LA-UR-20-25822
Approved for public release; distribution is unlimited.

Title: Machine learning models for real-time forecasting of shale gas production

Author(s): Mudunuru, Maruti Kumar
Hoover, Karla Fay

Intended for: 2020 DOE-FE MLEF Summer Internship Technical Forum, 2020-08-03/2020-08-07 (Washington DC, Virginia, United States)
Report

Issued: 2020-08-03

Machine learning models for real-time forecasting of shale gas production

Author: Karla Hoover, Penn State University, Email Correspondence: karlahoover1@gmail.com

Mentor: Maruti Kumar Mudunuru, LANL, Email Correspondence: maruti@lanl.gov; **Co-Mentor:** Velimir V. Vesselinov, LANL; **Collaborators:** MSEEL team at LANL.

Technical report prepared as a part of DOE-FE MLEF Summer Internship Program, 2020.

Abstract: The goal of this project is to develop a machine-learning (ML) based model for real-time forecasting of shale gas production at the Marcellus Shale Energy and Environment Laboratory (MSEEL). The ML-model is based on Random Forest Regressor and is trained, tested, and validated on synthetic data that is representative of a MSEEL site. The production data for this project is simulated using fast and accurate physics of gas flow and transport. The associated multi-physics high-fidelity model for gas flow and transport is solved using PFLOTRAN multi-physics code. Multiple realizations (~3000) are generated that encompass a range of possible field-site characteristics found at the MSEEL site. A site-behavior library is then created using this PFLOTRAN synthetic production data. The simulated data in the site-behavior library is then used to create a ML-model to help refine key site parameters (e.g., fracture network statistics, matrix permeability, matrix porosity) and then forecast future gas production rate. The forecasts include total cumulative production, production rate, stage-specific production, and spatial evolution of the quantities of interest. The Random Forest-based ML-model predictions are compared to the PFLOTRAN simulated data. The predictions of the ML-model have a R^2-score greater than 90% when compared to the ground truth. Moreover, the time needed to train, test, and run the proposed ML-model is very low (~10^5 faster) compared to running a single high-fidelity PFLOTRAN simulation. These results instill confidence in our proposed ML-workflow that our developed models are fast, accurate, and reliable methods to estimate shale gas production.

Introduction: The data collection field-site for this MLEF project is the Marcellus Shale Energy and Environmental Laboratory (MSEEL) in Morgantown, West Virginia. MSEEL is a Department of Energy (DOE) field lab for unconventional shale gas reservoirs. The reservoir was developed in 2011 with new wells added in 2015. MSEEL is a long-term field site used to develop and validate new technology (e.g., pressure management) and offers an opportunity to field-validate strategies for improved gas production (Mudunuru et al., 2020). The goal of the MSEEL project is to improve shale gas recovery while minimizing environmental impacts. Creating ML-based models for real-time forecasting of shale gas production is one of the most important aspects of the MSEEL's overreaching project goals.

High-fidelity numerical simulations based on physics models are needed to create a data library for ML-model development. However, because they are computationally intensive (e.g., takes hours to days to run a single simulation), it is not feasible to use them for real-time forecasting of shale production. Machine learning (ML) models are generally quicker to develop and implement. The question is whether they are fast, accurate, and reliable for real-time forecasting of shale production. To help answer this question, first, a site behavior library is constructed. The library contains synthetic production-data created using PFLOTRAN multi-physics code (https://www.pflotran.org/) and is based on a range of possible MSEEL site characteristics. Then, a ML-based model (Random Forest) is generated using the site behavior library to refine key site parameters and then estimate the future gas production rate, including total cumulative production, total gas production, stage-specific production and spatial evolution (residual gas, P). Next, comparisons are made between the simulated physics model (PFLOTRAN) predictions and the Random Forest model predictions. Finally, computational costs of running a single high-fidelity numerical simulation are compared to the computational costs of running a ML-model.

Machine Learning Workflow (Methods):

To begin the process of creating a ML-model for the forecasting of shale gas production, a range of possible site characteristics were collected from the MSEEL field-site (Table 1). These field characteristics were the basis on which we created the synthetic production data that went into the site behavior library. The Discrete fracture network (DFN) (Hyman et al., 2015) and upscaled model was used to simulate the physical behavior of the fractured reservoir at MSEEL-I, starting with stage 10 of MIP-3H. Natural fracture planes connect to the hydraulic fractures. These fractures were upscaled and form a continuum model in which the upscaled permeability and porosity is obtained. Then, the PFLOTRAN model was solved on this upscaled continuum model. The PLOTRAN model produced the synthetic production data found in the data library. The statistics that are used to generate the library include aperture minimum, aperture maximum, aperture mean, aperture log variance, matrix permeability, and matrix porosity (Table 1).

