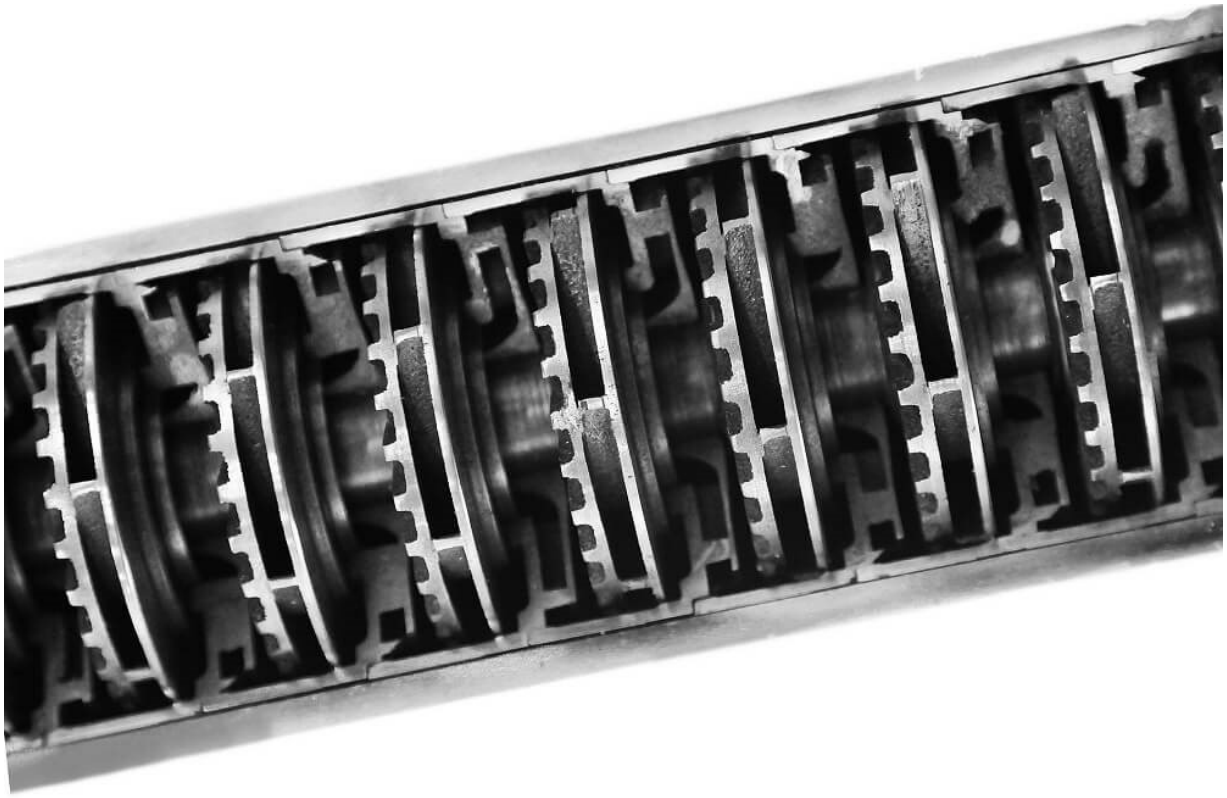
WARMINGUP

Innovatief Duurzaam Warmtecollectief



Model-based monitoring of geothermal assets

Case study: Electrical submersible pumps

Report by:

R. Octaviano, S. Dussi, H. de Zwart, P. Shoeibi Omrani, V. van Pul-Verboom, K. Elewaut, B. van Schravendijk

2022-05-31

Model-based monitoring of geothermal assets, case study: electrical submersible pumps

R. Octaviano, S. Dussi, H. de Zwart, P. Shoeibi Omrani, V. van Pul-Verboom (TNO), K. Elewaut (ECW Energy), B. van Schravendijk (Aardyn)
31 May 2022

Dit project is uitgevoerd als onderdeel van het Innovatieplan WarmingUP. Dit is mede mogelijk gemaakt door subsidie van de Rijksdienst voor Ondernemend Nederland (RVO) in het kader van de subsidieregeling Meerjarige Missiegedreven Innovatie Programma's (MMIP), bij RVO bekend onder projectnummer TEUE819001.
WarmingUP geeft invulling aan MMIP-4 – Duurzame warmte en koude in gebouwde omgeving en levert daarmee een bijdrage aan Missie B – Een CO_2-vrije gebouwde omgeving in 2050.

Projectnummer
060.43190

Keywords
Geothermal doublets, ESP, production optimization

Jaar van publicatie
2022

Meer informatie
Pejman Shoeibi Omrani
T 061 188 82 55
E Pejman.shoeibiomrani@tno.nl

Table of contents

Summary

The operation of the geothermal systems deals with several challenges caused by the brine chemistry, high volumetric flow rates and operational conditions. Often these challenges, such as scaling, corrosion, NORM and pump failures are interdependent and multi-dimensional meaning that there are several parameters causing these issues to occur and mitigating one challenge could enhance the other one. Thus, making a decision in the operational time-scales (up to real-time) could be a challenging task and leading to a sub-optimum decision. The current state-of-the-art is to take a reactive decision or in other words the decision is made when it is too late. The main question to address is how operational decisions can be supported with real-time data, enabling operators to go from reactive to proactive operation?

A generic model-based monitoring framework was developed for the geothermal applications. This framework enables an integrated monitoring of the geothermal wellbore and top-side facilities performance with real-time data and predictive models. This workflow could be applied for a single component to the whole system and plant. In order to demonstrate the workflow, a case study needed to be selected. A thorough survey with several geothermal systems' operators was performed to gain insights on the impact and frequency of various operational risks and failure registers and determine which failure mode could be best demonstrated with the developed workflows. Several operational challenges and failure modes were investigated and the case study was selected based on three criteria; data availability, operators support and partners consensus. Based on this survey, the monitoring of the electrical submersible pumps (ESP) was selected as a case study.

To monitor ESP performance and assist with operational decisions, the developed workflow consisted of physics-based and data-driven models. This physics-based model monitored pressure and temperature in real time from dynamic Inflow Performance Relationship (IPR), Vertical Lift Performance Relationship (VLP) and ESP pump performance curve. Other monitoring variables (e.g. motor temperature, vibration, ESP power, etc.), whose trends are difficult to capture using physics-based approaches, either in terms of availability of internal ESP data or high computational time, were modelled using Machine Learning. In particular, a random forest regressor and a neural network were employed as methods. In addition, an explainable AutoEncoder was also used for early anomaly detection based on sensors data. All these models could then be implemented in an integrated environment to provide operators with real-time Key Performance Indicators (KPIs) for decision support. Finally, a transfer learning approach was tested to demonstrate how the monitoring workflow can be improved in accuracy and reliability.

Three Dutch geothermal wells were selected with different types of ESP failure and different ESP lifetime (<1 year, between 1-2 year, >2 years). The model was calibrated (history-matched) at the beginning of the ESP operational life (assuming no ESP performance degradation occurred yet), and then was used to predict future ESP behavior. The model error is used to monitor early pattern of ESP failure. Physics-based models were able to monitor an increasing error between measured sensor data and predicted value (1-10% absolute error) around 1 month (well 1), 3 months (well 2) and 6 months (well 3) before failure. Data-driven models showed similar error trends for well 2 and well 3, whereas not for well 1 due to a limited operational variations present in the training dataset. The third model also showed that the AutoEncoder has similar error trends with the other models.

The Reconstruction error increased close to ESP failure and insights on the most relevant sensor causing the anomaly were obtained.

Overall, the proposed monitoring tool equipped with early warning flags coming from these 3 complementary models can help operators to do preventive maintenance or changing operational conditions before ESP failure. For future works, the workflow will be tested on a larger well database to estimate how long the ESP will fail from the detection of mismatch, providing therefore predictive capability to the monitoring tool. The developed workflow was generic enough to be deployed for other operational challenges.

1 Introduction

Geothermal energy has great potential to contribute to the heat transition and to make the heating of the horticulture sector and built environment in the Netherlands free from natural gas. Apart from the investment decisions to drill the wells at a right location in a cost-effective manner, the operational challenges can have a big impact on the business case and CO_2 footprint of the geothermal systems and consequently impact the uptake of geothermal systems. The operational issues such as scaling, corrosion, NORM and pump failures occur in many geothermal plants (Shoeibi Omrani, et al. 2021) which either lead to a decrease in the productivity and injectivity of the doublets (Wasch, et al. 2019) or eventually it cease the heat production.

The operation and maintenance of the systems can be in three manners, reactive, preventive and predictive (Figure 1). In the reactive strategy, the system is being operated until the equipment (nearly) fails and all the components are used to their limit. This strategy could lead to a very long and costly downtime and restart. In order to take a robust strategy, some of the operators often take the preventive measures meaning that they perform regular and planned maintenances. The disadvantage of this conservative approach is that the cost of replacement of components in the lifetime of the system is sub-optimum since there is a high chance that some of the components are changed unnecessarily. In addition, in case of off-normal cases, the operators might still face the unforeseen failures which will be a reactive strategy. If the performance of the system and components can be predicted, then the operation will be more proactive. It is important to note that the prediction accuracy of the model is crucial in deployment of the predictive strategy.

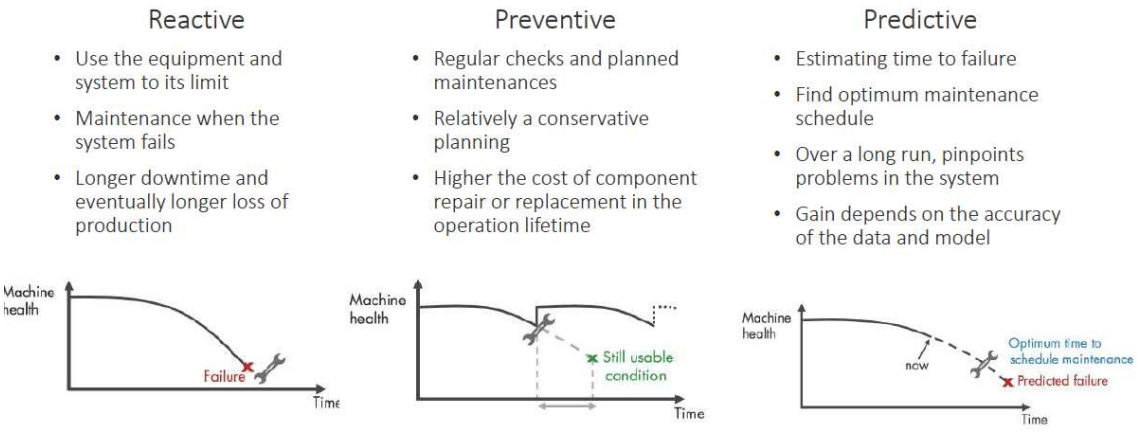


Figure 1. Three different operation and maintenance strategies, from reactive to preventive and predictive. Adapted from MATLAB

The main objective of this study is to demonstrate a model-based monitoring and optimization technology to enable short-term production optimization in geothermal systems up to real-time scales and support operators for operational decisions.

The next chapter describes the criteria to select the case study for the demonstration of the technology. The developed methodology is presented in the chapter 3. The datasets used in the study were from three geothermal wells in the Netherlands which are further described in chapter 4. The results of applying the developed methodologies on three wells are demonstrated in the chapter 5 which is followed by the conclusion chapter.

2 Case study

In order to demonstrate the workflow, a case study needs to be selected. A survey was setup to collect information on different operational challenges in geothermal systems, the frequency of their occurrence and operational cost estimates for mitigation or remediation of them. Firstly, each part/component of the geothermal system was discussed with several operators, from producer to the injector well according to Figure 2. The following operational challenges were discussed in all the component depicted in Figure 2; scaling, corrosion, NORM, pump reliability and failure, clogging and erosion.

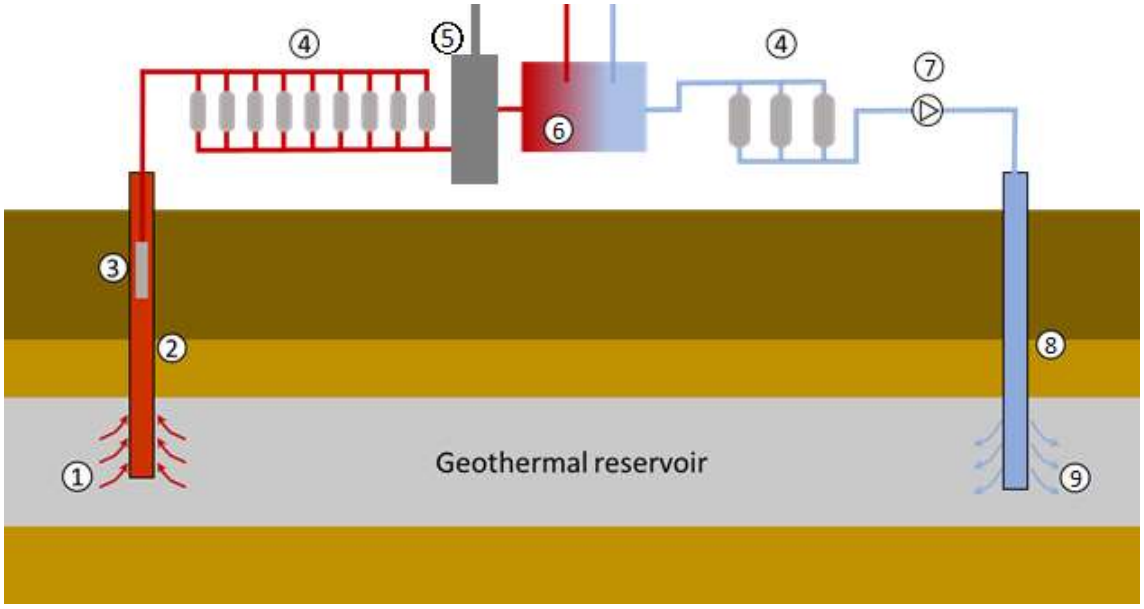


Figure 2. Schematic of a geothermal doublet, (1) inflow zone, (2) producer well, (3) ESP. (4) filters, (5) degasser/separator, (6) heat exchanger, (7) booster/injector pump, (8) injector well and (9) inflow to the reservoir

Based on the collected cost figures, the CAPEX and annual OPEX of two different type of systems were compared. The results are demonstrated in Figure 3. Operator 1 case is an averaged cost value of several systems with a similar design. It can be seen that on average the annual OPEX is around 8% of the CAPEX. The second operator was using a different design and more durable materials which led to a higher CAPEX. However, the percentage of the annual OPEX was way lower compared to the first case. This comparison suggests that an increase in the CAPEX in this case led to a saving in an annual OPEX and a lower overall cost in the lifetime of the geothermal system.

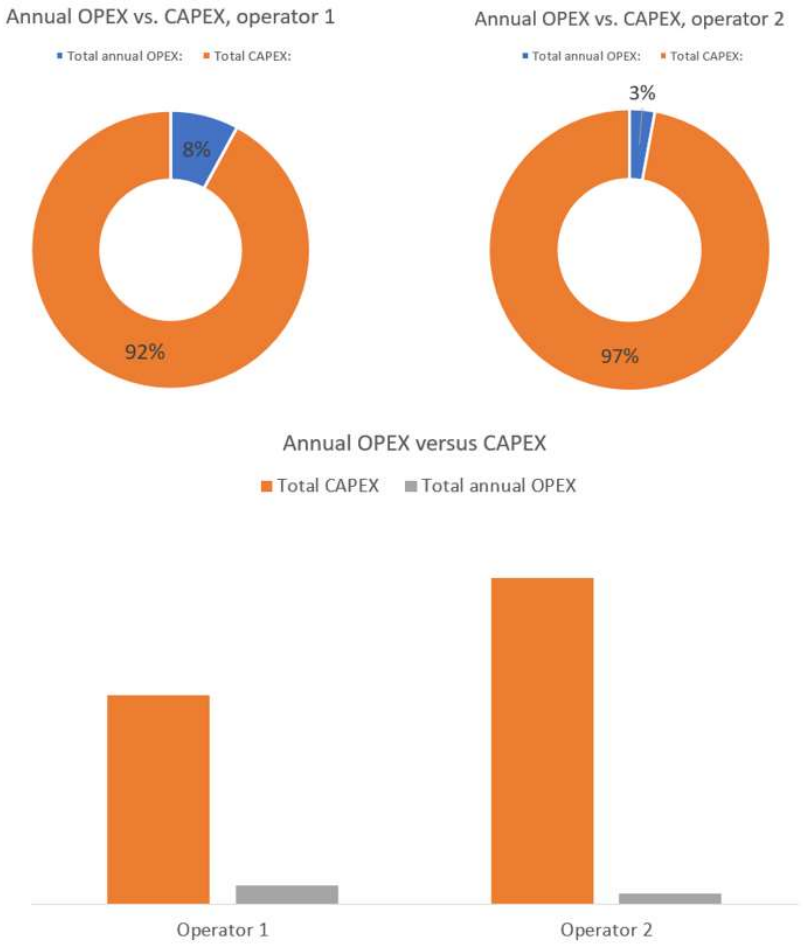


Figure 3. Annual OPEX versus CAPEX for two different type of geothermal systems

For the same cases, the contribution of each component in the geothermal system to the OPEX is shown in Figure 4. In both cases, it is evident that ESP inspection, replacement and associated plant downtime is one of the most dominant component in the annual OPEX. Another highlighted item is the OPEX related to the production and injector wells which drops significantly when more durable and higher grade materials and equipment was used. The surface facilities and filters see a relative increase but in practice these differences are marginal. The details of this study can be found in a separate report by Hidde de Zwart (de Zwart 2021).

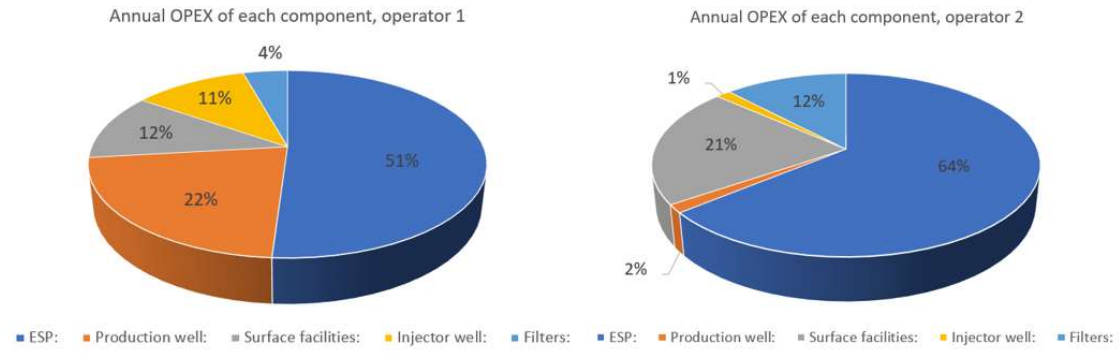


Figure 4. Contribution of each component in geothermal system to the OPEX for two different wells, the well on the right has a higher CAPEX and more durable materials in the wells

Based on the collected data and information, several case studies were formulated to demonstrate the added value of production and operation monitoring of geothermal systems. The case studies were:

- ESP and injection pump performance monitoring
- Monitoring of clogging in the filters and topside facilities
- Scaling monitoring
- Corrosion monitoring
- Handling of the bi-produced gasses
- Monitoring of the sand production in the wells and top-side facilities

These case studies were evaluated based on three criteria: consensus of the project partners, availability of data and support of operators. Figure 5 shows the ranking of consortium interest in different case studies which were presented. It can be seen that ESP, scaling/clogging and corrosion monitoring found to have the highest interest among the case studies. This interest list is also linked to the availability of the data in order to allow for an accurate model-based monitoring of the proposed processes.

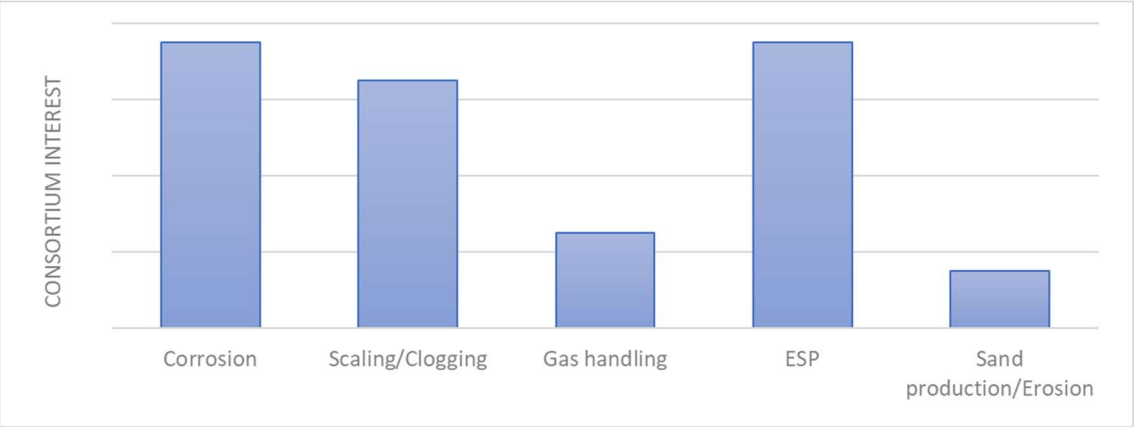


Figure 5. The ranking of consortium interest for different case studies

For all the case studies, there are some general data required which are listed below:

- Production data (time-series of production rates, pressure, temperature, …)
- Well completion, trajectory and top-side process diagram
- Fluid composition
- If available, result of a well test

For the specific case studies the following data will be required:

- ESP and pump monitoring; ESP and injection pump curves and specifications, pump monitoring data, maintenance logs.
- Corrosion monitoring; coupon sampling results or other corrosion monitoring such as ER If available, inhibitor type and dosage
- Clogging/scaling monitoring; filter specifications, surface installation measurement (pressure and temperature), logs of filter exchange or heat exchanger interventions, if available precipitated sample analysis

All the case studies were compared against the three criterion and summarized in Table 1. From the list, it is evident that ESP and injection pump performance monitoring checked all the criterion since it is an important factor in OPEX and reliability of geothermal energy production and it is one of the well instrumented component in the geothermal systems (including several downhole sensors such as motor temperature, vibrations, pressures, etc.). Thus, ESP performance monitoring was selected as the case study to be pursued in the rest of the project.

