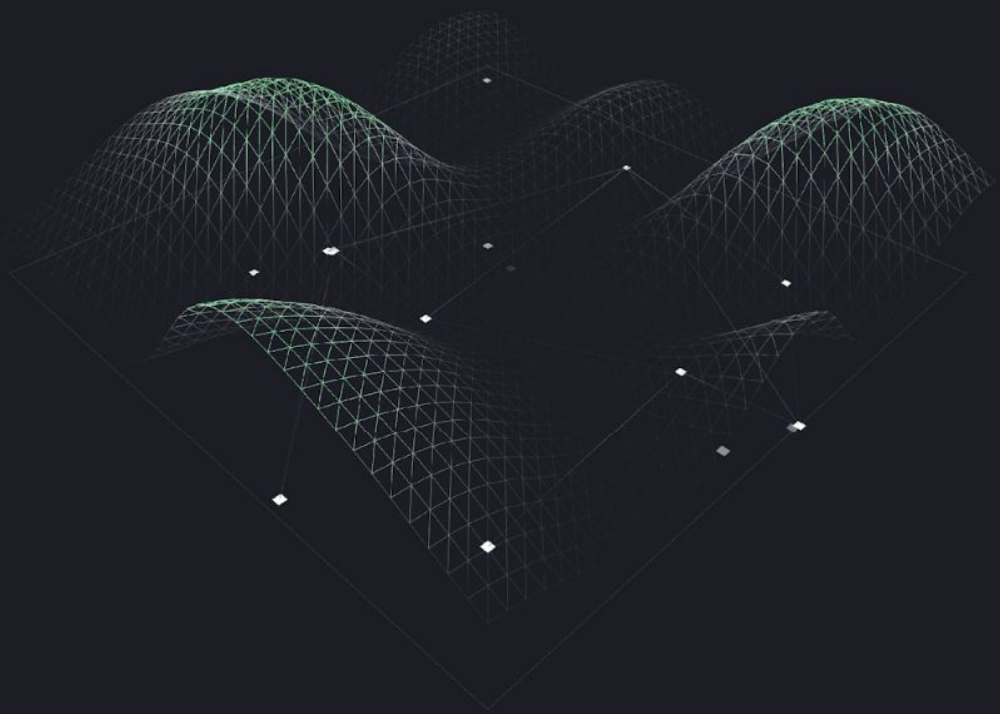


AI Flow Solutions

Delivering AI-powered, best value flow rates for all

Solution Seeker

contact@solutionseeker.no || solutionseeker.no

Want to be in on the secret of how we develop and deliver virtual flow rate estimates for hundreds of production wells? Read on.

Looking at the number of pages and thinking "it's too much, I'm not gonna bother"? We promise it is worth your time! A complete read-through of this document costs you about 1.5 hours, but, if time is of the essence and you would like to browse only the **juicy** bits, we recommend jumping to Section 4 - NeuralCompass Virtual Flow Meter and Section 5 - FlowFusion. These sections will likely take you about 20 minutes.

1 INTRODUCTION

The topic of the time is all about *artificial intelligence (AI)* and *machine learning (ML)*. With the recent and rather disruptive entrance of ChatGPT, the subject gathers a newfound interest for its seemingly limitless potential. Although certainly in the wind, AI has been around for a while and we have long spoken about its accomplishments in many real-life applications. For instance, DeepMind's AlphaGo Zero program applies reinforcement learning to achieve superhuman proficiency in the game of Go (Silver et al. 2017). In our everyday lives, Google Lens, which was first launched in 2017, helps us translate text and detect objects from an image on our phone, with an impressive accuracy through the use of neural networks (Shapovalov 2020). Despite the numerous success stories in various industries, there are particular sectors where AI technology is conspicuous by its absence. Ironically enough due to their vast amount of historical and real-time sensor data, process industries such as petroleum production and chemical manufacturing have been slow to follow the trend (Clark et al. 2022). Of course, we might come across a piece or two in the literature (Balaji et al. 2018), yet many of these are from a purely experimental perspective and struggle to realize the promised potential of AI - a challenge many companies can relate to (Fontaine et al. 2019). In this paper we propose AI-powered technologies that exempts us from this statistic. So buckle up and listen in as we divulge our secrets on data-driven virtual flow metering and production allocation within the oil and gas industry.

Accurate flow rate estimates or measurements are crucial to efficiently operate a petroleum production asset. In their simplest form, they can be used for monitoring, providing situational awareness for better decision making between intermittent well tests. They are also part of many important downstream applications, enabling for instance allocation, production optimization and reservoir modeling. There exist many devices that deliver continuous flow rate estimates such as multiphase flow meters (MPFMs) and virtual flow meters (VFMs). With regard to the latter, today's industry standard is mechanistic VFMs, e.g. Prosper, LedaFlow, and FlowManager. Although mechanistic VFMs have proven successful in many scenarios, there are also some pitfalls with this type of VFM. Due to uncertain subsurface fluid properties and complex multiphase flow dynamics it can be difficult to achieve the required precision in real-time using first principles (Jansen, 2015). Consequently, maintaining sufficiently accurate estimates from a mechanistic VFM is often just as resource demanding as developing them. High maintenance solutions are also expensive making it challenging for smaller assets to invest.

Emergent on the market for continuous flow rate estimation are data-driven VFMs, such as our commercially available *NeuralCompass Virtual Flow Meter (NC-VFM),* which offers a counterpart to mechanistic VFMs. A particularly appealing property of data-driven VFMs is their independence of fluid properties and dynamics, typically reducing the required maintenance hours and thus making them a cost-efficient solution compared to many mechanistic VFMs (AL-Qutami et al. 2018). That does not mean, however, that the AI-leveraged alternatives have no challenges of their own. They are for example extremely data dependent and can suffer if the data quantity and quality is poor (Grimstad et al. 2021). Our VFM, which we introduce in later sections, is based on a specialized technology that tackles these data challenges: NeuralCompass. NeuralCompass is based on more than 10 years of heavy research and experience in the field.

10 years

Heavy research and experience in the field

Since no MPFM or VFM is perfect in practice, the estimated or measured flow rates for the individual wells rarely add up to the more accurate topside measurements of the total production. This impacts production allocation, which again negatively affects reservoir simulation, drawdown strategy, and placement of new wells. Hence, to mitigate measurement errors and draw up a justifiable revenue split between stakeholders, the observed imbalance should be distributed to appropriate wells in proper proportions (Kanshio 2020). Nevertheless, the tricky part is not the concept, it is the execution. Similar to the field of virtual metering, the most common approaches within allocation today, such as pro-rata allocation, fail to take advantage of the high-quality information that lies in the production data - that is, quantifiable uncertainties, well tests, redundant flow rate data, and detectable MPFM or VFM errors to mention a few (Badings and van Putten 2020). It sure is less of a hassle to allocate proportionate to a well's contribution, but there are more occasions where the assumptions behind this strategy fall apart than not. In this paper we introduce *FlowFusion (FF)*, an industry tested allocation framework that optimally exploits the production data to assign the imbalance to where it most likely belongs and facilitate the detection of gross errors.

FlowFusion is an industry tested allocation framework that optimally exploits the high-quality information that lies in production data

Lastly, we presume that the time is ripe to address the elephant in the room; how is data-driven technology supposed to harmonically co-exist and interoperate with processes and systems that are already in use. AI is well-known to facilitate automated and streamlined workflows, but it needs to be fully adopted in existing structures to thrive (Davenport & Ronanki, 2018). From our experience, the petroleum industry has many well-established traditions that require low-intrusive and flexible technology. Our semi-automated framework involves the whole pipeline from recording well tests to training NeuralCompass VFM to fusing it all together with FlowFusion. However, the attentive reader may have noticed the word "*semi-automated*". Due to the complexity of the production system and the decades of knowledge acquired by industry professionals, we believe it is crucial to keep humans in the loop to successfully apply AI in the petroleum industry. Admittedly, not everything can be solved or detected by a fancy AI-leveraged algorithm. Therefore, our technology platform, *ProductionCompass AI (PCAI)*, strives to automate as much as possible in a seamless and low-intrusive fashion while involving industry professionals to exploit the powerful foundation of knowledge only they have.

Our vision is to set the standard with AI-powered, best value flow rates for all. How? You will have to continue to be in the know.

In the following section (Section 2) we describe some common data challenges found in the oil and gas industry and how it affects AI-leveraged algorithms. We then proceed in Section 3 by explaining how we tackle some of these using our platform, *ProductionCompass AI*, for data processing. Sections 4 and 5 delve into our AI Flow Solutions, firstly with our product for data-driven flow rate estimates, NeuralCompass VFM, in Section 4, before Section 5 deep-dives into our allocation framework, FlowFusion. We wrap it all up with a summary and our sources in sections 6 and 7.

2 DATA CHALLENGES IN THE O&G INDUSTRY AND HOW IT AFFECTS AI-LEVERAGED ALGORITHMS

As mentioned in the introduction, AI algorithms do have their challenges. Because AI, or machine learning in general, learns from observations of the world, many of the challenges are tied to the data foundation that the algorithms have access to. It is well known that many AI success stories in other industries, such as self-driving cars, have algorithms that are learnt from an impressive data foundation with billions of examples. In a sense, we can compare machine learning with a toddler learning to walk. The toddler tries out different leg and body positions, and is awarded with an outcome. Through experience, the toddler can eventually learn the relationship between different leg and body positions and the movement of walking. The more experience, the better the toddler will be at walking.

