exchange problem*. Remarkably, even subtle density contrasts can generate substantial movement, underscoring the importance of these flows in natural and engineered

systems. Density currents can dramatically alter flow dynamics, inducing sharp changes in velocity and generating regions of intense turbulence. These flows are prevalent across a broad spectrum of natural and human-influenced environments, often playing a pivotal role in the transport and dispersion of materials. In this sense, sedimentation promoted by particle-laden flows can mold the seabed, producing different geological structures such as canyons, dunes, and ripples. They typically develop strong turbulence, which directly impacts the particles' ability to move relative to the carrying fluid, to settle, or to be re-entrained. Depending on what prevails, settling or resuspension, the current and its turbulent structures might evolve in an entirely different manner, and consequently, those flow changes affect the transport of particles..

Understanding such phenomena and being able to reproduce them is crucial. One of the challenges is the high computational cost of evaluating a high-fidelity model (HFM), or full-order model (FOM), names used interchangeably throughout this text. The present work aims to build surrogate models to mitigate this drawback. We have previously studied surrogate models for the prediction of the fluid concentration in the Lock-exchange in the 2D case, see [1]. Now, we extend the methodology for a 3D problem of sediments being released on a channel-basin configuration. We vary two input parameters, the velocity of injection and the reference sediment concentration. As for the output, i.e., the quantities of interest, we measure the percentage of deposited material, the maximum longitudinal length the deposited material reaches, and the anisotropy ratio. We vary both input parameters, and we create FOM simulations for 60 input parametric pairs. We then apply two different strategies to assess the quality of our methodology, a Random Forest and a Neural Network. We measure the error of both techniques for the three quantities of interest and draw conclusions on the entire methodology.

2 Methodology

Here, we describe the physical setup, the data generation, and the machine learning models analyzed for the creation of the surrogate models.

2.1 Physical Modeling: phenomenological equations

This section presents the governing equations for turbidity currents developed by combining mass and momentum balances with rheological phenomenological models. Sediments are modeled as a continuum and described by the volumetric concentration. The spatial domain in which the flow takes place along the time interval $[0, t_f]$ is denoted by $\Omega \subset \mathbb{R}^{n_{dim}}$, where n_{dim} is the number of spatial dimensions, and Γ the boundary of Ω. A velocity-pressure non-conservative form of the Navier-Stokes (NS) equations describes the incompressible turbulent fluid flows carrying sediments in suspension. We assume that particle inertia and particle-particle interactions are negligible. Moreover, we also apply the Boussinesq hypothesis, which accounts for the fluid-particle interaction using a forcing term proportional to the local difference in the fluid density due to the presence of sediments. Fluid motion drives the sediment particles, but they are also endowed with extra mobility modeled by their settling velocity u_S, related to grain size sediment in the gravity direction, $\mathbf{e}^g$. Thus, an advection-diffusion equation models the sediment transport. The motion of each grain size, embedded in the mixture, is mapped to the fields $c = C/C_0$, the scaled concentrations, expressing the volume fraction occupied by each particle size. C and C_0 are, respectively, the actual concentration and the initial reference concentration or normalization value, the latter typically taken as the total initial volume fraction of the particles. Diffusion of the sediment is supposed to be small. Motivation for its inclusion in the modeling is often by numerical reasons [6]. Accordingly, the dimensionless equations that govern the particle-laden flow are:

$$\frac{\partial \mathbf{u}}{\partial \mathbf{t}} + \mathbf{u} \,.\nabla \mathbf{u} = -\nabla p + \frac{2}{\nu_f \sqrt{Gr}} \nabla.(\nu_m(c) \nabla^s \mathbf{u}) + \; c\, \mathbf{e}^g \;\; \text{in} \;\; \Omega \times [0, t_f] \tag{1}$$

$$\nabla \,.\, \mathbf{u} = \mathbf{0} \;\; \text{in} \;\; \Omega \times [0, t_f] \tag{2}$$

