URTeC: 5059

A Physics-Guided Deep Learning Predictive Model for Robust Production Forecasting and Diagnostics in Unconventional Wells

Syamil Mohd Razak*[1], Jodel Cornelio[1], Young Cho[1], Hui-Hai Liu[2], Ravimadhav Vaidya[2], Behnam Jafarpour[1]; (1 University of Southern California, 2 Aramco Americas).

Abstract

Forecasting the future performance of unconventional wells is essential for reserves estimation, production optimization and to support decision-making. In this paper, we present a physics-guided deep learning predictive model that forecasts production profiles given an input of completion, formation, and fluid properties. The predictive model accounts for prediction errors originating from potentially missing physical phenomena, simulation inputs, and incomplete reservoir description by correcting the production profiles obtained from a physical simulator. Using the multi-task learning concept, the predictive model creates a shared representation that can also be used to forecast production behavior, estimate reserves, and as a diagnostic tool to understand well performance under varying well properties. Compared to purely data-driven predictive models, it also allows for improved prediction for when well properties (as input) are outside the range of the original training data. The developed method is tested using field data from the Bakken Shale Play in North Dakota. A physics-based simulator is used to obtain simulated production responses using relevant completion, formation, and fluid properties from wells in Bakken. The deep learning architecture is jointly trained with the field data and simulated data. For any unseen input well properties (test data), the learned feature space (i.e., shared representation) provides a reliable diagnosis of potential reserves, oil rate decline trend, and other well performance metrics. The multi-task predictive model also provides robust production forecasts when compared to a purely physics-based/data-driven model or individual (single-task) predictive models.

Introduction

Hydrocarbon production from hydraulically fractured unconventional reservoirs involves complex flow and transport processes that are not well understood. Existing simulation methods are based on theories and mathematical models rooted in the flow and transport phenomenon of conventional reservoirs and may introduce substantial errors when directly applied for unconventional resources (Trangenstein et al., 1989; Aziz et al., 1979). Modeling flow behavior and physical processes that take place during production from extremely tight formations with complex fracture networks is an active research area. To accurately represent the physical relationship between well properties (i.e., formation, completion, and fluid properties) and production responses, a simulation model must be able to capture the wide range of uncertainty in model parameters, such as the interaction between natural and induced fractures, coupled with complexities in modeling geomechanical interaction that can adversely affect predictions obtained from a simulator.

Notwithstanding these complexities, conventional reservoir simulation models remain a popular choice for modeling unconventional reservoirs (Altman et al., 2020) as certain aspects of simulation models may still be useful and relevant for describing the production behavior in unconventional wells and ultimately obtain production forecast to facilitate field development and management decisions. Alternatively, data-driven analytics can provide predictive models based on statistical relations identified from collected field measurements that may be incomplete, noisy, and erroneous. However, statistical models, cannot extrapolate outside the training data set and do not offer the opportunity to incorporate domain constraints. In data-driven modeling, a trained model extracts salient features from the input to provide predictions as the output.

Under the concept of multi-task learning, a statistical predictive model can be designed to create a shared feature representation for multiple tasks (i.e., to provide multiple outputs). Training a predictive model to solve multiple related tasks simultaneously using labeled data can improve the generalization power of the model as well as reduce the expensive cost of data acquisition and curation (Wu et al., 2020). In our application, the tasks of predicting simulation errors, cumulative oil production, and the likelihood of success for any given well are similar. In a multi-task predictive model, the domain-specific information contained in the training data serves as an inductive bias to help other tasks be learned better when compared to multiple separate predictive models working independently (Ruder, 2017).

