Advances in Geo-Energy Research

Vol. 8, No. 2, p. 71-75, 2023

Perspective

Intelligent modeling with physics-informed machine learning for petroleum engineering problems

Chiyu Xie[1,2], Shuyi Du[1,2], Jiulong Wang[2,3], Junming Lao[1,2], Hongqing Song[1,2]*

[1]*School of Civil and Resource Engineering, University of Science and Technology Beijing, Beijing 100083, P. R. China*

[2]*National & Local Joint Engineering Lab for Big Data Analysis and Computer Technology, Beijing 100190, P. R. China*

[3]*Computer Network Information Center, Chinese Academy of Sciences, Beijing 065007, P. R. China*

Keywords:
Physics-informed machine learning
petroleum engineering
data-driven
embedding mechanism

Cited as:
Xie, C., Du, S., Wang, J., Lao, J., Song, H. Intelligent modeling with physics-informed machine learning for petroleum engineering problems. Advances in Geo-Energy Research, 2023, 8(2): 71-75.
https://doi.org/10.46690/ager.2023.05.01

Abstract:
The advancement in big data and artificial intelligence has enabled a novel exploration mode for the study of petroleum engineering. Unlike theory-based solution methods, the data-driven intelligent approaches demonstrate superior flexibility, computational efficiency and accuracy for dealing with complex multi-scale, and multi-physics problems. However, these intelligent models often disregard physical laws in pursuit of error minimization, which leads to certain uncertainties. Therefore, physics-informed machine learning approaches have been developed based on data, guided by physics, and supported by machine learning models. This study summarizes four embedding mechanisms for introducing physical information into machine learning models, including input data-based embedding, model architecture-based embedding, loss function-based embedding, and model optimization-based embedding mechanism. These "data + physics" dual-driven intelligent models not only exhibit higher prediction accuracy while adhering to physic laws, but also accelerate the convergence to improve computational efficiency. This paradigm will facilitate the guide developments in solving petroleum engineering problems toward a more comprehensive and efficient direction.

1. Introduction

Petroleum engineering is an important field of engineering considering problems such as seismic exploration, well logging, production development, etc., which is essential for providing energy resources. Petroleum engineering problems are usually across multiple scales, with multi-physics coupling and multiple fluids. Researchers have developed many techniques to tackle the complex flow problems in petroleum engineering, such as molecular dynamics at the microscopic scale (Karplus and Petsko, 1990; Karplus and McCammon, 2002), Monte Carlo simulations (Metropolis and Ulam, 1949; Rubinstein et al., 2016) and lattice Boltzmann methods (Chen and Doolen, 1998) at the mesoscopic scale, computational fluid dynamics (Versteeg and Malalasekera, 2007; Hughes, 2012) at the continuous scale, and reservoir simulations at the reservoir scale (Peaceman, 2000). In recent years, breakthroughs in artificial intelligence and big data technologies (Liu et al., 2023) are providing a new mode to further enhance the study of petroleum engineering problems.

With the deep integration of big data, artificial intelligence, and petroleum engineering, the research methods are expanded into 4 models (Fig. 1): experimental investigation, theoretical analysis, numerical simulation, and digital intelligent modeling. Experimental investigations use all kinds of measurement approaches to obtain real and reliable data, then deduce and explain the flow phenomena and laws. Theoretical analysis is the process of analytically solving the mass conservation equations, energy equations, and state equations constructed by classical fluid mechanics theory under certain initial and boundary conditions. Numerical simulations solve the fluid motion either by the Euler or by the Lagrange approaches, which usually consume huge computational resources and rely on many assumptions. Digital intelligent modeling offers a

Yandy Scientific Press

*Corresponding author.
E-mail address: chiyuxie@ustb.edu.cn (C. Xie); 15702449699@163.com (S. Du); jlwang@cnic.cn (J. Wang); junminglao@xs.ustb.edu.cn (J. Lao); songhongqing@ustb.edu.cn (H. Song).

Received February 17, 2023; revised March 1, 2023; accepted March 9, 2023; available online March 12, 2023.

