

Developing a Hybrid Data-Driven, Mechanistic Virtual Flow Meter - a Case Study

M. Hotvedt* B. Grimstad** L. Imsland*

* *Engineering Cybernetics Department, NTNU, Trondheim, Norway*
(e-mail: {mathilde.hotvedt, lars.imsland}@ntnu.no)

** *Solution Seeker (e-mail: bjarne.grimstad@solutionseeker.no)*

Abstract: Virtual flow meters, mathematical models predicting production flow rates in petroleum assets, are useful aids in production monitoring and optimization. Mechanistic models based on first-principles are most common, however, data-driven models exploiting patterns in measurements are gaining popularity. This research investigates a hybrid modeling approach, utilizing techniques from both the aforementioned areas of expertise, to model a well production choke. The choke is represented with a simplified set of first-principle equations and a neural network to estimate the valve flow coefficient. Historical production data from the petroleum platform *Edvard Grieg* is used for model validation. Additionally, a mechanistic and a data-driven model are constructed for comparison of performance. A practical framework for development of models with varying degree of hybridity and stochastic optimization of its parameters is established. Results of the hybrid model performance are promising albeit with considerable room for improvements.

Keywords: hybrid modeling, virtual flow metering, petroleum production systems

1. INTRODUCTION

For a petroleum asset to succeed economically, the operators have to make crucial decisions regarding optimization of the asset. Knowledge regarding the multiphase flow rates in the asset is therefore of high importance. The flow rates may be obtained with deduction well testing, test separators and multiphase flow meters (MPFM), however, these methods are costly and MPFMs call for well intervention upon failure (Marshall and Thomas, 2015). An alternative is virtual flow meters (VFM) that take advantage of measurements to describe the input-output relationship of a system with a mathematical model (Toskey, 2012).

There are several types of VFM models. Dependent on the amount of available process data and prior knowledge of the system, the types may be placed on a scale ranging from mechanistic models (M-models) derived from first-principles, to data-driven models (DD-models), which are generic mathematical models fitted to input-output data (Stosch et al., 2013), see Fig. 1. Often, the two extremes are called white-box and black-box models, with reference to the extent of prior knowledge about the system, for example physical interpretation of parameters and relationship between variables. The models in between are hybrid models (H-models) or gray-box models, which utilize modeling techniques from both fields and have a mixture of physical and non-physical parameters.

In this research, an H-model of a well production choke is developed using historical production data from the petroleum platform *Edvard Grieg* (Lundin Norway, 2019). In addition, an M-model and a DD-model are developed for comparison of performance. A practical framework

facilitating development of models with varying degree of hybridity and stochastic optimization of model parameters is constructed and conveniently enables future research into the field of hybrid modeling. Background into VFM modeling and the contributions of this research is given in Section 2, the three model types of the production choke is presented in Section 3, the practical framework is outlined in Section 4, the Edvard Grieg case study is presented in Section 5, before simulation results and a conclusion is given in Sections 6 and 7.

2. BACKGROUND

2.1 Virtual flow meter modeling approaches

The most common way to model VFM in today's oil and gas industry are with M-models, where some well known commercial VFM are Olga, K-Spice and FlowManager (Bikmukhametov and Jäschke, 2020). A great advantage with M-models is their way of representing prior knowledge through the use of first-principles, which leads to interpretable parameters and usually good extrapolation abilities. However, in order for M-models to be computationally feasible, model simplifications are usually a necessity and plant-model mismatch is unavoidable (Solle et al., 2016). Additionally, in complex processes, unknown physical relations are oftentimes present and difficult to capture. VFM with DD-models have shown promising performance suitable for real-time monitoring, without the need of prior knowledge about the system (AL-Qutami et al., 2018). Further, unknown phenomena may be captured if reflected in the process measurements. However, DD-models are data hungry (Fig. 1), they struggle with

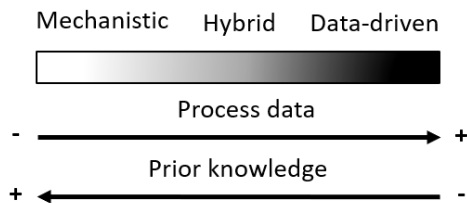


Fig. 1. Range of VFM models from mechanistic to data-driven, white-box to black-box.

