IMPROVED ACCURACY AND EXPLAINABILITY OF MACHINE LEARNING IN MULTIPHASE FLOWRATE ESTIMATION USING PHYSICS-AWARE ALGORITHMS

Timur Bikmukhametov[a], Johannes Jäschke[a*]
[a]Dept. of Chemical Engineering, Norwegian University of Science and Technology
NO-7491, Trondheim, Norway

Abstract Overview

Machine learning algorithms for multiphase flow estimation are an attractive alternative to the first principles modeling approach and traditional hardware metering devices. One of the drawbacks of these methods is that they are often used as a black-box solution with unexplainable behavior, which is one of the reasons why petroleum engineers commonly prefer conventional flow metering methods. In this paper, we create machine learning models for each part of the production system which are more aware of the physical behavior of the system than the algorithms which use the raw data directly. In addition, we propose a simple strategy for combining the models using a linear meta-model which helps to explain the model behavior. We test this approach using several machine learning algorithms based on real field data. The results show that in addition to obtaining more physically meaningful machine learning algorithms, which are easier to interpret, the prediction accuracy of the models improves.

Keywords

Virtual Flow Metering, Explainable Machine Learning, Multiphase Flow, Oil Production Monitoring

Introduction

Multiphase flow estimation in oil and gas production systems is an important problem to solve to achieve efficient production optimization, improved oil recovery and robust operation with respect to flow assurance. An attractive alternative to traditional estimation methods of the production rates, such as multiphase flow meters (MPFMs) and test separators, is Virtual Flow Metering (VFM) which can be used as a standalone metering solution as well as a back-up system to the hardware devices. This approach is based on incorporating readily available field measurements such as pressure, temperature and choke opening into a mathematical model of the system. The mathematical formulation can be based on either first principles or machine learning (ML) models. The first principles VFM systems typically include mass, momentum and energy balances and consider the detailed physical representation of the production system to achieve high estimation accuracy (Holmås and Løvli, 2011). It results in a well-explainable behavior of the VFM system and good exploratory capabilities even when limited data is available. At the same time, it requires deep understanding of the complex multiphase flow behavior. Moreover, computational cost can be very high due to numerical solving procedure of the partial differential equations and the embedded non-linear optimization problem.

On the other hand, machine learning VFM systems typically do not consider the specifics of the process, for instance, well tubing geometry or flow regime, and flow estimation is performed based on the available data only.

* Corresponding author, e-mail: johannes.jaschke@ntnu.no

In most cases, feed-forward neural networks (NNs) were used for this task (Al-Qutami et al, 2018), while recurrent neural networks (Andrianov, 2018) and gradient boosting regression trees (GBRT) (Bikmukhametov and Jäschke, 2019) have also been recently tested. The advantage of ML VFM is that it is relatively easy to setup and perform inference when the model is trained. On the other hand, the produced results can be hard to explain because of the black-box nature of the machine learning algorithms.

In this paper, we propose an approach which combines the two VFM methods. To do this, we create input features for machine learning algorithms which are based on a physical relationship between the raw measurements and make algorithms to represent a certain production system part, for instance, a production choke. In addition, we create a linear meta-model which combines the created algorithm, improves the prediction accuracy and identifies the importance of each model. Using this approach, we: 1) create an explainable and physics-aware model; 2) improve the estimation accuracy by combining the algorithms in a clear and simple manner.

Methodology

Data

The data used in training and testing the algorithm performance is taken from a real field and based on measurements of sensors and a MPFM installed in a well in the North Sea. The system and the available data are shown in Figure 1. The data is relatively challenging for accurate estimation because the well is at the end of production period which means the fluid properties of the multiphase flow mixture change continuously, so that flowrate, which corresponds to certain pressures and temperatures now, may not be the same even after a short time period, which can influence the algorithm performance.

Physics-Aware Machine Learning Models

To create physics-aware machine learning models, first, we generate physically meaningful features instead of using the raw measurements directly. More specifically, we use choke and well tubing models. At the same time, we use simplified first principles models instead of more complex ones and exploit the machine learning algorithms to adjust the models such that they describe the data well.

For the choke model, we use a simple Bernoulli model with mixture fluid properties. For the tubing momentum equation model, we use a steady-state No-Pressure-Wave model which does not consider transient flow behavior and influence of viscosity and fluid acceleration. In addition, we do not use the temperature measurements directly, but rather the temperature drop over the choke and tubing. Since in a general case we do not measure the mixture density which is used in both models, we need to calculate

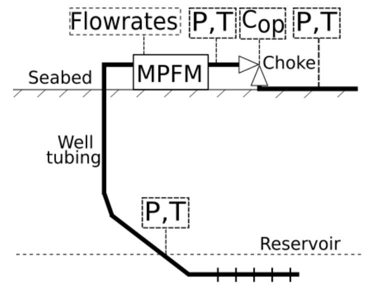

Figure 1 – Schematic representation of a well with available measurements.

it. To do this, we assume thermodynamic equilibrium conditions at the measurement points and use Soave-Redlich-Kwong Equation of State to compute phase volume fractions and densities and then compute the mixture density.

Case studies

We consider static and dynamic approaches for machine learning modeling. For the static models, we use feed-forward neural networks, gradient boosting with regression trees and random forest (RF) as VFM systems. For dynamic modeling, we use recurrent neural networks, more specifically Long-Short Term Memory (LSTM). Each algorithm is tested in the following case studies:

<u>Case 1:</u> ML model based on raw measurements.
<u>Case 2:</u> ML model based on the choke model.
<u>Case 3:</u> ML model based on the tubing flow model.
<u>Case 4:</u> ML model based on choke and tubing features.
<u>Case 5:</u> Meta-model based on linear regression and models from Case 2 and 3.