$$\frac{\partial c}{\partial t} + (\mathbf{u} + u_S \mathbf{e}^g).\nabla c = \nabla. \left(\frac{1}{Sc\sqrt{Gr}} \nabla c \right) \qquad \text{in} \;\; \Omega \times [0, t_f] \tag{3}$$

where $\mathbf{u}$, p and t are, respectively, non-dimensional velocity, pressure and time. Above, p, many times referred in the literature as the dynamic pressure, results after removing the hydrostatic component of the pressure. The rheological function $\nu_m(c)$ of the volumetric concentration, is the effective dynamic viscosity and u_S the particle

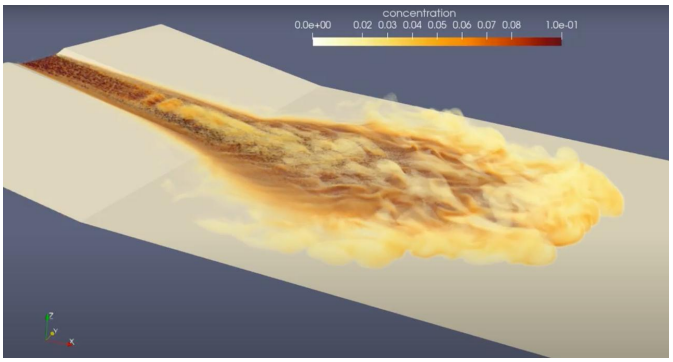


Figure 1. Domain and snapshot for a FOM simulation of the problem of interest, describing the deposition.

settling velocity acting in the direction of gravity $\mathbf{e}^g$. Gr is the Grashoff number that expresses the ratio between buoyancy and viscous effects given by:

$$Gr = \left(\frac{u_b H \rho_f}{\nu_f} \right)^2, \tag{4}$$

with ν_f and ρ_f are, respectively, dynamic viscosity and the fluid density, H is a characteristic length of the flow and the buoyancy velocity,

$$u_b = \sqrt{g H c_0 (\tilde{\rho}_p - \tilde{\rho}_f)/\tilde{\rho}_f},$$

where g stands for the gravity acceleration, $\tilde{\rho}_p$ and $\tilde{\rho}_f$ for, respectively, particles and fluid densities. The Reynolds number is such that $Re = Gr^2$. We assume that the different grains have the same density. A third dimensionless number, resulting from turning the governing equations into a non-dimensional form, is the Schmidt number, Sc, giving the ratio between diffusion and viscous effects:

$$Sc = \frac{\nu_f}{\kappa \rho_f} \tag{5}$$

where κ is the diffusion coefficient, supposed to be very small.

Essential and natural boundary conditions for Equation (1) are $\mathbf{u} = \mathbf{g}$ on $\Gamma_\mathbf{g}$ and $\mathbf{n} \cdot \left(-p\, \mathbf{I_d} + \frac{\nu_m}{\nu_f \sqrt{Gr}} \nabla \mathbf{u} \right) = \mathbf{h}$ on $\Gamma_\mathbf{h}$, where $\Gamma_\mathbf{g}$ and $\Gamma_\mathbf{h}$ are complementary subsets of the domain boundary Γ. Functions $\mathbf{g}$ and $\mathbf{h}$ are given, and $\mathbf{n}$ is the unit outward normal vector of Γ. A divergence-free velocity field $\mathbf{u}_0(\mathbf{x})$ is the initial condition for the velocity and $c_i(\mathbf{x}, 0)$ describing grain size composition and concentration of the suspended sediment in the beginning of the current have to be prescribed for the transport equation. For equation (3), boundary conditions modeling the transport of particles into and out the flow domain are $c = c_n$ on $\Gamma_n^{c_i}$, $\left(u_S \mathbf{e}^g\, c - \left(\frac{1}{Sc\sqrt{Gr}} \right) \nabla c \right) \cdot \mathbf{n} = 0$ on Γ_h^c, $\frac{\partial c}{\partial t} - u_S \nabla c \cdot \mathbf{n} = 0$ on Γ_{bottom} with $\Gamma = \Gamma_n^c \cup \Gamma_h^c \cup \Gamma_{bottom}$ and $\Gamma_n^c \cap \Gamma_h^c \cap \Gamma_{bottom} = \emptyset$.