Yet, despite the vast quantity of available production data in the petroleum industry, the data applicable for AI-leveraged algorithms is unfortunately typically just a fraction of the available data. Below, we describe the data challenges in the petroleum industry and how it poses a threat to AI-leveraged algorithms. Naturally, to be best at what we do, our data-driven NC-VFM and our allocation framework, FF, must handle these challenges. If you cannot wait to know how, you can skip to Sections 4 and 5.

Customarily, petroleum production systems are complemented with multiple real-time sensors, such as pressures, temperatures, and valve openings. They also typically measure the total production rate of the commingled wells in the system. Furthermore, most companies have some information about the multiphase flow rates from individual wells. Either through well-testing with test separator or deduction tests, or measurement devices such as multiphase flow meters (MPFMs) that provide real-time estimates. For systems without continuous monitoring of the wells flow rates, these can be deduced from the total production rate and allocation factors calculated with the well tests (Sadri et al. 2019). Despite the many available sensor readings, when analyzing typical available production data from petroleum assets, we often find four significant data challenges that, in our opinion, are some of the reasons why AI has not yet gathered momentum in the oil and gas industry: i) a low volume of applicable data, ii) a low degree of variety in the data, iii) poor data quality, and iv) a non-stationary underlying process (Grimstad et al. 2021).

The four major data challenges

Low volume
Low variety
Poor data quality
Non-stationary process

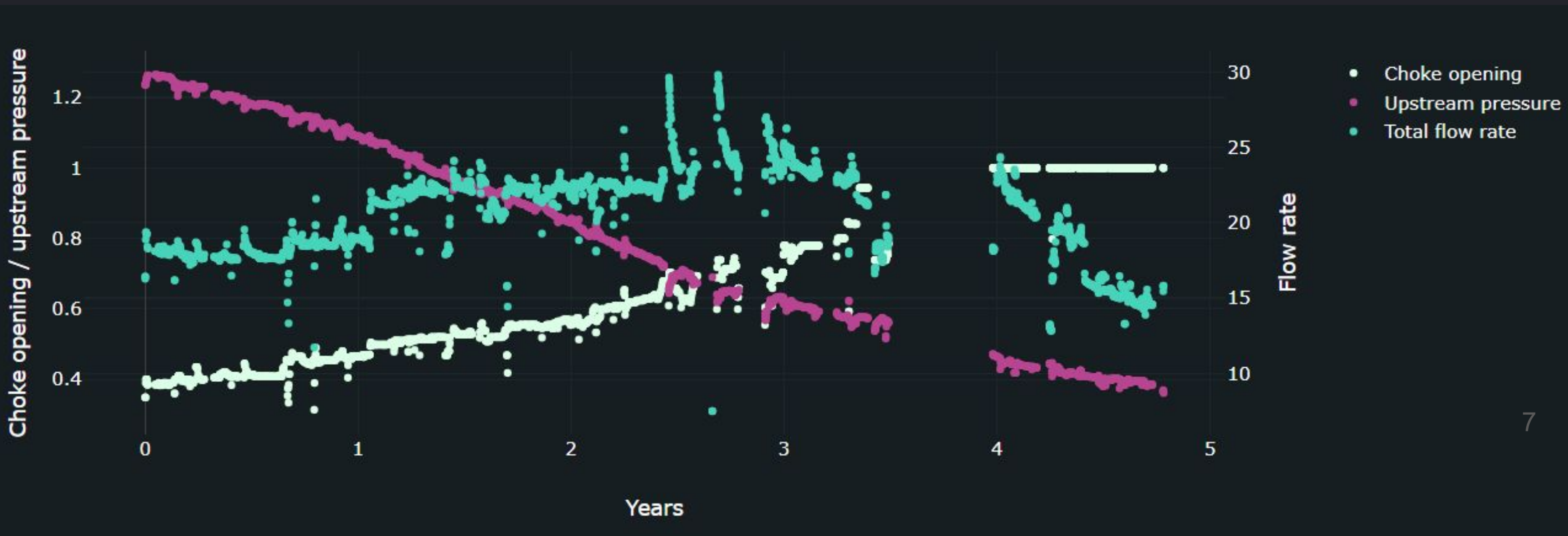
Data volume refers to the size of the data foundation that can be used to train the algorithm. For petroleum assets without continuous monitoring, the data volume tends to be low. For instance, development of VFMs on assets with infrequent well testing will typically suffer from low data volume as new data is obtained at most 1-2 times per month (Monteiro et. al. 2020). Consider a petroleum well with start-up in January 2015. By 2023 this well has, in a best case scenario, 192 observations to train the VFM with. Compared to the billions of examples that state-of-the art image classification algorithms use to obtain high accuracy, we can clearly see the data volume challenge. Naturally, low data volume with infrequent well-testing also greatly impacts allocation as the uncertainty and relevance of the allocation factors calculated from well tests often increases with the time between tests (Sadri et al. 2019).

For assets with continuous monitoring, the data volume can be higher but the degree of data variety can still pose a problem. By data variety we mean the range of values for the different measurements that we are using to build our machine learning models. Let us explain the challenge of low data variety by introducing the Weather Problem. Imagine that you want to build a machine learning model that can predict the weather in Norway. You collect weather data, such as temperature, during the summer months and train the model based on these data.

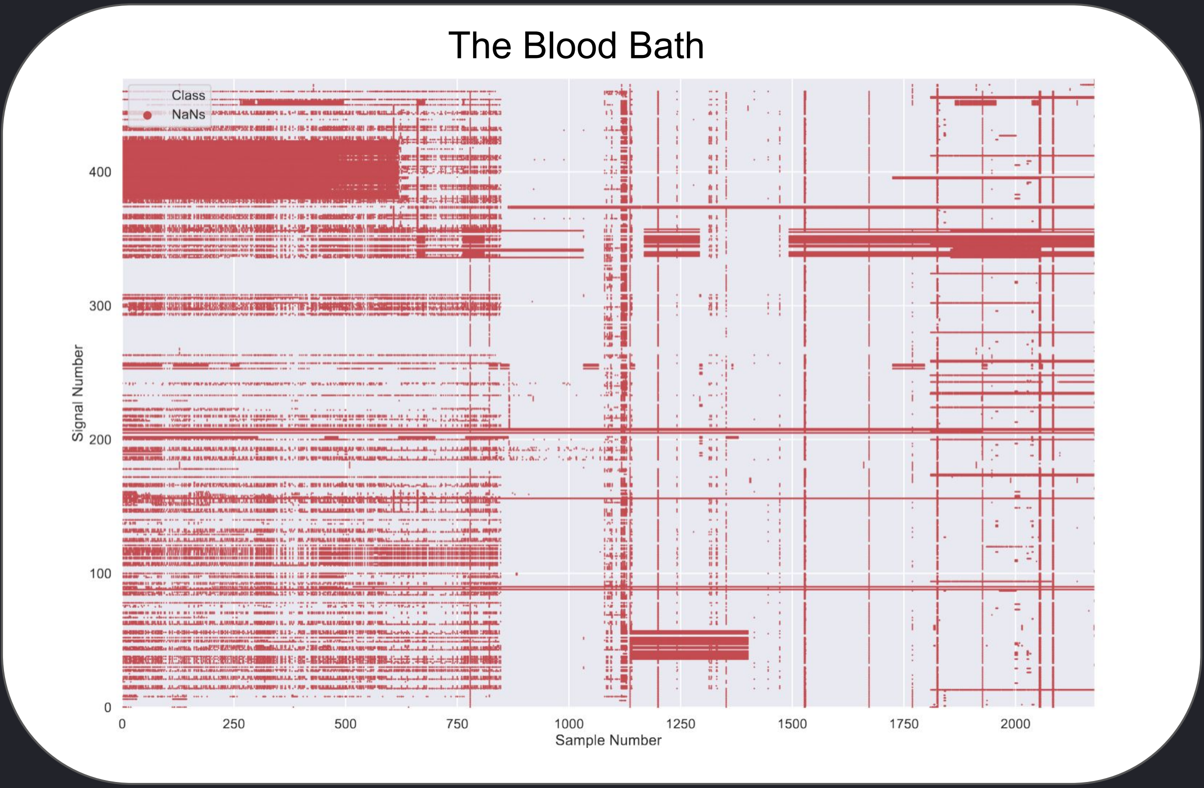
Now, in early autumn you might expect the model to do well, in particular if the summer temperatures persist, but would you expect the model to do equally well when winter arrives and the temperature drops below zero? Likely not. As the data you collected are from the summer months, the range of temperatures that the model has seen does not include negative degrees. How can you expect a model to predict something reasonable in the winter months if it has never experienced winter?

Let us return to the production data of a petroleum asset. In the figure below , we have plotted the choke opening, the upstream pressure of the choke, and the flow rate for one well as a function of years since the start of production.

The data is downscaled and does not represent the real values. A low degree of data variety is commonly seen due to the way petroleum assets are operated. In the early life of a petroleum well, the production is typically kept fairly constant by the operator, often limited by the processing capacity. As a consequence, typical measurements such as the choke opening and pressures upstream and downstream the choke valve are also fairly stable. As long as the measurements stay this way, we can expect a good performance of a machine learning model trained on these data. However, in later stages, when the field enters the decline phase, the operators often compensate for declining pressures and production rates by opening the choke valves, clearly seen in the figure. In practice, this means that for a certain point in time, the future process conditions and flow rates will not be similar to the historical and previously seen conditions and flow rates. Similarly to the weather problem, this poses a significant challenge for a data-driven VFM trained on historical data. This will also influence allocation as the uncertainty of measurements used in the allocation problem, for instance MPFMs, are affected by changing conditions (Corneliussen et al. 2005).