Table 1. The list of potential cases and the comparison of three criterion for the selection of the case study

Potential cases	Resulting in	Consensus of partners	Data availability	Operator support	Remarks
ESP and injection pump performance	1. Monitor performance degradation of pumps 2. Optimize operational envelope of pumps				Available data for all operators, Signficant part of OPEX
Monitoring of clogging in filters and topside facilities	1. Monitoring of performance vs KPI 2. Optimum time for filter replacement / cleaning				Not the largest part of OPEX, but high frequency in replacement of filters.
Scaling monitoring	1. Monitoring scale deposition 2. Optimum mitigation strategy (for instance degassing, inhibitor planning, dosing).				Not cosidered as the top priority. Lead scaling was mentioned.
Corrosion monitoring	1. Monitoring of corrosion rate 2. Optimum mitigation strategy (inhibitor planning and dosing, …)				A significant issue. Uncertainty in availability of relevant data

The electrical submersible pump (ESP) is one of the major artificial lift systems installed in the world. They are consisting of a motor, seals and pump and they are an important component in supplying the required flow rates in the geothermal systems. ESP can often suffer from performance degradation or failure due to changes in the system performance, wrong operational settings or an operator mistake. Apart from the production downtime due to ESP maintenance or replacement, there is a large cost associated with the replacement of ESP. A robust ESP operational decision support system is therefore needed to operate both well and ESP in an optimal way, and to enable operators to perform early failure detection/prediction for prolonging ESP run life. The average lifetime of ESPs in geothermal systems were found be around 2-3 years (Shoeibi Omrani et al. 2021). The main observed ESP failures, from the statistical analysis performed by the ESP RIFTS project for oil wells found to be:

- Electrical failures (motor, cables, phase imbalance)
- Well condition failures (scaling, corrosion, erosion, …)
- Component failures (seal, bearing, shaft and sensors)

There are several methods to perform a root-cause failure analysis (RCFA) of ESP based on the observed trends in the data. The availability of handbook guidelines could support operators to perform a RCFA, however this analysis is performed when the failure has already occurred. By deploying the model-based condition monitoring systems, which will be described in the next chapter, it is aimed that these decisions can be made in real-time.

3 Methodology

This section describes the methodology applied in this work, including brief descriptions of the theoretical background of the physical modelling and of the data-driven techniques.

3.1 Proposed workflow

Currently, there are two types of condition indicators which can be derived either from sensors or models, as shown in Figure 6. Condition-based monitoring extends beyond the use of sensor data to monitor the processes and their performances, known as signal-based condition indicators. The use of model-based condition monitoring/indicators can extend the signal-based monitoring capabilities by employing fast (up to real-time) models to monitor derived parameters (e.g. efficiency, performance) and/or parameters which are not continuously measured (e.g. corrosion rates, scaling deposition, …). This can provide additional information to the operators for an in-time decision.

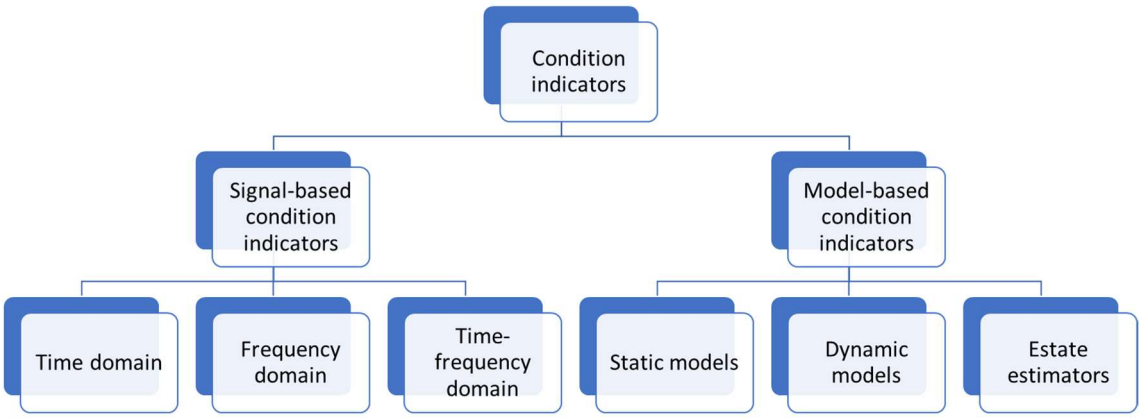


Figure 6. Different type of condition monitoring/indicators mainly signal-based or model-based.

One of the basis of the model-based condition indicators or monitoring is the system models which can describe the behaviour of the component(s) under various operational conditions. The type of model depends on several factors including the dynamics of the process, domain knowledge to model the intended process and data availability. These models can be either physics-based or data-driven models. In order to have accurate models, field data from the signal-based condition monitoring systems are required. This data can be used to calibrate/train the models and also act as an event detector in case of off-normal behaviours in the system. The model-based condition indicators could provide several layers of analytics; from descriptive to diagnostics, predictive and prescriptive. For the current application within this study, only descriptive and diagnostics analytics were derived.

A proposed workflow is demonstrated in Figure 7. The first step of the workflow is about a set of models to describe the behaviour of the system and its components. These models are incorporated within the monitoring suite. The developed models need to be trained/history matched (HM) using the historical data which are in the historic database. In the model training step, the averaged errors need to be evaluated and registered which is an important step during the model-based monitoring to distinguish between model error or system performance degradation. After the training step, the

models will be connected to the real-time database. The models will enable detecting mismatch between the predicted values and measured values. The output of the models can be integrated into a visualization layer to provide insights to the operators. In case of any performance mismatch an alarm can be triggered and sent to the operators/expert for validation and afterwards saved into the event database. By utilizing this workflow overtime, more and more events are being detected and validated and the event database becomes more accurate. In the next section, the modelling approaches are further explained.

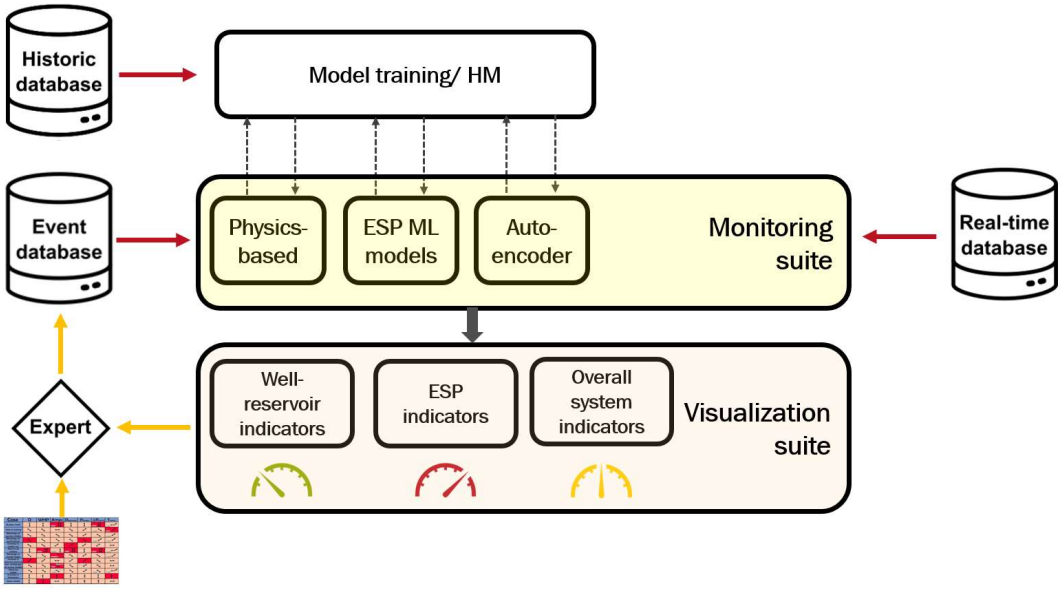


Figure 7. Schematic of the proposed workflow for the model-based monitoring and the integration method with the historical and real-time production database

3.2 Physics-based modelling

In this section, we will describe our approaches using model-based condition indicators to monitor ESP performances. This simplified and fast model can be implemented in real-time to give additional insights to the engineer.

3.2.1 Bottomhole pressure model

The idea of monitoring downhole pressure during operation is to know in advance if there is additional resistance due to scaling, sand or skin that can occur at different locations such as in the formation, in the near well-bore region, or well tubing. The additional resistance can lead to a decline of production or a higher workload for the ESP.

Inflow Performance Relationship (IPR)

IPR is defined as well flowing bottomhole pressure (P_{wf}) as a function of measured production rate (Q). P_{wf} is defined in the pressure between average reservoir pressure (P_{res}) and atmospheric pressure. An example of a typical IPR graph is shown in Figure 8.

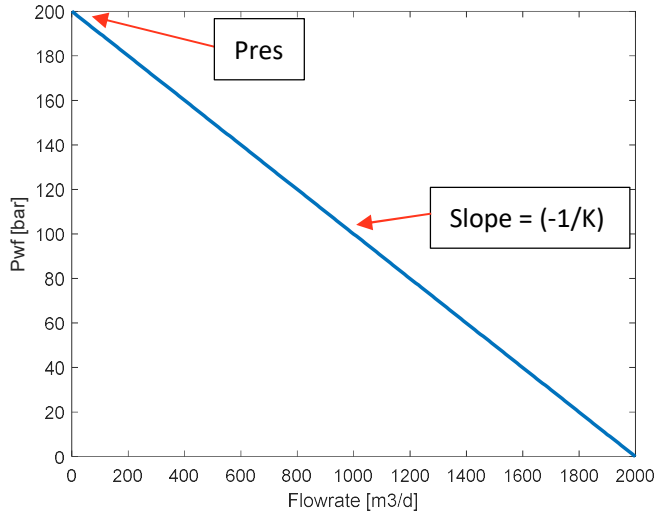


Figure 8. Example Inflow Performance Relationship graph

The y-intercept is defined as reservoir pressure when there is no flow. The negative slope (K) is also called by Productivity Index (PI), a ratio between flowrate and well drawdown.

$$P_{wf} = P_{res} - \left(\frac{Q}{K}\right)$$

$$K = \frac{Q}{P_{res} - P_{wf}}$$

Where,

P_{res}	reservoir pressure [bar]
P_{wf}	flowing bottomhole pressure [bar]
Q	flowrate [m3/d]
K	productivity index [m3/d/bar]

Vertical Lift Performance Relationship (VLP)

VLP is defined as the bottomhole pressure as a function of flowrate in the tubing. The VLP depends on the well depth, well trajectory, tubing size, water cut, GOR, or PVT fluid properties. The boundary condition in case of there is no ESP is the wellhead pressure (P_{wh}) or the intake pressure of ESP if there is ESP installed. In order to calculate the pressure drop along the tubing from topside, we use 2 correlations based on fluid phase: single phase or two-phase.

For a single phase flow (Techo, Tickner en James 1965), the total pressure loss govern by gravitational and frictional pressure drop. The gravitational pressure drop is calculated as function of local gravity and well tubing inclination.

$$\Delta P_{grav} = \rho_l g \sin\theta$$

Where,

ΔP_{grav}	gravitational pressure drop [bar]
ρ_l	liquid local density [kg/m^3]
g	local acceleration due to gravity [m/s^2]

θ tubing inclination [rad]

The frictional pressure drop is calculated proportional to the square of the flow velocity and inversely proportional to the pipe diameter know by Darcy-Weisbach equation.

$$\Delta P_{fric} = \lambda \frac{1}{2} \rho_l \frac{u^2}{D}$$

Where,

λ friction factor or flow coefficient
u mean velocity [m/s]
ρ_l liquid local density [kg/m^3]
D pipe diameter [m]

In turbulent flow, many equations have been given for determining friction factor λ as a function of Reynolds number (Re). The implicit equation may be solved by the Newton-Raphson iterative technique or may be approximated by following equation

$$\lambda = \left[0.86859\ ln\left(\frac{Re}{1.964\ln(Re) - 3.8215}\right)\right]^{-2}$$

$$Re = \frac{uD}{\nu}$$

Where,

Re Reynolds number
u mean velocity [m/s]
D pipe diameter [m]
ν kinematic viscosity [m^2/s]

Thus total pressure loss is calculated by

$$\Delta P_{total} = \Delta P_{fric} + \Delta P_{grav}$$

$$P_{wf} = P_{wh} + \Delta P_{total}$$

For two-phase flow in inclined pipe, there is additional parameter besides two-phase friction factor called liquid holdup. The liquid holdup has a dependency on angle. The correlation of liquid holdup divided into 3 horizontal flow patterns, e.g. Segregated, Intermittent and Distributed. More detailed equations are presented in (Beggs en Brill 1973).

Nodal Analysis

Nodal analysis uses both IPR and VLP correlation. The intersection between these two lines called operating point where the actual flowrate is produced by the well for a given operating condition (Figure 9).

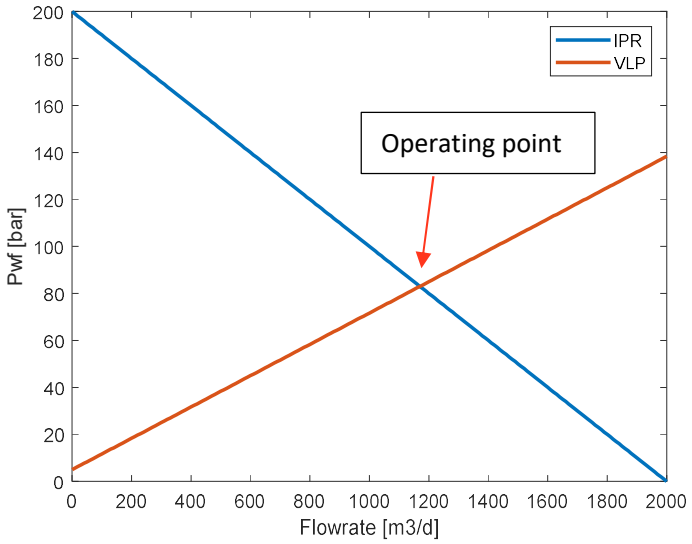


Figure 9. Example of operating point between IPR and VLP

The calculated flowrate using nodal analysis can be found by minimizing P_{wf} calculated from IPR and P_{wf} calculated from VLP.

$$\min_{Q}(P_{wf,1} - P_{wf,2})^2$$

For the system without ESP, we can use well head pressure (P_{wh}) as topside boundary condition. If the wellhead pressure is not available, we also can use tank pressure or pipeline pressure with additional pressure drop. For the system with ESP, we can use the intake pressure as boundary condition to calculate nodal analysis (Figure 10).

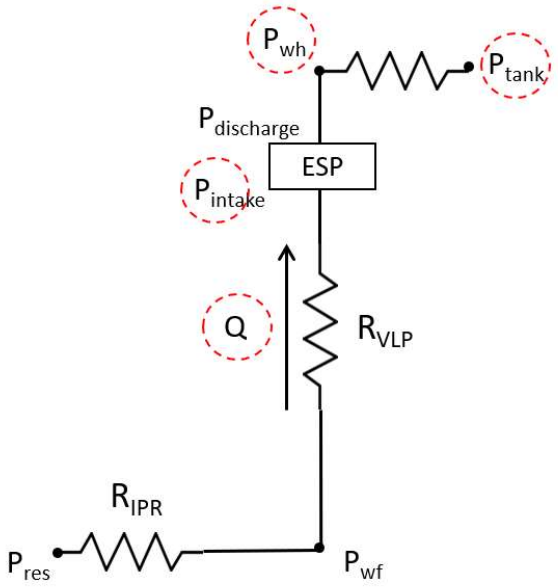


Figure 10. Schematic of well-reservoir nodal system

This model can be used in real-time to monitor P_{wf} if there is any download pressure measurement. We can see if there is any an increasing error between calculated P_{wf} and measured P_{wf}. In case of there is no downhole sensor, we can monitor calculated flowrate from nodal analysis with measured

flowrate. If the measured flowrate decreases, it means that there is additional resistance in downhole.

3.2.2 ESP pump curve model

ESP performance will degrade over time during operation, either it requires more power or more pump head to deliver the required flowrate. To monitor the ESP performances from all measured sensor around ESP (e.g. pressure, frequency, efficiency, power), we can use simplified model theoretical ESP pump curve from the vendor. This theoretical pump curve is assumed as an ideal ESP where there is no ESP degradation due to wear or scaling. When there is systematic error between calculated value with measured value, it means there is ESP performance degradation.

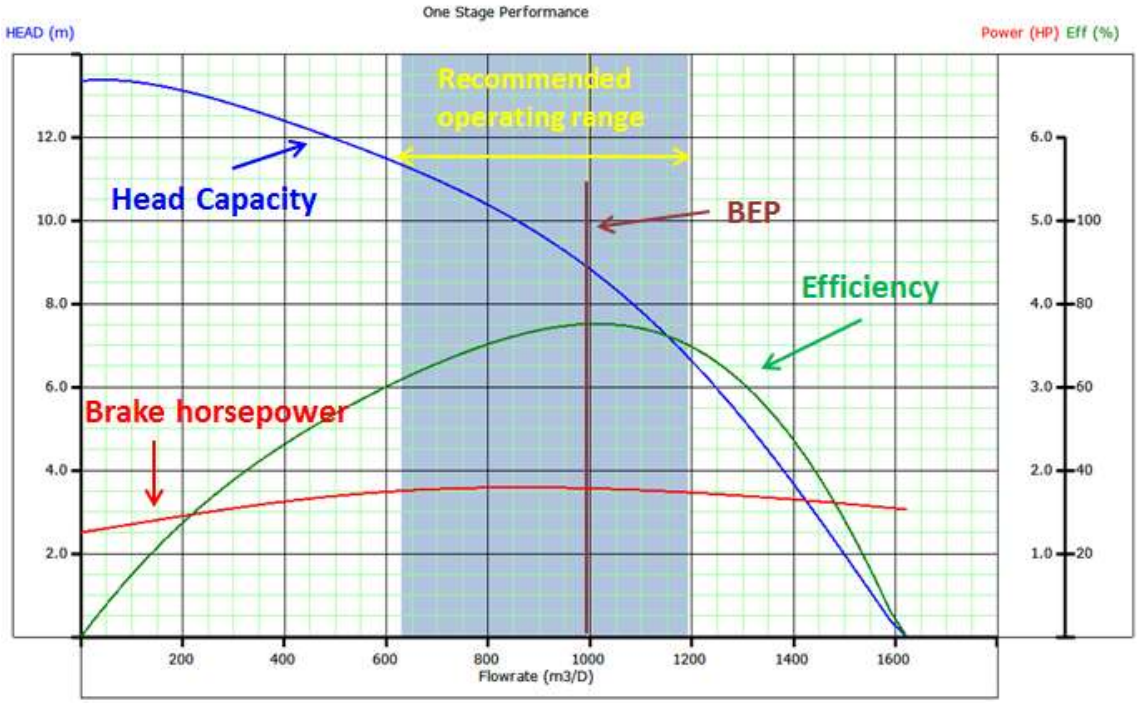


Figure 11. Example of one stage ESP pump curve at 60 Hz from the vendor

As shown in Figure 10, the pump head can be monitored by comparing measured pump head (discharge pressure - intake pressure) with ESP pump head capacity curve. However, most of geothermal system in the Netherlands, there is no sensor measuring discharge pressure. Thus, we need to calculate discharge pressure using VLP from topside pressure.

In a typical situation, the ESP pump curve is delivered by the vendor as a table for each frequency. To avoid discontinuity, either we develop a look up table with an interpolation or we can also develop a fit function from one stage 60Hz line. In this case, we use 5th order polynomial approach.

$$\Delta H = \text{stage}\left(\frac{f}{60}\right)^2 (c_0 + c_1 Q + c_2 Q^2 + c_3 Q^3 + c_4 Q^4 + c_5 Q^5)$$

$$BHP = \text{stage}\left(\frac{f}{60}\right)^3 (d_0 + d_1 Q + d_2 Q^2 + d_3 Q^3 + d_4 Q^4 + d_5 Q^5)$$

Where,

ΔH Pump head [bar]

BHP	Power [kW]
stage	Number of stages
f	ESP frequency [Hz]
Q	Flowrate [m3/d]
c	pump head coefficient
d	brake horsepower coefficient

3.3 Data-driven ESP model: Random Forest Regressor and Neural Networks

To construct a data-driven model for the ESP, two different supervised learning techniques were employed that will be briefly described below. Supervised learning consists in training/calibrating the model based on known input-output pairs, before using the model on 'unseen' data. In both cases, (part of) the sensor data associated to the ESP were used either as model input or as output. The goal was to use only a subset of the available quantities to predict another quantity associated to the ESP using the data-driven algorithm. The timeseries were split in two periods, one for training the models and one for testing their predictions. The training period should consist of data associated to the 'healthy' period of the ESP, meaning that no degradation or malfunctioning occurred yet. In the testing period, the mismatch between the values predicted from the model(s) and the values measured by the sensors can be used for monitoring purpose. A large mismatch would indicate that the quantity of interest does not follow the expected model behavior (that should correspond to the healthy behavior of the ESP). This can also give insights into potential causes of ESP failures, as will be discussed in Chapter 5.

3.3.1 Random Forest Regressor

Random forest is an ensemble learning method, meaning that it is a technique that combines the predictions of multiple models, trained on the same dataset, into one. The final prediction is expected to be more accurate than the one obtained from an individual model. The random forest model can be used either for classification or for regression tasks. Here we used it for regression. The random forest regressor (RFR) algorithm creates a random sample of multiple decision trees and merges their predictions, as schematically shown in Figure 12(a). More specifically, RFR is based on the bootstrap aggregation (or bagging) of several decision tree models. Decision trees are built by recursively splitting the input into sub-sets based on rules learned during the training process. Each tree makes a prediction. RFR uses averaging and cross-validation to improve the final prediction accuracy and to control over-fitting. In this way, RFR is more robust to the bias and the variance of the dataset compared to a single decision tree. In this study, we typically used n=100 trees.

3.3.2 Neural Networks

Neural networks (NN) are popular machine learning models with a range of possible architectures and suitable for numerous tasks. Here we used a feed forward NN as schematically shown in Figure 12(b) for a regression task. The inputs of the NN are the values of the selected variables (some of the ESP sensors) at a given time. The output to be predicted is the value of another ESP-related quantity (not included in the input set) at the same time of the input. The relationships between the input and output is approximated by the NN via the several parameters (e.g. weights of predefined

functions) associated to each node (also called neuron. We typically used an architecture composed of two hidden layers with size double of the input size. A relu activation function is used for the NN nodes, except for the last node where a linear activation function is used. The conventional adam optimizer with a loss based on the mean squared error is used for training the NN. Early stopping is implemented through cross-validation based on 15% of the dataset (of the training period). The results were robust against different choices of hyperparameters. The trained NN is then used on the testing dataset (typically a portion of the timeseries towards the end of the ESP run life) to predict the output quantity and compared it to the measured value.