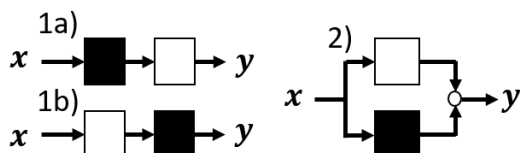


Fig. 2. Illustration of hybrid model variants. Serial H-models (type 1) and parallel H-models (type 2).

extrapolation in unseen operational settings, parameters generally lack physical interpretation and incorporating process constraints may be challenging, although existing dependent on the DD-method (Pitarch et al., 2019). Several industrial and academic M- and DD-models are reported in Mokhtari and Wlatrich (2016); Balaji et al. (2018); AL-Qutami et al. (2018); Bismukhametov and Jäschke (2020) and references therein.

An in-between solution designed to utilize the best of both worlds are H-models. First, notice that the expression "H-models" is widely used in literature for other concepts than combinations of M- and DD-models. Further, one should differ between a hybrid model development procedure and a hybrid model in application. To clarify, most M-models use real data for parameter estimation. Thus, these models are hybrid in their development procedure, however, after development, parameters are fixed, and the model in application is an M-model. Likewise, a DD-model trained on generated data from an M-model would be hybrid in development, although not in application. Therefore, in this article, we define an H-model as follows:

Definition A hybrid model combines equations from first-principles with generic mathematical structures, both in model development and application.

Following the definition, an H-model is fundamentally categorized in two ways, serial or parallel, see Fig. 2. Examples of serial models are online (that is, at each new state) parameter estimation with a DD-model (1a), a DD-model to capture unknown physical phenomena or modeling errors (1b) and physical equations utilized to construct specialized features as input to the DD-model, called feature engineering (1b). A parallel H-model (type 2) would be achieved if a composition of M- and DD-submodels are connected or used in an ensemble model. Naturally, combinations of the two fundamental ways will also be an H-model. Expectantly, compared to an M-model, the H-model should have an increased ability to capture unknown phenomena, yet have better interpretability than a DD-model through the inclusion of prior knowledge and physical parameters. Generally, the DD-

part in the H-model will be smaller (in terms of number of parameters) than in a DD-model and should thus require fewer data samples to obtain a satisfactory approximation of the process (Psychogios and Ungar, 1992), see Fig. 1.

2.2 Hybrid models in literature

Some of the earliest reported H-models are within the field of chemistry (Psychogios and Ungar, 1992; Kramer et al., 1992). However, H-models for VFM are rare although some examples exist in literature. For instance, Xu et al. (2011) used feature engineering in a neural network for wet gas metering. Although feature engineering has shown to boost DD-models, choosing appropriate features is challenging (Sutton and Barto, 2018). Al-Rawahi et al. (2012) estimated the mixture density of multiphase flow using a neural network. However, the neural network required the underlying primary measurements from a MPFM, which may not be as readily available as other measurements. Additionally, MPFM are known to require frequent calibration and may yield high measurement error in-between calibrations (Falcone et al., 2001). Although not a VFM, Baraldi et al. (2014) used an ensemble H-model to detect degradation of production choke valves.

2.3 Contributions

The contributions of this research are two-fold:

- A practical and convenient framework to facilitate development of models with varying degree of hybridity and stochastic optimization of the model parameters.
- A hybrid VFM model for production chokes, developed and validated utilizing real historical production data with readily available measurements such as pressures, temperatures and choke openings.

It must be specified that the main ambition of this research has been to establish a convenient framework for development and utilization of hybrid models. In addition, this research attempts to highlight that H-models may offer advantages over M- and DD-models. Therefore, only one type H-model (type 1a) with parameter estimation using a neural network in an existing M-model has been developed. However, a notable feature with the framework is that, regardless of the hybrid model structure, the model may be trained requiring only measurements of the *output* variable.