We propose a deep convolutional neural network architecture for enhancing simulation predictions that take well properties (completion, formation, and fluid properties) as the input to predict the simulation error (i.e., the difference between simulated responses and field observations). Once trained, the model is applied to forecast the production response of a well by augmenting physics-based predictions with the learned prediction errors from the deep learning model. Additionally, the neural network architecture branches out from the learned shared feature representation layer to fully connected layers that output other metrics such as cumulative oil production and the likelihood of success. The data we use for our model consists of (1) well formation, completion and fluid properties (denoted as $\boldsymbol{x}$), (2) the corresponding simulated production responses as $\boldsymbol{d}_{sim}$ and observed field production data as $\boldsymbol{d}_{field}$ that are used to calculate the simulation error $\boldsymbol{d}_{err}$ (3) cumulative oil produced as scalar values (denoted as y) and (4) class label (denoted as $\boldsymbol{c}$) consisting of 3 classes (i.e., low, mid, and high performing wells) represented as one-hot vector encodings. In the remainder of this paper, we present the model in more detail and a series of examples to demonstrate its application to field data.

Methods

Artificial neural networks are abstract mathematical functions modeled loosely after the operations of a human brain and consist of interconnected nodes to mimic the information processing behavior of neurons. They can be trained to recognize complex patterns and relations from data and use them to generate multiple types of prediction for classification and/or regression tasks. Variants of neural networks have been applied in the petroleum engineering domain for reservoir history matching (Mohd Razak and Jafarpour, 2020a; 2020b), log interpretations and facies classification (Bressan et al., 2020), production prediction using multi-disciplinary data (Bhattacharya et al., 2019), fracture characterization (Malallah and Nashawi, 2005) and production optimization (Zhong et al., 2020). Recent studies on production prediction for unconventional reservoirs include the application of random forest regression to forecast shale gas production (Xue et al., 2021) and deep neural networks to forecast production in Bakken shale reservoirs (Wang et al., 2019). The integration of physical knowledge and insight into neural networks to improve their predictive capability for physical systems is currently an active research area (Raissi et al., 2019; Goswami et al., 2020; Klie and Florez, 2020; Park et al., 2020; Zhang et al., 2018).

The schematic of the proposed predictive model is illustrated in **Figure 1**, where components I_θ, O_ζ and O_ω are composed of multiple fully-connected layers with nonlinear activations (i.e., leaky-ReLU function). The component I_θ takes well input properties $\boldsymbol{x} \in R^{N_x}$ where N_x denotes the number of well properties available/collected. The component O_ψ consists of a one-dimensional (1D) convolution operation, leaky-ReLU non-linear activation function and a one-dimensional up-sampling function. For time series, one-dimensional convolutional operation extracts local temporal features in the time series. In this work, successive convolution and up-sampling operations gradually reconstruct the simulation error $\boldsymbol{d}_{err} \in R^{N_t \times N_f}$ where N_t denotes the data timesteps and N_f denotes the number of features (i.e., simulation error for oil, water, and gas phases). The last fully-connected layer of the component O_ω outputs a scalar value $y \in R^1$ that corresponds to the total cumulative oil production. The component O_ζ outputs a vector $\boldsymbol{c} \in R^{N_c}$ where N_c denotes the number of class labels (i.e., one of low/mid/high performing).