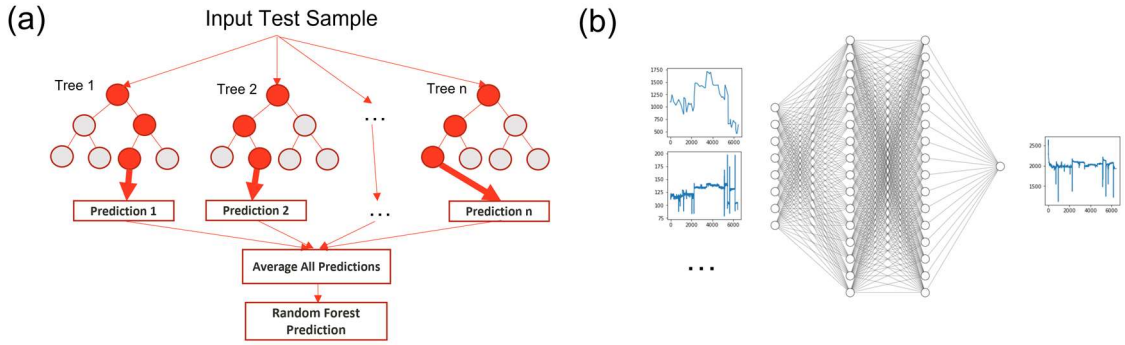


Figure 12. (a) Schematic of Random Forest Regression algorithm. (b) Schematic of a Neural Network with several input nodes, two hidden layers, and a single output. Inputs and output are from the timeseries associated to the different ESP and well sensors.

3.4 Deep learning for system monitoring

To monitor the entire system simultaneously, instead of a single variable at a time as explained in the section above, a deep learning method based on autoencoder was deployed.

An autoencoder is a neural network that aims to replicate the input quantities. The typical network architecture of an autoencoder consists of a 'bottleneck' made of one or more hidden layers that are smaller than the input and output layer (that have both the same size), see Figure 13. The goal of having such an architecture is to first encode the input, meaning that a representation of the input data based on a smaller number of variables is learnt. This is achieved by the layers from the input to the middle one where the dimension is progressively reduced. In this way, the model is forced to construct (learn) only the most important relationships between the features, neglecting insignificant correlations and noise in the data. To ensure that the representation of the input data is a valid one, in the second part of the autoencoder the layers have increasing dimension, such that the information is 'decoded'. The autoencoder is trained to minimize the reconstruction error between the original input data and the model output. Typically, for our analysis we used a symmetric autoencoder with one or three hidden layers, with the middle one of size 3 or 4 node (depending on the number of input variables considered), with an exponential linear unit (elu) activation function. The conventional adam optimizer with a loss based on the mean squared error is used during training.

Autoencoders can be used either in an unsupervised or in a semi-supervised mode. The semi-supervised mode was chosen, the timeseries of the sensor data was split in a training and testing period. The autoencoder is then trained using the data from the first period. Afterwards, the trained model is deployed in the testing period and the reconstruction error can be monitored. An 'anomaly' in the data corresponds to a large error between the measured (sensor) value and the model prediction. If the anomaly is associated to only one or very few consecutive datapoints, it might be considered as a 'false warning', caused for example by an inaccuracy in the model. If it is however

for a longer period, the warning should be investigated to check if it can represent a potential cause for ESP degradation, malfunctioning or eventually failure. It is therefore crucial to understand which model features are the most important, as will be discussed in the following section.

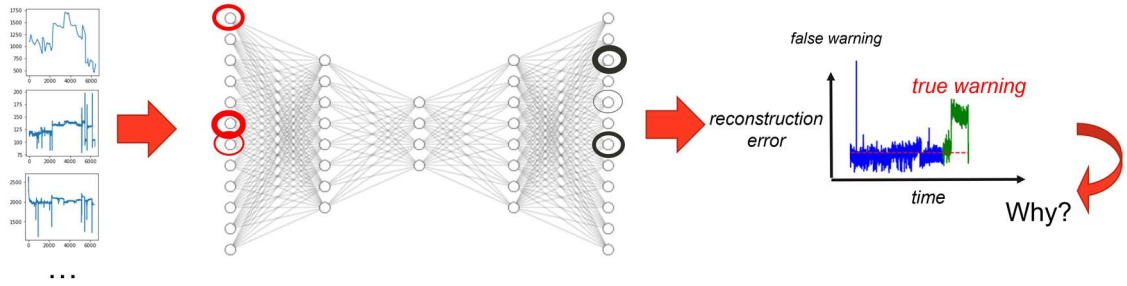


Figure 13. Schematic workflow on how to use an autoencoder for system monitoring. Timeseries data are fed into the autoencoder that tries to reconstruct the input, while it is being forced through a bottleneck (smaller hidden layer) to capture the most relevant correlations describing the system. After a training interval (blue period in the right panel), it is possible to use the autoencoder reconstruction error to monitor the mismatch between model expectations and actual measurements ('anomaly' detection). When the reconstruction error is large for a long period, the cause of the potential true anomaly should be investigated. By performing a SHAP analysis on some or all output variables (black circles), it is possible to pinpoint the most important model features (red circles, input side of autoencoder) that can give insights into the cause of the anomaly.

3.5 Model explainability through Shapley values

The machine learning algorithms introduced above will be able to capture complex interdependencies between the variables and provide predictions for the quantities of interest. However, they can be viewed as black-box methods since due to the underlying model complexity it is difficult to gain a clear interpretation of why the model is making a certain prediction.

To provide an explanation to the predictions of the data-driven models, a framework based on Shapley values to identify the most important model variables was used. Shapley values were first defined in the context of cooperative game theory to determine the contribution to each player in a coalition (Shapley 1953). Applied in the context of machine learning, it became a method to identify the most and least relevant features for a given model. The Shapley value is the average marginal contribution of a certain feature among all the possible feature combinations.

Depending on the complexity of the machine learning model (for example in case of non-linear models), such values can be computationally very expensive to calculate, since often require the evaluation of many combinations. In our study we used the SHAP (Shapley Additive exPlanations) library (Lundberg and Lee 2017), that can handle different machine learning models and has implementations of ad-hoc algorithms to calculate the Shapley values. Given a particular prediction, a value is assigned to each model feature. A ranking among the different features can be established for example looking at the average absolute value of several predictions.

SHAP analysis was applied to the results obtained with the Random Forest Regressor and with the Autoencoder. For the implementation in the latter case, a procedure described in (Antwarg, et al. 2021) was used.

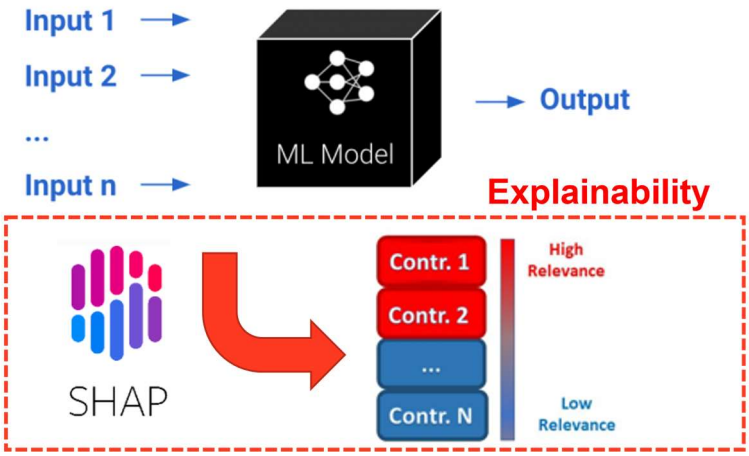


Figure 14. Typical machine learning approaches are regarded as black-box methods. Looking at Shapley values (using the SHAP Python library), it is possible to rank the most important parameters contributing to the model and provide an explanation to the model predictions.

3.6 Transfer learning

Transfer learning is a machine learning approach that consists in exploiting the knowledge acquired while solving a problem and using it to solve a different but related problem, as schematically shown in Figure 15. It is becoming popular in the context of deep learning techniques (such as neural networks), where specific elements (e.g., weights, layers…) of a model that has been trained to accomplish a certain task are transferred to a second model that needs to perform a second task, often more specific than the first task. In this project, transfer learning is used in Section 5.4 for a case of ESP monitoring when only limited data are available. The monitoring method is based on using a NN for regression as described in Section 3.3. For transfer learning, a pre-trained NN (model 1) is repurposed by performing another training process on a new smaller set of data (dataset 2). Specifically, the knowledge transfer consists in using the weights of the pre-trained NN (model 1) as the initial weights for the training process. In this way, we can obtain a calibrated model (model 2) that exploits the previous information stored in model 1.

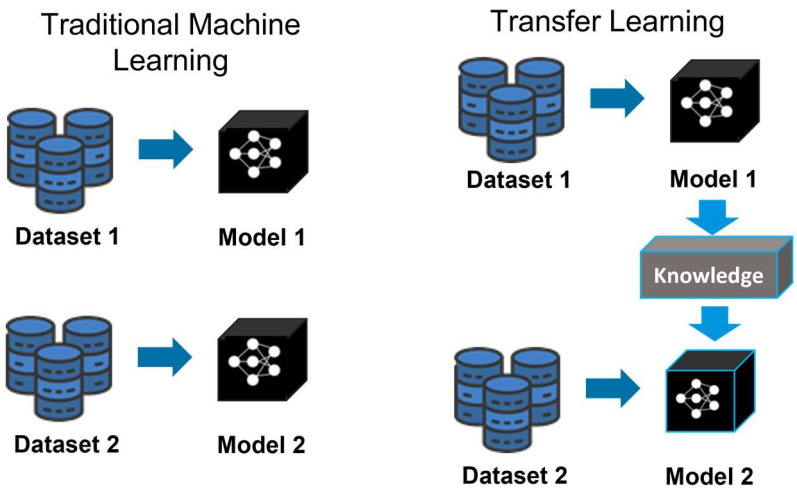


Figure 15. Traditional machine learning approaches consist in constructing an independent model for each dataset available. On the contrary, transfer learning is based on exploiting the knowledge acquired during the construction of the first model and repurpose it for the second dataset.

4 Dataset

In this chapter the field cases used in the study are described. Operators supplied data from 3 geothermal wells in the Netherlands. Each well has different ESP run life, e.g. less than 1 year, between 1-2 year and more than 2 years. The data are divided into time-series data from sensor measurement with different time interval and static data regarding well schematic, ESP properties, log book and fluid properties.

4.1 Received data

4.1.1 Well 1

The operator provided a 2-years ESP dataset from 1-1-2019 to 1-3-2021. Well 1 has 1 ESP failure due to electrical problem on January 2021 (Figure 16), then the engine was replaced. The installed ESP is a 14-stage ESP (Figure 17) and located 700 m from the surface. We also received time series data from several sensor with interval 15 minutes and 1 hour. List of available sensors are presented in Table 2. There is no missing sensor value in this well except wellhead pressure in the first several weeks during start-up.

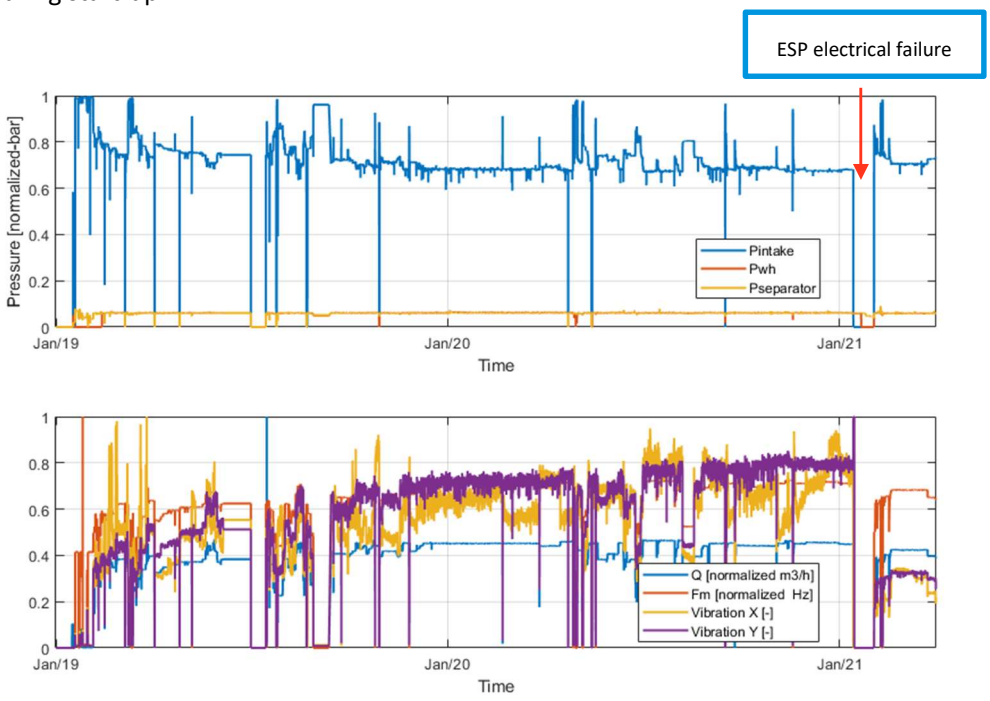


Figure 16. Normalized time series sensor data of well 1

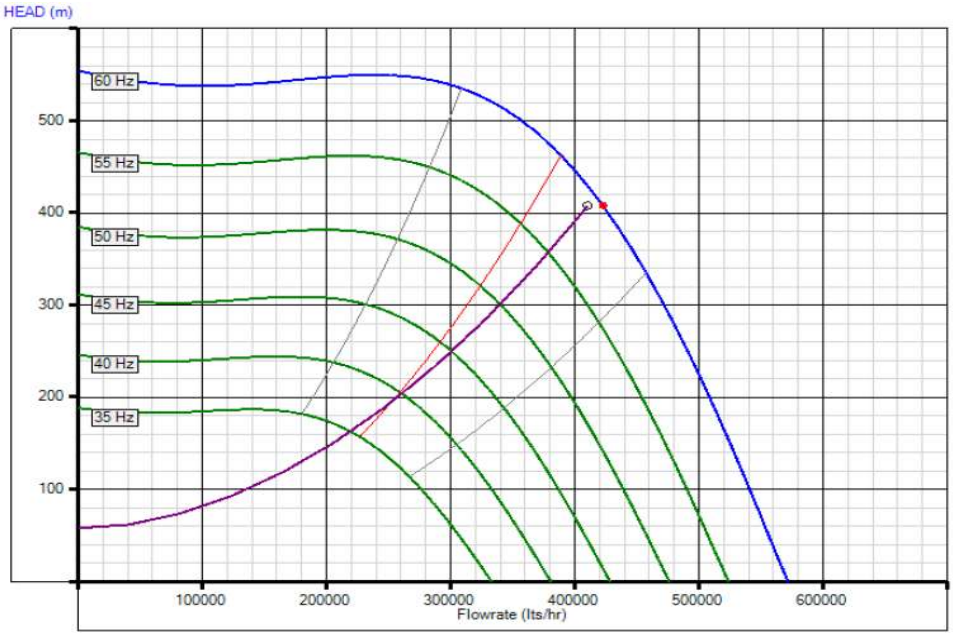


Figure 17. Pump normal curve 14 stages ESP (van 't Spijker and Ungemach 2016).

4.1.2 Well 2

Operator provided a 2.5 years ESP dataset from 1-10-2018 to 1-5-2021 (Figure 18). There were 3 ESP failures on September 2019, April 2020 and November 2020. The ESP is a 12-stage ESP (Figure 19) and located in depth of 770 m. This well has several missing sensor data between May-November 2020. As a result analysis was performed only in time where all the data is available which is ESP 1st failure due to mechanical problem. However, the run life of this ESP is limited less than 1 year.

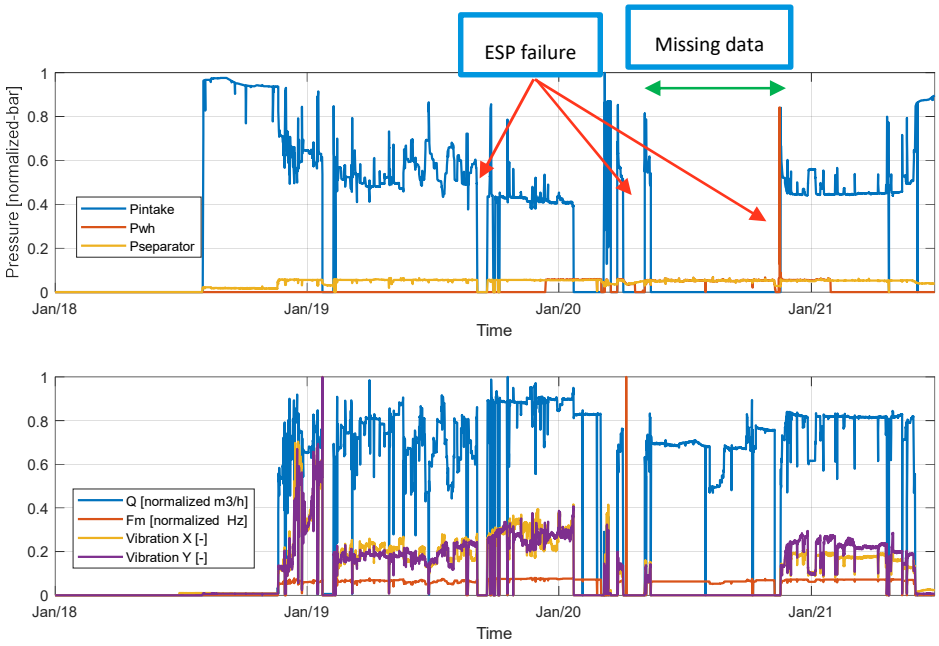


Figure 18. Normalized time series sensor data of well 2

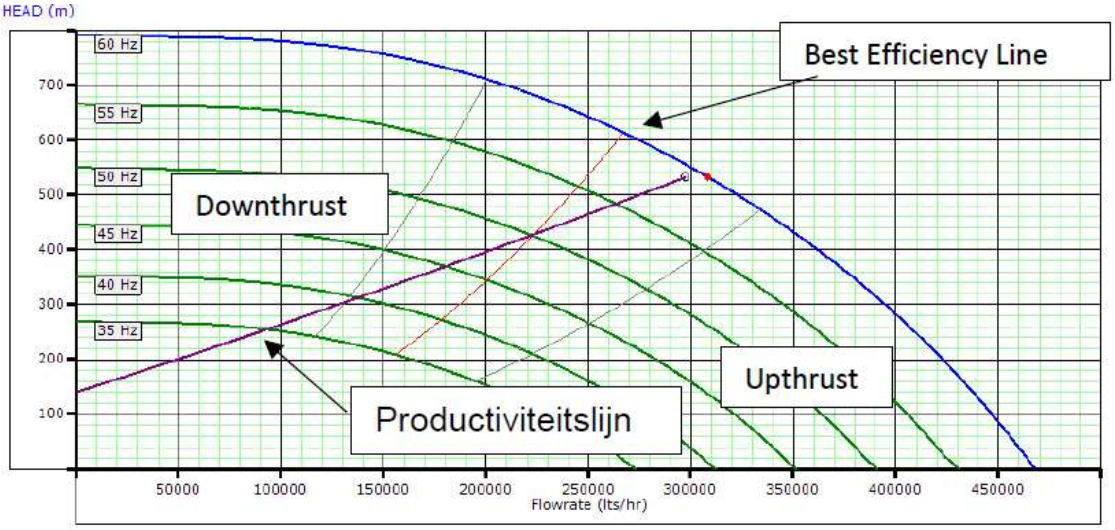


Figure 19. Pump normal curve 12 stages ESP

4.1.3 Well 3

Operator provided 4 years of ESP data from 1-10-2016 to 1-1-2021. There is 1 ESP failure on December 2019 (Run life 3 years). The ESP is a 20-stage ESP (Figure 21) and it is located 650 m below the surface. Despite long period of data, several sensors are missing (Figure 20). Firstly there is no reliable vibration measurement data before October 2020, thus no data-driven modeling for this indicator could be performed. Secondly there is no wellhead temperature sensor, data pre-processing from temperature sensors at 5 heat exchangers locations needed to be performed to use the data.