3. CHOKE MODELS

A production choke may be illustrated as in Fig. 3, where the volumetric oil flow rate, Q_o , will be estimated using nearby measurements; pressures (p), temperatures (T) and choke opening (z). Three model types have been developed, M-, H- and DD-model. In short notation, these are represented with $\hat{y}_\xi = f_\xi(\mathbf{x}_\xi; \boldsymbol{\theta}_\xi)$ where $\xi \in \{m, h, dd\}$, $\hat{y}_\xi = Q_o^e$ is the estimated oil flow rate, f_ξ are the set of model equations, \mathbf{x}_ξ are the input measurements and $\boldsymbol{\theta}_\xi$ are learnable model parameters. In the model development, also called training procedure, an optimization algorithm finds the $\boldsymbol{\theta}_\xi$ that minimizes the deviation between estimated (Q_o^e) and existing measurements (Q_o^m) of the volumetric oil flow rate, see Section 4. The following Sections (3.1, 3.2, 3.3) briefly explain the three model types

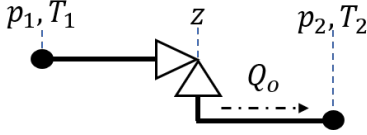


Fig. 3. Illustration of the well production choke

and Table 1 gives an overview of inputs and parameters to the three models and highlights the difference between the M- and H-model, in this case the form of the C_v -curve.

3.1 Mechanistic model

The M-model is from Kittelsen et al. (2014), chosen for its simplicity, and described with the equations (1)-(7). More widespread choke models exist (e.g. Sachdeva, Hydro, Al-Safran, see Haug (2012) and references therein) and should be experimented with in future work.

$$Q_o = \frac{w_o \dot{m}}{\rho_o, ST} \left[\frac{Sm^3}{h} \right] \quad (1)$$

$$\dot{m} = NC_v(z) \sqrt{Y^2 \rho_m (p_1 - p_2)} \quad \left[\frac{kg}{h} \right] \quad (2)$$

$$Y^2 = \left(1 - \frac{1}{3} \frac{x_{lim}}{x_{TP}}\right)^2 \frac{x_{lim} p_1}{p_1 - p_2} \quad [-] \quad (3)$$

$$\frac{1}{\rho_m} = \frac{w_g}{\rho_g} + \frac{w_o}{\rho_o} + \frac{1 - w_g - w_o}{\rho_w} \quad \left[\frac{m^3}{kg} \right] \quad (4)$$

$$\rho_g = \frac{M_w p_1}{z_g R T_1} \quad \left[\frac{kg}{m^3} \right] \quad (5)$$

$$x_{lim} = \min(x_P, x_{TP}) \quad [-] \quad (6)$$

$$x_P = \frac{p_1 - p_2}{p_1}, x_{TP} = \frac{p_1 - p_2}{p_1} \Big|_c \quad (7)$$

Two important assumptions are those of frozen flow and incompressible liquid; the mass phase fractions, $\mathbf{w} = [w_g, w_o, w_w]$ and liquid densities ρ_o and ρ_w are constant in a given operating point (ST for standard conditions). For this model, the valve flow coefficient; $C_v(z)$, is determined with linear interpolation between a given set of test points, which are usually from lab-experiments with water, yet calibrated to the multiphase flow once in place. Further nomenclature may be found in Kittelsen et al. (2014). The learnable model parameters are chosen to be $\boldsymbol{\theta}_m = [\rho_o, \rho_w, a]$, where a allows the $C_v(z)$ to be shifted; $C_{v,new}(z) = a C_{v,old}(z)$.

3.2 Data-driven model

The DD-model is a fully-connected, feed forward neural network (NN) with the Rectified Linear Unit (ReLU) as activation function on each layer. See e.g. Balaji et al. (2018) for description of neural networks. The learnable parameters are the weights and biases on each layer, $\boldsymbol{\theta}_{dd} = [\mathbf{W}_{dd}, \mathbf{b}_{dd}]$.

3.3 Hybrid model

The H-model (type 1a, Fig. 2) is represented with the same equations as for the M-model (1)-(7), but with the C_v obtained from a fully-connected, feed forward, NN with ReLU as activation function on each layer. The mass fractions were included as inputs to the NN in an attempt to have the C_v -curve reflect well-specific properties. Thus, the learnable parameters are $\boldsymbol{\theta}_h = [\rho_o, \rho_w, \mathbf{W}_h, \mathbf{b}_h]$.