The first condition describes the quantity of sediment entering in the flow domain. The second and third boundary conditions are enforced to reproduce physical mechanisms of particle motion through the remaining boundary, either by diffusion or advection. Sedimentation is allowed at the bottom on Γ_{bottom}. This last condition implies a loss of sediment but does not take into account any modification of the bottom geometry by deposition.

2.2 Data Generation and Processing

The solver for the sediment transport coupled problem employed here relies on a weak formulation based on the Residual Based Variational Multiscale Method (RBVMS) introduced within the context of Finite Element Stabilized Methods. RBVMS have been used with success in the simulation of turbulent flows [7] and [8], free-surface flows [9], and multi-transport [4], two fundamental aspects of the present problem. Details on how the governing equations are discretized can be found in [8]. Each simulation used 3.7 million mesh points, running over 96 CPU cores, during 23 hours, occupying approximately 23 GB. The domain of interest, as well as a snapshot of a FOM simulation, can be seen in Figure 1. Figure 2 shows how far in space can the deposits reach for different input parametric values given the same amount of volume injected. The angle of the basin is 4 degrees, while the channel one is 11 degrees. The inlet velocity was simulated in the range from 0.13 m/s to 0.37 m/s. Those correspond to a flow from 10.53 m^3/h to 30.0 m^3/h. The density used was of $2,650\ kg/m^3$ and $u_S = 1.797 \times 10^{-4}$ m/s. The range of the reference concentration was from 2.6% to 4.2%. Given both input parameters ranges, 60 simulations, pictured as parametric pairs in Figure 3, were generated.

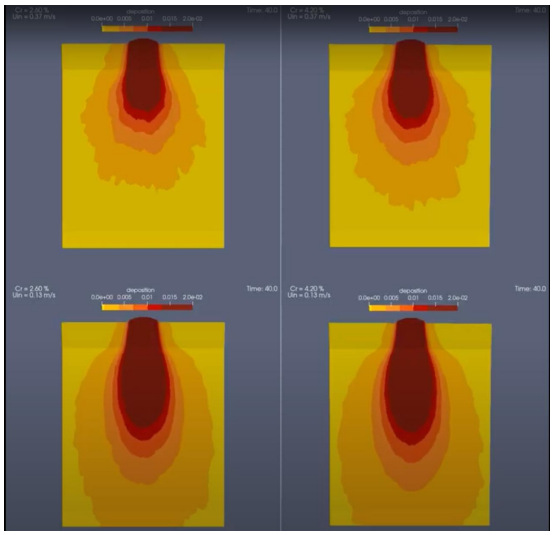


Figure 2. Examples of deposition for four different configurations of velocity of injection and reference concentration. They picture the different longitudinal and transversal lengths the material can reach, as well as its amount.

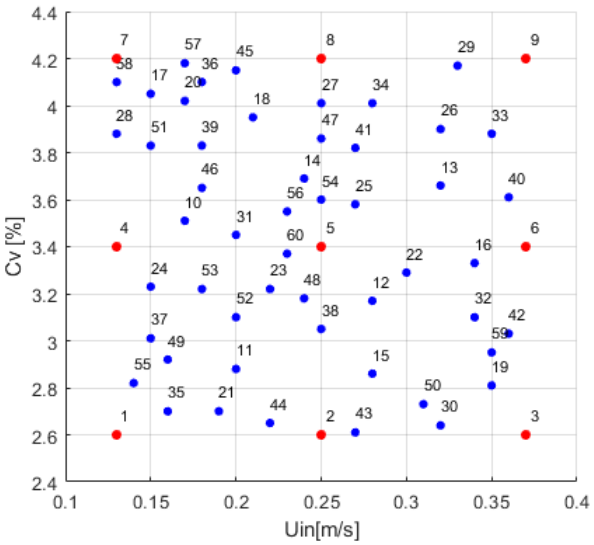


Figure 3. Grid points generated for uniformly distributed inlet velocity (Uin) ranging from 0.13 to 0.37 m/s and coefficient of variation (Cv) between 2.6% and 4.2%.

