Estimation of Long-term Power Demand of Oil and Gas Installations using Hybrid Models

Leif Erik Anderssona,\*, Adriana Reyes-Lúaa, Heiner Schümannb, Brage Rugstad Knudsena

aSINTEF Energy Research, Sem Sælandsvei 11, Trondheim, 7034, Norway

bSINTEF Industry, S.P. Andersens vei 15 B, Trondheim, 7031, Norway

\*leif.andersson@sintef.no

Abstract

A methodology to forecast power demand of oil and gas installations which uses publicly production available data, parametric models and data-driven Gaussian regression methods is presented. The methodology also captures the expected fuel gas consumption and energy ratio. The proposed methodology is tested on the Brage field on the Norwegian Continental Shelf. It is shown that the general oil and water production behaviour as well as fuel gas consumption trends can be predicted. However, the forecast inherits a significant uncertainty due to the publicly available dataset lacking metadata and a complete description of the energy sinks.

**Keywords**: Gaussian Process Regression, Forecast, Oil and Gas, Power demand

* 1. Introduction

The need for decarbonization and the energy transition will cause a decline in global fossil fuel demand in the years to come (IEA, 2023). Nevertheless, natural gas is expected to play an important role in the path towards decarbonization as a raw material to produce low-carbon hydrogen via steam reforming with carbon capture and storage (Ueckerdt et al, 2024). Extraction of oil and gas is energy intensive and most CO2 emissions from offshore petroleum activities are related to power generation to sustain offshore operations, with most of the offshore power production using natural gas as fuel (Voldsund et al, 2023). For instance, in 2022, 11.57 mill. ton CO2,eq were emitted in the Norwegian Continental Shelf (NCS), 80.73 % of which were generated by gas turbines and 7.14 % by engines (NDP, 2023). The relation between power demand and oil and gas production varies through the lifetime of a site. Energy intensity and emissions from power generation increase significantly towards the late life of fields due to the need of reservoir pressure support or gas lift to maintain operations. Also, equipment such as gas export compressors may operate far from their design point and thus, at lower efficiencies. There are high-profile studies aiming to predict global oil and gas production trends, which are driven not only by availability but also by economics (Bardi, 2019). At an installation level, it is possible to utilize engineering insights to estimate production and energy intensity (Masnadi and Brandt, 2017). Different machine learning tools have been used to analyze upstream operations (Koroteev et al, 2021), including deep-learning and hybrid models to predict production (Fan et al, 2021; Pan et al, 2023). Here, we analyze the system from the power demand perspective, proposing a methodology to forecast the long-term power demand using public available data for the NCS. Such tools will support operators to plan investment in low-carbon power supply for offshore sites and enable public authorities to monitor energy intensity and thus shape regulations for aging fields.

* 1. Methodology
		1. Data set

The data was retrieved from the Norwegian Diskos National Data Repository (DISKOS NDR, 2022). The data set contains monthly production data per field, installation and terminal of oil companies operating on the NCS. A challenge for creating a long-term forecast model is that additional metadata is not available. Therefore, the operational status or events, e.g., downtime, maintenance schedule, installation of new wells, upgrading or installation of new equipment or change of production, is not known. These factors influence the production and are visible as peaks and sudden large changes of trends in the time series data.

* + 1. Data pre-processing

A moving average filter is employed for outlier detection. If the data point $x$ in the time series deviates significantly from the moving average value $\hat{x} $using the standard deviation $s$ over the window, an outlier is detected and the data point $x$ is replaced with its moving average value $\hat{x}$, as in Eq. (1), where $δ$ is the Z-score.

|  |  |
| --- | --- |
| $δ = \left|\frac{x- \hat{x} }{\hat{x}}\right| > 2$,  | (1) |

* + 1. Gaussian process regression

Gaussian Process (GP) regression is used for the energy demand and production forecasts. GP is a machine learning method that works well with small data sets (Rasmussen and Williams, 2006), which makes it well suited for the data set used in this article. GPs are non-parametric, probabilistic kernel methods that aim to identify an unknown function $f$ from data. A zero-mean function with automatic relevance squared-exponential covariance is chosen. The log marginal likelihood is used to find the hyperparameters from a training data set.

* + 1. Forecast methodology

A hybrid forecast strategy is applied consisting of two parts, a data-driven model and a parametric model that is created based on expected production profiles of oil and gas fields. In case a parametric model is used it can be supplemented by a data-driven model that learns to estimate the difference between real data and parametric model.

* + - 1. Parametric models

Parametric models are included in the forecast framework to allow inclusion of expert knowledge into the forecast. Parametric models for oil, gas, and water production are included. The purpose is to include the knowledge that as oil and gas production declines, water production increases over the lifetime of a field. Exponential decay is used for oil and gas production while logistic growth is used for water production. These models are chosen since they usually fit well with the expected production profiles of and oil and gas field (Holt and Schümann, 2022). The data set is normalized before the parameter estimation step by least squares is employed.