Table 1. Overview of parameters and inputs and overview of the C_v -curve form

	M-model	H-model	DD-model
$\boldsymbol{\theta}$	ρ_o, ρ_w, a	$\rho_o, \rho_w, \mathbf{W}_h, \mathbf{b}_h$	$\mathbf{W}_{dd}, \mathbf{b}_{dd}$
\mathbf{x}	$[p_1, p_2, T_1, z, w_g, w_o]$	$[p_1, p_2, T_1, z, w_g, w_o]$	$[p_1, p_2, T_1, T_2, z, w_g, w_o]$
$C_v(\mathbf{x}')$	Linear interpolation $\mathbf{x}'_m = [z]$	NN $\mathbf{x}'_h = [z, w_g, w_o]$	n.a.

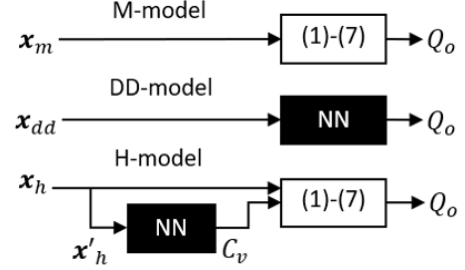


Fig. 4. The forward pass illustrated for the three model types

4. MODELING FRAMEWORK

To easily investigate different model types, a practical framework utilizing machine learning techniques is constructed¹. The framework enables a smooth transition between training a fully M-model to a fully DD-model. It consist of several parts and will be defined in the following.

4.1 Defining the model

This part enables a convenient way to implement models with varying degree of hybridity. Firstly, the model parameters must be defined, either as single, learnable parameters, as for the physical parameters, or as NN's with weights and biases. Thus, a model will effortlessly move on the gray-scale (Fig. 1) dependent on the defined parameters. Thereafter, the forward pass, the propagation of input data through the model, will be defined as a computational graph, enabling access to the model derivatives through automatic differentiation. The forward pass for the different models is illustrated in Fig. 4. A particularly appealing property with this framework is that measurements of the C_v are not required as the model is trained on the output, Q_o .

4.2 Defining the optimization problem

Once the model is defined, a general optimization problem to find the $\boldsymbol{\theta}_\xi$ that minimizes deviation between the model estimates $\hat{\mathbf{y}}_\xi = \mathbf{Q}_{o,\xi}^e = f_\xi(\mathbf{x}_\xi; \boldsymbol{\theta}_\xi)$ where $\xi \in \{m, h, dd\}$ and the measurements $\mathbf{y} = \mathbf{Q}_o^m$ may be set up as

$$\begin{aligned} \hat{\boldsymbol{\theta}}_\xi &= \arg \min_{\boldsymbol{\theta}_\xi} J(\boldsymbol{\theta}_\xi, \boldsymbol{\lambda}_\xi) \\ &= \arg \min_{\boldsymbol{\theta}_\xi} \left(\frac{1}{n} \sum_{i=1}^n \left(y^{(i)} - f_\xi(\mathbf{x}_\xi^{(i)}; \boldsymbol{\theta}_\xi) \right)^2 \right. \\ &\quad \left. + \frac{1}{n} \sum_{j=1}^p \lambda_{j,\xi} (\theta_{j,\xi} - \mu_{j,\xi})^2 \right) \end{aligned} \quad (8)$$

¹ We have utilized PyTorch, but other possibilities exist such as TensorFlow.

The first term in eq. (8) is the mean square error (MSE) and the second is an ℓ_2 -regularization term with regularization factors λ_i . For the physical parameters, the goal of regularization is to penalize deviation of the parameters from a prior (expected) value, μ_i , and maximum a posteriori (MAP) estimation has been set up to automatically calculate the λ_i factors, see Section 4.3. For the NN parameters, common practice is followed and μ_i is set to zero. If \mathbf{Q}_σ^m are available from different measurement sources, additional MSE terms may be added and weighted according to the uncertainty in the measurement source. In this research, only measurements from a MPFM has been utilized.