2.3 Description of the Technique

We collect data for three quantities of interest: the percentage of material deposited, the maximum longitudinal sediment deposit length reached by each simulation (for an equal amount of injected volume), and the anisotropy ratio. This last one, defined as the ratio between the maximum longitudinal sediment deposit length and the maximum transversal length. The collection and preprocessing procedure of such data is described in Figure 4.

We construct two different machine learning models for each of the quantities of interest. First, a random forest model using the Repeated K-fold technique. In this setting, all 60 points are utilized for both training and testing, with a random selection at each fold to determine the points allocated to either training or testing instances. The random forest regressor used 25 estimators with a maximum depth of 2. This configuration was the same for all three quantities of interest.

The second model was based on neural networks. We split the training (blue) and test (red) points as pictured in Figure 3. Such choice was driven by the desire of measuring the performance of the model over the center and boundary points of the considered parametric pairs space. We made use of the Scikit-learn library in Python and its Multilayer Perceptron routine. We chose a neural network with two hidden layers with 150 neurons each, a maximum number of iterations of 1000, and a tolerance of 1×10^{-4}. All other hyperparameters were the default

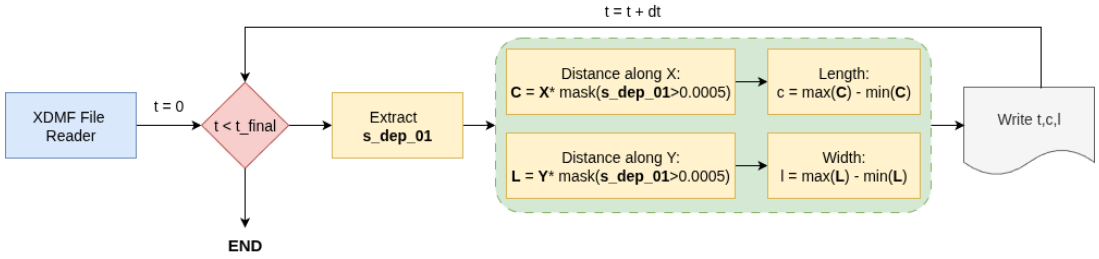


Figure 4. Diagram of data extraction and preprocessing.

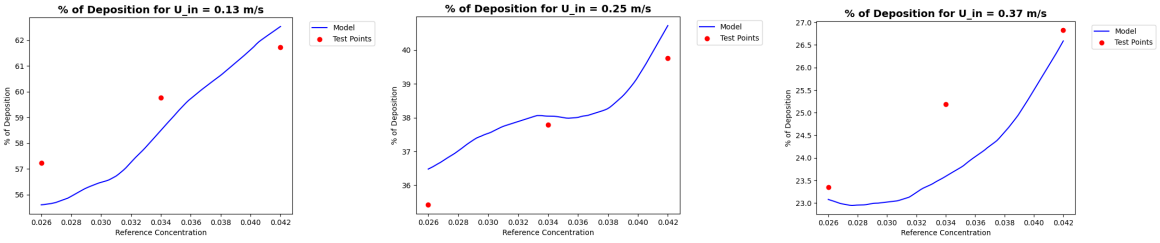
ones of such routine. We train the model with 51 parametric pairs and we test over the remaining 9. We perform the prediction of the model fixing one of the input variables (velocity of injection and reference concentration) while varying the other one. We show the prediction values over the entire range of the unfixed variable, over 100 points, while plotting the ground truth value of the test points. In such way, one can not only assess the absolute and relative errors over the test points but also regard the prediction over the entire range, observing the curve's shape the model predicts.