The third challenge is the quality of the available data. It is not uncommon for petroleum production data to suffer from poor quality due to high noise levels, causing imprecision in the sensor readings and systematic errors (Antonelo et al., 2009). Even recently and properly calibrated MPFMs have up to 5% error for gas, oil, and water rates, depending on the flow regime and conditions (Thorn et al. 2012). Some sensors can drift over time and some may even fall out for longer periods. There are also cases of equipment failures that lead to temporarily or permanently missing data. A study we did in-house in Solution Seeker looked at how often and how many of the sensors we are working with had missing data. A part of the outcome of the study is illustrated below. The figure has been nicknamed 'the bloodbath' for obvious reasons. It is clear that the data quality impacts both the data volume and the data variety as a large fraction of the data must be removed due to poor quality.



The last challenge relates to the non-stationarity of the underlying process in petroleum production systems, the reservoir. In time with the reservoir being depleted, the pressure declines. The declination leads to pressure declination in the rest of the production system and, as explained above, affects how petroleum systems are typically operated. As is, the nonstationarity will influence the data variety found in historical data. Furthermore, it will also introduce correlations between measurements that are not physical. For instance, we should expect an increase in the flow rate if we increase the choke opening. However, we have come across production data where there is a negative correlation between the choke opening and the flow rate due to the reservoir being depleted and the choke opening is increased to compensate for falling pressures. It is also typical to see a negative correlation between the choke opening and the oil fraction. On a daily basis this is not physical as it basically says that opening the choke would lead to less produced oil. However, considering time, it is of course a consequence of the reduced total production rate. Trying to isolate the effect of change in only one measurement to the change in the flow rate based solely on the data, we can falsely come to the conclusion that a reduction in the choke opening results in an increased flow rate. Generally, complex correlations between variables are difficult for a machine learning model to learn and can reduce the accuracy. Furthermore, the properties and composition of the produced fluid will change in time. This can cause a change in the general physical behavior of the well that the machine learning model has not observed before and the model can struggle with predictions.

Data challenges

Low data volume

A consequence of infrequent well testing

Low data variety

A consequence of how typical production wells are operated

Poor data quality

Failing and drifting sensors with high noise levels. Will impact both data volume and variety.

Non-stationary process

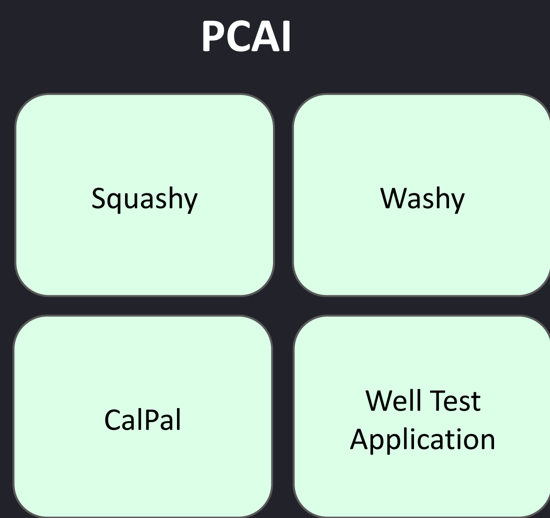
A consequence of depleting the reservoir. Will give faulty correlation between variables and time-varying properties.

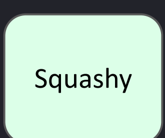
The four data challenges are crucial to overcome in order to successfully and sustainably implement AI-powered solutions in the oil and gas industry. As mentioned above, NC-VFM and FF do handle these challenges. Common for both of them, and a crucial back-bone, is the utilization of PCAI, which processes the production data to retrieve only high-quality data and thereby targeting challenge three. How NC-VFM and FF handle the other challenges is revealed in Sections 4 and 5, respectively.

3 PCAI - DATA PROCESSING

As explained in the preceding section, the data challenges in the petroleum industry imposes certain requirements along the lines of data processing and washing. Due to low-quality data and low amounts of suitable data, it is crucial to both filter out erroneous data that may destabilize an ML training session and ensure that we capitalize fully on the available data to squeeze out every last bit of potential. Good data management - gathering, cleaning, labeling, storing, and optimizing data so it suits the job at hand - is simply indispensable in this industry. Of course, this is not unique to the oil and gas industry. In fact, most data scientists, regardless of sector, find that they often use 80% of their time on data handling and 20% on analyzing the performance. This phenomenon is so well-known and widespread in the data engineering community that it is often called the “80/20 rule” (Gabernet and Limburn 2017).

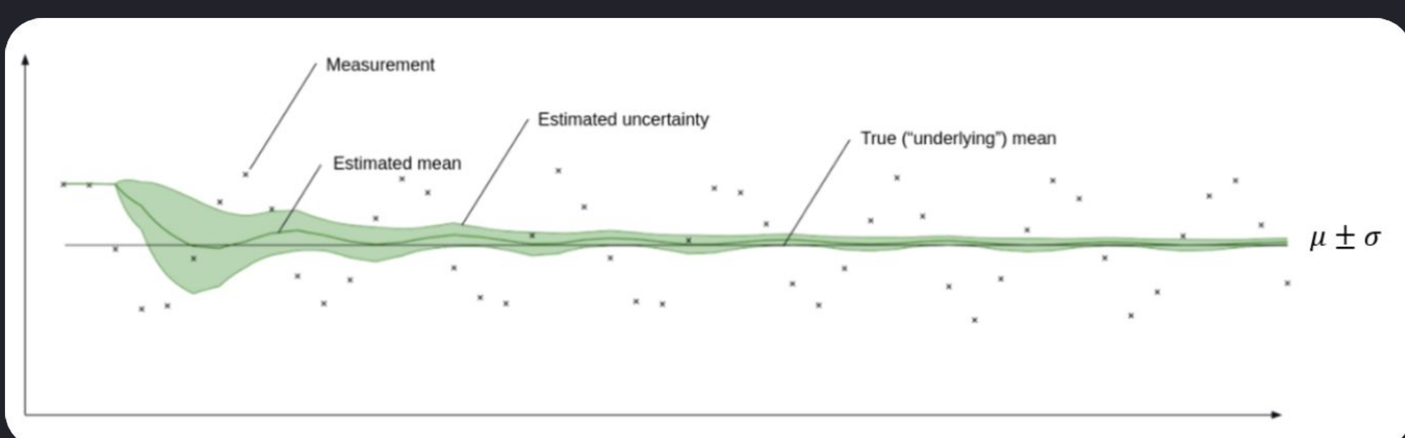
Within PCAI, we have built several data processing systems that work back-end in real-time. They serve as the backbone for all our data-driven services, including NC and FF. The most important are Squashy for compressing production data, the Well Test Application for test data bookkeeping, Washy for quality assurance, and lastly CalPal to virtually calibrate biased rate measurements. If reading about the fundamental building blocks of our services seems a bit tedious and you are really here for the tenderloin, namely NC-VFM and FF, skyrocket through this section and land in Sections 4 and 5. Of course, if you are interested in a deep-dive into PCAI, please do not hesitate to contact us for more information.





Squashy is an algorithmic framework that selects data intervals of interest, and extracts statistical information to represent the raw data in a compressed format for multiple downstream tasks. The framework can detect any type of interval or event in real-time. A typical use-case is to detect and extract system operating points or well tests.