The framework solves the optimization problem in eq. (8) using iterative gradient-based optimization. The update formula may be stated as follows

$$\boldsymbol{\theta}_\xi^{k+1} = \boldsymbol{\theta}_\xi^k - \alpha^k \mathcal{M}(\mathbf{x}_\xi^{k'}; \boldsymbol{\theta}_\xi^k, \boldsymbol{\lambda}_\xi) \quad (9)$$

where α^k is the learning rate (or step-size), $\mathbf{x}_\xi^{k'}$ is a subset of the data samples and \mathcal{M} is the set of equations calculating the step direction. Different algorithms may be selected, such as stochastic gradient descent (SGD), Adam among others (Bottou et al., 2018). Stochastic gradient-based optimization algorithms has the advantage of being well suited for large scale models, either in terms of large data-sets or many parameters, where other optimization algorithms utilizing linesearch may be to computationally expensive (Bengio, 2012). In SGD, $\mathcal{M} = \nabla_{\boldsymbol{\theta}} \tilde{J}(\mathbf{x}; \boldsymbol{\theta}^k, \boldsymbol{\lambda})$, where $\nabla \tilde{J}$ may be calculated with different number of samples (batch size). Knowing which optimization algorithm yields the best result is challenging as it might be problem dependent. Therefore, the framework promote investigation of different optimization algorithms. In this research, Adam is used for all models.

4.3 Calculation of regularization parameters

The λ_i regularization factors for the physical parameters may be automatically calculated through MAP estimation. If one assumes a model of the form

$$y = f(x; \theta) + \epsilon \quad \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2) \quad (10)$$

the MAP estimation may be set up as follows utilizing Bayes' rule, where (X, y) is the collection of data points

$$\hat{\theta}_{MAP} = \arg \max_{\theta} (\log p(y|X, \theta) + \log p(\theta)) \quad (11)$$

If one additionally assumes independent Gaussian priors of the parameters $\theta_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, the MAP estimation will result in, after some rearrangements,

$$\hat{\theta}_{MAP} = \arg \min_{\theta} \left(\sum_{i=1}^n (y^{(i)} - f(x^{(i)}; \theta))^2 + \sum_{i=1}^p \frac{\sigma_\epsilon^2}{\sigma_i^2} (\theta_i - \mu_i)^2 \right) \quad (12)$$

Dividing by n and setting $\lambda_i = \sigma_\epsilon^2 / \sigma_i^2$, the MAP estimation will be the same as the estimate in eq. (8). The σ_i may be determined based on physical bounds and if σ_ϵ is known, λ_i is automatically calculated. In practice, σ_ϵ must be tuned, however, the number of coefficients to determine decreases.

5. CASE STUDY - EDVARD GRIEG

Historical production data from *Edvard Grieg* has been utilized in the model development procedure and to analyze performance of the models. In addition to pressures, temperatures, and choke opening (see Fig. 3), measurements from a MPFM located upstream the choke restriction was used for training the model, keeping in mind that MPFM measurements may be faulty and require frequent calibration (Falcone et al., 2001). Future work should include well-tests which in general have higher accuracy than MPFM measurements. The production data are from 10 oil wells, yielding a total of 30 models, over a period of 1248 days. Consequently, the assumption of constant physical parameters may be a rough approximation and future work should consider updating the models at certain intervals in time to account for changes in the true process.

The data was preprocessed in two steps before performing modeling. First, the raw production data was processed by Solution Seeker's data squashing technology (Grimstad et al., 2016). The data squashing algorithm partitions the data into intervals of steady-state operation. The data in each interval is then compressed to mean values using statistics suitable for time-series data. The result is a compressed data set of steady-state operating points, suitable for steady-state modeling. In the second preprocessing step, samples considered invalid, such as samples with unrealistically large well head pressures or negative flow rates, were removed and some samples were slightly modified, for instance small negative flow rates, where measurement noise was the likely cause of error. The second step resulted in a variable number of samples per well, in the range 612-2175. Further, the mass fractions were calculated using MPFM flow rates and standard densities. In an industrial setting, the mass fractions are often calculated from sparse well-test samples, thus to mimic this setting, a mass fraction update time of 30 days was employed, using an average of the last 20 samples.