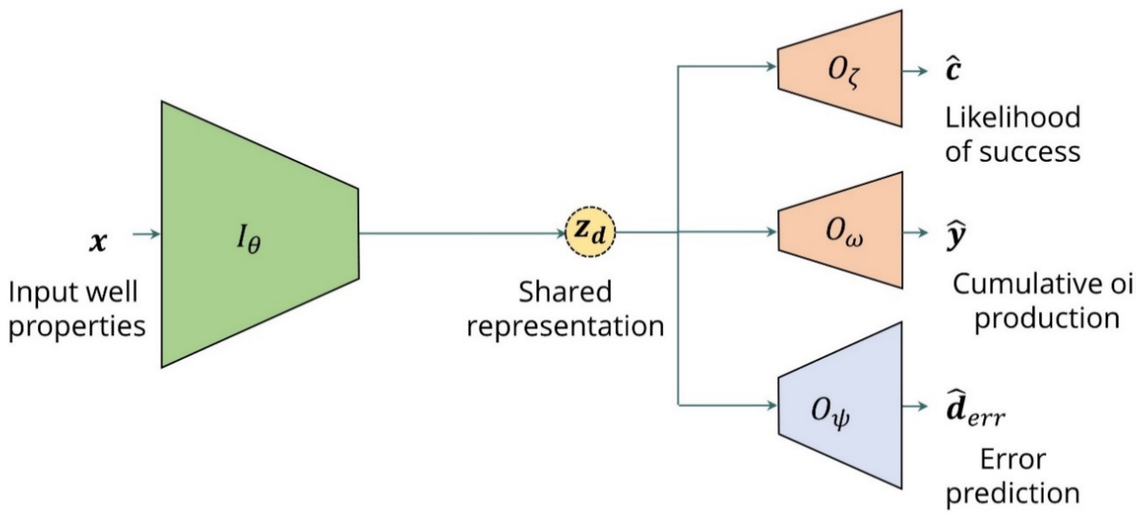

Figure 1. Schematic of neural network for the predictive models.

The predictive model is implemented with the deep learning library Keras (version 2.2.4) (Chollet et al., 2015) in the Python programming language. For more details on the mechanism of each function, we refer the readers to relevant literature in computer science (e.g., Ramsundar and Zadeh, 2018; Chollet et al., 2015). The salient information from the input well properties is formalized in an informative feature space (i.e., shared representation) that provides insights on potential reserves and production characteristics of any given well and can be used as a diagnostic tool. The loss function used for the regression outputs of O_ψ and O_ω is the mean-squared error (i.e., $\mathcal{L}_2$-loss) while categorical cross-entropy loss is used for the classification output of O_ζ after a soft-max operation is applied. The trainable weights in each component of the neural network are optimized with the following multi-task loss function (for any input tuple)

$$\mathcal{L}(\theta,\zeta,\omega,\psi) = \lambda_\psi \left\| \boldsymbol{d}_{err} - O_\psi\left(I_\theta(\boldsymbol{x})\right)\right\|_2^2 + \lambda_\omega \left\| y - O_\omega\left(I_\theta(\boldsymbol{x})\right)\right\|_2^2$$
$$+\lambda_\zeta \sum_i^{N_c} c_i \log\left(softmax\left(O_\zeta\left(I_\theta(\boldsymbol{x})\right)\right)\right)_i \qquad (1)$$

where λ_{ψ}, λ_{ω} and λ_{ζ} represent the relative importance of each loss function (i.e., tasks). Note that as the tasks are highly related to each other, a minimization update of one loss function typically results in a decrease in objective function value for another task.

Results

The Bakken field data (i.e., $\boldsymbol{x}$ and $\boldsymbol{d}_{field}$) used in this work is downloaded from the North Dakota Department of Mineral Resources web page. The input well properties $\boldsymbol{x}$ include the length of perforation interval, the volume of proppant injected, the weight of proppant, treatment pressure, treatment rate, number of stages, rate of penetration, and gamma-ray readings. We assume that there are no inter-well communications, and no distinction is made between wells that are producing from the Middle Bakken formation and the Three Forks formations. The corresponding simulated production response $\boldsymbol{d}_{sim}$ is obtained using a physics-based reservoir simulator where $\boldsymbol{d}_{err}$ is the difference between $\boldsymbol{d}_{field}$ and $\boldsymbol{d}_{sim}$ and is calculated for each fluid phase (i.e., oil, water, and gas). For each data point (i.e., well), the cumulative oil production is calculated by summing the oil production for the entire production period. The class label for the likelihood of success consists of three classes (i.e., low, mid, and high performing), and the label for each well is determined by thresholds of cumulative oil production and other subjective metrics.

In this experiment, the predictive model is trained until convergence using Equation 1. Once the model is trained, the predicted profiles are obtained by computing $\widehat{\boldsymbol{d}}_{err} = O_{\psi}\big(I_{\theta}(\boldsymbol{x})\big)$ followed by $\widehat{\boldsymbol{d}}_{field} = \boldsymbol{d}_{sim} + \widehat{\boldsymbol{d}}_{err}$. The normalized scatter plots in **Figure 2** show good prediction performance of the model on training and test datasets. The small prediction RMSE for the testing dataset indicates that the predictive model is effective and can generalize when used for unseen data points. The scatter points in Figure 2 does not fall perfectly on the unit slope due to substantial noise observed in $\boldsymbol{d}_{field}$. However, we notice that the predictive model can discern signal from noise as the generated predictions do not contain a similar type of noise observed in the training data.