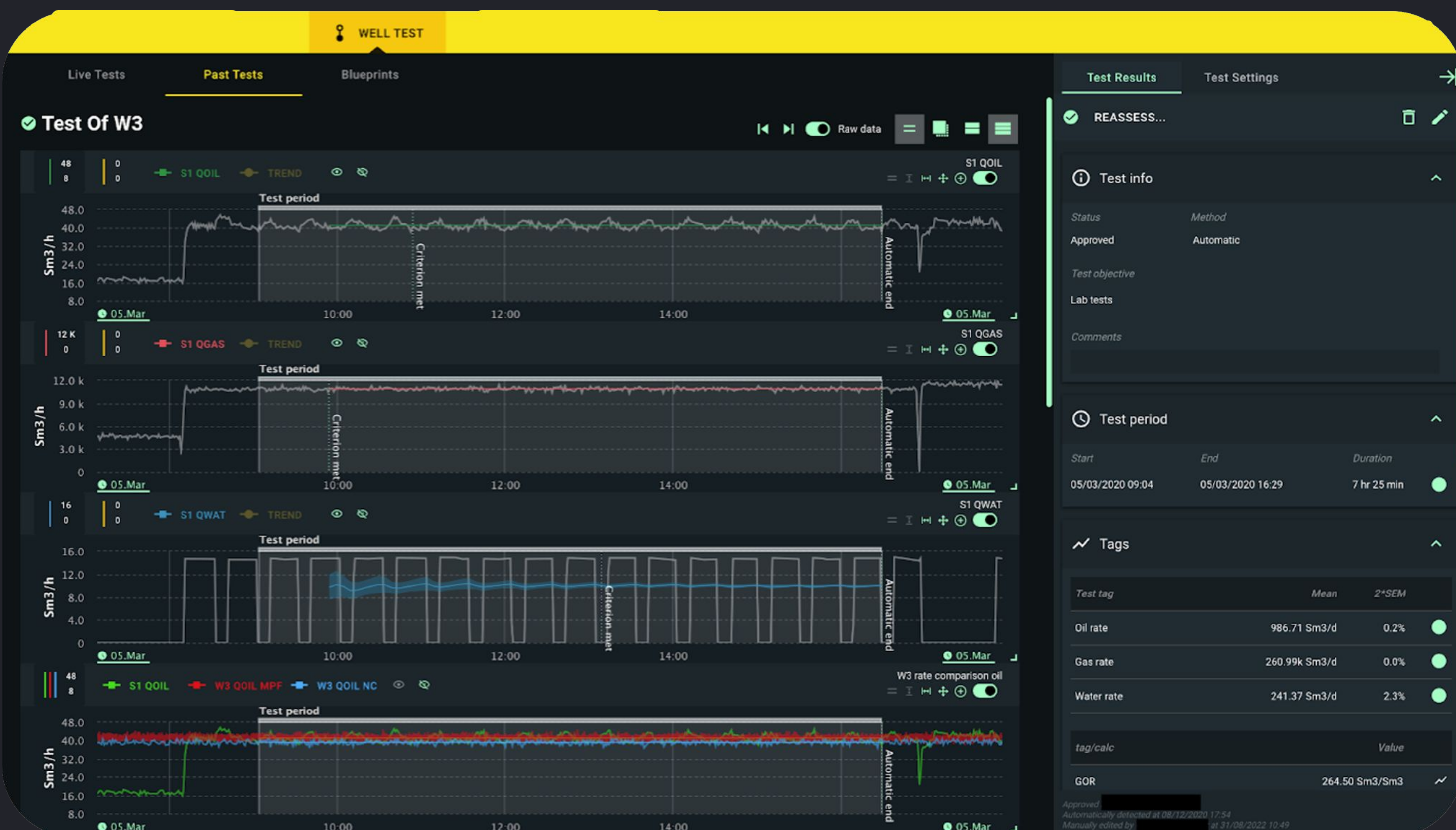
For this, Squashy removes noise by assuming the system to follow a weakly stationary process and computes the underlying data signal’s mean (μ) and standard deviation (σ) for the detected period. The more data samples that come in over time, the closer the estimate is to the true (“underlying”) mean of that interval, which is illustrated in the figure below. Squashy is always configured towards the relevant use case. In the context of NC-VFM, it is beneficial to train exclusively on intervals where the control settings are fixed due to the higher uncertainties related to transient periods. For FF it is often sufficient to split the continuous data stream into naïve 24-hour intervals for daily allocation. Hence, Squashy is often configured differently for these two products. Squashy is proprietary and protected by two patents. For details please read; Gunnerud et al. (2016) and Gunnerud et al. (2018).



Well Test Application

The Well Test Application, which is excerpted in the figure below, doubles as an interactive tool for automatically detecting (through Squashy), tracking and optimizing well tests and a historical database for retrieving any type of test related information depending on the task at hand.

Both NC-VFM and FF are seamlessly integrated with the Well Test Application. NC-VFM is, for example, trained on the production results and additional sensory data from historical tests, whereas FF applies derived information such as the observed MPFM and VFM errors, i.e., the deviations between the test results and the MPFM and VFM estimates, to update its belief of the uncertainties in the production system. With the application, high-quality test data becomes readily available in real-time, which makes for faster response times and less resource demanding calibrations. The Well Test Application - a stand-alone product in our portfolio - is therefore considered to be a vital part of our flow rate services. Note however that the application can also be set up to automatically read in tests from other systems for storing well tests in case a transition to a new system is premature or otherwise refrained. More can be read on the topic here: https://www.solutionseeker.no/products-and-solutions/well-testing/ .



Washy

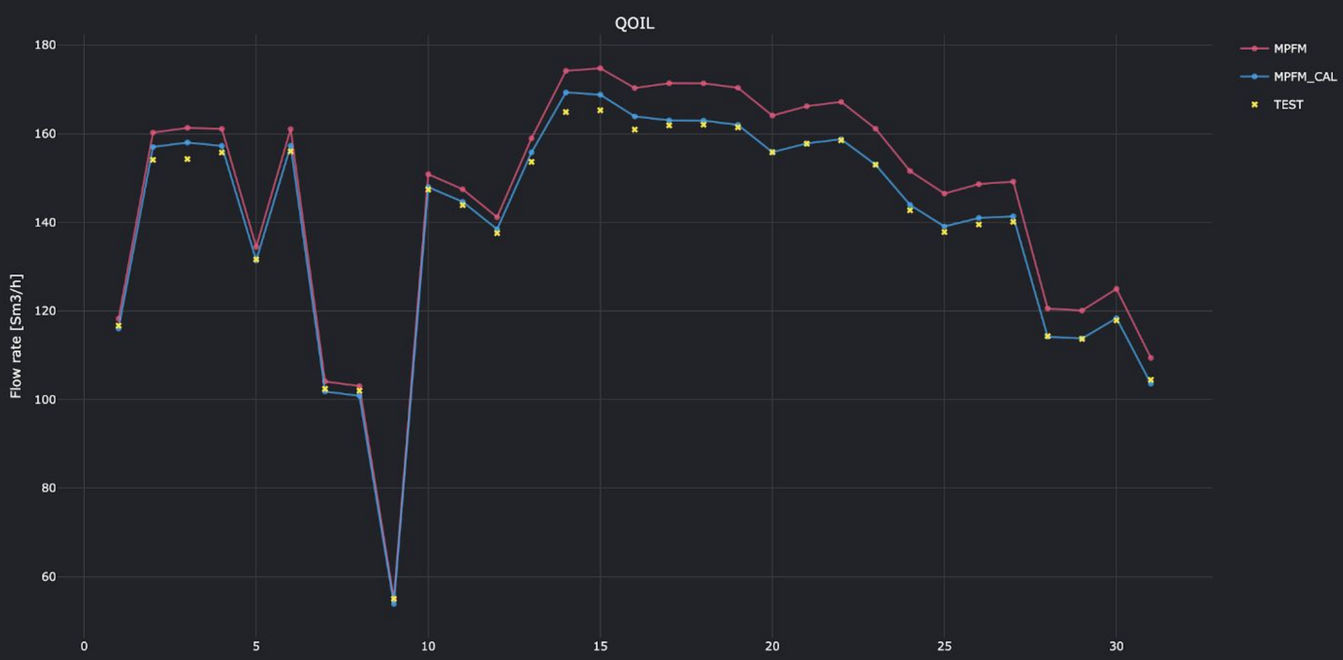
Washy, another algorithmic framework, typically follows downstream Squashy and the Well Test Application. It performs the filtering, imputation, and quality assurance of the statistical data, thereby producing the "end data foundation" - the high-quality cleaned and compressed production and test data that is optimized for the specific asset, well and use case it is intended for. Both NC-VFM and FF work on top of this type of data, although the two technologies often require different sets of working algorithms.

CalPal

Working with sensor data, we consider any measured or estimated sensor signal Y to deviate from the true underlying signal x in two ways: noisy and biased. This is described with the following mathematical model:

$$Y = (1 + a)x + b + \epsilon, \epsilon \sim N(0, \sigma^2)$$

where ϵ is the signal noise, ($1+a$) is a fixed error rate and b is a constant bias. Whereas Squashy is built to handle the noise of the signal, we often apply our technology for virtual calibration, CalPal, to account for the bias. Set in context, it is crucial to train NC-VFM on flow rate measurements that are as correct as possible. For FF, on the other hand, the reconciliation results will be invalidated if the flow estimates are biased. In CalPal, the biased measurements or estimates are adjusted with a multiplicative parameter a and an additive parameter b, per the model formulation stated above, that are learned to fit to historical samples of higher quality, often in the form of well test results. The technology is based on Bayesian linear regression, a probabilistic modeling approach that exploits Bayesian updating and quantifiable measurement uncertainties. The model parameters (a and b) have their own probability distributions, which are automatically updated once a new well test is recorded using Bayesian techniques to account for both recent and historical well test results. CalPal can be used for any flow meter, both virtual and physical, that is infrequently calibrated. An MPFM is therefore a prime example.



The figure above shows an example from a study we did on calibrating measurements from 25 MPFMs on four fields, where we found that the average error was reduced by 9.4%, 14%, and 18% for oil, gas, and water, respectively, and in the best case an error reduction of 57%, 53%, and 62%, for oil, gas, and water, respectively. The pink curve represents the original MPFM measurements, the yellow markers are the well test results for the well in question, and the blue curve represents the virtually calibrated MPFM measurements.

In the figure, we observe that the calibrated MPFM is much closer to the test results compared to the original MPFM. It can also be seen, in particular towards the middle of the plot, that the calibrated MPFM does not fine-tune itself immediately against a new test - a consequence of the Bayesian updating that accounts for the test history. The sensitivity of the calibration is flexible and can be modified to be more or less aggressive towards newer well tests. It is, however, vital that we balance this trade-off, often called the bias-variance trade-off in the fields of statistics and ML, correctly. After all, even samples from well tests can be erroneous.

57%

Best case error reduction with MPFM calibration for an oil rate measurement

4 USING MACHINE LEARNING TO ESTIMATE FLOW RATES IN REAL-TIME

Using machine learning to estimate multiphase flow rates in production systems, i.e. virtual flow metering, is not a new concept and there is much literature on the topic. Yet, even though studies show success, the adoption rate of data-driven virtual flow meters in the industry is low (Balaji et al. 2018), likely due to the problem of overcoming the data challenges introduced in Section 3. On the other hand, there has been an emergent trend in O&G companies of machine learning utilization to aid in estimation of multiphase flow rates in production systems, for instance, Kongsberg Digital (Knight 2019). Nevertheless, we claim to have the first commercially available, data-driven VFM: NeuralCompass VFM, a technology developed from years of heavy research on the topic. In this section, we will delve into the state-of-the-art machine learning in NeuralCompass and explain how it tackles the data challenges of virtual flow metering. We provide argumentation of why we believe NeuralCompass VFM can offer you the best value flow rates existing in the industry. The gory, detailed theoretical equations and theorems behind are spared as there are enough books on such topics, but we will provide some useful sources for the interested reader.