The data set of each well was divided into two, training (75%) and test (25%), where 15% of the latest training data was used as a validation set to decide upon the hyperparameters in the training procedure. An ambition was for the three model types to generalize well across all wells of the asset. Consequently, the same set of hyperparameters was used for a model type, instead of individual tuning of each model type for each well. However, one should expect a lower overall error by individual tuning due to dissimilar well operating conditions and variable sample numbers, and this should be considered in future work. The average root mean square error (RMSE) and average mean absolute error (MAE) of the 10 wells were monitored and the best set of hyperparameters was chosen based on the minimum obtained averages. However, if prominent overfitting occurred in a well for a set of hyperparameters, that is, if the validation error increased when the training error decreased towards the end of training, the next best set of hyperparameters was chosen. Practical recommendations from Bengio (2012) was followed in the tuning process.

For all models, the learning rate (α) was thoroughly experimented with as this often is the most important hyperparameter to tune (Bengio, 2012). Further, for the

M-model, the physical parameters had to converge within the specified bounds, thus, the number of epochs (E), that is, the number of loops through the training set, and σ_ϵ were tuned thereafter. For the H-model, both physical and NN parameters had to be found. However, E may be high and the NN architecture (width/depth) large without leading to overfitting of the NN as long as regularization of the NN parameters is applied (Bengio, 2012). Hence, E was set sufficiently high and σ_ϵ adjusted for convergence of the physical parameters within bounds, the width/depth was set to 20/2 and combinations of α and the NN regularization factor, $\lambda_{i,nn}$, were tested. For the DD-model, the same recommendations were followed. The E was set high and combinations of α and $\lambda_{i,nn}$ investigated. The width/depth was set to 70/2. Lastly, the batch size (B) is often tuned independently of the other hyperparameters (Bengio, 2012) and was thus tuned last. Even though considerable effort was put into fair tuning of the three models, a Bayesian optimization approach will be investigated in the future to avoid (non-intentional) advantage to either model.

An overview of the final hyperparameters are given in Table 2. Observe that the M-model required a larger σ_ϵ than the H-model for the physical parameters to converge within specified bounds, indicating that the H-model accounts for some of the measurements noise with the DD-part. Further, the best performance for the H-model was obtained with a low batch number, however, only small differences in average error lead to this choice.

Table 2. Overview of the final model hyperparameters

	M-model	H-model	DD-model
E	5000	2000	2000
B	150	32	150
α	0.01	0.01	0.01
σ_ϵ	25	10	-
$\lambda_{i,nn}$	-	0.01	0.001
width/depth	-	20/2	70/2

6. SIMULATION RESULTS

The simulation results are shown in Fig. 5, where the RMSE and MAE of the test set for the 10 wells are illustrated, and Fig. 6 which is a cumulative deviation plot (CDP) (Corneliusen et al., 2005) indicating the accuracy of the developed VFM models, that is, how many of the test points fall within a certain deviation from the measurement. There are several interesting observations to be made from the results. Firstly, notice the extreme outlier that is present in the DD-model performance in Fig. 5. The outlier is caused by one of the wells which had an operational setting very different from the setting in the training set. As mentioned in Section 2, DD-models may struggle with extrapolation in unseen operational settings which may explain the outlier. In that case, the results indicate that the H-model has preserved some of the extrapolation power of the M-model which do not have the extreme outlier. However, despite the H-model obtaining the lowest errors for some of the wells, the M-model performs better than the H-model due to less spread between the quartiles. Now, the only difference between the two models is the form of the C_v -curve. This indicates

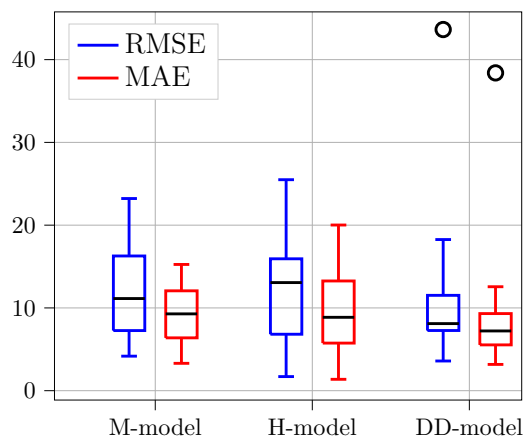


Fig. 5. Boxplot overview with median for the three model types across the 10 wells.