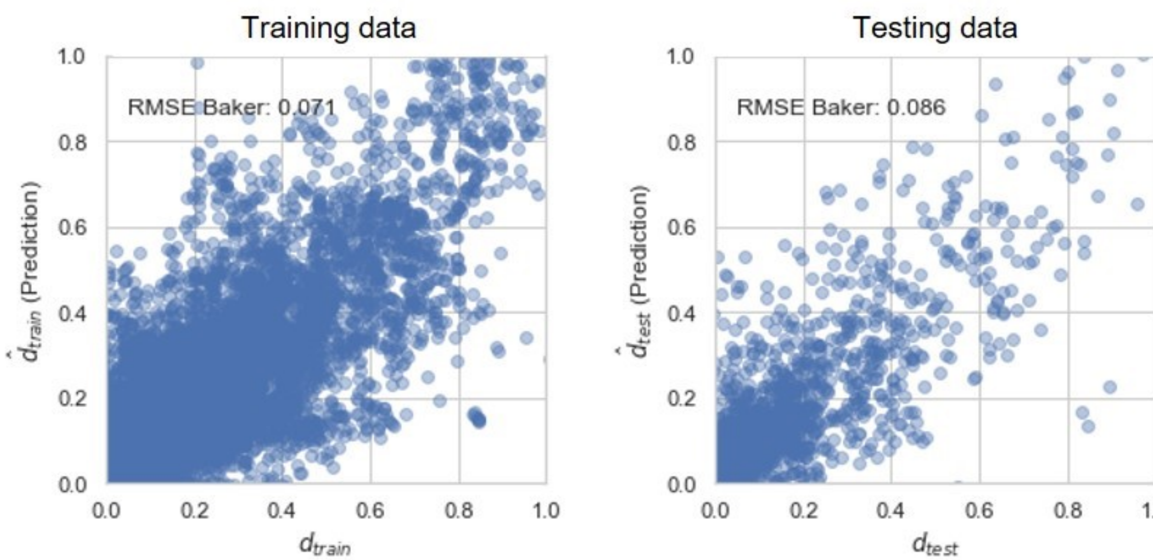

Figure 2. Scatter plot of data match of a predictive model for simulation error.

In **Figure 3**, two representative test cases are shown to illustrate the proposed approach. The scatter points represent the reference $\boldsymbol{d}_{field}$ and the stippled lines represent the $\boldsymbol{d}_{sim}$. Due to errors that may have come from undiscovered physics or imperfect description of unconventional reservoirs, $\boldsymbol{d}_{sim}$ almost always exhibit some discrepancies with $\boldsymbol{d}_{field}$. In Figure 3, these discrepancies, $\boldsymbol{d}_{err}$, are illustrated by the filled area between $\boldsymbol{d}_{field}$ and $\boldsymbol{d}_{sim}$ as under-estimations and over-estimations. The bold lines represent $\widehat{\boldsymbol{d}}_{field}$ and show better agreement with $\boldsymbol{d}_{field}$ when the predicted simulation errors in $\boldsymbol{d}_{sim}$ are accounted for. The predicted profiles shown for the two sample cases belonging to the testing dataset show a satisfactory match after the original physics-based simulation $\boldsymbol{d}_{sim}$ is corrected. The predictive model leverages the power of deep learning to account for systematic prediction inaccuracies due to incomplete knowledge about the reservoir model and the underlying flow processes.