NeuralCompass is state-of-the art machine learning developed in-house to tackle typical data challenges in machine learning. Our commercially available VFM is based on this technology.

Back-to-basic

Let us take a few steps back and explain some basic machine learning topics for those of you that are not as deep into it as we are. Feel free to skip this section if you are familiar with the topic. As already mentioned, machine learning models, or data-driven models, are learned solely based on data. We typically gather the data in a dataset where we differentiate between explanatory variables and target variables. The target variable is what we are trying to estimate. We also refer to this as the output of the machine learning model. For virtual flow metering, the output is typically the flow rate from individual wells. The explanatory variables are other measurements that are believed to be the cause of any change to the target. In a production system, these are choke openings, pressures, temperatures and so on. The explanatory variables are also called inputs. We refer to the training data when we talk about the dataset the model is trained on, typically, historical production data. The test data is data the model has not seen before which we will use the model to give estimations, for instance, future data.

There exist many types of data-driven models. NeuralCompass is a neural network based model trained using supervised learning algorithms. In short, a neural network is a high-dimensional nonlinear regression model. To understand what this means, recall back to the mathematics class and the equation of a straight line, or a linear graph if you like. This line has two parameters: *a* and *b*.

$$y = ax + b$$

The cost *y* of a taxi-ride can be described with this model. *b* is the cost of booking the taxi, while *a* is the amount you have to pay for each *x* kilometer you drive. In the taxi-ride problem, we know the *a* and *b* beforehand, but for other problems we do not necessarily know these. The taxi-ride problem is a one-dimensional regression problem because the cost is a function of only one variable, the number of kilometers. However, if you knew the problem to be dependent on more than one variable, there would be multiple *x*'s and we would need multiple *a*'s to describe the relationship.


The Taxi Ride

For virtual flow metering the problem is high-dimensional as we know the flow rate can be dependent on the pressures, temperatures, and choke opening. We also know that the relationship is nonlinear, in contrast to the taxi-ride problem, and a high-dimensional nonlinear regression model is therefore suitable. As you may have guessed, neural networks are of much higher complexity than the linear model for the taxi-ride problem. They have many parameters and it is generally hard to know the values for these beforehand. We therefore need to learn the parameters using the dataset that we have gathered. By using historical data of both the input and the output to find the parameters, we are learning the model using supervised learning. There exist other types of learning algorithms such as unsupervised where only the input data is used, but we will not delve further into these algorithms. For the interested reader, we recommend Goodfellow et al. (2016).

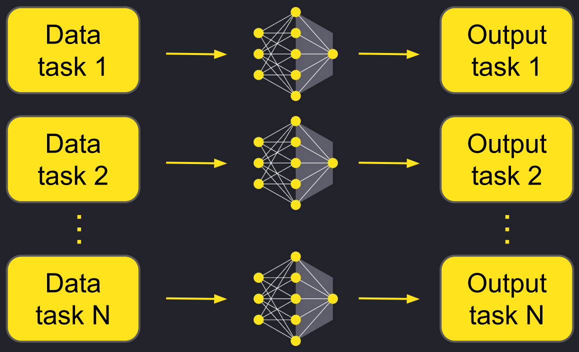
Due to the high complexity of neural networks, these models generally have high adaptivity and arbitrary complex, or even unknown, physical phenomena can be captured by the model as long as these are reflected in the training data. You can imagine that this is highly advantageous if we are trying to represent a complex process where the physics might not be fully understood. Furthermore, even for large models, the utilization in real-time for estimation is typically super quick and can easily provide estimations on a second or minutes basis. For those of you that have burrowed into the physical equations of multiphase flow in pipes and chokes know that to capture the details of the physical behavior, the equations become dirty very quickly and assumptions and simplifications are typically always necessary to obtain a model that can be used in real-time. Naturally, the simplifications impact the accuracy of the estimations.

Of course, the accuracy of the data-driven models depends on the quality and quantity of the training data. For high accuracy, the models must overcome the data challenges described in Section 2. As neural networks can adapt to arbitrary patterns in the data, it can also easily adapt to noise and faulty measurements. This is where our platform for data processing, PCAI, described in Section 3, becomes a crucial step towards NC-VFM providing high accuracy flow rates. Yet, we also need to deal with the three other data challenges. To do this, NC is not just the ordinary neural network you can find in most school books on machine learning, but a variant exploiting *multi-task* learning.

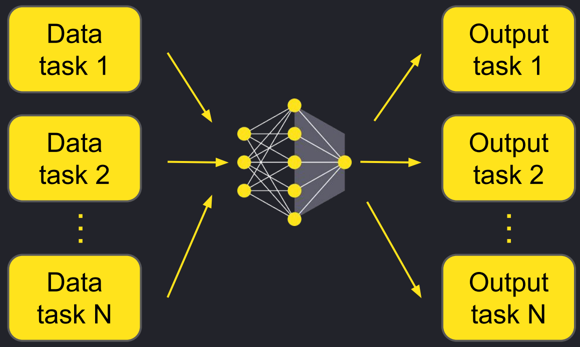
Multi-task learning

Multi-task learning is a machine learning concept where the model is allowed to learn from additional data from similar problems to what we are originally trying to learn. Let us consider the virtual flow metering problem. The typical approach in literature is what we like to refer to as single-task learning. This means, when developing a VFM for a particular well (with this concept, a so-called task), the machine learning model is trained on the historical data of this well only. This is illustrated in top figure to the right. However, you can imagine that for larger assets with more wells, there are other wells that are similar to the well we are developing the VFM for. Multi-task learning believes that data from similar tasks can be exploited to transfer knowledge to the original task. This is rooted in the belief that there is something common for similar tasks and that all tasks benefit from learning from each other. Rewinding to the Toddler problem introduced in Section 2, we can expect the toddler to learn to walk faster by studying other toddlers walking. A similar concept in the machine learning domain is transfer learning, often used in image classification problems. For instance, an image classifier trained to recognize trucks can obtain a higher accuracy by first learning to recognize cars as the tasks are quite similar in many aspects. The difference between transfer learning and multi-task learning is that instead of learning the tasks sequentially, i.e. first learning to recognize cars and then trucks, all tasks are learned simultaneously, as illustrated in bottom right figure.

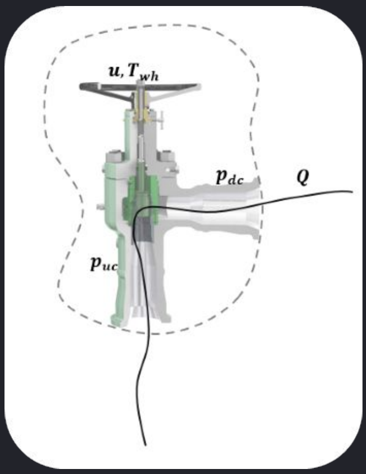
Single-task learning



Multi-task learning



Let us return to the virtual flow metering problem. Assume that we are developing a VFM for multiple wells at the same time. We do know that there are differences between the wells. For instance, they can be completed with different choke valves, or the length of the wellbore can be unequal. They can be located in different parts of a reservoir such that the produced fluids have different compositions of gas, oil, and water. On the other hand, we also know that there are common characteristics too, such as the physical phenomena occurring due to changes in the choke opening or pressures. We can compare this to a very trivial and simplified physical equation for the flow through a choke valve:

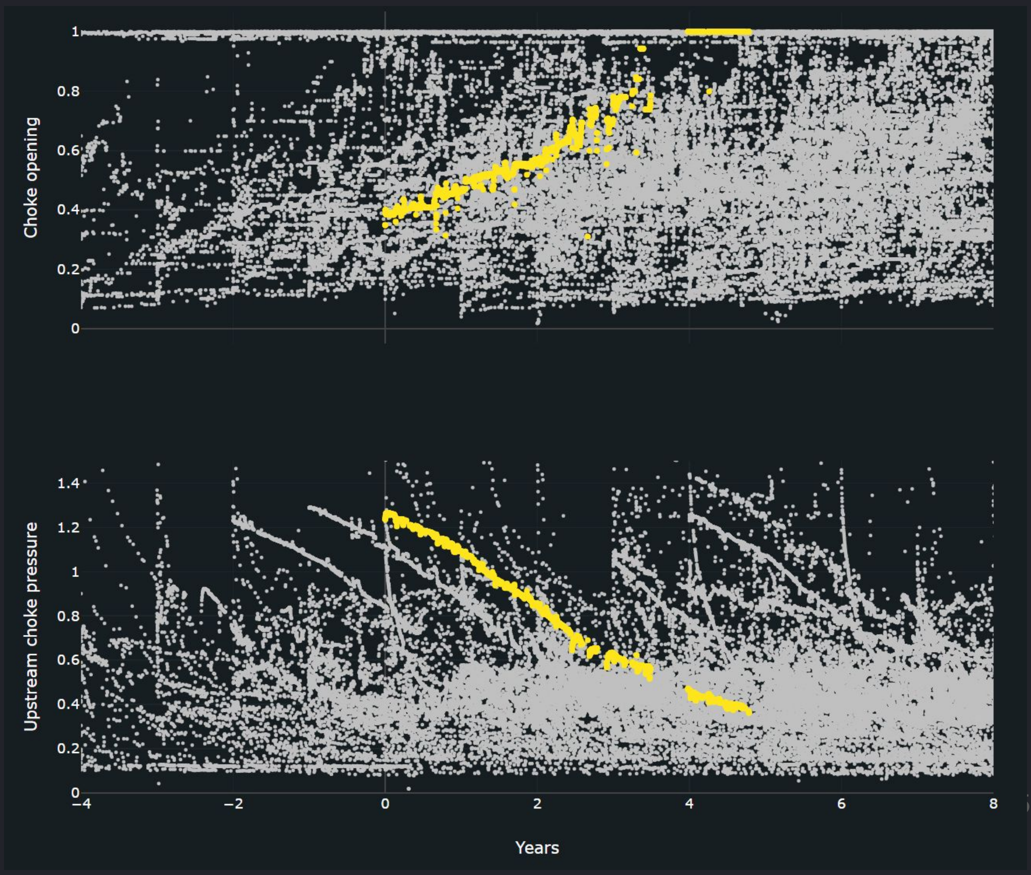


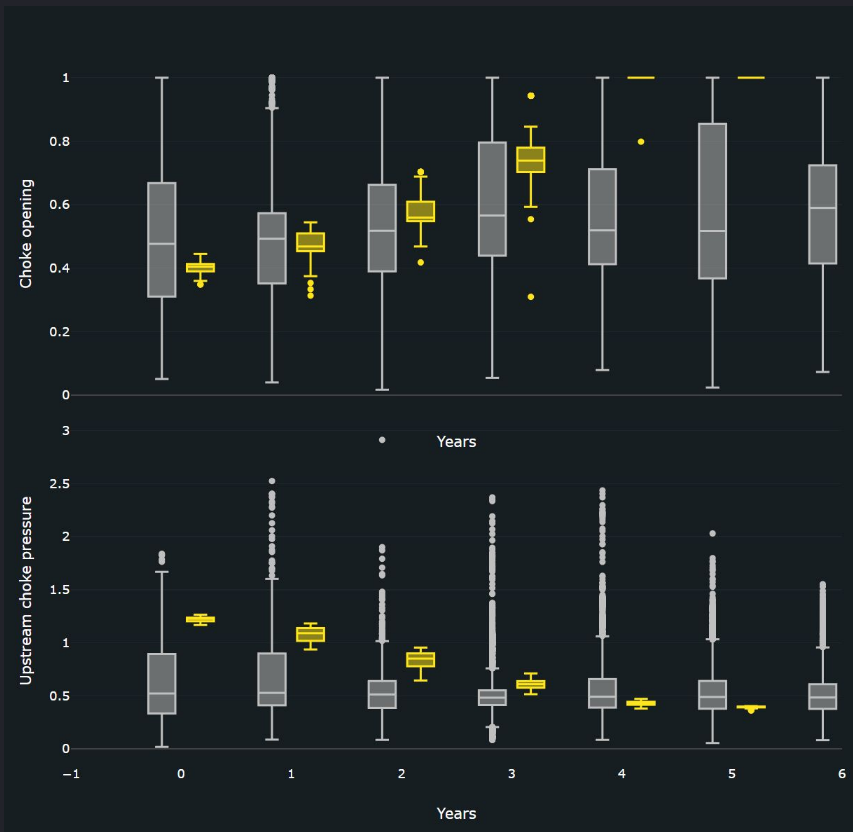
$$Q = CA(u)\sqrt{(\frac{p_{uc} - p_{dc}}{\rho})}$$

where Q is the multiphase flow rate through the choke, the C is a choke valve coefficient, A is the area of the choke, naturally a function of the choke opening u, p are pressures upstream (uc) and downstream (dc) the choke, and ϱ is the density of the produced fluid. The C and ϱ are parameters that are specific to the well we are trying to model. However, that the flow rate varies with the square root of the differential pressure is common for all wells. Using single-task learning, the data for a well must be able to reflect both the common and specific characteristics.

We know on the other hand that the common characteristics should also be reflected in the data for the other wells. By using multi-task learning, we allow all wells to contribute towards learning the common physics of wells. In a sense, considering the simplified physical equation above, all wells will contribute to learning that the flow varies with the square root of the differential pressure. There is no need to learn this separately for all wells. Naturally, the data for the specific well must be used to learn the specific characteristics of that well. Considering the choke valve equation, this could mean the C and the ϱ. Of course, a data-driven model does not have physical equations or parameters like the above equation. Therefore, to differentiate between common and specific characteristics, the MTL models have two sets of data-driven parameters. Going back to the linear regression model for the taxi ride introduced above we can consider trying to make a model that fits all taxi car types. For instance, all taxis can have the same booking cost (b) but the cost per kilometer ride (a) depends on the car type, e.g. limousine or a Ford Galaxy. The common parameter is then the b, and the specific parameter is the a.

Let us look at one example with real data to explain how the data challenges can be tackled by introducing more data into the machine learning problem. In the figure on the right we have plotted the upstream pressure and the choke opening of our well introduced in Section 2, page 7, in yellow, together with several other production wells in gray. Some of the other wells start their production before the yellow well and some start after. Two years into the well's life we have just over 1000 data points for this well. Considering that it is not uncommon to test a well with a frequency of once per month, our yellow well has quite a lot of data points. Remembering that state-of-the-art classification algorithms are trained on millions of data points, the data volume for this well is actually minor.

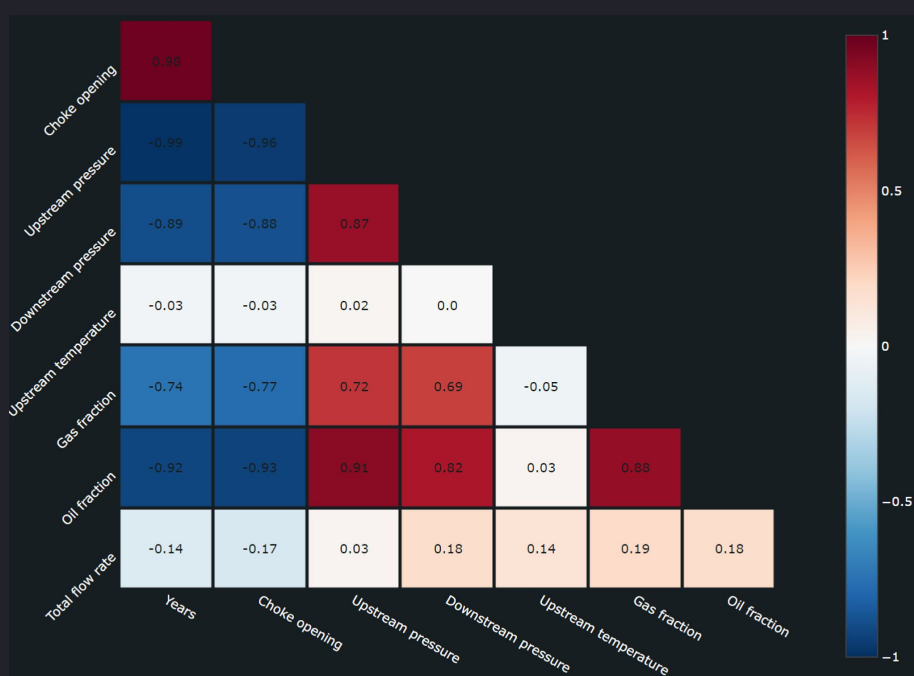




The data points for our yellow well are grouped together in a boxplot fashion on the left illustrating the distribution of the choke opening and the upstream pressure in the five years of the lifetime of our yellow well. The other producing wells during these years are grouped in the gray boxes. The figure shows that the choke opening and pressure data of our yellow well is not completely overlapping during the five years. In other words, trying to estimate the future flow rates for our yellow well, we will always find some input data configurations that our model has never seen before. If we develop a single-task machine learning model with this data, we can expect a relatively high error over time. On the other hand, if we consider the other wells too, we see that the yellow boxes are always contained within the gray boxes. In addition, using the gray wells the data volume based on historical data increases by a factor of 24! By letting our yellow well learn from gray wells in addition to its own data, it is clear that both the data volume and the variety increases significantly.

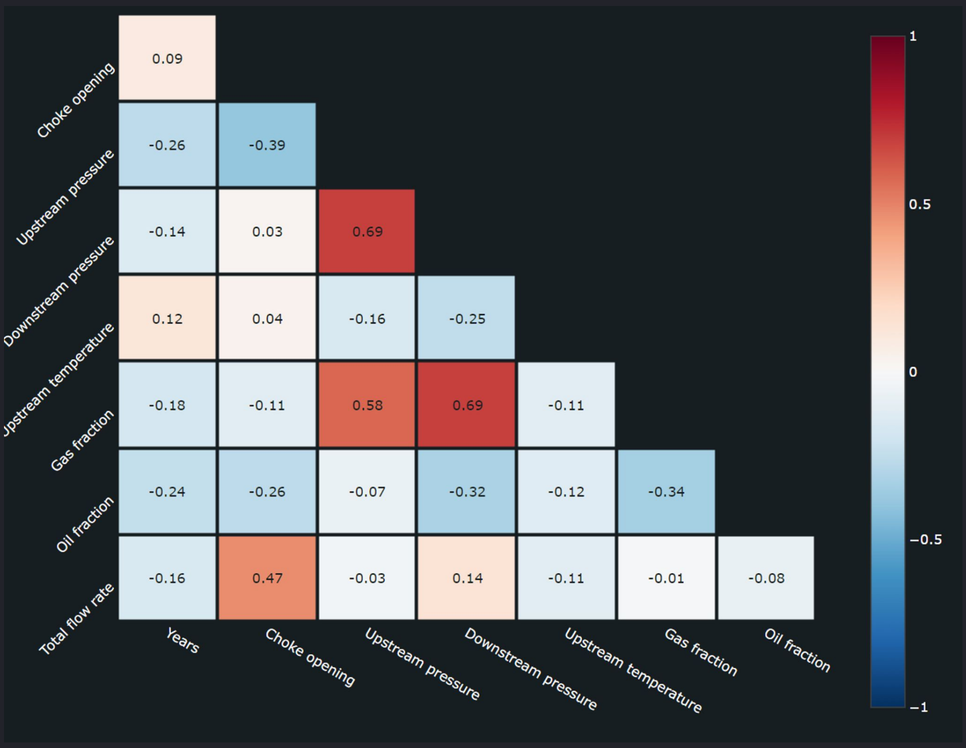
Now let us have a look at how data from several wells can positively influence the correlation between the variables which is caused by the fourth data challenge, nonstationarity. For this we have used our data to calculate a correlation matrix. The matrix will tell us how strongly two variables are correlated and if they are negatively or positively correlated. First, the figure below shows the correlation matrix for our yellow well. As mentioned in Section 2, we should expect an increase in the flow rate for increasing choke opening, e.g. a positive correlation.