Table-1: Statistics of realizations to generate site-behavior data library

Statistics of Realization	Minimum Value	Maximum Value
Aperture Mean	1.37e-4	1.51e-4
Aperture Log Variance	0.46	0.539524
Aperture Minimum	3.78e-06	1.47e-05
Aperture Maximum	7.82e-4	3.98e-03
Matrix Permeability	1.01e-19	2.00e-18
Matrix Porosity	5.00e-2	9.98e-2

The most important quantities of interest to be forecasted include cumulative production and its rate (Figure 1). These are estimated from the PFLOTRAN model realizations. This data was then used to create a Random Forest ML-based model in which the predicted production rate was the output.

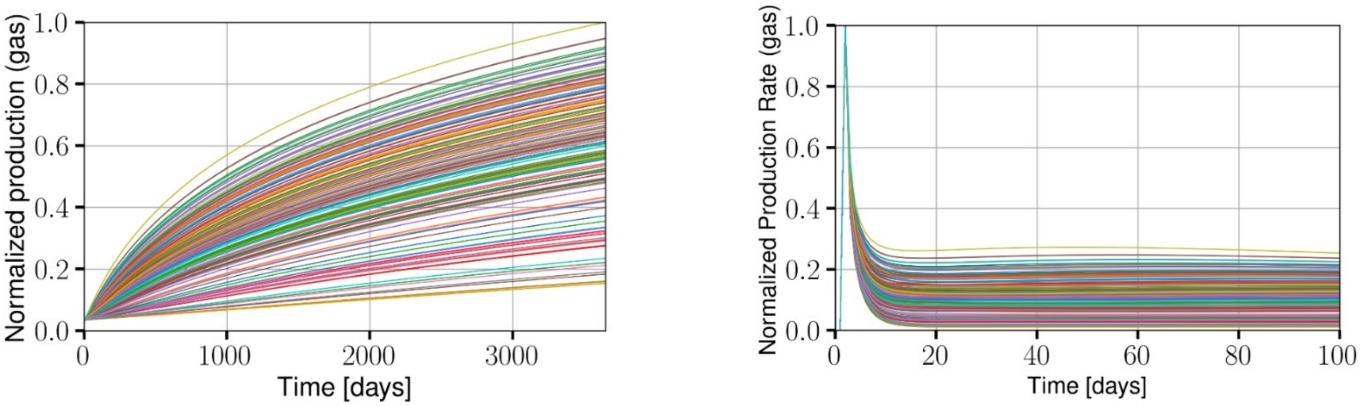

Figure-1: Data for ML-model development (Cumulative Production and Production rates). This includes training, testing, and validation.

The Random Forest model was created using 150 realizations. The training-data comprised of 70% of the data and the test-data comprised 30% of the data. The ML-model was based on 100 weak base estimators with a max depth decision tree estimator depth of 10. Computational costs (time) were recorded for the ML-model development, which were then compared to the computational costs of running the realizations of the PFLOTRAN simulated data. Figure 2 shows the entire workflow for data generation and the real-time forecasting.