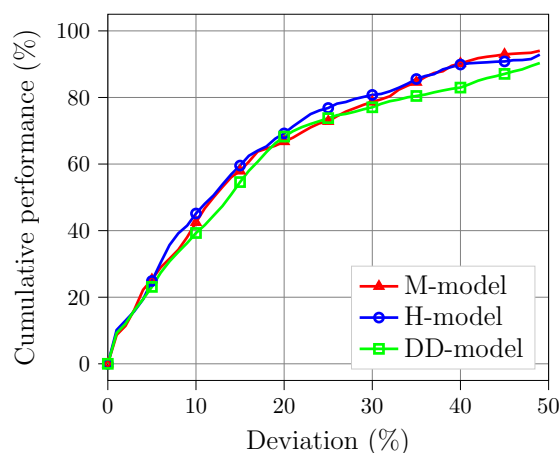


Fig. 6. Cumulative deviation plot of all test samples across the 10 wells.

that our prior belief of the C_v -curve in the M-model was good and that no flexibility was added to the H-model by having the C_v as an NN. Naturally, these results are preliminary and further investigations are necessary. In particular, different H-model variants may better leverage the advantages of both M- and DD-models.

Generally, the results show a higher error than expected. Other studies have reported almost 90% performance for 20% deviation in the CDP (e.g. AL-Qutami et al. (2018)), whereas in this paper about 70% performance for 20% deviation is achieved. There may be several causes for the large error. Firstly, preprocessing of the data could be improved by for instance further outlier removals. In addition, MFPM measurements was used for mass fraction calculation and in training despite a possibility of being faulty in between calibrations. Further, the mass fractions was updated every 30 days to mimic an industrial setting, however, in training, continuous mass fraction updates could be utilized. Hence, future work should include measurement sources with higher accuracy, such as well-tests, and analyze performance with continuous mass fraction updates. Secondly, the model types were generalized across all wells and improved results would most likely be achieved with individual tuning. Further, utilization of a more accurate mechanistic choke model

or optimization of additional physical parameters may decrease the error in the M- and H-model. Thirdly, the number of days for which the models are used in prediction should be taken into account. For some of the wells, the test set covered more than 200 days, whereupon the true process could have changed significantly and the models lack validity. Future work should consider online training of the models at regular intervals in time. Nonetheless, the main goal of this research was not to find exceptional models, but to illustrate that an H-model may offer advantages over M- and DD-models and to establish a framework for convenient future research.

7. CONCLUSION

Results in Section 6 indicate that hybrid modeling is promising and may offer advantages over both mechanistic and data-driven modeling. However, results are preliminary and there is considerable room for improvements. Future work should put more effort into preprocessing of the data set, analysis of the mass fraction calculation influence on performance and inclusion of well tests in training and validation. Further, individual tuning of each well should be investigated and the models should be tested with different data sets, for instance from other petroleum assets. Last but not least, future work should explore different hybrid model variants, for which the presented framework is convenient and highly suitable.

ACKNOWLEDGEMENTS

This research is a part of BRU21 - NTNU Research and Innovation Program on Digital and Automation Solutions for the Oil and Gas Industry (www.ntnu.edu/bru21) and supported by Lundin Norway AS.