Yet, we see the opposite correlation in the bottom row and second value. Naturally, this is due to the drop in production rate due to reservoir depletion, even though the choke opening is opened to try to compensate for the decreasing production. Furthermore, as mentioned in Section 2, the typical strong negative correlation between the choke opening and the gas and oil fraction. In other words, our data tells us that if you open the choke, you will produce less oil and gas. Also notice the strong negative correlation between the choke opening and the upstream pressure. This is a human induced correlation that happens as a consequence of the relationship between the choke opening and pressure. A last note is the very strong correlations between several of our explanatory variables and time, clearly indicating the issue with nonstationarity.



In the figure on the left, we show the correlation matrix for all our data points. First of all, notice the reduced correlation between our explanatory variables and time. We also notice the now positive correlation between the flow rate and the choke opening, and the reduced correlation between the choke opening and the upstream pressure, the gas fraction, and the oil fraction. With this figure, we clearly see the benefit of learning from more data when it comes to tackling the fourth data challenge.

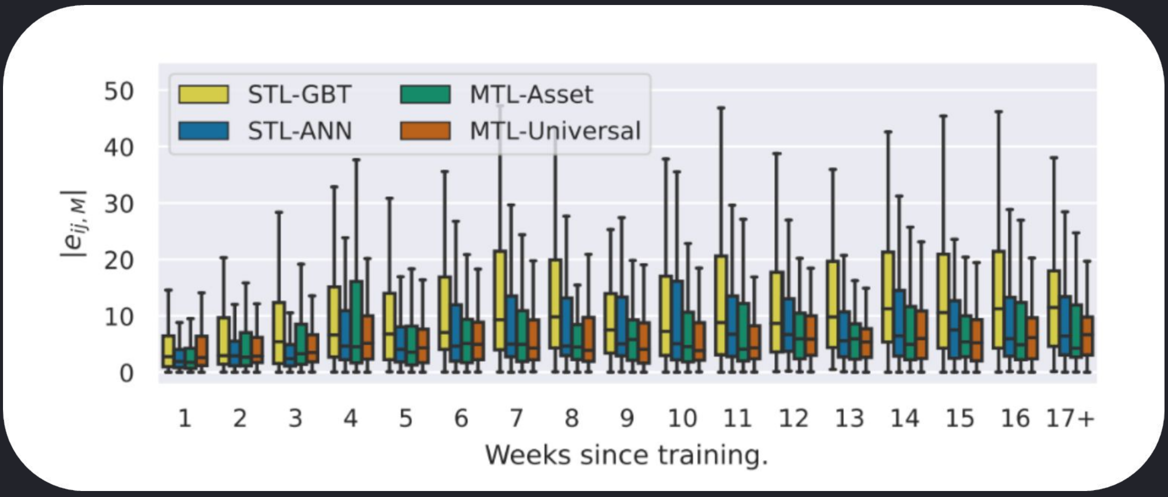
Now that we have demonstrated the positive impact that learning from more data will have on our machine learning model with regard to data challenges, it is time to also give some evidence.



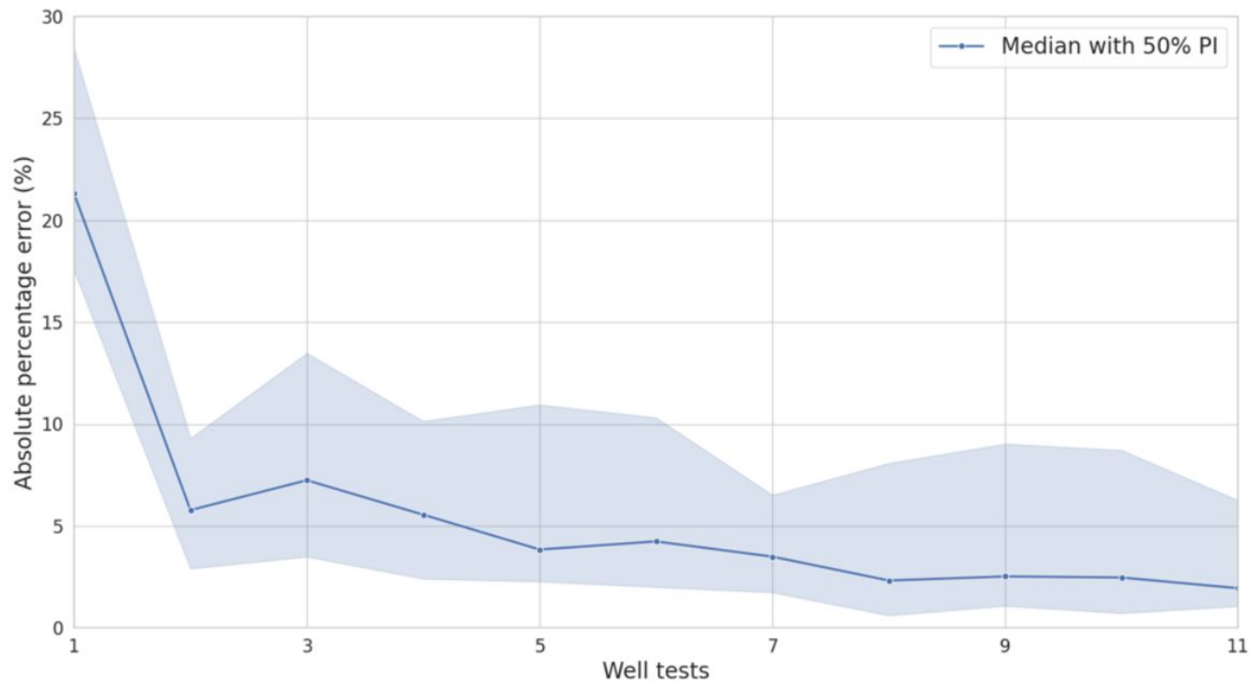
In a paper from Sandnes et al. (2021), two single-task learning models were compared to two multi-task learning models on the problem of virtual flow metering. Data from four different assets were used in the study. The table below summarizes the performance results in terms of the mean absolute percentage error (APE) across the 55 wells in the dataset.

Model	A.1	A.2	A-3	A.4
STL-GBT	15.6	13.5	10.4	18.3
STL-ANN	10.9	13.8	5.9	10.5
MTL-Asset	8.1	10.2	6.5	7.9
MTL-Universal	7.3	11.3	5.7	4.9

The STL-GBT is a single-task gradient-boosted-tree and the STL-ANN is a single-task artificial neural network. Both of these have been trained on data from one well at the time. The MTL-Asset is trained on data from wells within one asset while the MTL-Universal is trained on all data from all wells on the four assets. It is clear that the two MTL approaches are better than the STL approaches. The paper also looked into the predictive capabilities of the different models over time. This is shown in the figure below that illustrates the distribution of well MAPEs for the four different wells as a function of weeks since the model was trained. Also here, is it shown that the two MTL approaches are superior when time increases and is therefore better to handle the nonstationarity of the process.



We also want to demonstrate the effect that MTL can have on wells with few observations when developing VFMs. These can be infrequently tested wells or recently started wells. We did a study in-house where we looked at how many observations were needed for a well before the performance error increased to a low and acceptable level. The study used data from 80 wells. We used data from 60 wells to train our initial model, and the data from the 20 remaining wells to study the performance for wells with few observations. For the 20 wells, we introduced one new observation at a time. We used the model to calculate the error before we allowed it to learn from this new observation. The figure presents the results where the solid line is the median APE across the 20 wells and the shaded region is the 50% percentile interval, meaning that 50% of the wells have an error within this area. In the figure, for the first well test that is conducted the error shown is the error of an uncalibrated model. In other words, the model has never seen any observations for the well before. For the second well test, the model has been allowed to learn from one observation, namely the first well test, for each of the 20 wells. For well test number three the model has been trained on the two existing tests for the 20 wells and so on. The results show that already after one observation seen by the model it is able to offer a very good error, and that the error keeps decreasing as more tests are performed and trained on.