To compare the algorithms in a fair manner and accurately tune them at the same time, we perform Bayesian optimization of the hyperparameters and the algorithms' architectures using Expected Improvement acquisition function. The number of epochs and tree estimators is tuned using early stopping on the validation set.

Results and Discussion

Due to the limitations of the extended abstract length, we present an example of the results for the GBRT algorithm only, which are shown in Figure 2. We can see that all the cases with modified input features outperform the traditional machine learning VFM approach represented by Case 1. An interesting observation is that the choke model based algorithm in Case 2 is capable to capture more dynamic behavior and outlying flowrate values, and this result makes physical sense based on the model we created. Because the choke flow model is more dependent on the measured pressure change over the system (the tubing flow model also depends on the approximated head pressure loss), the pressure fluctuations

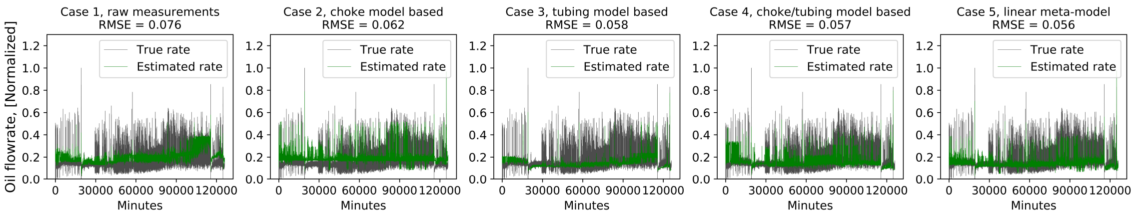

Figure 2 – Performance of gradient boosting regression tree algorithm using the proposed methods

over the choke helps to capture the dynamic flow fluctuations in a better way. Moreover, because the choke flow model describes the flow in a specific system location, the local change of the temperature also reflects high flow fluctuations well.

The tubing flow model (Case 3) in general better describes the regions where more stable flow regimes are observed. In addition, Case 4 and 5 which combine the two models of Case 2 and 3 show even a better performance. In case of the linear meta-model, it can be seen that some parts of the model estimates correspond to the behavior of the choke model, for instance, when estimating the outlying flowrate values, while in other parts it resembles the behavior of the tubing model. In general, we can see that the meta-model resembles more the behavior of the choke model which is in correspondence of the model weights which are 0.76 and 0.39 for the choke and tubing models respectively. As such, from this we can clearly see the advantages of using a linear meta-model which combines the model representatives of each part of the production system.

In general, explaining the algorithm performance based on the feature inputs is a hard task, however, in GBRT algorithm feature ranking is embedded into the training process, which makes it easier to interpret the algorithm outcomes. To emphasize the advantage of using the physics-based features, we show the feature importance for Case 1 (raw measurements) and Case 3 (tubing flow features) in Figure 3. In addition to the reduced feature space, we see that the created tubing mass flow feature has relatively high importance, which means that it helps the algorithm to describe the process behavior. In addition to the clear meaning, we can also control its importance by changing the feature complexity, for instance, applying a more comprehensive momentum balance equation. This is not the same in Case 1 where, apart from hardly explainable importance and meaning of the features, we cannot control its influence on the algorithm performance.

The obtained results with other algorithms show similar performance and emphasize the same advantages, however, in neural networks additional methods for analyzing feature importance in Cases 2, 3 and 4 are applied because it is not included in the training process.

Conclusions

Predictive accuracy of machine learning algorithms in multiphase flowrate estimation can be improved by incorporating physical relationships into the algorithm features instead of using raw measurements directly. It also helps to improve explainability of the algorithm performance which is of great importance in decision making process during real field operation. The proposed methods can be further developed by incorporating more complex first principle models into the machine learning algorithms to improve the performance.

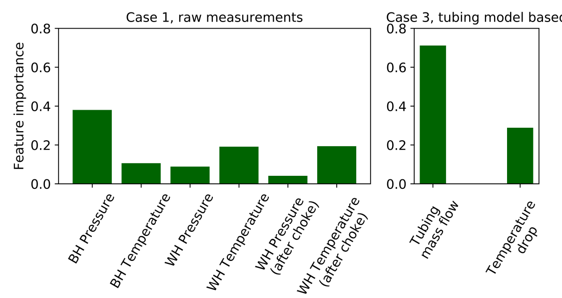

Figure 3 – Comparison of feature importance (BH – bottomhole, WH – wellhead)

Acknowledgments

The authors acknowledge SUBPRO, Research-based Innovation Centre within Subsea Production and Processing, which is financed by the Research Council of Norway, industry partners, and NTNU.

References

Al-Qutami, T., Ibrahim, R., Ismail, I., Ishak, M. (2018). Virtual multiphase flow metering using diverse neural network ensemble and adaptive simulated annealing. *Expert Systems With Applications, 93*, 72-85.

Andrianov, N. (2018). A Machine Learning Approach for Virtual Flow Metering and Forecasting. *3rd IFAC Workshop on Automatic Control in Oil and Gas Production.* Esbjerg, Denmark.

Bikmukhametov, T., Jäschke, J. (2019). Oil Production Monitoring using Gradient Boosting Machine Learning Algorithm. *To appear in 12th IFAC Symposium on Dynamics and Control of Process Systems, including Biosystems.* Florianópolis, Brazil.

Holmås, K., Løvli, A. (2011). FlowManager Dynamic: A mutiphase flow simulator for online surveillance, optimization and prediction of subsea oil and gas production. *15th International Conference on Multiphase Production* . Cannes, France.