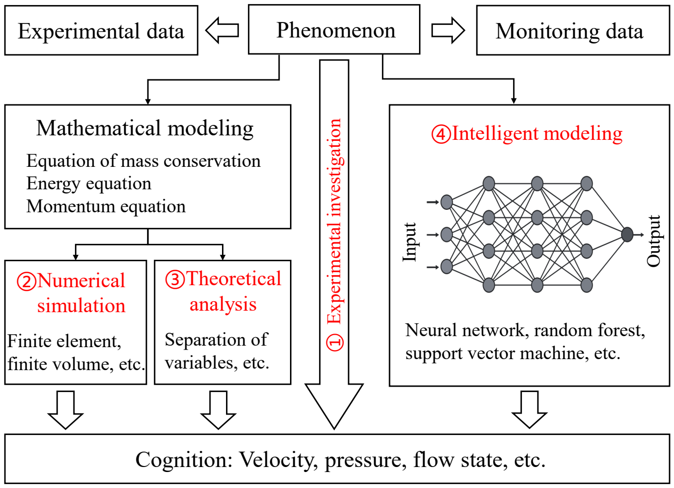

Fig. 1. Research paradigms of petroleum engineering.

new research paradigm, which learns the implicit mapping relationships between data and establishes an intelligent solver in a data-driven mode to achieve accurate flow prediction and flow field modeling without considering complex flow mechanisms and assumptions. Thus, it significantly improves the computational efficiency of numerical simulations by using parameter optimization and error correction. The three theory-driven approaches (Golparvar, 2018) described above can also provide physical guidance and data support for the intelligent model to enhance its prediction accuracy.

With the increasing maturity and rapid development of monitoring technologies in petroleum engineering, there have accumulated massive multi-source heterogeneous data bodies. Due to the complexity of the geological structure and flow mechanisms, it is difficult for traditional methods to adapt to the huge observation data. Data-driven methods have ultra-high computational efficiency, improving by several orders of magnitude compared to theory-based numerical simulations. However, these machine-learning methods exhibit a high dependence on data and uncertainty under certain conditions. This uncertainty is mainly reflected in the fact that the models can ignore the physical laws that the data itself needs to follow due to the pursuit of error minimization, which can lead to large errors with a butterfly effect-like effect on the inversion of subsequent parameters. To alleviate or even eliminate this phenomenon, “data + physics” dual-driven intelligent models have been developed by introducing physical information into machine learning. The physics-informed machine learning not only reduces the dependence of intelligent models on the data amount, but also improves prediction accuracy and computational efficiency through physical guidance. Therefore, the integration of a priori information and artificial intelligence to construct physics-informed machine learning is one of the inevitable trends for solving petroleum engineering problems.

2. Physics-informed machine learning

Physics-informed machine learning (Fig. 2) is based on data-driven machine learning models, employing direct or indirect methods to achieve the introduction of physical information for guiding the training of intelligent models. In general, these methods can be categorized into four mechanisms according to their different introduction modes: input data-based, model architecture-based, loss function-based, and model optimization-based embedding mechanisms. At present, a large number of researchers have constructed physics-informed machine learning models to solve complex problems such as flow field reconstruction, log interpretation, and production optimization in petroleum engineering.

2.1 Input data-based embedding mechanism

For physics-informed machine learning, input data-based embedding mechanism mainly consists of two types of methods: direct and indirect methods. The direct approaches are to introduce physical variables into the input data that are strongly correlated with the model output parameters through expert experience and existing physical knowledge, or to improve the quality of the input data through data filtering and cleaning so that it can show more powerful physical meaning. For example, Ling et al. (2016) explored that embedding invariance into the input features can significantly reduce the computer training cost in physical systems with symmetry or invarianceIn the process of reservoir development, Song et al. (2022b) used the physical equations of planar radial flow and spherical centripetal flow to complete the complementary input data including pressure and three-phase saturation to achieve accurate prediction of the remaining oil distribution. Du et al. (2023) performed outlier detection on production data of coalbed methane wells with the LOF-Xgboost framework, and the results revealed that the governance for the input data can significantly improve the robustness of the model for long-term production prediction.