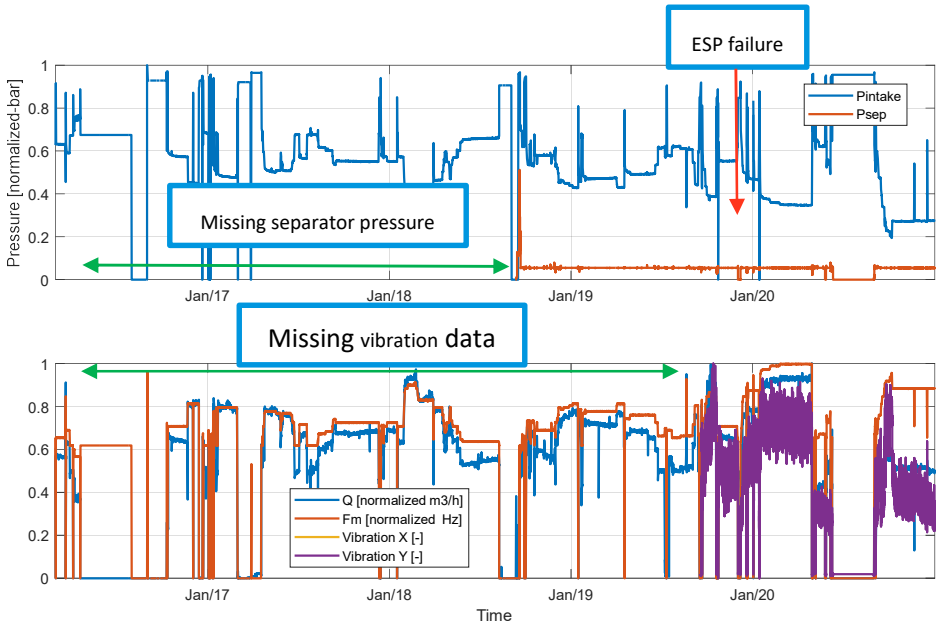


Figure 20. Normalized time series sensor data of well 3

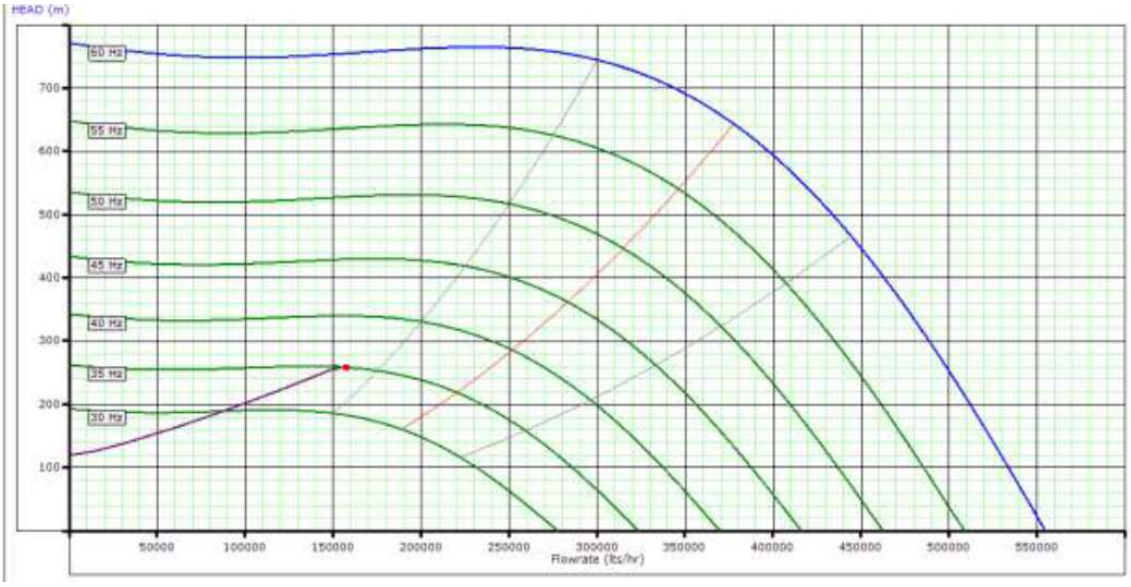


Figure 21. Pump normal curve 20 stages ESP

4.1.4 Available sensors

Operators provided for wells 1, 2 and 3 time-series data with interval of 1 hour and ESP related data with interval of 15 minutes. Overview of sensor measurements from each well are presented in Table 2. Unfortunately, not all of sensors have reliable data. Some sensors had missing values due to loss of communication with DCS server database or due to a sensor failure.

Table 2. Overview sensor availability for each well

Sensor	Well 1	Well 2	Well 3
Flowrate	available	available	available
Frequency	available	available	available
Intake pressure	available	partly	available
Discharge pressure	missing	missing	missing
Wellhead pressure	available	available	partly
Separator pressure	available	available	available
Intake temperature	available	partly	available
Wellhead temperature	available	partly	pre-processing
Motor temperature	available	partly	available
Voltage	available	available	available
Current	available	available	available
Vibration X	available	partly	partly
Vibration Z	available	partly	partly

4.2 Data pre-processing

As mentioned in the previous section, not all data is available from the sensors. And also some data has a bad reading (outlier). To avoid corrupted output results, pre-processing of input was required. Several steps of pre-processing are therefore employed.

4.2.1 Removing values outside operating range

The common error from the sensor data is sometimes the data lies outside operating conditions, for example flowrate measurement with a negative value or much higher flowrate compared to the average value. Removing data outliers was solved by clipping the sensor value with a certain range (Figure 22).

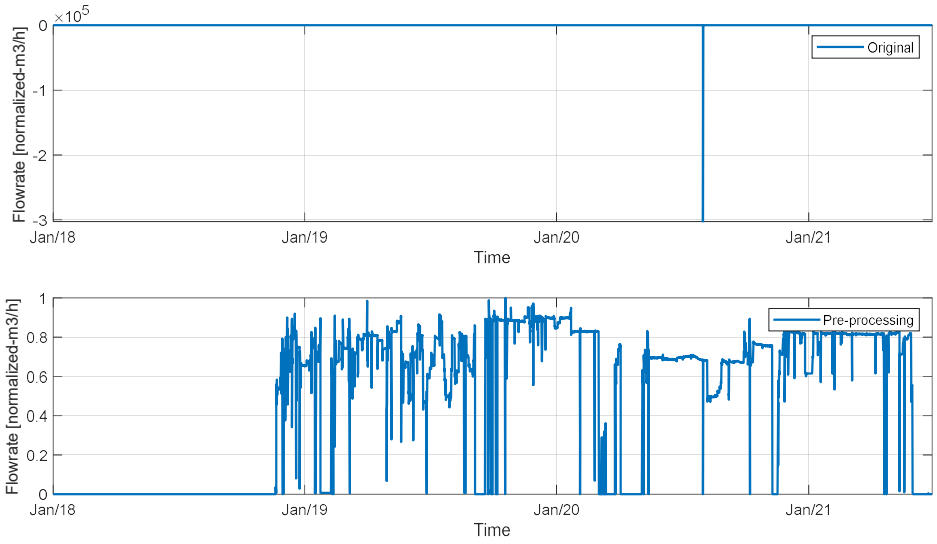


Figure 22. Pre-processing sensor data outside operating range. Top figure is original sensor data and bottom figure is the pre-processed data

4.2.2 Replacing missing values

Some sensors have data missing for several hours or not available due to communication issue. For the first problem, a gap filter was implemented that replaced sensor values that are missing for several hours with the last detected value. Replacing missing sensor data was limited to 1 day.

When there is a systematic missing value from sensor, an expert judgment to replace this with a corrected value is the only way forward. Not all missing sensor value can be corrected. In Well 3 case, the missing sensor value that can be replaced is the topside separator pressure. Since the pressure value is almost constant (Figure 23).

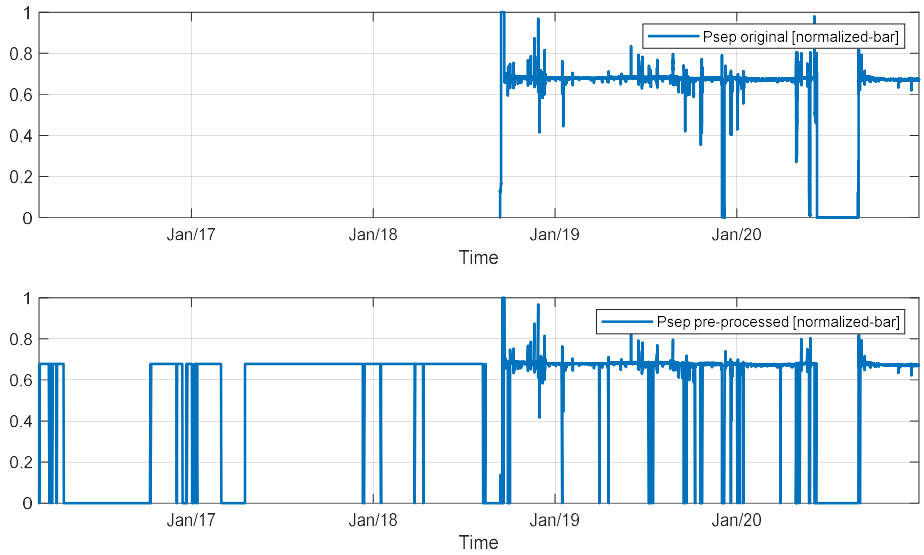


Figure 23. Pre-processing separator pressure with average value during flowing condition

4.2.3 Averaging noisy sensor data

For a noisy sensor data, for example temperature measurement, a moving average filter with window of 24 hours (Figure 22) was implemented.

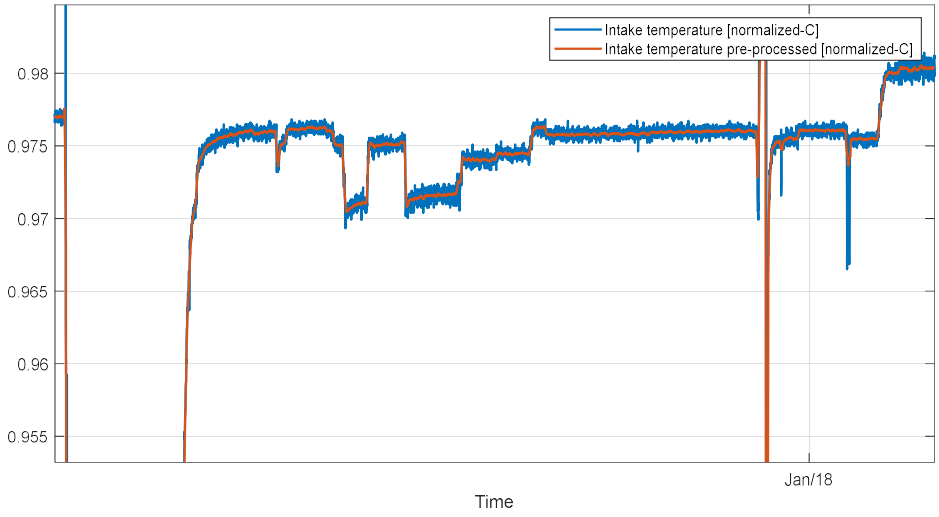


Figure 24. Pre-processing noisy sensor data using moving average value.

4.2.4 Deriving quantities from other sensor values

For well 3 case, there is no temperature measurement at the wellhead. There are only temperature sensor data from heat exchanger. However, there are multiple heat exchangers in parallel depends on the flowrate passing by. Since these are the closest temperature sensors to the wellhead, the maximum value from all these heat exchangers (Figure 25) was taken.

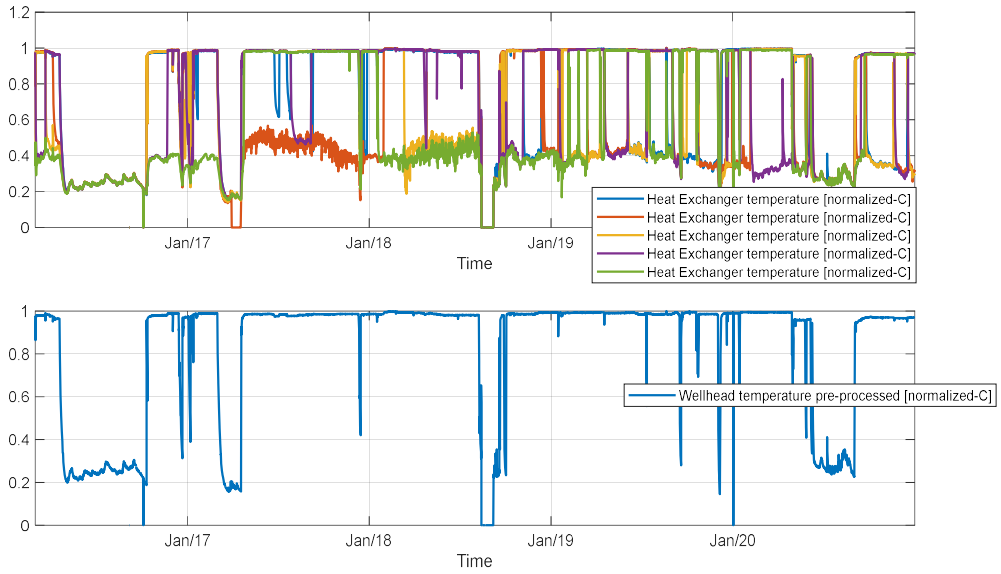


Figure 25. Pre-processing wellhead temperature from heat exchanger sensor data.

4.2.5 Calculate additional sensor values from model

In order to monitor ESP performance, discharge pressure and ESP power are needed. However, in the geothermal ESPs considered in this study there are no sensors at the adequate location to calculate the discharge pressure. Therefore, the VLP approach was used in order to calculate the discharge pressure from the wellhead pressure to the depth of the ESP (Figure 26). For ESP power, since there is no direct measurement, ESP power was obtained by multiplication of the voltage and the "current" sensor data. The computed ESP power was compared with the theoretical pump power curve.

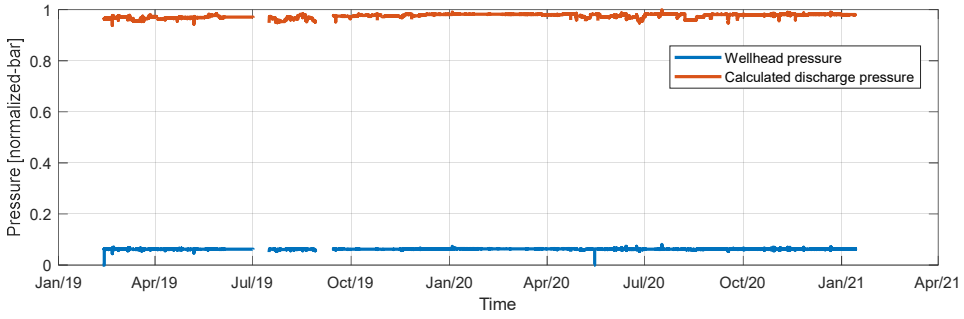


Figure 26. Pre-processing additional data using a model, for example discharge pressure using VLP from topside pressure.

5 Results

In this section the monitoring results using physics-based, data-driven and deep learning approaches are discussed separately for each of the 3 wells. The goal is to show the key components of a common workflow to be used for early detection of ESP failure. Furthermore, it is discussed how to deal with limited amount of data, e.g., how to monitor the ESP in the early phase of run life, and how to deploy the proposed workflow.

5.1 Well 1

As discussed in Section 4.1.1, the ESP run life is almost 2 years. We assume that the 1st year of operations (covering operating conditions during both the winter and the summer period) as the reference period where there is no degradation. Thus, we use this 1st year period for calibration (history matching) of our model both for physics-based model and data-driven model.

5.1.1 Pressure indicators from physics-based modelling

As explained in section 3.2, the model will monitor the downhole pressure and the pressure at the ESP location.

5.1.1.1 Downhole pressure monitoring

ESP Data from 1-1-2019 and 1-1-2020 was used to calibrate the Productivity Index of IPR. As can be seen in Figure 27, the relative error between Pwf from IPR and Pwf from VLP is around 0% with mean relative error -0.165% and standard deviation 1.33%

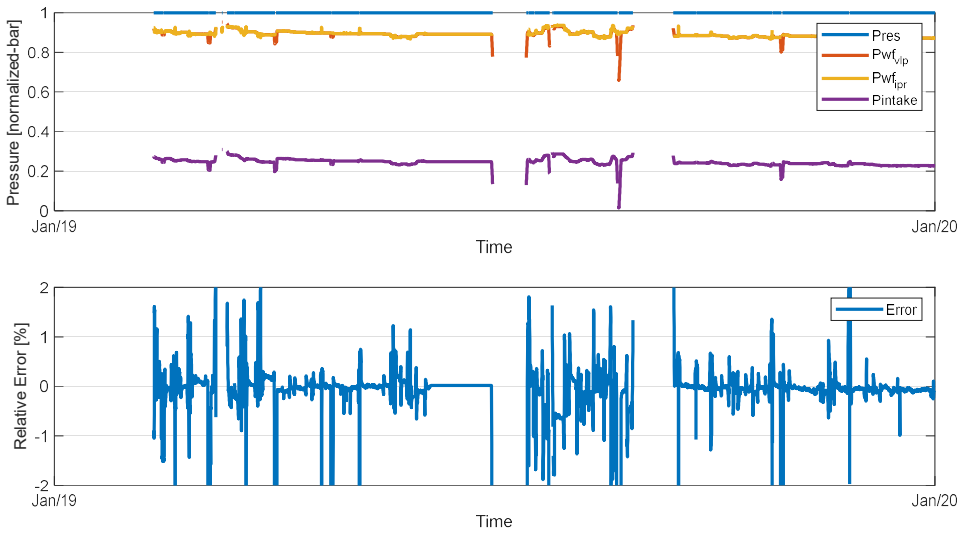


Figure 27. Calibration IPR and VLP model of well 1 from 1-1-2019 to 1-1-2020

After calibration, the model continued monitoring the downhole pressure until the ESP fails from 1-1-2019 to 1-2-2021 (Figure 28). From the picture, one can deduce that the relative error between two pressures is increasing starting 3 months before failure (October 2020). The flowing bottomhole pressure measure is lower than the calculated one, it means that there is an additional resistance

from the reservoir to near wellbore area. The mean relative error is increasing to -0.212% for the last 3 months.

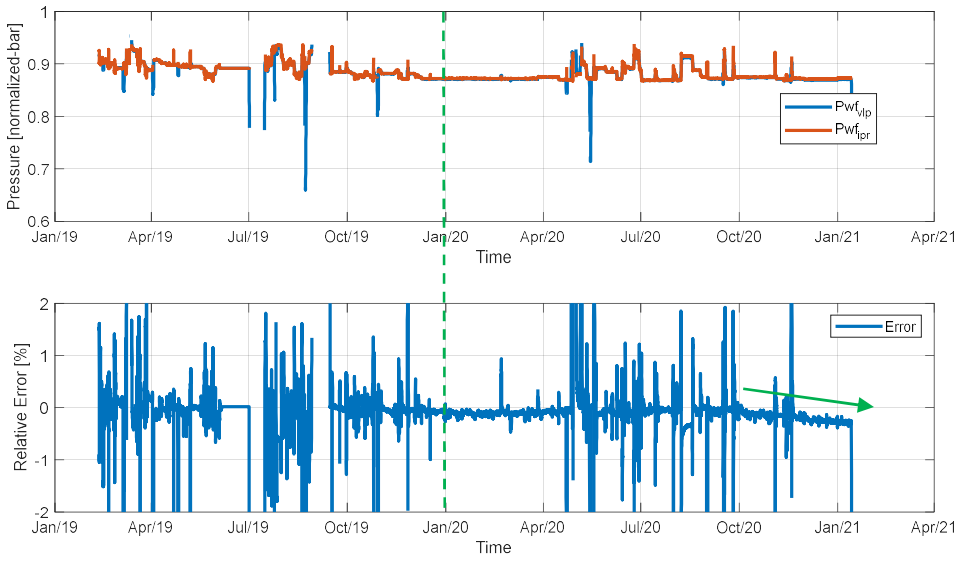


Figure 28. Flowing bottomhole pressure of well 1 for entire period. Left side dash green line is where the model is calibrated, the right side is where the data is being monitored

5.1.1.2 ESP pressure monitoring

Similar with downhole monitoring, the well friction factor model was calibrated with data from 1-1-2019 to 1-1-2020 (Figure 29). The relative error measured pump head compared to calculated theoretical pump head is -1.66% with standard deviation 12.47%

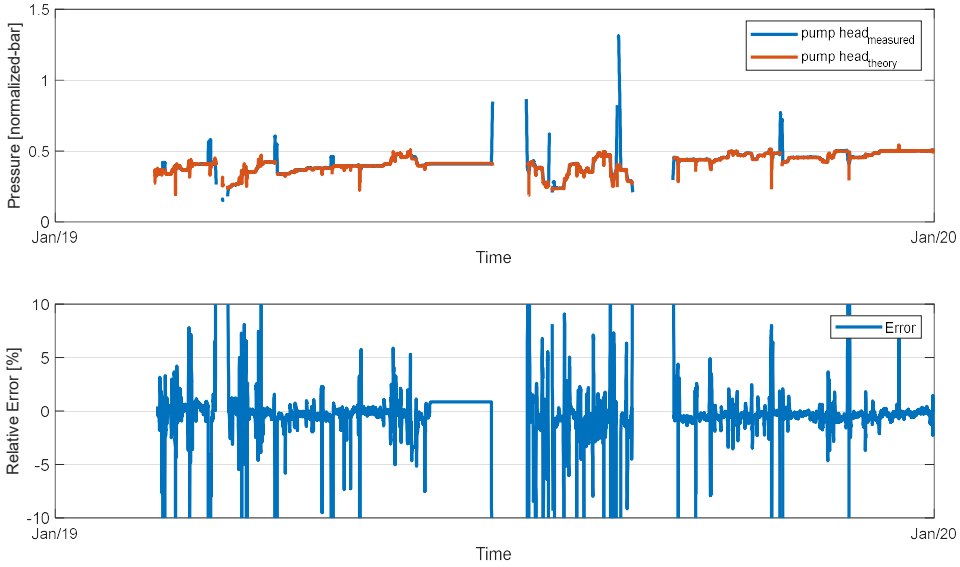


Figure 29. Calibration VLP model of well 1 to calculate discharge pressure from 1-1-2019 to 1-1-2020

Then, the monitoring model continued for the rest of the data. Similar with downhole monitoring, on October 2020 or 3 months before ESP failure, we see an increasing relative error of ESP pump head (Figure 30). The mean relative error is increasing to 1.31%.

Measured discharge pressure has a higher value compared to discharge pressure calculated using pump curve. Thus, in the last 3 months period before failure, the measured pump has lower pressure compared to theoretical value.

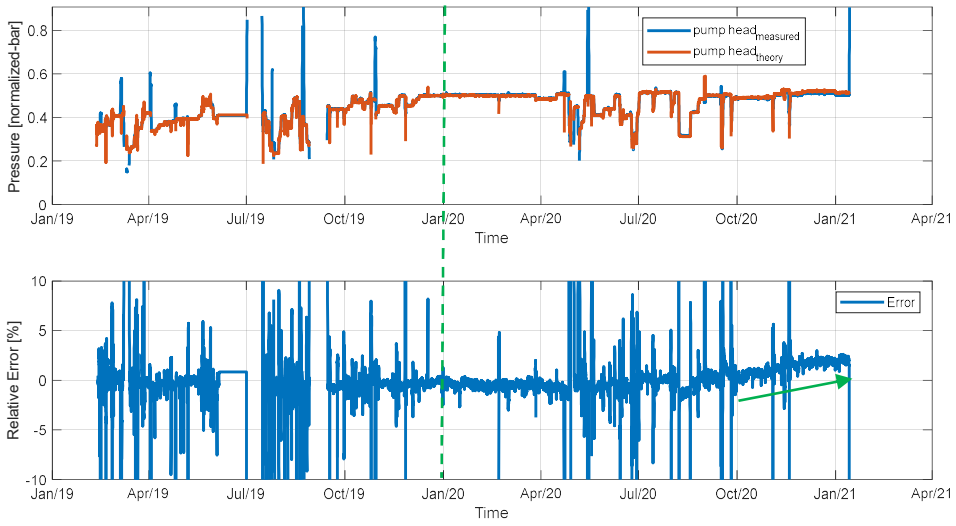


Figure 30. ESP pump head of well 1 for entire period. Left side dash green line is where the model is calibrated, the right side is where the data is being monitored

5.1.2 Data-driven ESP indicators

Data driven monitoring is used for the indicators which are difficult to model using the physics-based approach. Such indicators are ESP vibration sensor, motor temperature and electrical sensors. There are two data-driven approaches that are used here: Random Forest Regressor and Neural Network.