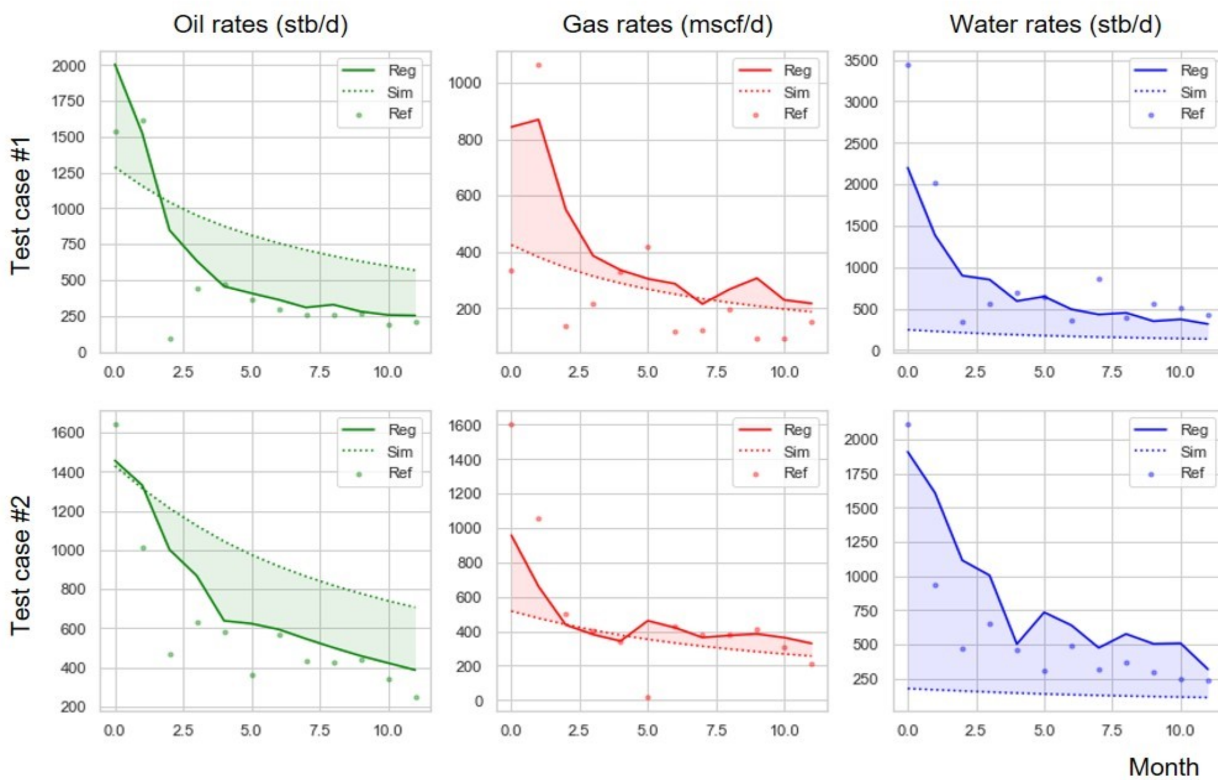

Figure 3. Match of restored profiles from a predictive model for simulation error.

The shared representation layer also includes information that can be used to predict the total cumulative oil production as well as the performance class label. The shared features represent salient information used in constructing regression planes for regression tasks and decision boundaries for classification tasks. The confusion matrices in **Figure 4** show ~95% accuracy for the classification task where each well in the training and testing dataset is confidently categorized as being a low, mid, or high performing well (by taking the class with the highest probability as the label). Note that the prediction vector $\hat{\boldsymbol{c}}$ represents the probability of a well belonging to each category and can provide a confidence measure of success. Misclassification of the wells observed in **Figure 4** may be attributed to inconsistent data and the complexity in mapping well properties to class labels. **Figure 5** shows the scatter plots of predicted cumulative oil production versus the reference values for the training and testing dataset. The scatter points fall on the unit slope indicating that the shared features can effectively predict the defined diagnostic metric and can generalize for unseen test data points, as observed by the scatter plot for the testing dataset.