The indirect approaches utilize reduced-order models or attention mechanisms to reduce or capture the heterogeneous data from multiple sources, allowing the model to extract and focus on the important physical information with limited resources indirectly. Shan et al. (2021) fused attentional mechanisms and convolutional neural networks to explore intelligent predictive models for complementary logging data. Zeng et al. (2020) developed an attention-based bidirectional gated recurrent unit model to achieve logging prediction as well as lithology identification. Input data-based embedding mechanism has achieved success in both theoretical and practical applications.

2.2 Model architecture-based embedding mechanism

Model architecture-based embedding mechanism contains two main approaches. The general approach involves introducing parameters with strong physical significance into the deep structure of the neural network, which enhances the learning capability of the model for those key physical parameters and reduces the adverse effects due to the differences in data dimensionality. For instance, Wu et al. (2018) introduced specific surface area and porosity parameters into a deep network, combining convolutional neural networks to rapidly predict permeability from two-dimensional digital core images. Additionally, Tang et al. (2022) incorporated

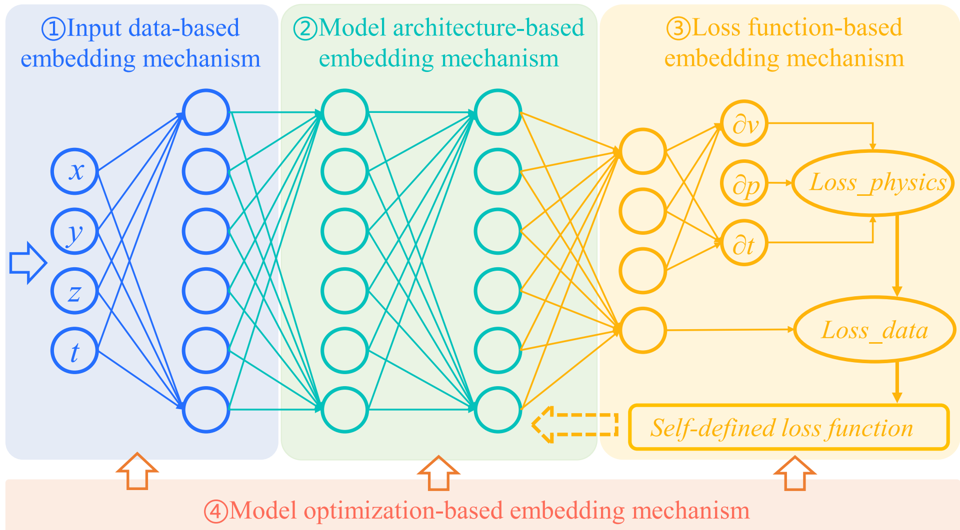

Fig. 2. Schematic diagram of the four embedding mechanisms for constructing physics-informed machine learning.

porosity and meander curvature into the Dense Block of the neural network and established a hybrid structure model to achieve three-dimensional core permeability prediction. Song et al. (2022a) extracted the static geological parameters of the reservoir and fracturing construction data into a deep network to establish an intelligent evaluation model of the fracturing effect, which can improve the prediction accuracy of fracture length by overcoming the problem that the network ignores the characteristics of static parameters due to the excessive number of dynamic parameter dimensions.

The other approach is to establish hybrid model architectures for the unique advantages possessed by different machine learning that can more fully extract temporal or spatial features in physical parameters. Choubineh et al. (2017) developed a hybrid neural network model to predict wellhead choke liquid critical-flow rates. Ashrafi et al. (2019) developed an integrated hybrid network architecture to predict penetration rates in drilling operations based on pump flow rate, pore pressure, bit rotational speed, density log, and shear wave velocity. Many other scholars have tapped the learning ability of convolutional neural networks for spatial features and long-short-term-memory networks for temporal data, creating ConvLSTM models to implement subsurface flow prediction, and reservoir development. Wei et al. (2022) exploited the ConvLSTM to accurately predict the future saturation distribution of carbonate reservoirs based on logging data and dynamic production data.