* + - 1. Data-driven models

GP regression is employed for the data-driven model. Alternatively linear regression is also tested to evaluate the advantage of using a nonlinear model. In both cases a (non) linear autoregressive model with exogenous inputs (ARX-type) was chosen for the model structure. All outputs of the previous time step are fed back to the model to predict the next time step.

* + 1. Performance parameters

The performance of the forecast models is evaluated calculating normalized root-mean square error (NRMS) on the test data set. We set as boundary the offshore facility and assume that the studied offshore systems are energetically self-sufficient. We use the energy ratio (ER), Eq. (2), to quantify the energy efficiency of the offshore facility:

|  |  |
| --- | --- |
| $$ ER =\frac{E\_{gas}^{exp}+E\_{oil}^{exp}}{E\_{gas}^{pow}}$$ | (2) |

where $E\_{gas}^{exp}$ and $E\_{oil}^{exp}$ correspond to the energy in the oil and natural gas exports to the onshore facilities and $E\_{gas}^{pow}$corresponds to the direct energy used to provide the power required for offshore operations, typically for natural gas via gas turbines.

* 1. Results

Here, we present the results of the proposed forecast methodology for the Brage field on the NCS. The order of the ARX-type model was varied. In addition, forecasts were created with only the parametric model, a hybrid model of parametric and data-driven models, and purely data-driven models, respectively. Figure 1 shows the training and test sets for oil and water production. Both parameters are captured well with the parametric model, and the forecast performance over the test set, to the right of the dashed vertical line, is very good.

|  |  |
| --- | --- |
|  |  |

Figure 1 Examples of parametric model fit for the Brage field. The (red) dashed vertical line indicates the split between training and test set.

In a first step, the one-step forecast performance of the forecast methodology is evaluated. Consequently, one month into the future is predicted. For the Brage field the linear regression models usually outperform the GP regression models in accuracy for all outputs. One exception is the hybrid model for oil production where the high order GP regression models perform the best. The GP regression models tend to follow more strongly the parametric models resulting in a smoother and more constant forecast than the linear regression models. This indicates that they rely less on the inputs from the previous time-step (auto-correlation part) to predict the next step. The normalized root-mean square error of all outputs is in the range of 8-15 %, where the oil production is predicted the worst.

In Figure 2, a forecast over about 60 months of the fuel gas consumption (energy consumption of the platform) and oil production for the Brage field is shown. The fuel gas consumption itself is predicted by a data-driven model since no parametric model was developed. However, some of the inputs to the fuel gas consumption model, e.g., the oil production, are forecasted by a parametric model. The variation in the fuel gas consumption is small, which is captured by the model. In fact, the strongest variations are present in the end of the data set (month 35-60), which is likely connected to the change from a primary oil to a primary gas-producing field, information that is not included in the meta-data. All prediction models varying the different parameters and prediction methods show a close to constant fuel gas consumption.

The oil production decreases for the parametric model and hybrid model while it increases for the purely data-driven model. This is not the case for all models, but it is observed that the parametric model alone usually outperforms any data-driven model (hybrid or purely data-driven) for the oil production. In a lesser extent this can also be observed for the water production, which is not shown in Figure 2.

|  |  |
| --- | --- |
|  |  |

Figure 2 Example of forecast using GP regression of fuel gas consumption (left) and oil production (right).

The fuel gas consumption is on a longer forecast horizons better predicted using a GP regression model with hybrid modelling. The NRMS of the fuel consumption is about 10%, which is the lowest error of all predicted outputs. It is followed by oil and water production with an error of about 15% and finally by gas production and water injection with an error of about 20%. Nevertheless, the trends are predicted correctly but the oscillation around the trend line is not predicted, which is to be expected.

To analyse the data-driven models, step responses of the identified models were tested. It reveals that some internal correlations are not as expected. According to the models, an increase in water injection decreases oil and gas production and decreases fuel gas consumption. On the other hand, a reduction in oil production also reduces gas production, fuel gas consumption and water injection. These two tests contradict each other. A reason is that each output is predicted with a separate model without considering output correlation. Moreover, the public data set is small and metadata is missing making the creation of data-driven models that capture the physics of the system challenging.

A 2030 outlook on the energy ratio of the Brage field is given in Figure 3. Since it is a forecast, this cannot be validated. The best models in the test set using varying input and output lag for the different regression methods are shown. The dashed lines show the best GP regression forecast models and the dash dotted lines the best linear regression forecasts. Moreover, one forecast (solid green line) for the ER is created using the best combination of prediction models for fuel gas, gas, and oil forecast models validated on the test set. It can be noted that the forecasts vary from an efficiency of about 5.5 - 10.0 in 2030. The combination of the best performing models over the test horizon predicts an efficiency of about 7.0 in 2030.