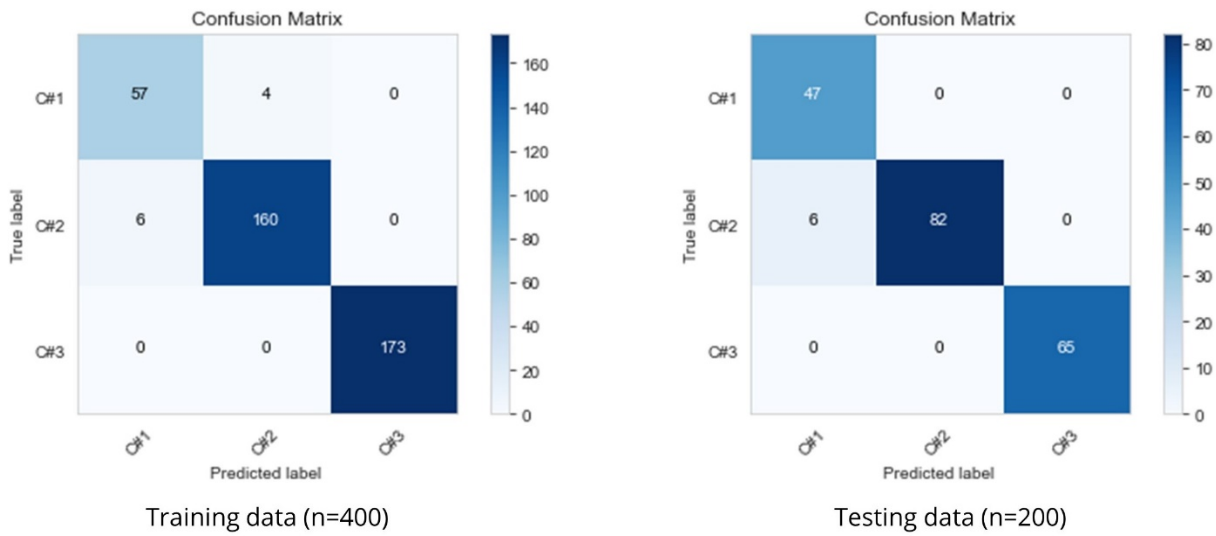

Figure 4. Confusion matrix for the training and testing dataset.

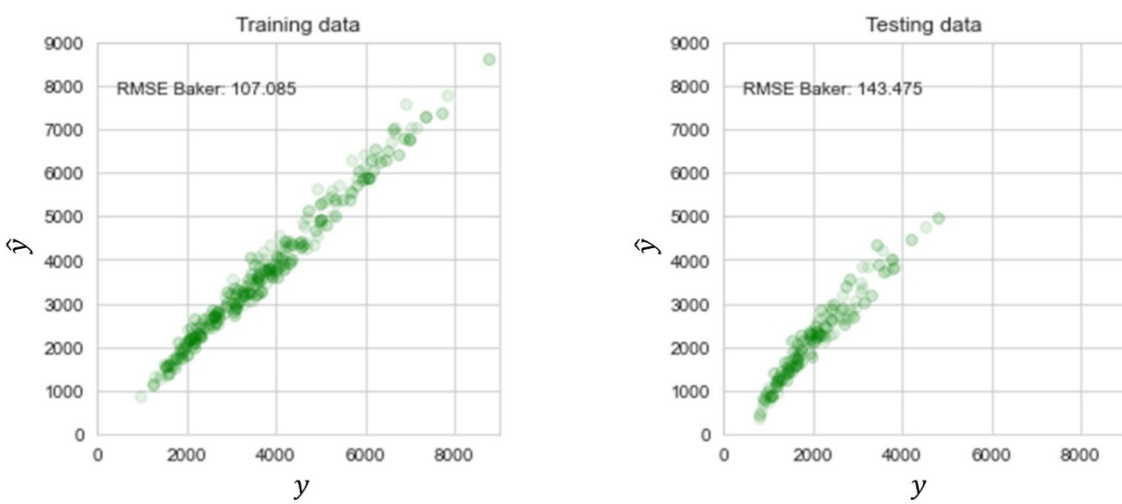

Figure 5. Scatter plots of predicted cumulative oil production versus reference for the training/test dataset.