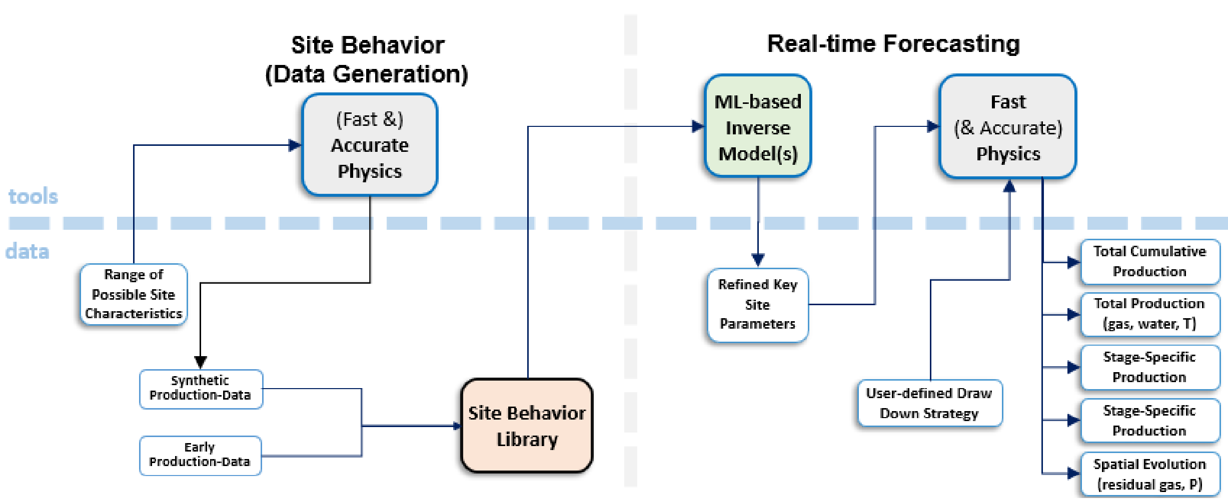

Figure-2: ML-Workflow utilizes a library of data on site characteristics to inform forecasting models

Results: The PFLOTRAN simulated data built in the site behavior library was found to be representative of the MSEEL field production data. The Random Forest-based ML-model gas production rate predictions were compared to the PFLOTRAN simulated data gas productions rates (Figure 3). The ML-model predictions had a R^2-score > 90% when compared to the simulated data. The root mean square error (RMSE) was 0.283 for the training data and 0.366 for the test data. The Random Forest model results based on 100 trees, indicates that it predicts production rates very well and shows promise as a fast, accurate, and reliable method to estimate shale gas production.

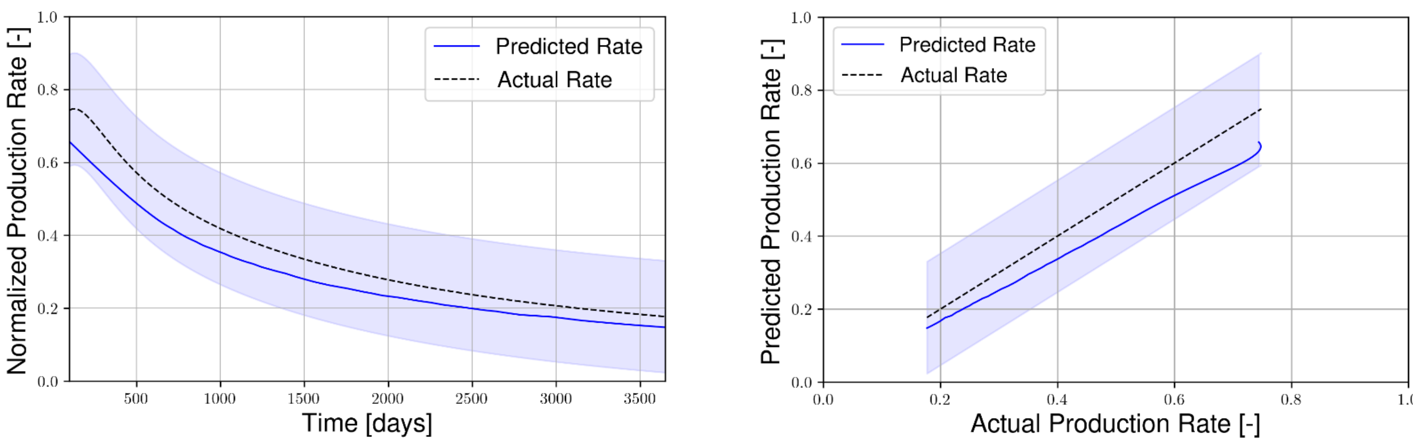

Figure-3: Random Forest-based ML-model predictions on unseen data. The ML-model consists of an ensemble of weak base regressors, which is 100 trees.

It took approximately 4 hours to create the ML-model and only 11.65 minutes to run it in its entirety on a single-core processor. In comparison, it took anywhere between 6 and 72 hours to run a high-fidelity PFLOTRAN model realization (Figure 4).