2.3 Loss function-based embedding mechanism

The loss function-based embedding mechanism is to introduce the governing equations, boundary conditions, initial conditions, and even expert experience in the physical laws into the neural network model by reconstructing the loss function to guide the training of the machine learning model. The key to a neural network's ability to self-learn is that it can iteratively update the weights and thresholds in the network through a backpropagation algorithm based on the error gradient calculated by the loss function. Therefore, the loss function is crucial for every training of the neural network. This mechanism is mainly derived from the physics-informed neural network (PINN) proposed by Raissia et al. (2017) and Raissia et al. (2019), which embeds the control equations into the loss function to solve the forward and inverse solution problem in fluid flow modeling. A large number of neural network models have been explored for flow field modeling via the loss function-based embedding mechanism. Yan et al. (2022) reconstructed the loss function by combining the pressure gradient operator and established a gradient-based neural network approach to achieve accurate prediction of pressure and saturation fields in multiphase flows under different geological conditions. Wang et al. (2020) considered engineering controls and expert experience to develop theory-guided neural network models that can rapidly predict the subsurface flow.

Besides, a great number of improved PINN algorithms have been applied to different fields such as biofluids and subsurface fluids. Chiu et al. (2022) coupled numerical differentiation and automatic differentiation to construct the can-PINN framework providing a potential alternative for fluid flow simulation. Zhang (2022) utilized physics-informed deep convolutional networks to predict transient Darcy flow under heterogeneous reservoir conditions. More PINN-based derivative frameworks such as XPINN, cPINN have been proposed for the solution of multi-scale flow problems (Kharazmi et al., 2019; Shukla et al., 2021). It is worth noting that the loss function-based embedding mechanism is mainly used in theoretical studies at present, and less for practical applications.

2.4 Model optimization-based embedding mechanism

Model optimization-based embedding mechanism is used to improve the learning ability of machine learning models for physical properties by introducing optimization algorithms. On the one hand, the embedding of the optimization mechanism can automatically adjust the hyperparameters, which are predefined parameters for the model. Some hyperparameters can change the structure of the model such as the number of hidden layers in the neural network and base learners in the random forest, while others do not change the architecture of the model but affect the learning efficiency (Sultana et al., 2022) such as learning rate, regularization parameters, etc. Therefore, many optimization algorithms have been integrated with machine learning to solve classical fluid mechanics problems, including stochastic search, genetic algorithms, particle swarm optimization, and simulated annealing. Kaydani et al. (2011) integrated particle swarm optimization and neural networks to construct an efficient intelligent model to achieve the prediction of minimum miscibility pressure in carbon dioxide (CO_2) injection. Zhang's group combined optimization algorithms and machine learning for the optimization of well location (Qi et al., 2022), real-time production (Wang et al., 2022), and injection (Xue et al., 2022). A coupled architecture (Owoyele et al., 2022) combining machine learning and genetic algorithms was employed to solve the fluid flow prediction problem.

The embedding of optimization algorithms can also optimize the weights or coefficients of the whole model. For example, Bayesian neural network models are developed, which utilized Bayesian algorithms to change the original weights in the neural network into a Gaussian distribution obeying a mean of μ and a variance of σ. Accurate prediction of liquid-phase diffusion coefficients based on Bayesian neural networks was performed by Mariani et al. (2020). Yue et al. (2011) employed Bayesian regularized back propagation networks to predict the oil-gas drilling costs. Sun et al. (2020) designed a Bayesian optimization neural network model based on physical constraints for reconstructing fluid flow in sparse and noisy data.

3. Conclusions

Intelligent modeling approaches are important for the future study of petroleum engineering. They have been successfully applied to many complex multi-physics and multi-scale problems. The physics-informed machine learning is the inevitable trend of intelligent modeling, which is a new research paradigm based on data, guided by physics, and supported by artificial intelligence models. This paper provides a comprehensive overview of four embedding mechanisms for introducing physical information into machine learning, including the input data-based embedding mechanism, model architecture-based embedding mechanism, loss function-based embedding mechanism, and model optimization-based embedding mechanism. These methods not only have higher prediction accuracy by following physical laws, but also accelerate convergence to significantly improve computational efficiency.