Figure 3 - 2030 forecasts of energy ratio (ER).

* 1. Discussion

In some tested cases the parametric models' forecasts are the best compared with the data-driven and hybrid models. While the gas, oil, and water production forecasts are improved using parametric models, they do not necessarily have a positive effect on the forecast for other outputs like fuel gas consumption, for which no parametric model was formulated. In fact, a purely auto-regression model of the fuel gas consumption has a similar performance as the model using additional inputs. For the 2030 forecast of the ER, however, the inclusion of the parametric models is essential since the gas and oil production forecast improves significantly, which is part of the ER.

The outputs of the GP regression model are the mean and standard deviation. The standard deviation has not been used in this paper since several models that have also a purely deterministic output were combined. The GP regression models could also be used to create stochastic forecast trajectories. A weakness is that during training it is assumed that the inputs are deterministic and not uncertain while in the long-term forecast uncertain inputs are used. A result is that the uncertainty in the forecast may be underestimated.

When creating the forecast models, physical correlations are not necessarily inferred from the data set, e.g., increase in water injection increases fuel gas consumption. A reason is that a change in one input, e.g., the water injection also influences other variables so the interaction might be more complex. Another reason is that the data set does not account for all consumers on a platform. Furthermore, the data set is small since it has just about 300 data points where just 75% are used for training. It is expected that the forecast methodology improves significantly with a more comprehensive data set. Additionally, the metadata (e.g. installation of new equipment, new well installation, downtime, maintenance time) could be included to avoid wrongly inferred correlations from data.

In this work, just two parametric models were developed, one for oil and gas production and one for the water production. A challenge is to fit a start-up of the production into these models. At the start of a field’s lifetime a production ramp-up can usually be observed. Moreover, a few years constant high production is achieved before the decay starts. These periods cannot be represented by the current models. A practical approach could be to exclude the first years of a field's lifetime for the training of parametric models. In addition, parameter estimation is dominated by the first period of production because of the significantly larger production volume and relatively larger error in the model fit compared to production at later periods. However, the goal is to create a model that predicts well the behavior in the future so periods at the end of the training set should be weighted more strongly. A challenge is, however, to find suitable weights. Another challenge that was observed for the Brage field is production increases due to events that are not part of the meta-data, such as installation of new wells or upgrade of equipment. This can increase production for some time, which can lead to an over-prediction of future production by the parametric model.

* 1. Conclusions

In this paper a forecast methodology was presented that uses publicly available production data from oil and gas fields. It was shown that the general trends can be captured by the forecast models, which helps to estimate future energy consumption on the platform. The parametric models help to support the data-driven models to capture the expected trends for oil and gas productions. In the future, the methodology can be tested on more oil and gas fields. With different types of oil and gas fields, additional parametric models for other production variables may need to be developed.

**Acknowlegdment**

Support from the Research Council of Norway and partners of PETROSENTER LowEmission (project code 296207) is gratefully acknowledged.

References

C.E. Rasmussen and C.K. Williams , 2006, Gaussian processes for machine learning, MIT Press.

D. Fan, et al., 2021, Well production forecasting based on ARIMA-LSTM model considering manual operations. Energy 220, 119708.

D. Koroteev and Tekic, Z., 2021, Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future. Energy and AI 3, 100041.

F. Ueckerdt, et al., 2024, On the cost competitiveness of blue and green hydrogen, Joule, 8,1–25.

International Energy Agency (IEA), World Energy Outlook 2023, 2023. www.iea.org/reports/world-energy-outlook-2023

M.S. Masnadi, A.R. Brandt, 2017, Climate impacts of oil extraction increase significantly with oil field age, Nature Climate Change, 7, 551-556

M. Voldsund et al., 2023, Low carbon power generation for offshore oil and gas production. Energy Conversion and Management: X 17, 100347.

Norwegian Petroleum Directorate, The DISKOS production database (DISKOS NDR), 2022. www.npd.no/en/diskos/production/

Norwegian Petroleum Directorate, Greenhouse gas emissions, 2023. www.norskpetroleum.no/en/environment-and-technology/emissions-to-air

Pan, S. et al., 2023, Oil well production prediction based on CNN-LSTM model with self-attention mechanism. Energy 284, 12870.

U. Bardi, 2019. Peak oil, 20 years later: Failed prediction or useful insight? Energy Research & Social Science 48, 257–261.

T. Holt and H. Schümann, 2022, Energy Efficient Operation of Petroleum Production Plants. SPE Norway Subsurface Conference.