3 Results

For the random forest model, we run 15 different repetitions (K-fold technique) of the dataset and we assess the average (over the parametric pair points) percentage relative error, obtaining the values of 5.7%, 2.6%, and 1.8% for the percentage of sediment deposition, the maximum longitudinal sediment deposit length, and the anisotropy ratio, respectively.

For the neural network model, for the percentage of sediment deposition, one can regard the predictions in Figures 5 and 6. As for the maximum longitudinal sediment deposit length, results are shown in Figures 7 and 8. Finally, for the anisotropy ration, results are presented in Figures 9 and 10.

The produced surrogate models are able to perform predictions in less than a second, contrary to the FOM model in the order of almost a day.



((a)) Prediction and test points fixing the velocity of injection to 0.13 m/s.

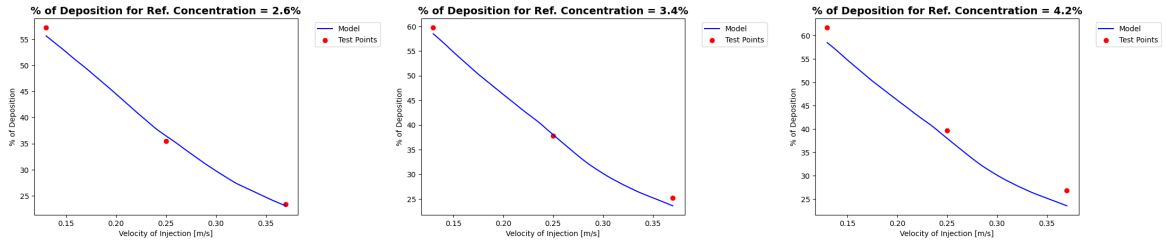
((b)) Prediction and test points fixing the velocity of injection to 0.25 m/s.

((c)) Prediction and test points fixing the velocity of injection to 0.37 m/s.

Figure 5. Prediction and ground truth for the percentage of deposition fixing the velocity of injection and varying the reference concentration.

4 Conclusions and Future Works

We observe both machine learning models provided predictions with different levels of error depending on which quantity of interest was addressed. Concerning the neural network case, we observe for all the three quantities of interest that the case fixing the reference concentration had a better agreement between predictions and test points, i.e., smaller error, while compared to fixing the velocity of injection. The model for the prediction of the maximum longitudinal length was the one with smaller errors, followed by the percentage of deposition model, and finally the anisotropy ratio one. This pattern is not followed by the random forest models. This could be by the fact in the neural network case we mainly test over extrapolation (in the parametric domain) points, while in the random forest case we average out results over the entire set of parametric points. Based on previous experiments with an smaller number of parametric points, we could observe the reduction of errors as the number of parametric points increase - what is expected under the theory of machine learning and suggests one should create a larger

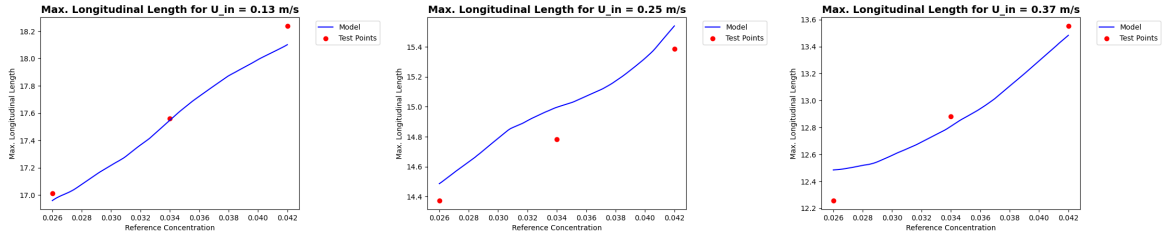


((a)) Prediction and test points fixing the reference concentration to 2.6%.