5.1.2.1 Random Forest Regressor

Random Forest Regressor was used to predict the outputs. Typical number of trees that was used is 100 with the maximum depth of tree is 6. Similar with physics-based approach, the period between 1-1-2019 and 1-1-2020 was used to train the data-driven model and use the model to monitor the rest of the data.

ESP Vibration

In Figure 31, the vibration sensor of the ESP is monitored. The inputs of the model are Flowrate, Intake pressure, wellhead pressure, ESP power, ESP frequency, intake temperature, wellhead temperature and motor temperature. The model predicts vibration well in the first 4 months, then it sees a systematic increasing relative error up to 15% started 6 months (June 2020) before failure.

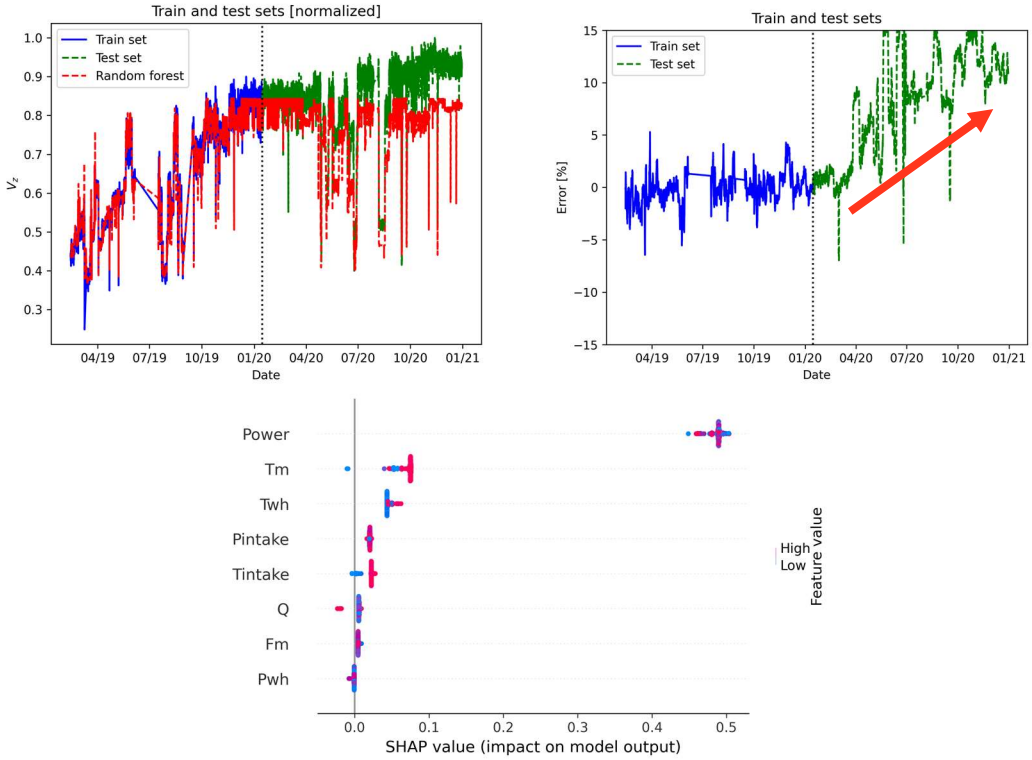
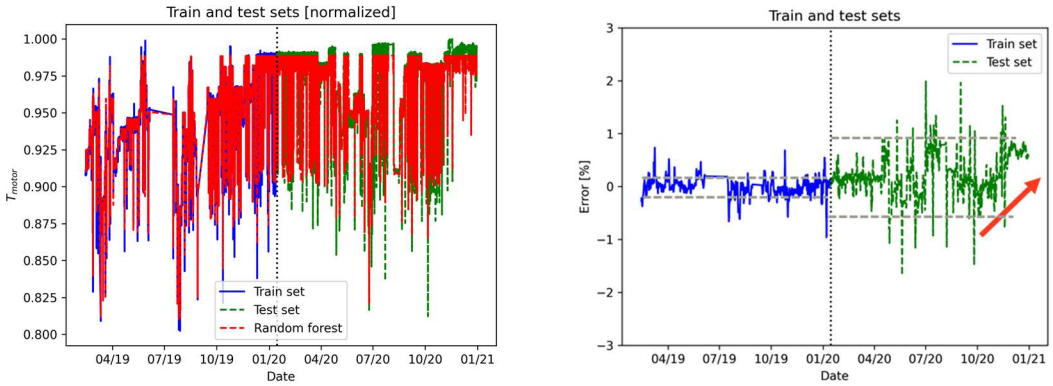


Figure 31. Random Forest Regressor for predicting ESP vibration of well 1

SHAP value was used to measure the input features' contribution to individual predictions. SHAP stands for "SHapley Additive exPlanations." Shapley values are a widely used approach from cooperative game theory, described in Section 3.5. For vibration prediction, ESP Power is the main contributor, followed by motor temperature, wellhead temperature, intake pressure and intake temperature.

ESP Motor Temperature

The monitoring of the ESP motor temperature was set up in an identical way as the ESP vibration. The prediction model has a low relative error up to the first 4 months in 2020 and then shows a bigger error fluctuation starting June 2020. The measured motor temperature keeps increasing in last 3 months before failure, but the ESP didn't fail yet in August when the mismatch is in the similar level.



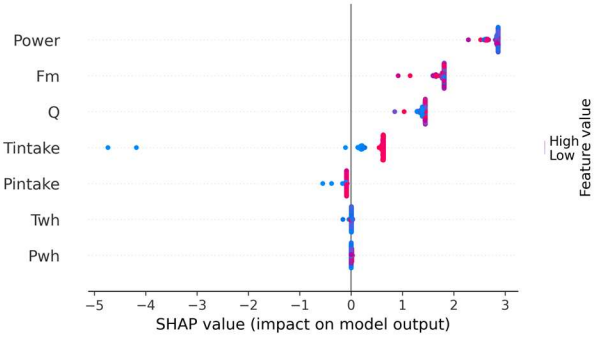


Figure 32. Random Forest Regressor for predicting ESP motor temperature of well 1

The top 3 main contributor input parameters for motor temperature are ESP power, frequency, and flowrate.

ESP Power

With a similar setup as before, the ESP power was monitored. Related trends were found that having bigger error fluctuation starting on June 2020 and an increased power mismatch in the last 3 months.

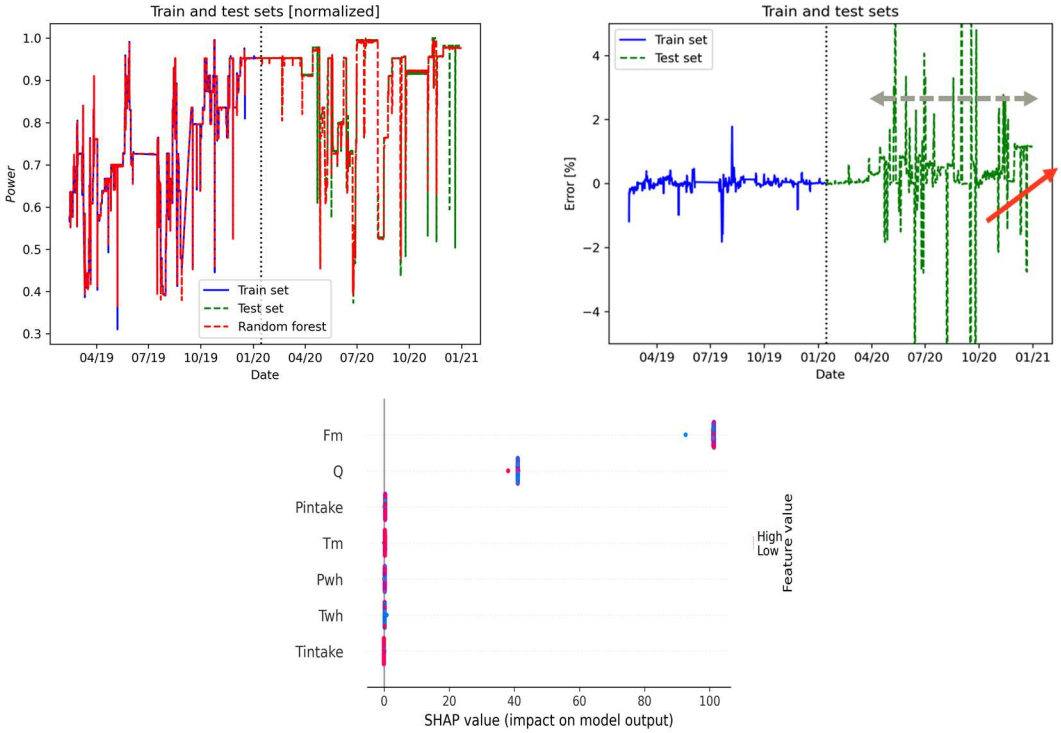


Figure 33. Random Forest Regressor for predicting ESP power of well 1

Frequency and flowrate are the 2 main contributors for predicting the ESP power.

5.1.2.2 Neural Network

To corroborate the analysis performed with the random forest regressor (RFR), a monitoring approach based on a neural network (NN) was used, as described in Section 3.3. In Figure 34(a), the trend for predicting the vibrations along the z-direction (Vz) with the associated absolute error can

be seen. Identical training/testing intervals as in the previous section were applied. Different combinations of input for the NN model, ranging from only 4 features (ESP current, voltage, motor temperature, and vibrations in x), shown in Figure 34(a), till to the 8 features used for the RFR analysis (flowrate, intake pressure, wellhead pressure, ESP power, ESP frequency, intake temperature, wellhead temperature and motor temperature). In all cases, the trend of Vz is very well captured in the training period, and the mismatch between model prediction and measured Vz is evident in the testing period. The qualitative increasing trend in the error and the quantitative mismatch (up to 15-20%) are consistent with the RFR analysis. Identical analysis was performed for the motor temperature (Tm), as shown in Figure 34(b). The nine quantities used as model input in this case are ESP power, frequency, vibrations in x and z, flowrate, intake pressure and temperature, separator pressure. We observe that the model can predict with sufficient accuracy Tm in a large portion of the training period, where the absolute error averages around zero with small fluctuations (on the order of couple percent). In the testing period, the mismatch between model and field data exhibits significantly larger fluctuations. Such an increase for the testing period is consistent with the RFR analysis, suggesting that larger mismatches are due to some intrinsic to the data and not dependent on the algorithm. However, differently from the RFR analysis, the error trend towards the end of ESP run life appears to be slightly downward for the NN case.

The analysis was repeated for some other combinations of input features, where some of the variables above were not included. Satisfactory and consistent results in several cases were found, except when the intake temperature was not part of the input set. In that case the model was not able to properly capture the finer variations in the motor temperature trend. Intake temperature is an important variable when predicting motor temperature. This in line with the observation from the SHAP analysis on the RFR (Figure 32) in with Tintake is ranked quite high in the feature importance.

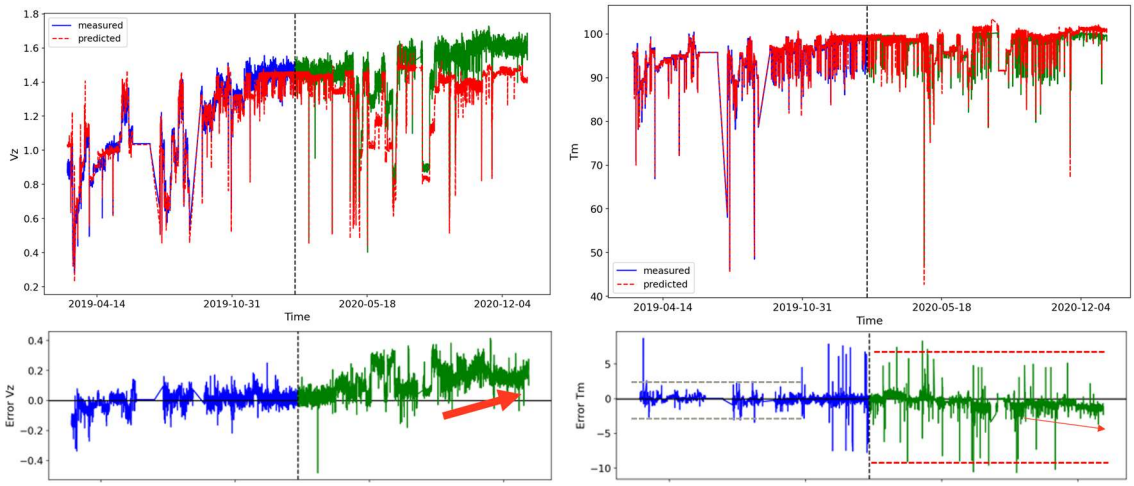


Figure 34. Monitoring ESP indicators using a neural network. Prediction for the vibrations along z-direction Vz (left) and motor temperature Tm (right). The vertical dashed line indicates the split between the training and testing period. Associated errors are also shown on the bottom. Trends are consistent with the analysis performed with the random forest regressor.

Both monitoring approached based on the mismatch between measured data and model predictions presented for the ESP of Well 1 are robust to the choice of algorithm, being RFR or NN, especially for the vibrations.

5.1.3 System level monitoring with deep learning

An analysis on the entire system using a deep learning approach was performed based on the autoencoder methodology described in Section 3.4. The dataset was split in training and testing period using the same intervals of the previous sections. In Figure 35 the total reconstruction error of the trained autoencoder model as a function of time is reported. For this model, only 5 features are considered: ESP current (Im), voltage (Vm), vibrations along the x-direction (Vx) and along the z-direction (Vz), and motor temperature (Tm). From the graph, it can be clearly observed that the reconstruction error is higher towards the end of the ESP run life, indicating an overall mismatch from the model prediction (that has been trained on the healthy data) and the sensor values. Overall, the system is therefore behaving anomalously compared to the supposedly healthy response expected by the model. A SHAP analysis based on the entire testing period (green interval in Figure 35, on the right of the vertical dashed line) ranked the 5 input features with the ESP current (Im) being the most important. This indicates that a cause of the overall large reconstruction error could be related to some anomaly associated to the ESP current.

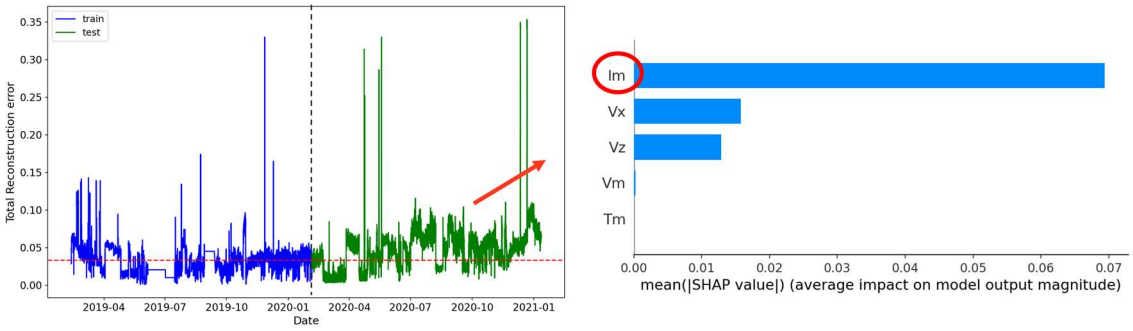


Figure 35. Left: total reconstruction error as a function of time. The vertical dashed line indicates the split between the training and testing period, whereas the horizontal line is a guide-to-the-eye (indicating the average of the reconstruction error during the training period). Right: summary of the SHAP analysis for the entire testing period. The ranking indicates that the ESP current (Im) is the most important model parameter.

It is also possible to examine in more detail the trend of the reconstruction error during the testing interval, for example identifying periods with similar signatures and investigate the cause of e.g., an increase of the reconstruction error. This would allow to gain some insights in the possible anomaly, that could be further combined with the other data-driven and physics-based models to coherently assess if a true malfunctioning is happening. In Figure 36 we show the SHAP analysis for two intervals in which the reconstruction error increased and stayed large for some time. For both cases, the same trained autoencoder was utilized, but the SHAP analysis was performed only for the regions indicated in the figure. In the first case, the most important model parameter is identified in the vibrations Vz, whereas the ESP current Im is the most relevant for the period just before the ESP failure. This could suggest that something associated to the current is indeed the ultimate cause of ESP failure. In contrast, the 'anomalous' behavior detected earlier (where Vz is the main cause) might be due to for example a change in operation conditions that however is not responsible of the final failure.

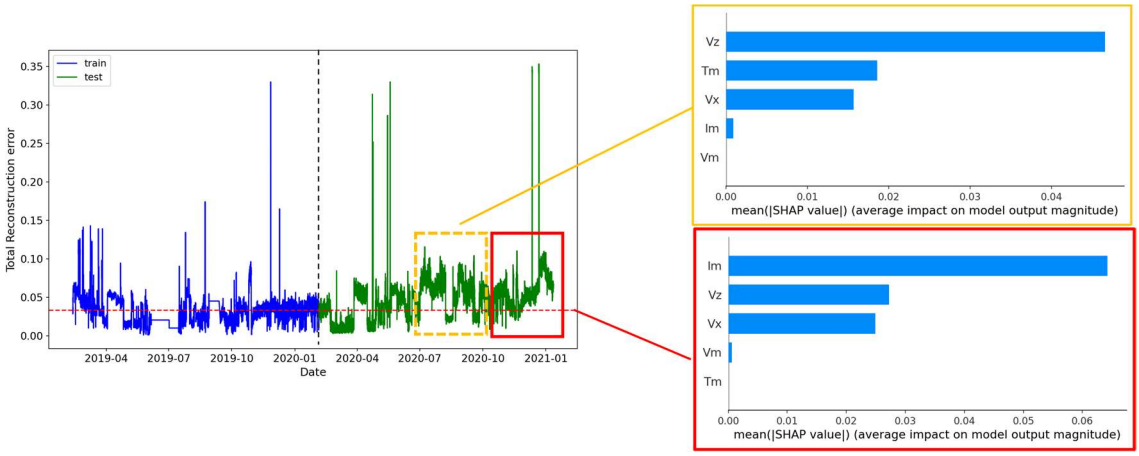


Figure 36. The SHAP analysis performed on two different periods indicates different ranking among the model input. From July to October 2020, the vibrations in z (Vz) resulted the most important parameter causing the large reconstruction error in that period. However, in the last months of the ESP run life, the ESP current (Im) is the most important indicator, suggesting problems related to it as possible cause of ESP failure.

5.1.4 Summary of observations

Using the 3 approaches for monitoring the ESP of Well 1, the following can be observed:

- Using 1 year of data, that covered the variability of the flowrate (geothermal summer and winter operating conditions), was sufficient for calibrating the physics-based model and for training the data-driven models
- In all cases, the mismatch between the model and monitored quantity is visible at least 3 months before the ESP failure with an increasing trend before failure.
- In the last 3 months, the predicted intake pressure and discharge pressure are lower than the measured ones, pointing to an increasing resistance in downhole (decrease inflow).
- This trend is also seen when monitoring motor temperature and ESP power where the measured value is higher than the predicted value.
- The mismatch between model and measurement of the vibration data is seen already 6 months before the failure, and it is captured equally by the two monitoring approaches (RFR and NN).
- The reconstruction error from the autoencoder also increased significantly in the last 3 months, with the ESP current being the most relevant cause of such a large mismatch between model and measurements.

5.2 Well 2

The dataset associated to well 2 contains several ESP failures. As explained in Section 4.1.2, depending on the period considered, several sensors can contain many missing values. The largest data availability is associated to the 1st ESP. However, this ESP fails within less than a year. Having such limited set of data could present a challenge for the data-driven approaches.

5.2.1 Pressure indicators from physics-based modelling

An identical exercise to well 1 was performed for well 2. Both pressure at the downhole conditions and ESP pump head pressure will be monitored.

5.2.1.1 Downhole pressure monitoring

Since the run life is 7 months, the IPR and VLP are calibrated with a shorter data period around 10 weeks from 1-2-2019 to 14-4-2019. Only with several flowrate variation, the model can predict flowing bottomhole pressure well with mean relative error -0.144% and standard deviation 1.37% (Figure 37)

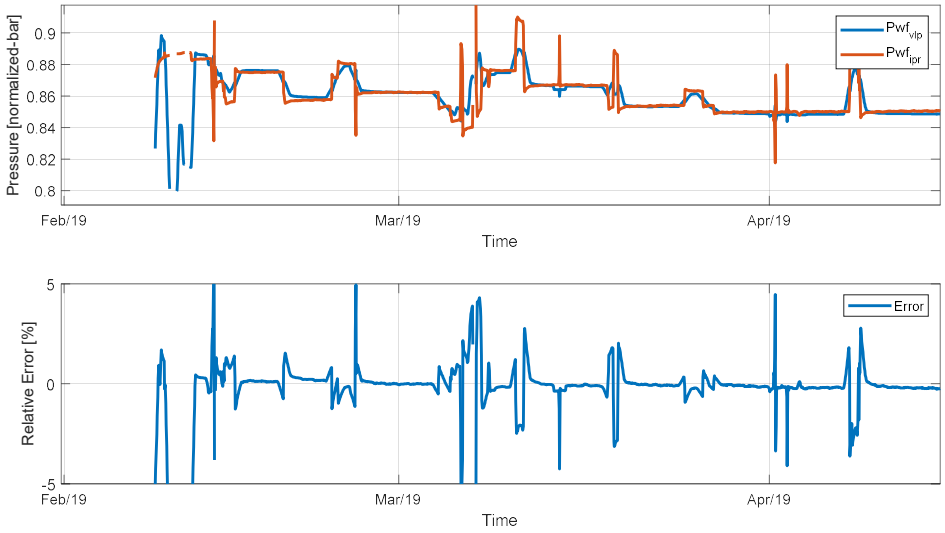


Figure 37. Calibration IPR and VLP model of well 2 from 1-2-2019 to 14-4-2019

Then, the model monitored downhole pressure and ESP pressure for the rest of the data period. No increasing error pattern several weeks before ESP failure like well 1 can be observed. However, the model captures a relative error mismatch around 2% started beginning of May up to mid of July (green box in Figure 38), indicating an increase in inflow performance that was recovered again in August. In the beginning, it was believed to be due to a mismatch in the extrapolated flowrate range, because the flowrate when the mismatch started is higher and not in calibration data set (yellow box). However, in August, when the error mismatch is recovered, the flowrate range during this month is similar with the flowrate during June-July. Thus, the extrapolation error assumption is not correct. There is a changing system dynamic downhole.