In the future, the modeling for complex hydrodynamic problems will be diversified, taking full advantage of the strengths of different research methods. The fusion modeling combining theory-driven and data-driven approaches will become the new research paradigm, which will accelerate the study of petroleum engineering problems and promote the theoretical system toward a more comprehensive direction continuously. It holds both powerful flow evolution capability and physical interpretability from theory-driven models, and the improved computational efficiency from data-driven approaches under the premise of guaranteed accuracy.

Acknowledgements

The authors gratefully acknowledge the financial support of this work from the National Natural Science Foundation of China (Nos. 52274027 and 11972073). This work was supported by China Postdoctoral Science Foundation (No. 2022M713204).

Conflict of interest

The authors declare no competing interest.

References

Ashrafi, S. B., Anemangely, M., Sabah, M., et al. Application of hybrid artificial neural networks for predicting rate of penetration (ROP): A case study from Marun oil field. Journal of Petroleum Science and Engineering, 2019, 175: 604-623.

Chen, S., Doolen, G. D. Lattice Boltzmann method for fluid flows. Annual Review of Fluid Mechanics, 1998, 30(1): 329-364.

Chiu, P. H., Wong, J. C., Ooi, C., et al. CAN-PINN: A fast physics-informed neural network based on coupled-automatic–numerical differentiation method. Computer Methods in Applied Mechanics and Engineering, 2022, 395: 114909.

Choubineh, A., Ghorbani, H., Wood, D. A., et al. Improved predictions of wellhead choke liquid critical-flow rates: Modelling based on hybrid neural network training learning based optimization. Fuel, 2017, 207: 547-560.

Du, S., Wang, J., Wang, M., et al. A systematic data-driven approach for production forecasting of coalbed methane incorporating deep learning and ensemble learning adapted to complex production patterns. Energy, 2023, 263: 126121.

Golparvar, A., Zhou, Y., Wu, K., et al. A comprehensive review of pore scale modeling methodologies for multiphase flow in porous media. Advances in Geo-Energy Research, 2018, 2(4): 418-440.

Hughes, T. J. R. The Finite Element Method: Linear Static and Dynamic Finite Element Analysis. New York, USA, Courier Corporation, 2012.

Karplus, M., McCammon, J. A. Molecular dynamics simulations of biomolecules. Nature Structural Biology, 2002, 9(9): 646-652.

Karplus, M., Petsko, G. A. Molecular dynamics simulations in biology. Nature, 1990, 347: 631-639.

Kaydani, H., Mohebbi, A., Baghaie, A. Permeability prediction based on reservoir zonation by a hybrid neural genetic algorithm in one of the Iranian heterogeneous oil