((b)) Prediction and test points fixing the reference concentration to 3.4%.

((c)) Prediction and test points fixing the reference concentration to 4.2%.

Figure 6. Prediction and ground truth for the percentage of deposition fixing the reference concentration and varying the velocity of injection.

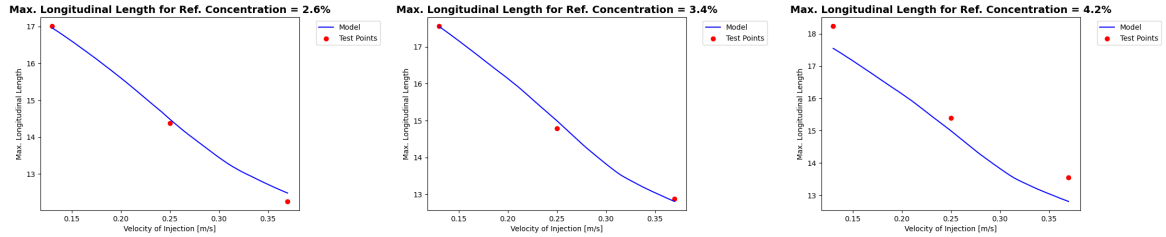


((a)) Prediction and test points fixing the velocity of injection to 0.13 m/s.

((b)) Prediction and test points fixing the velocity of injection to 0.25 m/s.

((c)) Prediction and test points fixing the velocity of injection to 0.37 m/s.

Figure 7. Prediction and ground truth for the maximum longitudinal length fixing the velocity of injection and varying the reference concentration.

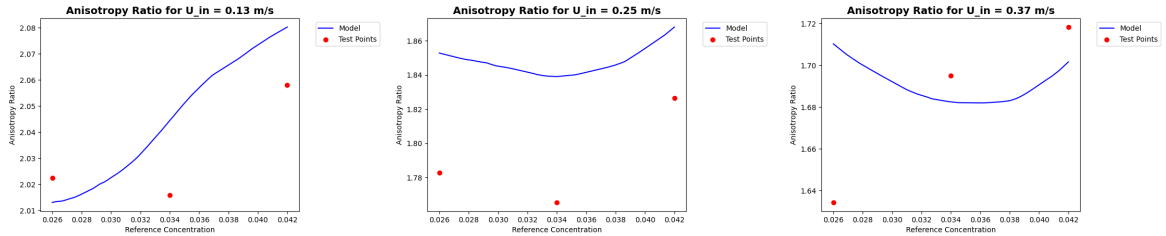


((a)) Prediction and test points fixing the reference concentration to 2.6%.

((b)) Prediction and test points fixing the reference concentration to 3.4%.

((c)) Prediction and test points fixing the reference concentration to 4.2%.

Figure 8. Prediction and ground truth for the maximum longitudinal length fixing the reference concentration and varying the velocity of injection.

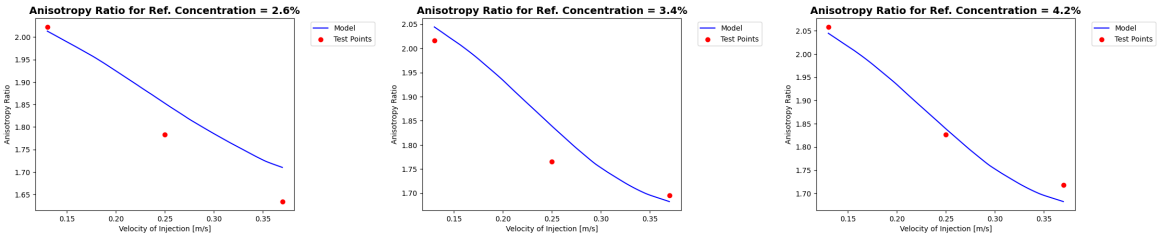


((a)) Prediction and test points fixing the velocity of injection to 0.13 m/s.