Discussions and Conclusions

Flow simulation models for tight formations with complex fracture networks are still in need of long-term fundamental research. More understanding of the physical processes that take place in these complex systems is necessary to sufficiently represent the flow and transport processes. In the meantime, the fast development phase of unconventional resources necessitates new methods of modeling well performance. The proposed predictive model combines the strengths of data-driven methods and physics-based methods. Purely data-driven models have limited ability to generalize beyond the range of data used in their training but can capture complex hidden patterns. A purely physics-based model may be based on the imperfect physical model but provides causal predictions for any range of input parameters.

The deep learning predictive model learns the discrepancy between simulated and observed production data and uses it to enhance the accuracy of simulation-based predictions. Additionally, with multi-task learning, the predictive model can be designed to predict other metrics of success where the salient information is extracted into a shared representation feature space. Our predictive model comprises additional components tailored for classification and regression tasks, all of which are trained simultaneously. In **Figure 6**, we compare the performance of the predictive model in multi-task and single-task mode on the testing dataset. For the single-task mode, each of the models is trained separately until convergence without knowledge of other tasks. For all the tasks, we consistently observe that multi-task learning helps improve the prediction performance on the testing dataset in terms of the root-mean-square error and classification accuracy. This shows that multi-task learning can improve the generalization power of the model and potentially reduce the number of training data needed.

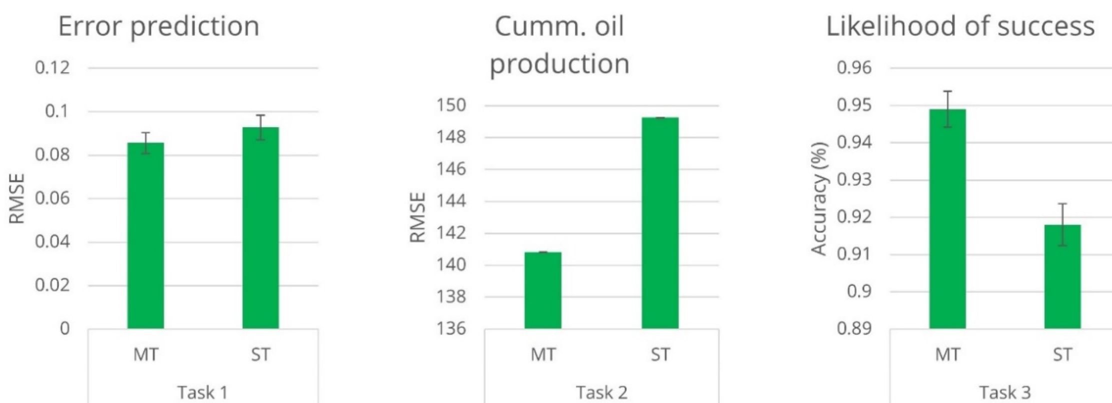

Figure 6. Performance comparison between multi-task (MT) and single-task (ST) learning for each task.

The combination of physics-based and deep learning models provides an opportunity to enhance the extrapolation capability of data-driven models using physics-based simulation data. The preliminary results from our experiments show that a hybrid multi-task predictive model can provide robust production forecast and performance diagnostics for unconventional wells, especially when presented with unseen test datasets. In this work, we tested the proposed approach using limited data from the Bakken Shale Play, however, a larger training and testing dataset are needed to obtain more concrete conclusions. Additionally, the performance of the proposed workflow needs to be validated for scenarios where inter-well communications may exist and for when well stimulation or intervention activities are done.

The rapid development of unconventional resources and the lack of an in-depth understanding of the flow mechanism and physical processes in fractured tight formations motivate the need for modeling methods beyond the traditional tools. However, the availability of abundant data from many unconventional wells that typically have limited interference enables the development of data-driven tools for predicting production performance. The hybrid multi-task workflow we propose takes advantage of the main strength of physics-based models that use causal relations to provide predictions for any ranges of input parameters and data-driven models to enhance the predictions. Our proposed hybrid multi-task approach leverages recent advances in deep learning to combine the existing understanding of flow physics in the unconventional domain with abundant data collected from unconventional wells in the past decade.