reservoirs. Journal of Petroleum Science and Engineering, 2011, 78(2): 497-504.
Kharazmi, E., Zhang, Z., Karniadakis, G. E. Variational physics-informed neural networks for solving partial differential equations. ArXiv preprint, 2019: 1912.00873.
Ling, J., Jones, R., Templeton, J. Machine learning strategies for systems with invariance properties. Journal of Computational Physics, 2016, 318: 22-35.
Liu, P., Zhang, K., Yao, J. Reservoir automatic history matching: Methods, challenges, and future directions. Advances in Geo-Energy Research, 2023, 7(2): 136-140.
Mariani, V., Pulga, L., Bianchi, G. M., et al. A Bayesian neural network methodology to predict the liquid phase diffusion coefficient. International Journal of Heat and Mass Transfer, 2020, 161: 120309.
Metropolis, N., Ulam, S. The monte carlo method. Journal of the American Statistical Association, 1949, 44(247): 335-341.
Owoyele, O., Pal, P., Vidal Torreira, A., et al. Application of an automated machine learning-genetic algorithm (AutoML-GA) coupled with computational fluid dynamics simulations for rapid engine design optimization. International Journal of Engine Research, 2022, 23(9): 1586-1601.
Peaceman, D. W. Fundamentals of Numerical Reservoir Simulation. Houston, USA, Elsevier, 2000.
Qi, J., Zhang, K., Xue, X., et al. An evolutionary sequential transfer optimization algorithm for well placement optimization based on task characteristics. SPE Journal, 2022, in press, https://doi.org/10.2118/212870-PA.
Raissi, M., Perdikaris, P., Karniadakis, G. E. Physics informed deep learning (Part I): Data-driven solutions of nonlinear partial differential equations. ArXiv preprint, 2017, 1711.10561.
Raissi, M., Perdikaris, P., Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 2019, 378: 686-707.
Rubinstein, R. Y., Kroese, D. P. Simulation and the Monte Carlo Method. Hoboken, USA, John Wiley & Sons, 2016.
Shan, L., Liu, Y., Tang, M., et al. CNN-BiLSTM hybrid neural networks with attention mechanism for well log prediction. Journal of Petroleum Science and Engineering, 2021, 205: 108838.
Shukla, K., Jagtap, A. D., Karniadakis, G. E. Parallel physics-informed neural networks via domain decomposition. Journal of Computational Physics, 2021, 447: 110683.
Song, H., Du, S., Yang, J., et al. Evaluation of hydraulic fracturing effect on coalbed methane reservoir based on deep learning method considering physical constraints. Journal of Petroleum Science and Engineering, 2022a, 212: 110360.
Song, H., Zhu, J., Wei, C., et al. Data-driven physics-informed interpolation evolution combining historical-predicted knowledge for remaining oil distribution prediction. Journal of Petroleum Science and Engineering, 2022b, 217: 110795.
Sultana, N., Hossain, S. Z., Abusaad, M., et al. Prediction of biodiesel production from microalgal oil using Bayesian optimization algorithm-based machine learning approaches. Fuel, 2022, 309: 122184.
Sun, L., Wang, J. X. Physics-constrained bayesian neural network for fluid flow reconstruction with sparse and noisy data. Theoretical and Applied Mechanics Letters, 2020, 10(3): 161-169.
Tang, P., Zhang, D., Li, H. Predicting permeability from 3D rock images based on CNN with physical information. Journal of Hydrology, 2022, 606: 127473.
Versteeg, H. K., Malalasekera, W. An Introduction to Computational Fluid Dynamics: The Finite Volumn Method. London, UK, Pearson Education, 2007.
Wang, N., Zhang, D., Chang, H., et al. Deep learning of subsurface flow via theory-guided neural network. Journal of Hydrology, 2020, 584: 124700.
Wang, Z., Zhang, K., Chen, G. D., et al. Evolutionary-assisted reinforcement learning for reservoir real-time production optimization under uncertainty. Petroleum Science, 2022, in press, https://doi.org/10.1016/j.petsci.2022.08.016.
Wei, C., Huang, R., Ding, M., et al. Characterization of saturation and pressure distribution based on deep learning for a typical carbonate reservoir in the Middle East. Journal of Petroleum Science and Engineering, 2022, 213: 110442.
Wu, J., Yin, X., Xiao, H. Seeing permeability from images: Fast prediction with convolutional neural networks. Science Bulletin, 2018, 63(18): 1215-1222.
Xue, X., Chen, G., Zhang, K., et al. A divide-and-conquer optimization paradigm for waterflooding production optimization. Journal of Petroleum Science and Engineering, 2022, 211: 110050.
Yan, B., Harp, D. R., Chen, B., et al. A gradient-based deep neural network model for simulating multiphase flow in porous media. Journal of Computational Physics, 2022, 463: 111277.
Yue, Z., Songzheng, Z., Tianshi, L. Bayesian regularization BP Neural Network model for predicting oil-gas drilling cost. Paper Presented at 2011 International Conference on Business Management and Electronic Information, Guangzhou, China, 13-15 May, 2011.
Zeng, L., Ren, W., Shan, L. Attention-based bidirectional gated recurrent unit neural networks for well logs prediction and lithology identification. Neurocomputing, 2020, 414: 153-171.
Zhang Z. A physics-informed deep convolutional neural network for simulating and predicting transient Darcy flows in heterogeneous reservoirs without labeled data. Journal of Petroleum Science and Engineering, 2022, 211: 110179.