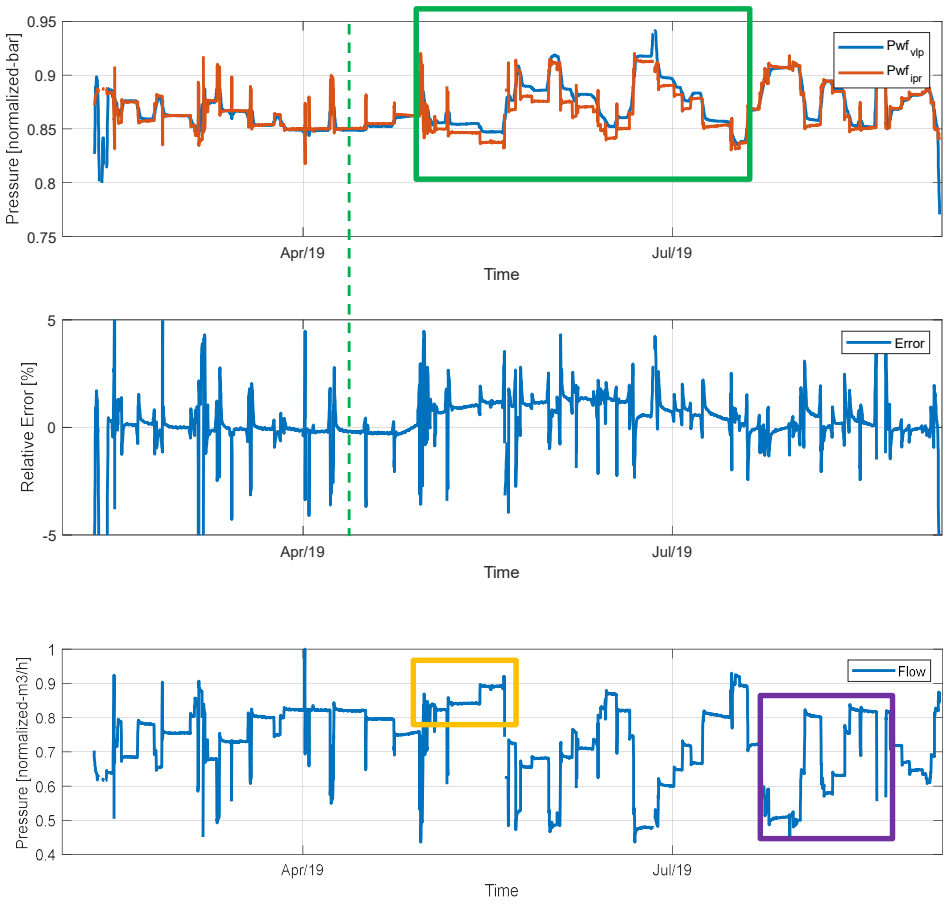


Figure 38. Flowing bottomhole pressure of well 2 for entire period. Left side dash green line is where the model is calibrated, the right side is where the data is being monitored

5.2.1.2 ESP pressure monitoring

Similar with the previous section, it was verified that the measured pump head using the discharge pressure calculated from the tank pressure matches with the analytical pump head pressure from the pump curve during first 10 weeks of operations. The mean relative error is around 0.356% with standard deviation is 3.48%, However the relative error is a higher during start-up phase due to flowrate fluctuation (Figure 39).

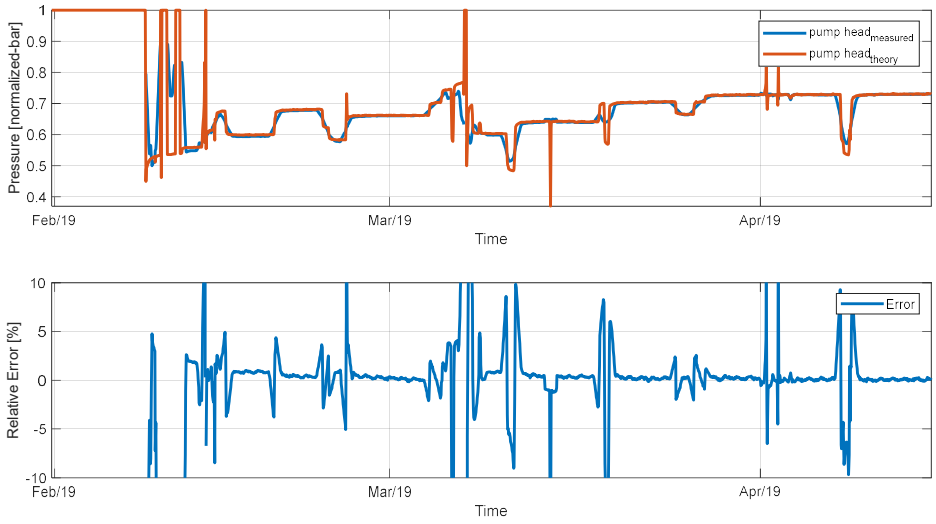


Figure 39. Calibration VLP model of well 2 to calculate discharge pressure from 1-2-2019 to 14-4-2019

The model is used to monitor ESP pump head until November 2019. Around end of May, the measured pump head is having a lower discharge pressure (or lower pump head) until the rest of dataset. This mismatch is delayed by 3 weeks compared to the mismatch happening in the downhole.

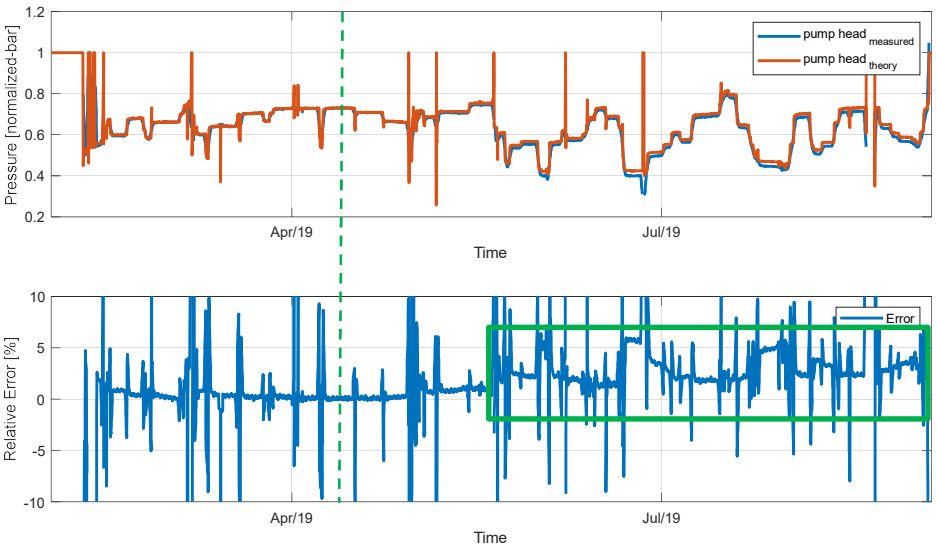


Figure 40. ESP pump head of well 2 for entire period. Left side dash green line is where the model is calibrated, the right side is where the data is being monitored

5.2.2 Data-driven ESP indicators

5.2.2.1 Random Forest Regressor

Similar approach to well 1 was used to monitor ESP power, vibration and motor temperature using random forest regressor. However since the data is limited due to short ESP run life, variations of the training period for the model were explored by using 3 months and 5 months of data, as presented in Figure 31, Figure 32, and Figure 33. Unfortunately for none of the indicators a good model for

monitoring could be obtained because the prediction error is suddenly bigger when using a new data. In fact, this can happen due to an extrapolation error, that is when the model 'overfits' the training dataset. Increasing the training period is also not necessarily a good approach because it could be that the ESP degradation before failure is also included when training model. Thus, no increase in the mismatch is observed for the motor temperature. Other observations are an increasing error trends for vibration and power, but it is not clear if it is due to model overfitting or extrapolation error since the error is directly increasing during monitoring period. The conclusion is that the input data should contain more variations to cover bigger ESP operating range (for example summer and winter) and reliably train the model.

ESP Motor Temperature

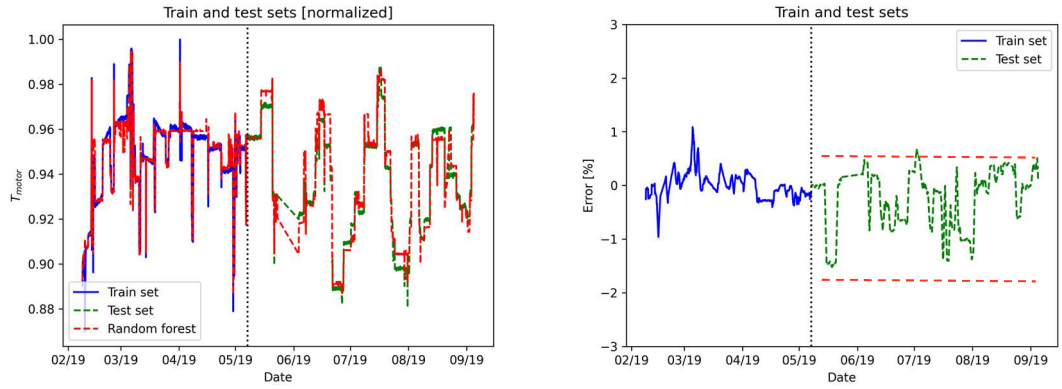


Figure 41. Random Forest Regressor for predicting ESP motor temperature of well 2 with 3 months training period

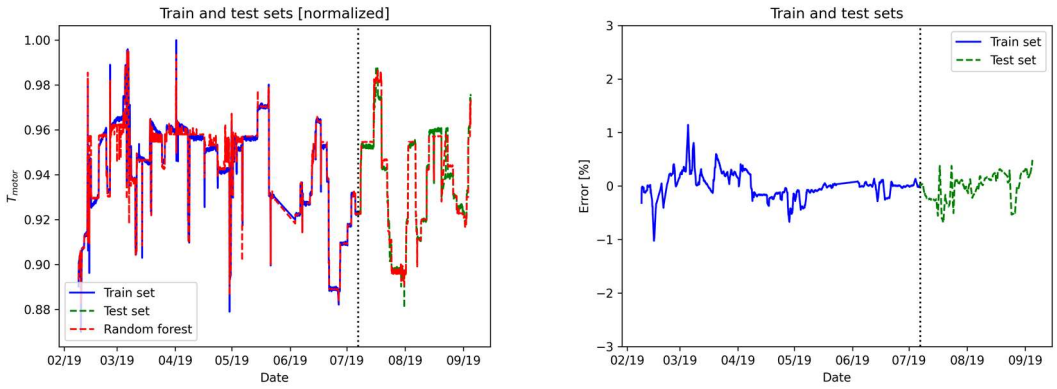


Figure 42. Random Forest Regressor for predicting ESP motor temperature of well 2 with 5 months training period

ESP Vibration

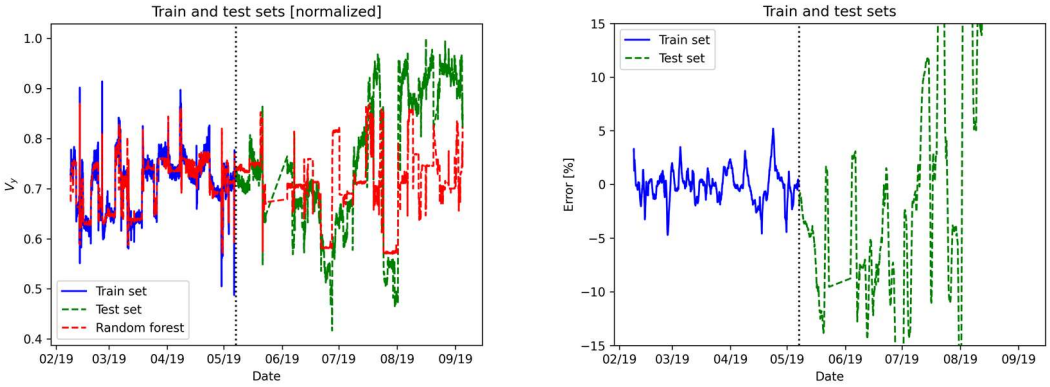


Figure 43. Random Forest Regressor for predicting ESP vibration of well 2 with 3 months training period

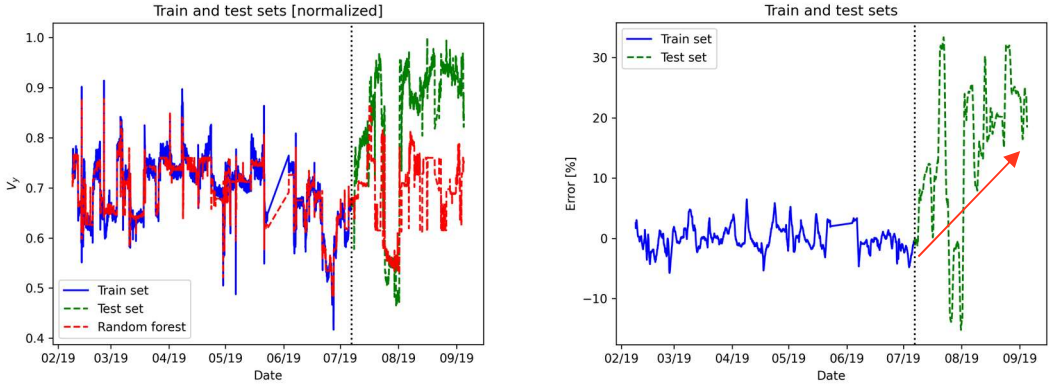


Figure 44. Random Forest Regressor for predicting ESP vibration of well 2 with 5 months training period

ESP Power

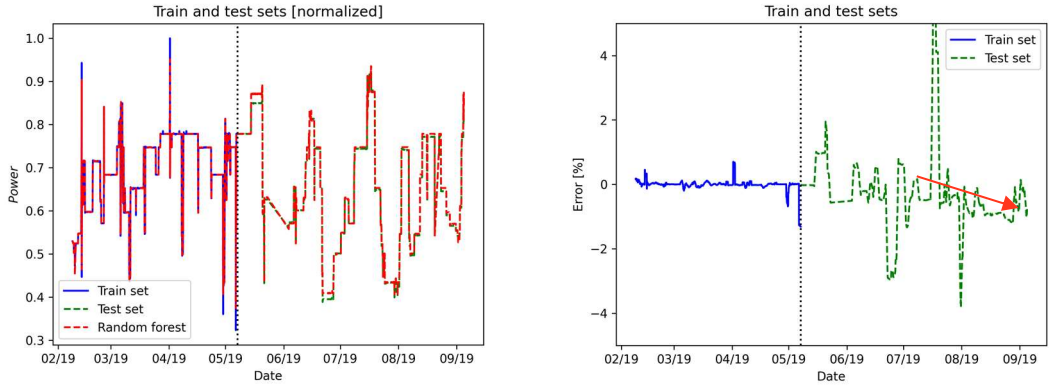


Figure 45. Random Forest Regressor for predicting ESP power of well 2 with 5 months training period

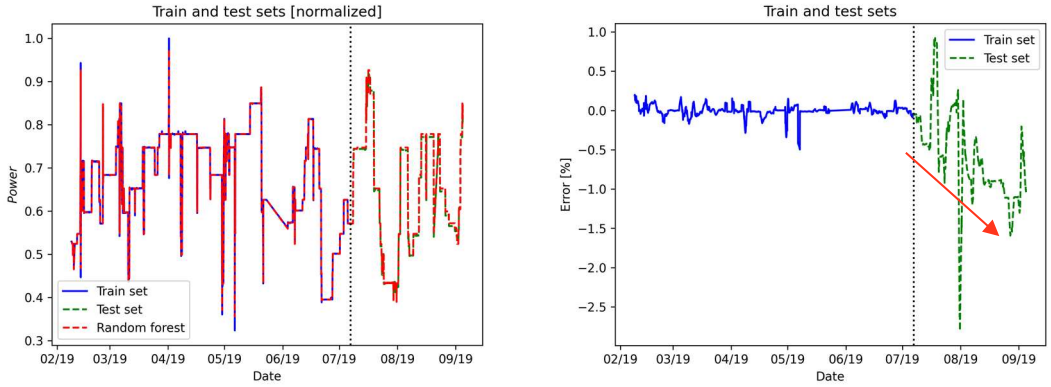


Figure 46. Random Forest Regressor for predicting ESP power of well 2 with 5 months training period

5.2.2.2 Neural Network

The analysis is not shown since no satisfactory results have been achieved for the same reasons as in RFR. The data in the training period are not significant enough, meaning that they do not cover a sufficiently different range of operating conditions to properly fit the model and be confident that the potential mismatch observed in the testing dataset is robust and not merely due to extrapolation errors of the uncalibrated model.

5.2.3 System level monitoring with deep learning

Analogously, we refrain from presenting the analysis based on the (explainable) auto-encoder for monitoring at the system level the well 2. Like for the other data-driven methods, the ESP data for well 2 (in the training period) is not sufficient to reliably train an autoencoder and meaningfully look at the trend in the reconstruction error in the testing period. In fact, we observed large variations in the results when small changes to the hyper-parameters of the model were done despite keeping fixed the training/testing intervals, therefore further suggesting that a suitable model could not be constructed.

5.2.4 Summary of observations

Using the 3 approaches for monitoring the ESP of Well 2, the following can be observed:

- The physics-based model indicates that there is an increased inflow for 3 months during mid operations as the measured intake pressure is higher than the predicted value and is back to normal 2 months before ESP failure.
- The ESP pump head mismatch is visible 1 month after the inflow problem that was found in the downhole condition. This suggests that there might be a delayed effect in relation to the well condition and the ESP performance.
- Due to the limited data, lacking enough variation in the operating conditions, it was difficult to apply the data-driven approaches to this case.

5.3 Well 3

The ESP run life of Well 3 is 3 years and 3 months. This is the longest dataset, compared to the previous 2 wells. The 1st year operational data (1-10-2016 to 1-10-207) will be used as training dataset of our model, then let the model runs until December 2019 when there is ESP failure.

5.3.1 Pressure indicators from physics-based modelling

Similar analysis was performed as for well 1 and well 2. Both downhole pressure and ESP pump head pressure will be monitored.

5.3.1.1 Downhole pressure monitoring

The IPR and VLP was calibrated for the 1st year of operation, the mean relative error is -0.007% with standard deviation of 0.91%. We see that after long shut-ins (green circles in Figure 47) there is some well dynamics for around 1-2 weeks before it disappears. The measured flowing bottomhole pressure from VLP is higher than the predicted value, suggesting a possible increase in the PI or a higher reservoir pressure.

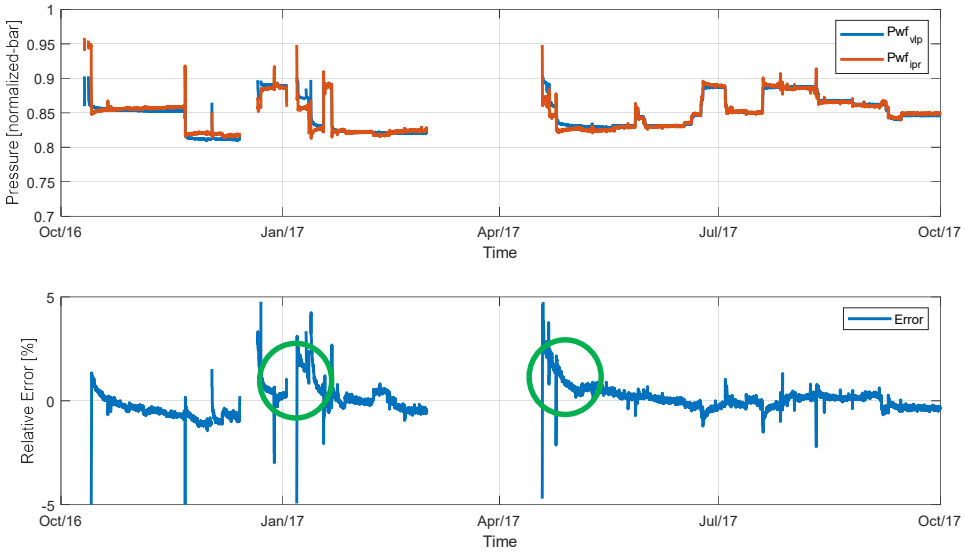


Figure 47. Calibration IPR and VLP model of well 3 from 1-10-2016 to 1-10-2017

The model is then used to monitor the rest of the data. The predicted value is still having a good match until August 2018 before a long shut-in. After the well starts re-producing the flow, the relative error mismatch is started to increase. However, after 1-2 weeks, it does not go back to the desired value. In the calibration set, it can be seen that P_{wf} is higher than predicted value. This is the other way around where P_{wf} calculated from measured intake pressure is lower than the predicted value. It could be that there is additional resistance in the wellbore or tubing (decrease PI) or lower reservoir pressure. This mismatch is happening for a year until October 2019 where the flowrates is increased then there is a well shut-in. After this shut-in the mismatch is disappeared. Thus, lower reservoir pressure is not the case.

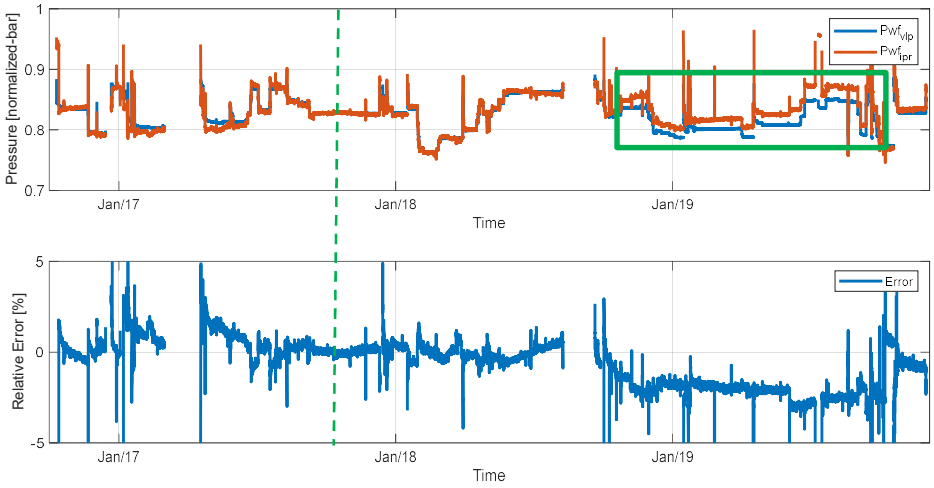


Figure 48. Flowing bottomhole pressure of well 3 for entire period. Left side dash green line is where the model is calibrated, the right side is where the data is being monitored

It can be hypothesized that this additional resistance could be due to a higher density content in the liquid (for example due to sand). For this reason, the brine density was increased from 1089 kg/m^3 with additional sands around 2% to 1100 kg/m^3. The relative error of is improving until June 2019 and then there is more resistance (green square in Figure 49). After a high flowrate on beginning October, the additional resistance is gone. Now, the measured P_{wf} from VLP is higher than the predicted value form IPR. A possibility is therefore that the sand has been flushed away (orange square).