Acknowledgments

The USC authors acknowledge the Energi Simulation support of the Center for Advanced Reservoir Characterization and Forecasting at USC.

References

Altman R., Tineo R., Viswanathan A., Gurmen N. (2020) Applied Learnings in Reservoir Simulation of Unconventional Plays. *Society of Petroleum Engineers*.

Aziz K., Settari A. (1979) Petroleum Reservoir Simulation.

Bhattacharya S., Ghahfarokhi P. K., Carr T. R., Pantaleone S. (2019) Application of predictive data analytics to model daily hydrocarbon production using petrophysical, geomechanical, fiberoptic, completions, and surface data: A case study from the Marcellus Shale, North America. Journal of *Petroleum Science and Engineering*. 176. 702 – 715.

Bressan T. S., Souza M. Kehl de, Girelli T. J., Junior F. C. (2020) Evaluation of machine learning methods for lithology classification using geophysical data. *Computers & Geosciences*.

Chollet F., others. (2015) Keras.

Goswami S., Anitescu C., Chakraborty S., Rabczuk T. (2020) Transfer learning enhanced physics informed neural network for phase-field modeling of fracture. *Theoretical and Applied Fracture Mechanics*.

Klie H., Florez H. (2020) Data-Driven Prediction of Unconventional Shale-Reservoir Dynamics. *Society of Petroleum Engineers*. 25.

Malallah A., Nashawi I. S. (2005) Estimating the fracture gradient coefficient using neural networks for a field in the Middle East. *Journal of Petroleum Science and Engineering*. 49, 3. 193 – 211.

Mohd Razak S, Jafarpour B. (2020a) Convolutional neural networks (CNN) for feature-based model calibration under uncertain geologic scenarios. *Comput Geosci*. https://doi. org/10.1007/s10596-020-09971-4.

Mohd Razak S, Jafarpour B. (2020b) History Matching with Generative Adversarial Networks. *ECMOR XVII*. 1–17.

Park J., Datta-Gupta A., Singh A., Sankaran S. (2020) Hybrid Physics and Data-Driven Modeling for Unconventional Field Development – Onshore US Basin Case Study. *Unconventional Resources Technology Conference.*

Ramsundar B., Zadeh R. B. (2018) TensorFlow for deep learning: from linear regression to reinforcement learning.

Raissi M., Perdikaris P., Karniadakis G. E. (2019) Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*. 378. 686 – 707.

Ruder S. (2017) An overview of multi-task learning in deep neural networks. *arXiv preprint* arXiv:1706.05098.

Trangenstein J. A., Bell J. B. (1989) Mathematical structure of the black-oil model for petroleum reservoir simulation. *SIAM Journal on Applied Mathematics*. 49. 749–783.

Wang S., Chen Z., Chen S. (2019) Applicability of deep neural networks on production forecasting in Bakken shale reservoirs. *Journal of Petroleum Science and Engineering*. 179. 112 – 125.

Wu S., Zhang H., Re C. (2020) Understanding and Improving Information Transfer in Multi-Task Learning. *International Conference on Learning Representations (ICLR) 2020.*

Xue L., Liu Y., Xiong Y., Liu Y., Cui X., Lei G. (2021) A data-driven shale gas production forecasting method based on the multi-objective random forest regression. *Journal of Petroleum Science and Engineering*. 196.

Zhang Y., He J., Yang C., Xie J., Fitzmorris R., Wen X. H. (2018) A Physics-Based Data-Driven Model for History Matching, Prediction, and Characterization of Unconventional Reservoirs. *Society of Petroleum Engineers*.

Zhong Z., Sun A. Y., Wang Y., Ren B. (2020) Predicting field production rates for waterflooding using a machine learning-based proxy model. *Journal of Petroleum Science and Engineering.*