((b)) Prediction and test points fixing the velocity of injection to 0.25 m/s.

((c)) Prediction and test points fixing the velocity of injection to 0.37 m/s.

Figure 9. Prediction and ground truth for the anisotropy ratio fixing the velocity of injection and varying the reference concentration.



((a)) Prediction and test points fixing the reference concentration to 2.6%.

((b)) Prediction and test points fixing the reference concentration to 3.4%.

((c)) Prediction and test points fixing the reference concentration to 4.2%.

Figure 10. Prediction and ground truth for the anisotropy ratio fixing the reference concentration and varying the velocity of injection.

dataset of simulations (new parametric points). Regard the current dataset took about a month to be generated while using a high performance computing infrastructure.

One may expect future versions of such models, with more data, will be able to produce predictions with smaller errors and also give us the possibility of studying characteristics of the deposition profile for unseen values of the input variables, with speedups larger than 10^5 times. We also draw attention to the fact that, since all simulations for the parametric points are saved in disk, one can create new machine learning surrogates for other desired metrics over the deposition map, thus creating fast predictors without running one more time the entire set of simulations associated to the parametric points of interest. This also open the possibility of studying correlation between metrics, being it through the use of data from the simulated parametric pairs (fewer points but with full fidelity) or via the surrogate models (unlimited number of points but with reduced fidelity).

Acknowledgements. This study was financed in part by CAPES, Brazil - Finance Code 001. This work is alsopartially supported by FAPERJ, CNPq, and Petrobras.

Authorship statement. The authors hereby 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] R. M. Velho, A. M. Côrtes, G. F. Barros, F. A. Rochinha, and A. L. G. A. Coutinho. Advances in data-driven reduced order models using two-stage dimension reduction for coupled viscous flow and transport. *Finite Elements in Analysis and Design*, vol. 248, 2025.

[2] V. K. Birman, B. A. Battandier, E. Meiburg, and P. F. Linden. Lock-exchange flows in sloping channels. *Journal of Fluid Mechanics*, vol. 577, 2007.

[3] T. Spychala, J. T. Eggenhuisen, M. Tilston, and F. Pohl. The influence of basin setting and turbidity current properties on the dimensions of submarine lobe elements. *SEDIMENTOLOGY*, 2020.

[4] A. M. Cortes, E. F. Lins, G. M. Guerra, R. M. Silva, J. L. Alves, R. N. Elias, F. A. Rochinha, and A. L. Coutinho. Edgecfd: a parallel residual-based variational multiscale code for multiphysics. *International Journal of Computational Fluid Dynamics*, vol. 34, n. 7-8, pp. 529–548, 2020.

[5] E. Meiburg and B. Kneller. Turbidity currents and their deposits. *Annual Review of Fluid Mechanics*, vol. 42, n. 1, pp. 135–156, 2010.

[6] M. I. Cantero, S. Balachandar, M. H. García, and D. Bock. Turbulent structures in planar gravity currents and their influence on the flow dynamics. *Journal of Geophysical Research: Oceans*, vol. 113, 2008.

[7] Y. Bazilevs, V. Calo, J. Cottrell, T. Hughes, A. Reali, and G. Scovazzi. Variational multiscale residual-based turbulence modeling for large eddy simulation of incompressible flows. *Computer Methods in Applied Mechanics and Engineering*, vol. 197, pp. 173 – 201, 2007.

[8] G. M. Guerra, S. Zio, J. J. Camata, F. A. Rochinha, R. N. Elias, P. L. Paraizo, and A. L. Coutinho. Numerical simulation of particle-laden flows by the residual-based variational multiscale method. *International Journal for Numerical Methods in Fluids*, vol. 73, n. 8, pp. 729–749, 2013.

[9] E. F. Lins, R. N. Elias, F. A. Rochinha, and A. L. G. A. Coutinho. Residual-based variational multiscale simulation of free surface flows. *Computational Mechanics*, vol. 46, n. 4, pp. 545–557, 2010.