REFERENCES

- AL-Qutami, T., Ibrahim, R., Ismail, I., and Ishak, M.A. (2018). Virtual multiphase flow metering using diverse neural network ensemble and adaptive simulated annealing. *Expert Systems With Applications*, 93, 72–85.
- Al-Rawahi, N., Meribout, M., Al-Naamany, A., Al-Bimani, A., and Meribout, A. (2012). A neural network algorithm for density measurement of multiphase flow. *Multiphase Science and Technology*, 24, 89–103.
- Balaji, K., Rabiei, M., Suicmez, V., Canbaz, C., Agharzeyva, Z., Tek, S., Bulut, U., and Temizel, C. (2018). Status of data-driven methods and their application in oil and gas industry. *Society of Petroleum Engineers*.
- Baraldi, P., Mangili, F., Gola, G., Nystad, B., and Zio, E. (2014). A hybrid ensemble-based approach for process parameter estimation and degradation assesment in offshore oil platforms. *International Journal of Perfromability Engineering*, 10, 497–509.
- Bengio, Y. (2012). *Practical Recommendations for Gradient-Based Training of Deep Architectures*. In: *Montavon G., Orr G.B., Mller KR. (eds) Neural Networks: Tricks of the Trade. Lecture Notes in Computer Science.*, volume 7700, 437–478. Springer, Berlin, Heidelberg.
- Bikmukhametov, T. and Jäschke, J. (2020). First principles and machine learning virtual flow metering: a literature review. *Journal of Petroleum Science and Engineering*, 184.
- Bottou, L., Curtis, F., and Nocedal, J. (2018). Optimization methods for large-scale machine learning. *Society for Industrial and Applied Mathematics*, 60(2), 223–311.
- Corneliussen, S., Couput, J.P., Dahl, E., Dykesteen, E., Frøysa, K.E., and Malde, E. (2005). Handbook of multiphase flow metering. *Norwegian Society for Oil and Gas Measurement (NFOGM), 2. revision*.
- Falcone, G., Hewitt, G., Alimonti, C., and Harrison, B. (2001). Multiphase flow metering: current trends and future developments. *SPE Annual Technical Conference and Exhibition*.
- Grimstad, B., Gunnerud, V., Sandnes, A., Shamlou, S., Skrondal, I.S., Uglane, V., Ursin-Holm, S., and Foss, B. (2016). A Simple Data-Driven Approach to Production Estimation and Optimization. In *SPE Intelligent Energy International Conference and Exhibition*. Society of Petroleum Engineers.
- Haug, R. (2012). *Multiphase Flow Through Chokes*. Master's thesis, NTNU, Trondheim, Norway.
- Kittelsen, P., Fjalestad, K., and Aasheim, R. (2014). Stabilized and increased well production using automatic choke control. *Society of Petroleum Engineers*.
- Kramer, M.A., Thomsom, M.L., and Bhagat, P.M. (1992). Embedding theoretical models in neural networks. *Proceedings of American Control Conference*, 1, 475–479.
- Lundin Norway (2019). Norway - edvard grieg. <https://www.lundin-petroleum.com>.
- Marshall, C. and Thomas, A. (2015). Maximising economic recovery - a review of well test procedures in the north sea. *Society of Petroleum Engineers*.
- Mokhtari, K. and Wlatrich, P. (2016). Performance evaluation of multiphase flow, models applied to virtual flow metering. *WIT Transactions on Engineering Sciences*, 105, 99–111.
- Pitarch, J.L., A., S., and de Prada, C. (2019). A systematic grey-box modeling methodology via data reconciliation and sos constrained regression. *Processes*, 7(3), 1–23.
- Psychogios, D. and Ungar, L. (1992). A hybrid neural network-first principles approach to process modeling. *AIChE Journal*, 38(10), 1499–1511.
- Solle, D., Hitzmann, B., Herwig, C., Remelhe, P.M., Ulonska, S., Wuerth, L., Prata, A., and Steckenreiter, T. (2016). Between the poles of data-driven and mechanistic modeling for process operation. *Chemie Ingenieur Technik*.
- Stosch, M., Oliveria, R., Peres, J., and Azevedo, S. (2013). Hybrid semi-parametric modeling in process system engineering: Past, present and future. *Computers and Chemical Engineering*.
- Sutton, R. and Barto, A. (2018). *Reinforcement Learning, an introduction*. The MIT Press, Cambridge, Massachusetts, London, England.
- Toskey, E. (2012). Improvements to deepwater subsea measurements rpsea program: Evaluation of flow modeling. *Offshore Technology Conference*.
- Xu, L., Zhou, W., Li, X., and Tang, S. (2011). Wet gas metering using a revised venturi meter and soft-computing approximation techniques. *IEE transactions on instrumentation and measurement*, 60, 947–956.