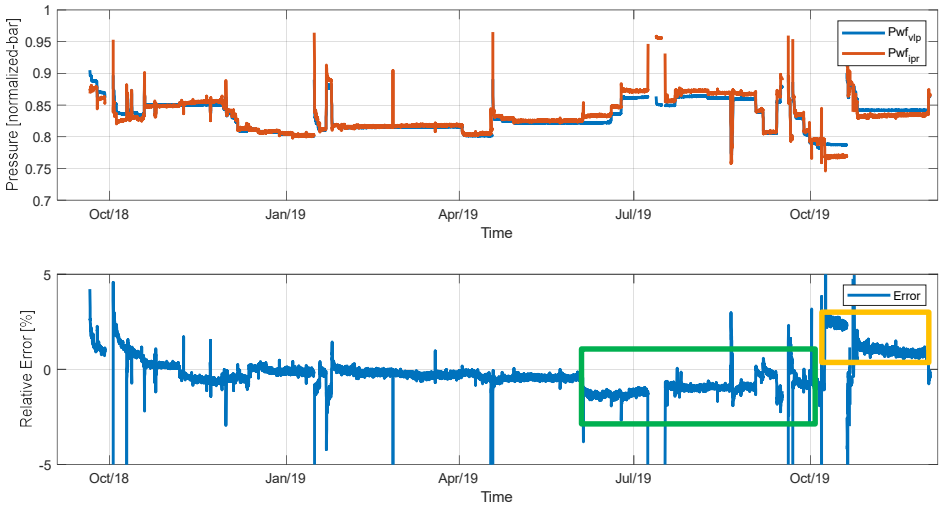


Figure 49 Flowing bottom hole pressure of well 3 with an increased density to 1100 kg/m^3

5.3.1.2 ESP pressure monitoring

The VLP model was calibrated from topside to calculate discharge pressure of well 3 as shown in Figure 50. The mean relative error is 0.052% with standard deviation of 1.41%.

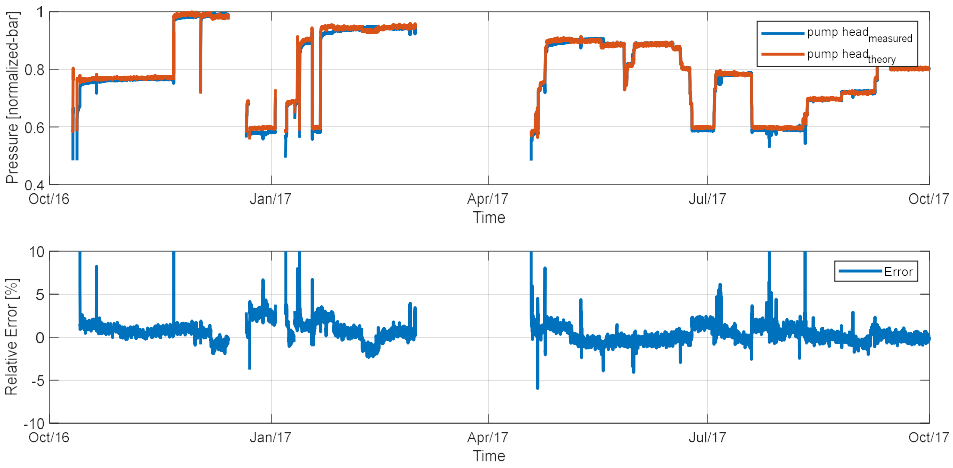


Figure 50. Calibration VLP model of well 3 to calculate discharge pressure from 1-10-2016 to 1-10-2017

The model was used to monitor pump head after green dash line in Figure 51. The model shows a good match for the next 2 years. However, since October 2019 (or 2 months before ESP failure), the measured pump head has a higher value compared to theoretical pump head. The mean relative error is around -5%.

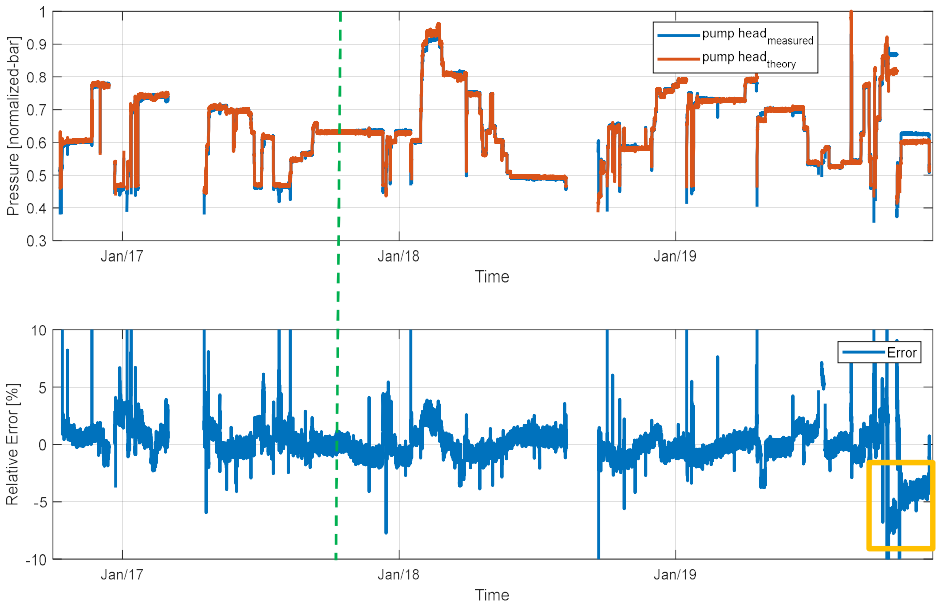


Figure 51. ESP pump head of well 3 for entire period. Left side dash green line is where the model is calibrated, the right side is where the data is being monitored. Yellow square shows an increasing error between measured pump head and theoretical pump head.

5.3.2 Data-driven ESP indicators

5.3.2.1 Random Forest Regressor

Due to unavailability of vibration sensor data, only r motor temperature and power was monitored for data-driven approach.

ESP motor temperature

Similar approach is also implemented for Well 3. However, the training period of the model is increased to 1.5 years to cover worst winter period in 2018. ESP was monitored from the summer until it fails. On December 2018, a model mismatch is observed since the measured motor temperature is lower than the predicted value. Afterwards, the error is again becoming small 2 months before failure. This behavior is also witnessed during the monitoring of the downhole pressure using physics-based model.

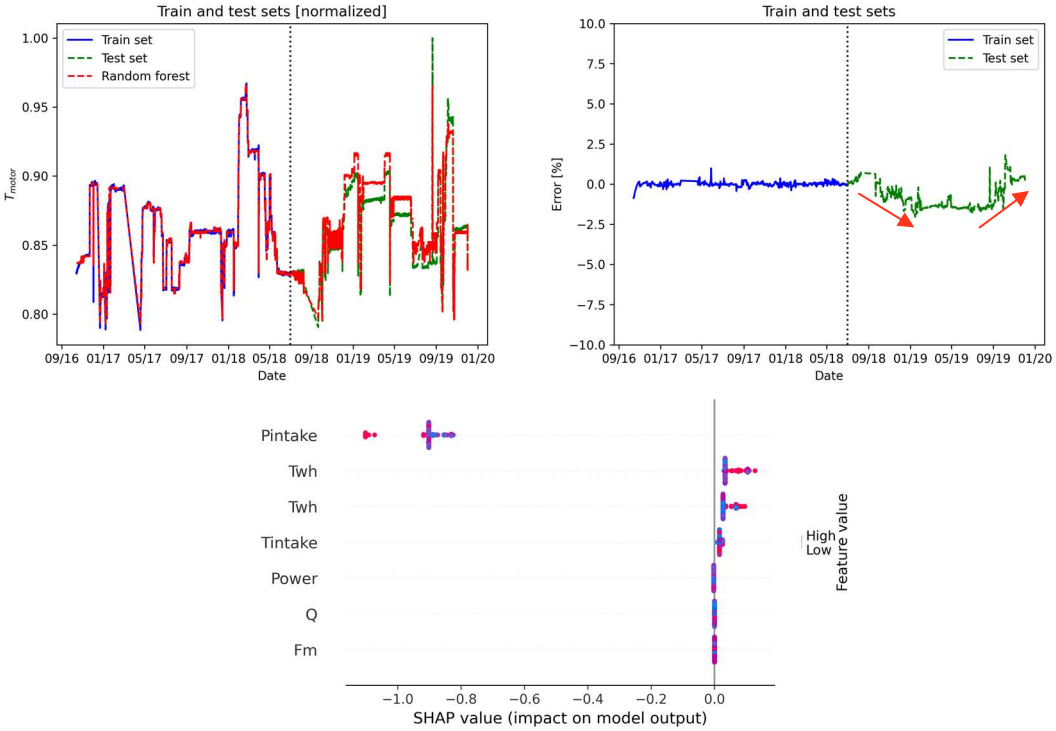
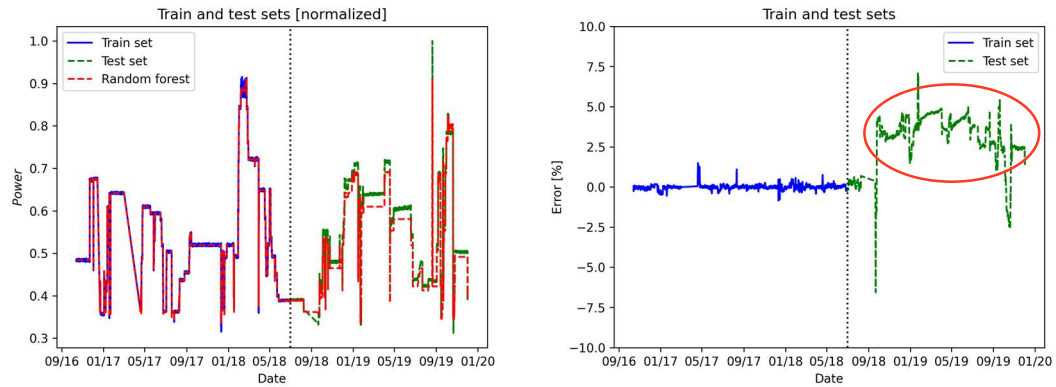


Figure 52. Random Forest Regressor for predicting ESP motor temperature of well 3

ESP Power

Similar trends are observed for ESP power, but the other way around. The predicted power is lower than the measured value. The relative error is around 4% and then the error is decreasing 2-3 months before ESP failure. Similar with Well 1, the main contributor for predicting ESP power is the frequency.



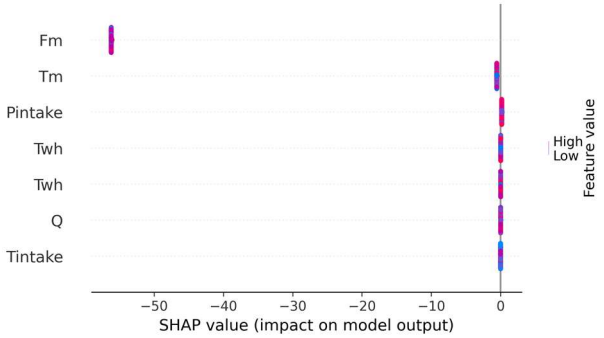


Figure 53. Random Forest Regressor for predicting ESP power of well 3

5.3.2.2 Neural Network

Also for the well 3, the ESP monitoring was performed using the NN approach. The timeseries are split in training and testing period according to the intervals mentioned earlier and indicated in Figure 54 with a vertical dashed line. In the figure, both the trend for the motor temperature (Tm) and for the ESP power (Pow) are shown. For the first case the model inputs are ESP power, frequency, flow rate, intake pressure and temperature, well-head temperature. In the second case, the ESP power (that is the output to be predicted) is replaced by the motor temperature. In both cases, the trends are consistent with the RFR analysis. A very good match for Tm during the training period and the first months of the testing period can be observed. Afterwards, the mismatch appears with the model overestimating the temperature for several months and rematching again towards the end, as observed previously with the RFR. Analogously, for the power a clear mismatch is observed several months before ESP failure with the model underestimating the power, closely following the trend captured with the RFR.

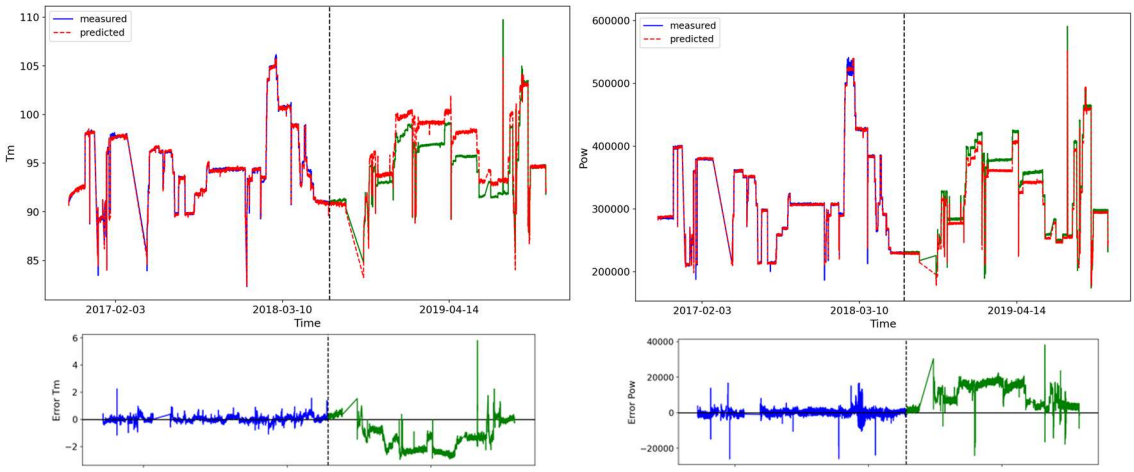


Figure 54. Monitoring ESP indicators using a neural network. Prediction for motor temperature Tm (left) and ESP power (right). Associated errors are also shown on the bottom. Trends are consistent with the analysis performed with the random forest regressor.

The monitoring approach presented for the ESP of Well 3 is robust to the choice of data-driven algorithm, being RFR or NN.

5.3.3 System level monitoring with deep learning

In this section, the use of the autoencoder for monitoring well 3 is described. For this model 7 features are considered: ESP power (Pow), frequency (Fm), flowrate (Q), motor temperature (Tm), intake pressure (Pintake) and temperature (Tintake), and well-head temperature (Twh). The model is trained in the interval until July 2018. In Figure 55 the trend of the reconstruction error as a function of time is reported. Approximately for the last year of ESP run life, the reconstruction error is significantly higher than other periods. In addition, despite steadily decreasing around the end of 2019, the last two months (Nov2019-Jan2020) exhibit a new steady increase of the error. Such a final rise could be an early detection of the anomaly causing the final ESP failure. Careful inspection of the trend also reveals that the reconstruction error is relatively high immediately after shut-in periods, that can be identified in the graph as time interval with missing datapoints. A SHAP analysis on the entire test dataset was performed and observed that the ESP power is ranked as most significant feature.

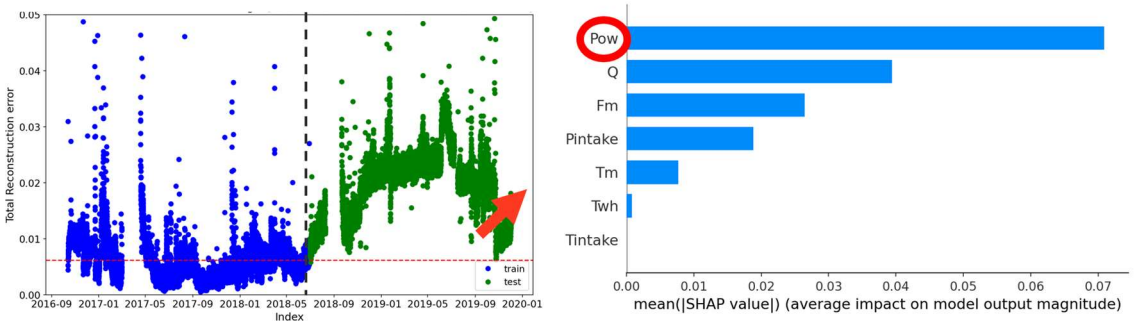


Figure 55. Left: total reconstruction error as a function of time. The vertical dashed line indicates the split between the training and testing period, whereas the horizontal line is a guide-to-the-eye (indicating the average of the reconstruction error during the training period). Right: summary of the SHAP analysis. The ranking indicates that the ESP power (Pow) is the most important model parameter when considering the entire testing period.

Subsequently, we looked into different periods for the SHAP analysis, and we particularly focused on the initial rise of the reconstruction error in the testing period (after July 2018) and the last increase just before ESP failure. The results of these two cases are shown in Figure 56. Interestingly, in the first case the flowrate (Q) is the most important parameter, suggesting that the novel behavior detected by the autoencoder might not be an anomaly but rather a change in operating conditions. On the contrary, the cause for the last increase before failure is mainly due to the ESP power and it can be an indication to support a more detailed root cause failure analysis.

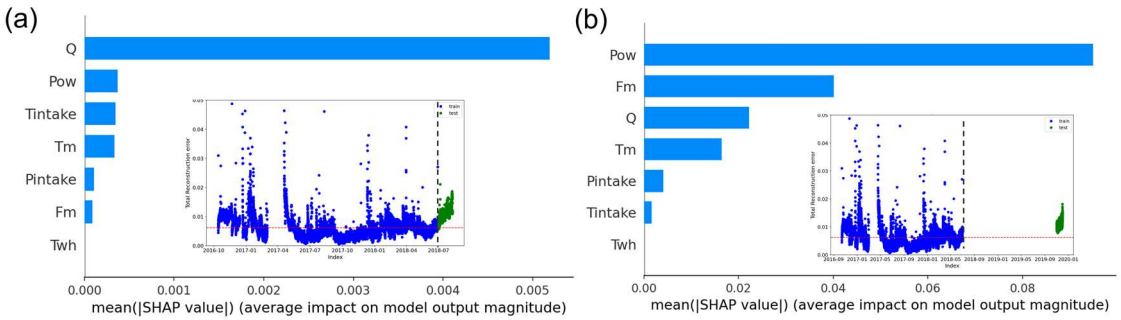


Figure 56. SHAP analysis performed on (a) the period corresponding to the initial rise of the reconstruction error (July – September 2018) and (b) the period corresponding to the increase of the reconstruction error just before ESP failure (November – end December 2019). The leading model parameter is different, with the flow rate (Q) being the most important in (a) and the motor power (Pow) being the most important in (b), suggesting a possible cause for the ESP failure.

5.3.4 Summary of observations

Using the 3 approaches for monitoring the ESP of Well 3, the following can be observed:

- A significant decrease in the inflow is observed after a long shut-in. From the physics-based models there are indications that it could be caused by a blockage or sand production near the wellbore area. The downhole pressure mismatch disappears 1 month before ESP failure, just after a high flowrate flush.
- 2 months before ESP failure, it is observed that the measured pump head is higher than the value predicted by the model.
- Similarly, a mismatch is detected for the motor temperature and the ESP power where the measured values are higher than the model predictions. Both data-driven approaches (RFR and NN) indicate the same trends.
- The reconstruction error from the autoencoder is very large for a long period. Closer look to the final increase, just before ESP failure, indicates that the motor power is the most relevant parameter.

5.4 Learning from small data

When sufficient training data are not available, as for example in the case of well 2, data-driven approaches can give unreliable predictions. This can be a problem both in the case of ESP failing in unexpectedly short time and when the monitoring workflow is immediately deployed at the beginning of ESP operations. It should be stressed that it is not only the amount of available data that make a training dataset suitable for the data-driven approaches but also the variations in the operation conditions contained in such a dataset. The main problem of an uncalibrated model (due to inadequate training data) is to generate a large mismatch between the predicted and measured quantity, that would trigger a (false) warning that would erroneously induce a stop to the operations. In this section, we show a possible solution to the problem of a small dataset. Using transfer learning, we demonstrate how it would be possible to trust the proposed data-driven workflow for ESPs operating since a short time, if a large database of ESP data would be available. In particular, a case for which an ESP is running since only two months is considered, and where the last month is to be predicted. We compare (i) the 'traditional' case where the training is based on the available data for the ESP under consideration and (ii) the case where, thanks to transfer learning, previous knowledge based on a common database of ESPs is deployed for the ESP of interest.

First step for the comparison is to construct an extended database based on the timeseries available. Note that as shown in Table 2, only a sub-set of the different variables is common to all datasets, e.g., not all wells have available data for the same type of sensors. Furthermore, only periods where there is certainty that the ESP is in a "healthy" operation, are considered. For these reasons, 11 months of data from Well 1 and 22 months of data from Well 3 are combined. In addition to data directly measured by the sensors, also calculated variables are included. The variables considered in the combined dataset are flowrate (Q), frequency (Fm), ESP power (Pow), motor temperature (Tm), intake temperature and pressure, ESP head (or alternatively the calculated discharge pressure Pdis), ESP flow resistance (or pressure drop over flowrate), and ESP label id (to capture the different manufacturer specs), and they are shown in Figure 57 (in arbitrary units).