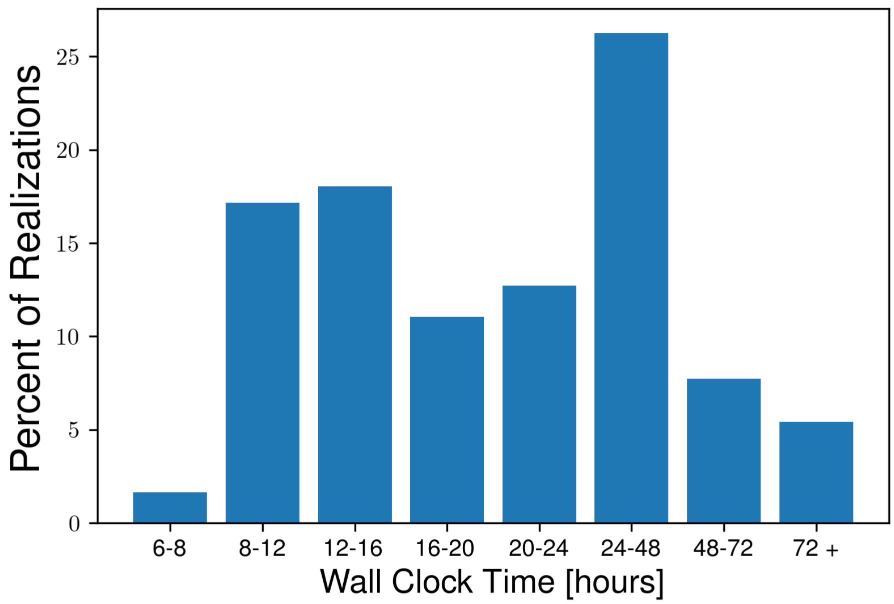

Figure-4: Data generation time to run high-fidelity PLOTRAN Model. Analysis is performed for 3000 realizations.

Conclusions: In summary, a first-cut of the site behavior library, which is representative of the MSEEL site, was created. A ML-model (Random Forest) was built using the site behavior library for real-time forecasting. Comparisons were made between the PFLOTRAN simulated model predictions and the Random Forest model predictions. The Random Forest model predictions closely matched those of the PFLOTRAN simulated data. The cost of computation is greatly reduced when using the ML-model in comparison to the PFLOTRAN high-fidelity simulations. The next steps involve comparing predictions from the random forest ML-model built on the 3000-11000 PFLOTRAN model realization library to the MSEEL field production data (similar to Figure 5).

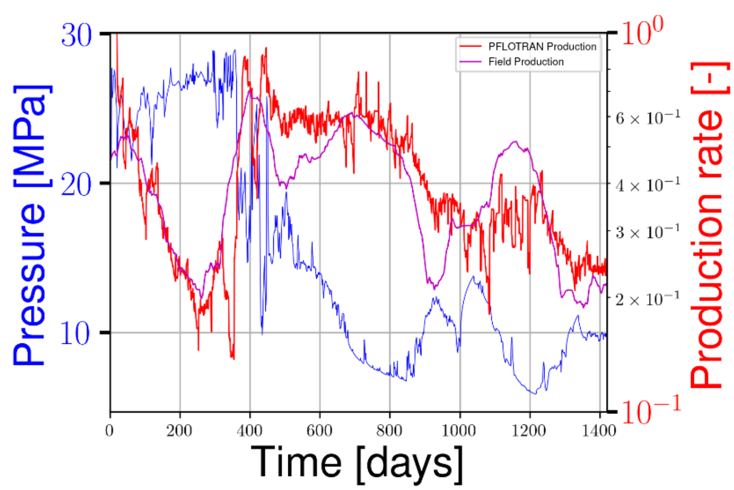

Figure-5: Comparison of gas production field-scale data at MSEEL site and PFLOTRAN simulation shows promise.

References:

Belyadi, H., Fathi, E. and Belyadi, F., 2019. *Hydraulic fracturing in unconventional reservoirs: theories, operations, and economic analysis*. Gulf Professional Publishing.

Géron, A., 2019. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media.

Hyman, J.D., Karra, S., Makedonska, N., Gable, C.W., Painter, S.L. and Viswanathan, H.S., 2015. dfnWorks: A discrete fracture network framework for modeling subsurface flow and transport. *Computers & Geosciences*, *84*, pp.10-19.

Mudunuru, M.K., O'Malley, D., Srinivasan, S., Hyman, J.D., Sweeney, M.R., Frash, L., Carey, B., Gross, M.R., Welch, N.J., Karra, S. and Vesselinov, V.V., 2020. Physics-informed Machine Learning for Real-time Unconventional Res-ervoir Management. AAAI 2020 Spring Symposium on Combining Artificial Intelligence and Machine Learning with Physics Sciences.

PFLOTRAN: https://www.pflotran.org/