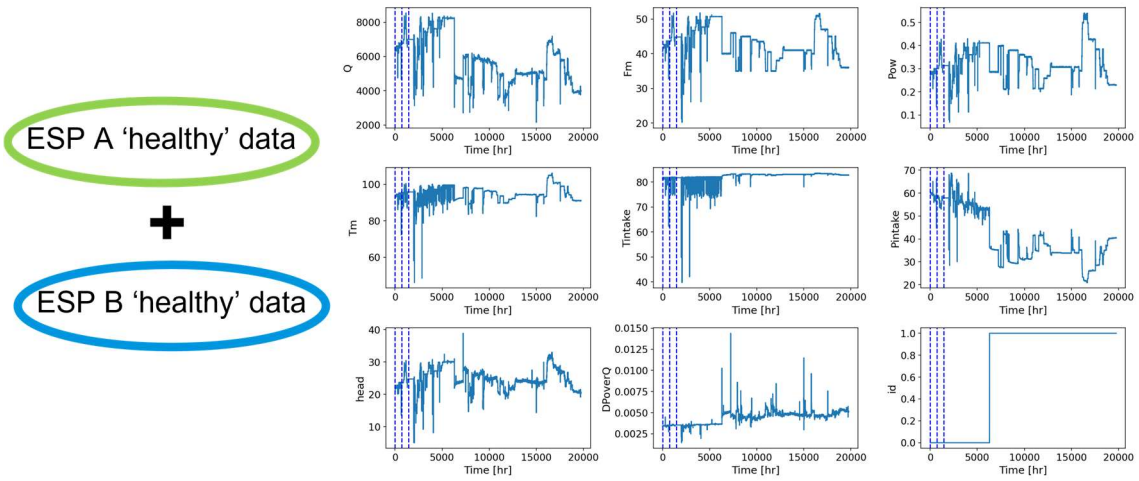


Figure 57. A database consisted of only the data from the 'healthy' portion of the timeseries for two ESPs is created. Nine variables, including original sensor data and calculated quantities, are considered. Period 1 is the first month of the timeseries (in between the first two dashed vertical lines), period 2 is the second month, period 3 is the remaining time (after the third vertical dashed line).

Therefore, the total combined dataset, consisting of 33 months of hourly data associated to healthy operations, is further divided in different periods depending on the test performed, as described in the following.

Test (i), traditional approach:

Period 1 (1 month ESP A) is used for training
Period 2 (1 month ESP A) is used for testing

Test (ii), transfer learning:

Period 3 (22 months ESP A+B) is used for training
Period 1 (1 month ESP A) is used for re-training
Period 2 (1 month ESP A) is used for testing

In both cases a neural network (NN) to monitor key ESP quantities is used. For a good model, no mismatch between the predicted and measured quantities is expected, since data from only healthy periods is used. Test (i) follows the same procedure used in the sections above. For test (ii), the schematic description of the workflow is shown in Figure 58. It consists in training the NN on the combined database that would allow to learn generic relationships between the variables for a sufficiently varying range of ESP operations and well properties. The obtained pre-trained NN can then be re-trained on the actual data of interest (first month of ESP operations) such that the generic relationships between variables (i.e., the NN weights) are calibrated to the ESP of interested. Finally,

the fully trained model is used to make predictions in the testing period (one month of ESP operations).

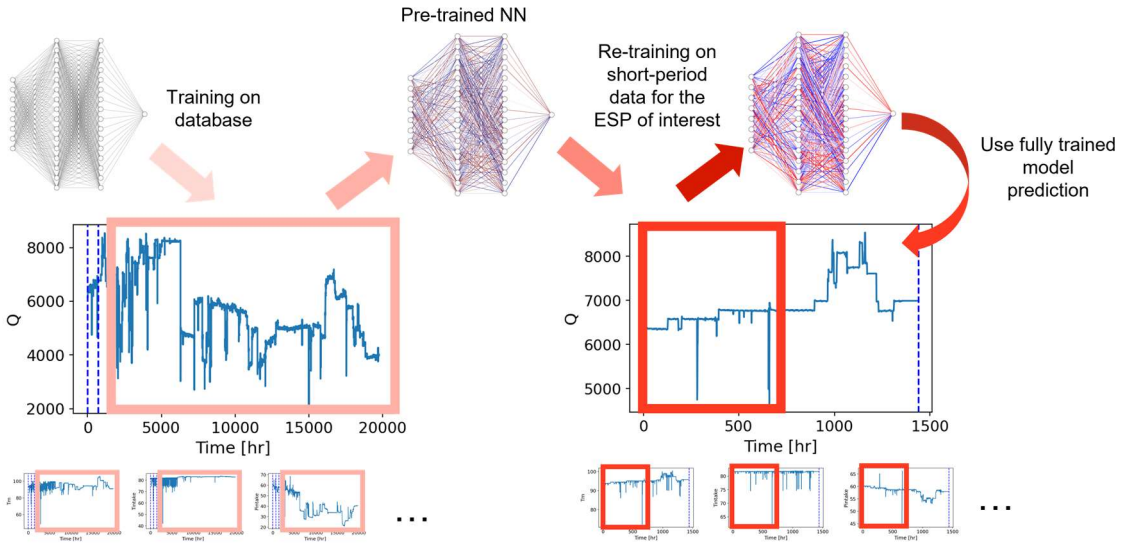


Figure 58. Workflow for transfer learning approach to ESP data. The neural network (NN) is first trained "offline" on data previously stored. The pre-trained NN is then used as starting point for the monitoring of the ESP of interest, where little data are available. The NN is re-trained ("calibrated") on the short-period data and the fully trained model is used for the prediction.

The results of both tests are reported in the last figures of this section, where only the trend for the testing dataset (period 2) is shown where no mismatch between a good model and measured data is expected. In Figure 59, the results of test (i) for the ESP power are reported. The model mistakenly predicts a smaller ESP power than measured suggesting therefore that there might be some anomaly in the ESP operations. However, this mismatch is due to the uncalibrated model that likely cannot recognize the change in operation conditions that lead to an increase in ESP power. In Figure 60, the results of test (i) for the motor temperature is shown. Also in this case some mismatch is observed since the temperature is underestimated for the corresponding period as the large mismatch in the ESP power was detected.

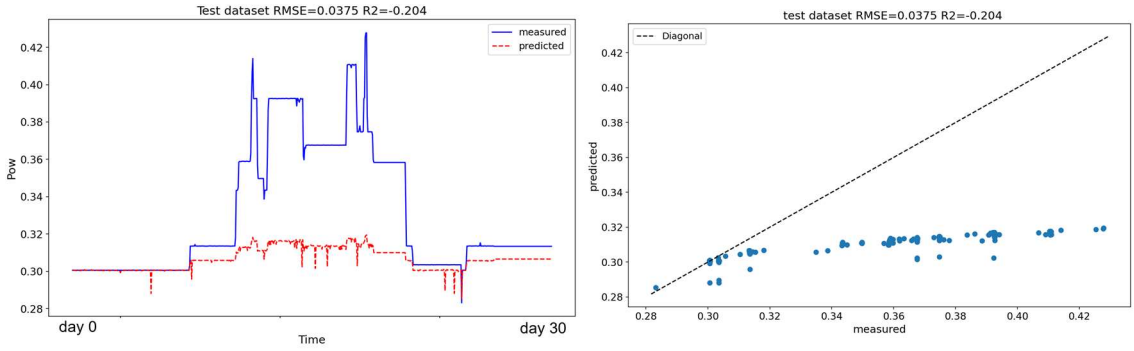


Figure 59. Results for test (i), traditional machine learning approach using only the limited data available of the ESP under consideration. Predictions for the ESP power Pow (arbitrary units) are shown for the test dataset. Despite no mismatch is expected (healthy data), clear difference between the (erroneous) model and the measured data is observed.

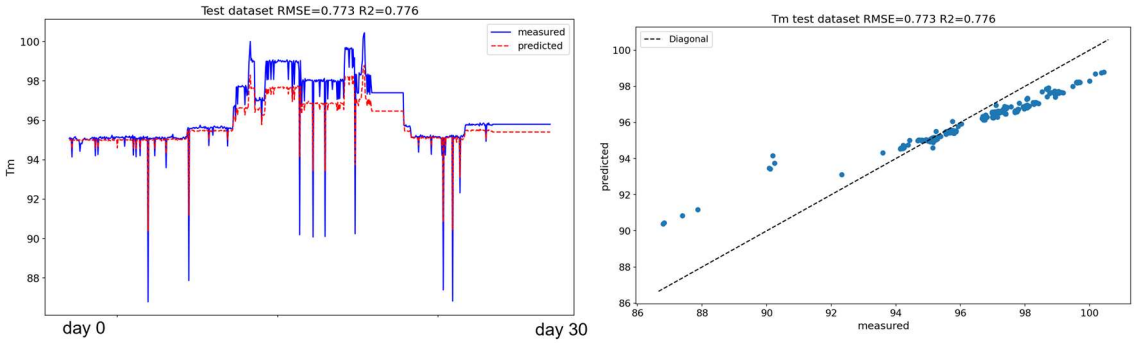


Figure 60. Results for test (i). Predictions for the motor temperature (Tm) are shown for the test dataset. The model underestimated for a large period the measured temperature.

In Figure 61, we report the results for test (ii) where the transfer learning approach is used. Compared to the previous test, it is evident that now the model captures the measured trend significantly better. Similarly, Figure 62 shows the predictions for the motor temperature and it can be observed that also in this case that the model accuracy improved compared to test (i). For both figures it can be concluded that in test (ii) a well calibrated model can be obtained, and no mismatch is detected as is expected from healthy operation conditions.

This shows that the problem of limited data, that undermines the confidence on data-driven approaches, can be overcome using a transfer learning approach. These results demonstrate that the data-driven workflow can be deployed also at the beginning of ESP operations, providing the existence of a general database containing enough varying ESP data.

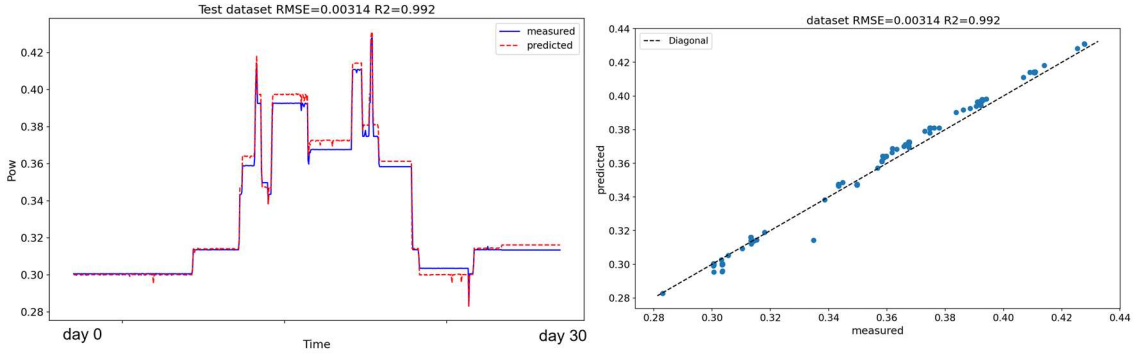


Figure 61. Results for test (ii) when using the transfer learning approach. The model now correctly predicts the system behavior. Results are shown for the ESP power (arbitrary units).

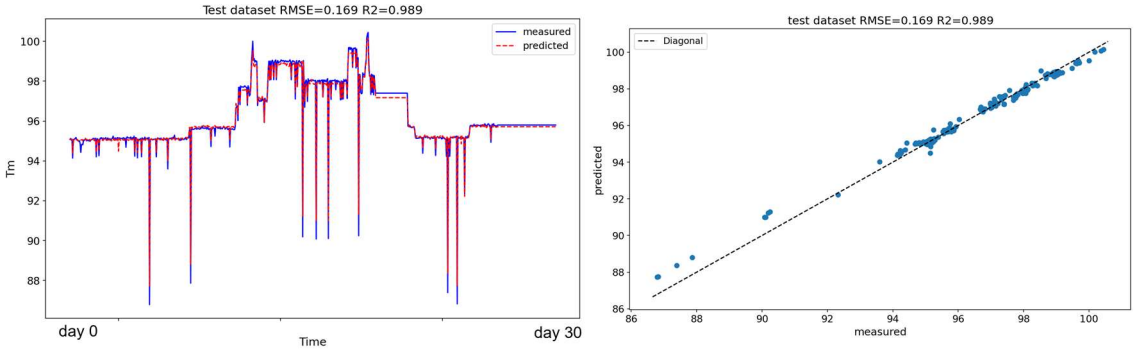


Figure 62. Results for test (ii) shown for the motor temperature. Also in this case, higher prediction accuracy compared to the traditional machine learning approach is achieved.

6 Conclusions and recommendations

6.1 Conclusions

In this report a workflow was demonstrated to perform operation optimization of geothermal assets up to real-time. The intention of the workflow was to act as a decision support tool to the operators of geothermal systems to improve the operational excellence, minimize the operational costs and expenditure and eventually make a transition towards a proactive operation. The selected case study for the workflow demonstration was ESP performance monitoring which was selected based on the three criteria of operators' priorities, data availability and partners consensus. An extensive survey was performed to provide insights in these criteria. ESP performance monitoring was selected due to high costs associated with ESP inspection and replacement, high frequency of failure occurrence and most importantly availability of data since ESP is one of the most instrumented component in the geothermal assets.

The developed workflow was a model-based condition monitoring and it consisted of several physics-based and machine learning models. The models have been applied on three geothermal assets in which several ESP failures were observed. The physics-based models had the capabilities to determine offsets and mismatch in the pressure (and/or flow rates, depending on the quantity of interest) across the geothermal wellbores, such as bottomhole, wellhead, intake and discharge of the ESP. The integrated analysis of the mismatches showed insights on the potential root-cause of the performance degradation, e.g. due to inflow restriction, pump efficiency degradation and/or tubing and casing restrictions. The physics-based models were capable of showing some mismatches in three wells in the bottomhole pressure estimations and pump heads up to few months before the failure.

Apart from the production conditions (pressure, flow rates) which provide lots of insights in ESP failure, the ESP parameters (such as motor temperature, current, voltage, power and vibrations) monitoring is also of a great importance. For a reliable and accurate model-based monitoring of ESP parameters in real-time, a set of machine learning models were trained to predict the ESP parameters. Among all the tested parameters, motor temperature, vibration and ESP power found to be relevant parameters to indicate a trend towards pump degradation and failure. Due to the availability of the sensor data for the ESP parameters, a direct comparison of the model predictions with the measured parameters will be possible. The mismatch in the ESP indicators in terms of mismatch and an increase in the fluctuations of the predicted values were found to be a good indicator for performance degradation.
Furthermore, a deep learning algorithm, called auto-encoder, was deployed to provide insights on the performance of the overall systems and translate it into a single KPI, known as reconstruction error. This indicator provides insights to the changes in the system behaviour from the normal behaviour (as introduced by the user). This approach has been applied to the geothermal wells for monitoring and to showcase how to identify the cause of failure.

Finally, transfer learning can be used to overcome problems arising from small amount of data that do not contain enough variation in the operating conditions. Using the monitoring method based on neural networks, it can be concluded that that when only data from short periods (couple of months) are available, the traditional training approach resulted in a poor model. In fact, a large mismatch in the monitored quantities was observed even for healthy ESP behavior, with the model therefore

suggesting false warnings. By applying the transfer learning procedure, consisting in pre-training the model on a larger general dataset of ESPs and afterwards deploying it for the ESP of interest, no false mismatch was detected, and the model was considered adequate for the monitoring purpose.

The bullets below lists the highlights and lowlights of the conclusions:

- The developed framework is sufficiently generic to be applied to monitor several critical processes in the geothermal plants, upon the data availability. This workflow was capable of detecting the indications towards the failure from 1 to 6 months prior to the failure.
- All the inputs of the model should be reliable and accurate. This means a thorough data pre-processing is required to remove outliers and noisy data and fill the missing data.
- A thorough analysis of the ESP performance degradation and failure requires data from the production and ESP to be combined, as the root cause of the ESP performance might be linked to the processes upstream and downstream of the ESP.
- Combination of all the methods (physics-based and machine learning) is necessary to provide a full picture of the processes leading to the degradation and failure of ESP.
- During the training and history matching of the models, the error of the prediction needs to be evaluated. This step will enable differentiating the mismatches due to prediction accuracy or an anomalous behaviour. As it was shown, the reconstruction error using auto-encoder approach during the training phase could indicate the error threshold.
- The variability in the data or in other words the information density of the data should be sufficient to allow for a reliable prediction. In this case, the data for all methods needs to cover all possible operating condition (on summer-winter) to minimize the extrapolation error.
    - Data driven workflow cannot be applied when data is limited (ESP short failure)
    - Using the data from similar wells and ESPs could be used to pre-train the data-driven models (application of transfer learning)

6.2 Recommendations for the future work

- The current workflow is required to be tested on the cases in which the failure time is not known to perform a blind test and evaluate the predictability of the workflow.
- It is recommended to always monitor both the well production data and the ESP itself. If manufacturing challenges can be overcome, having a sensor for monitoring the ESP discharge pressure would be beneficial for a direct comparison with the models.
- To overcome difficulties in model calibration for a single ESP, it would be advisable to perform several tests (e.g., at beginning of operations) to cover larger variations of operating conditions. Future work should therefore be dedicated to study and define a suitable procedure for such testing (e.g. what is the minimum range and number of variations required).
- Create a common database by gather a larger amount of information on well and ESP during operations. This will greatly improve the workflow accuracy and reliability since it will be possible to benefit from the transfer learning approach discussed here and improve the data-driven models.
- More test cases are necessary to standardize the workflow procedure. Standardization would range from the set of data pre-processing operations needed (e.g., deal with the data acquired after shut-ins) to indications on how to quantitatively use the indicators during monitoring. For example, it will be necessary to carefully estimate the model error during the training / history matching period. Based on such estimate, a threshold should be

defined and used during monitoring to quantitatively assess if the mismatch between model and sensor measurements is a false warning (mismatch below threshold) or a true indicator of degradation/malfunctioning/failure.

- As the next step, the workflow needs to be tested for predictive and prescriptive analytics in order to forecast the behaviour of the system and eventually propose optimum controls to mitigate the unwanted processes and failures.
- The existing domain-knowledge on ESP guidelines for root-cause analysis should be directly incorporated in the workflow, in order to assist in real-time the operator in distinguishing the possible cause of malfunctioning and preventing failure.
- Developing of an ESP database, by collecting all the relevant data from geothermal ESPs, could improve the knowledge on the performance degradation of the ESP and eventually enable a sector-wide learning by sharing data.

7 References

Antwarg, Liat, Ronnie Mindlin Miller, Bracha Shapira, and Lior Rokach. 2021. "Explaining anomalies detected by autoencoders using Shapley Additive Explanations." *Expert Systems with Applications* 115736.

Beggs, D.H., and J.P. Brill. 1973. "A study of two-phase flow in inclined pipes." *Petroleum Technology* 25: 607-617.

de Zwart, Hidde. 2021. "Cost reduction of geothermal systems in the Netherlands." Internship presentation within WarmingUP program (confidential).

Lundberg, Scott M., and Su-In Lee. 2017. "A unified approach to interpreting model predictions." *Advances in Neural Information Processing Systems* 30.

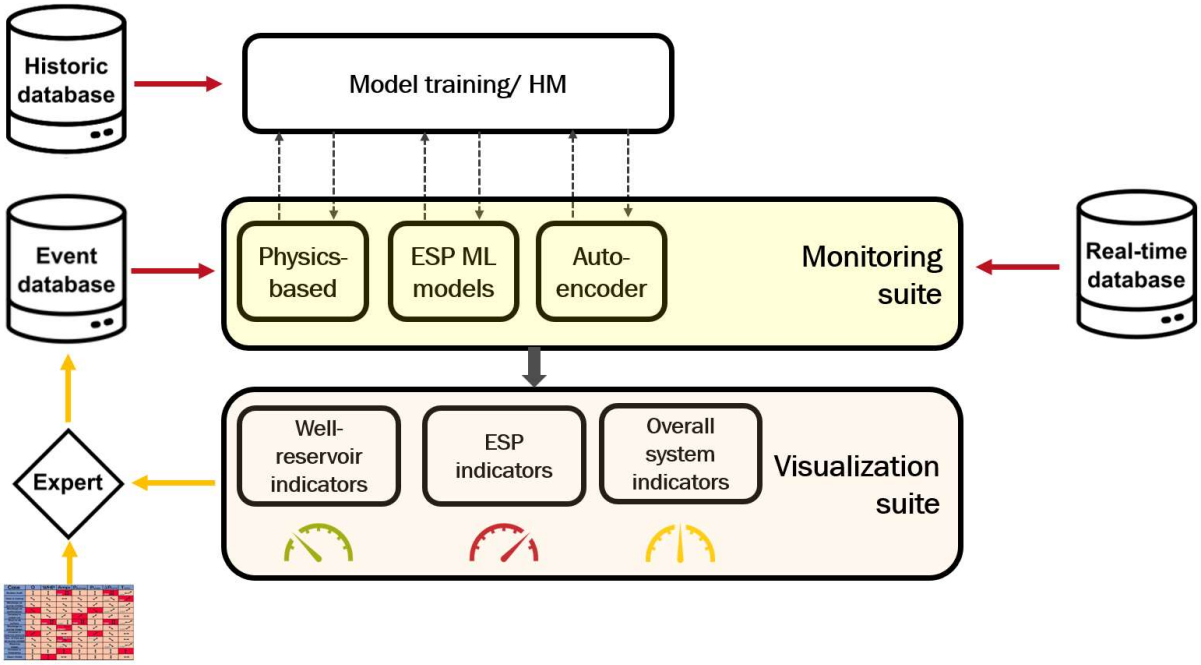
Shapley, Lyod S. 1953. "A value for n-person games." *Contributions to the Theory of Games* 307-317.

Shoeibi Omrani, Pejman, Kay Van der Valk, Wim Bos, Eduard Nizamutdinov, Laurens Van der Sluijs, Joren Eilers, Hajo Pereboom, Koen Castelein, and Frank Van Bergen. 2021. "Overview of Opportunities and Challenges of Electrical Submersible Pumps ESP in the Geothermal Energy Production Systems." *SPE Gulf Coast Section Electric Submersible Pumps Symposium.* Texas, USA. doi:10.2118/204524-MS.

Techo, R., R.R. Tickner, and R.E. James. 1965. "An accurate equation for the computation of the friction factor for smooth pipes from the Reynolds number." *Applied Mechanical* 32: 443.

van 't Spijker, Hans, and Pierre Ungemach. 2016. "Definition of Electrosubmersible pump (ESP) design and selection workflow."

Wasch, Laura, Raymond Creusen, Florian Eichinger, Tanya Goldberg, Claus Kjoller, Simona Regenspurg, Troels Mathiesen, Pejman Shoeibi Omrani, and Viola van Pul-Verboom. 2019. "Improving Geothermal System Perfomance Through Collective Knowledge Building and Technology Development." *European Geothermal Congress.* Den Haag, the Netherlands.



Adres
Princetonlaan 6
3584 CB Utrecht

Postadres
Postbus 80015
3508 TA Utrecht

Telefoon
088 866 42 56

E-mail
contact@warmingup.info

Website
www.warmingup.info