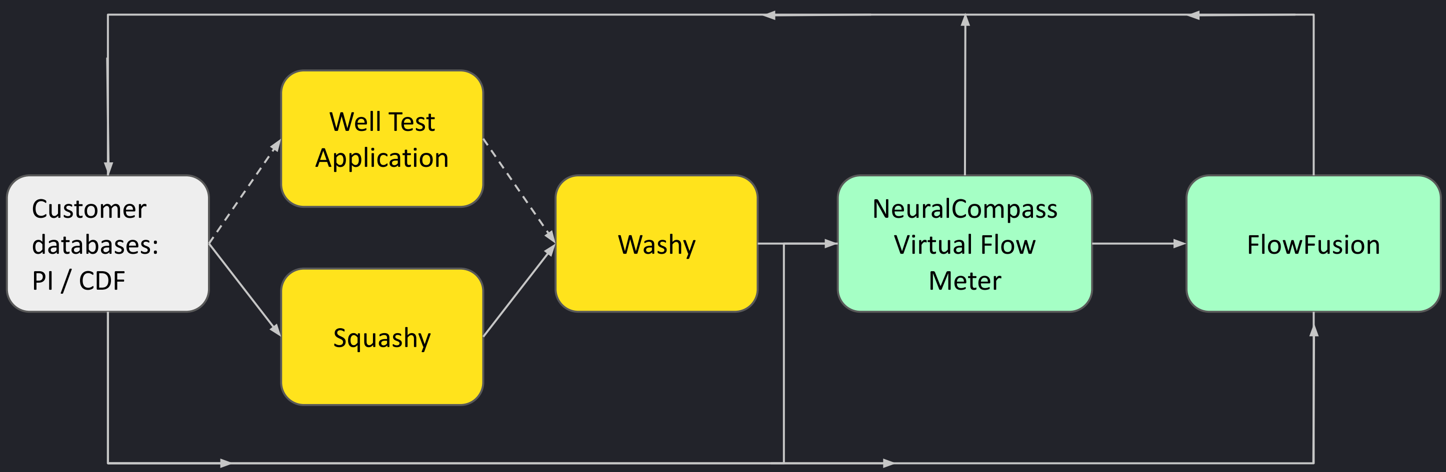


As you have seen, the MTL is superior to STL for virtual flow metering and is a critical approach to tackle the data challenges introduced in Section 2. However, theory and controlled studies in research is one thing, a completely different aspect is how to go from theory to practice when applying machine learning in real-time industrial scenarios. The next Section will describe NC-VFM, our real-time service that offers AI-powered continuous estimates of the flow rates for production wells.

NeuralCompass Virtual Flow Meter

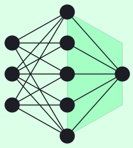
Based on the two preceding sections, you might have guessed that NC-VFM applies a multi-task learning approach to deliver virtual flow metering of production wells. Yet, to deliver flow rates of the best value to you over time, it is not enough to deliver a one-time trained machine learning model and leave you to figure out the rest. No, NC-VFM is an AI-as-a-service with high interoperability that offers frequently and automatically calibrated flow rate estimates with high quality over time. Let us explain.

With PCAI, whose parts are introduced in Section 3, we connect to your databases to read the required measurements for training and maintaining our machine learning model, and to calculate real-time continuous flow rate estimates. As explained in Section 3, if you choose to use our Well Test Application, NC-VFM will seamlessly be integrated with this application. Through PCAI Squashy and Washy, we make sure that we extract high-quality data when developing the VFM. The calculated flow rates are written back to the databases and can be provided at a high frequency, for instance, on a per minute basis. If desired, PCAI can also offer the flow rates at a lower frequency, for instance every sixth hour, or as a daily average to be used in allocation. Every second week, we do maintenance of NC-VFM with data quality checks and automatic calibration extracting the most recent data we have for our wells. Our maintenance framework ensures up to date models and increases the volume and variety of our data foundation. In addition, we ensure to capture any new well behavior, for instance, due to the nonstationarity of the process. The pipeline from data streams to flow rates is illustrated below. Here we also imply that the flow rates from NC-VFM can be used further into FF for allocation, which will be explained in the next section.

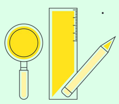


With the setup illustrated, you can see that there is little action required from you in order for us to deliver flow rates from NC-VFM, besides granting access to read relevant data streams. NC-VFM is not dependent on either PVT tables, fluid properties, or geometric data for different wells. This, along with the beauty of the MTL architecture allowing us to model several wells at the same time, NC-VFM has high scalability and is very quickly set up for many wells. In fact, once the required data is in place, it is a matter of hours before we can write back the flow rate for a well to you. With the automatic calibration routine enabled by our AI-powered algorithms, NC-VFM is also highly maintainable, and it can deliver calibrated, high-quality flow rates for hundreds of wells with zero maintenance efforts from your side. As NC-VFM is offered as-a-service, there will be no surprise bills if it turns out that some wells require a bit more love and care in maintenance than the standard well. We also want to point out that NC-VFM is continuously improving for every new data point we achieve for customers. We would claim that with the high scalability and maintainability, NC-VFM is a cost-efficient solution to virtual flow metering. For those of you who are interested in numbers, we have delivered flow rates for 2000 wells to 13 operators.

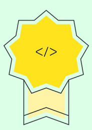
NC-VFM as-a-service



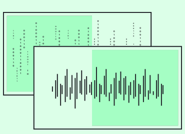
STATE OF THE ART
One single physics inspired machine learning model able to learn from data gathered across the world



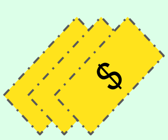
SCIENTIFIC EXCELLENCE
Application is built on ten years of scientific research



CONTINUOUSLY IMPROVING
Every data point from any customer improves every future estimate for every customer



HIGH INTEROPERABILITY
Fetching data to automatically calibrate NC-VFM gives you high accuracy, up to date, flow rate estimations every day with 90% less effort



COST-EFFICIENT
As a highly scalable and low maintenance solution, NC-VFM is a cost-efficient solution to VFM. Offered as-a-service, there are no surprise bills.

NC-VFM real-life results

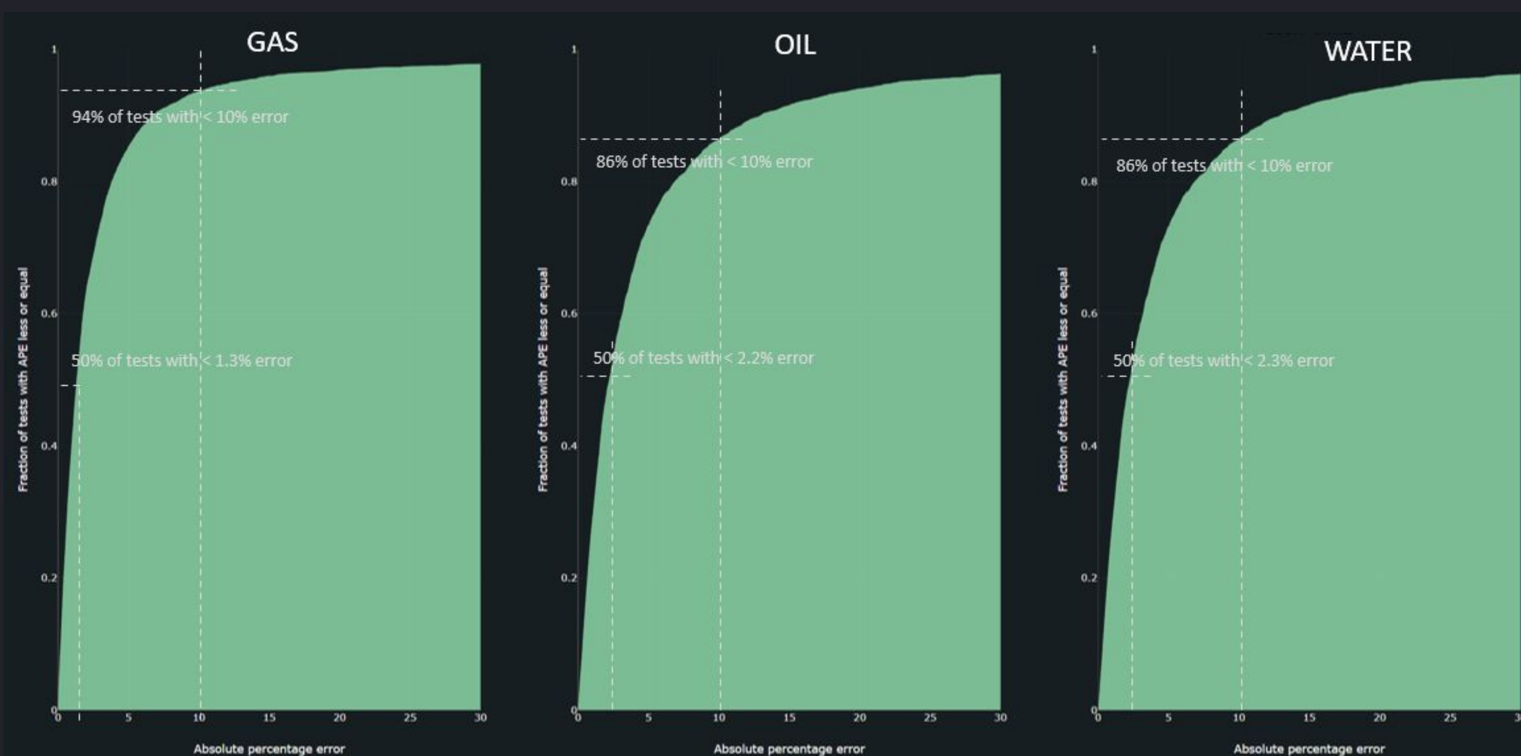
94%

NC-VFM average performance for oil flow rate

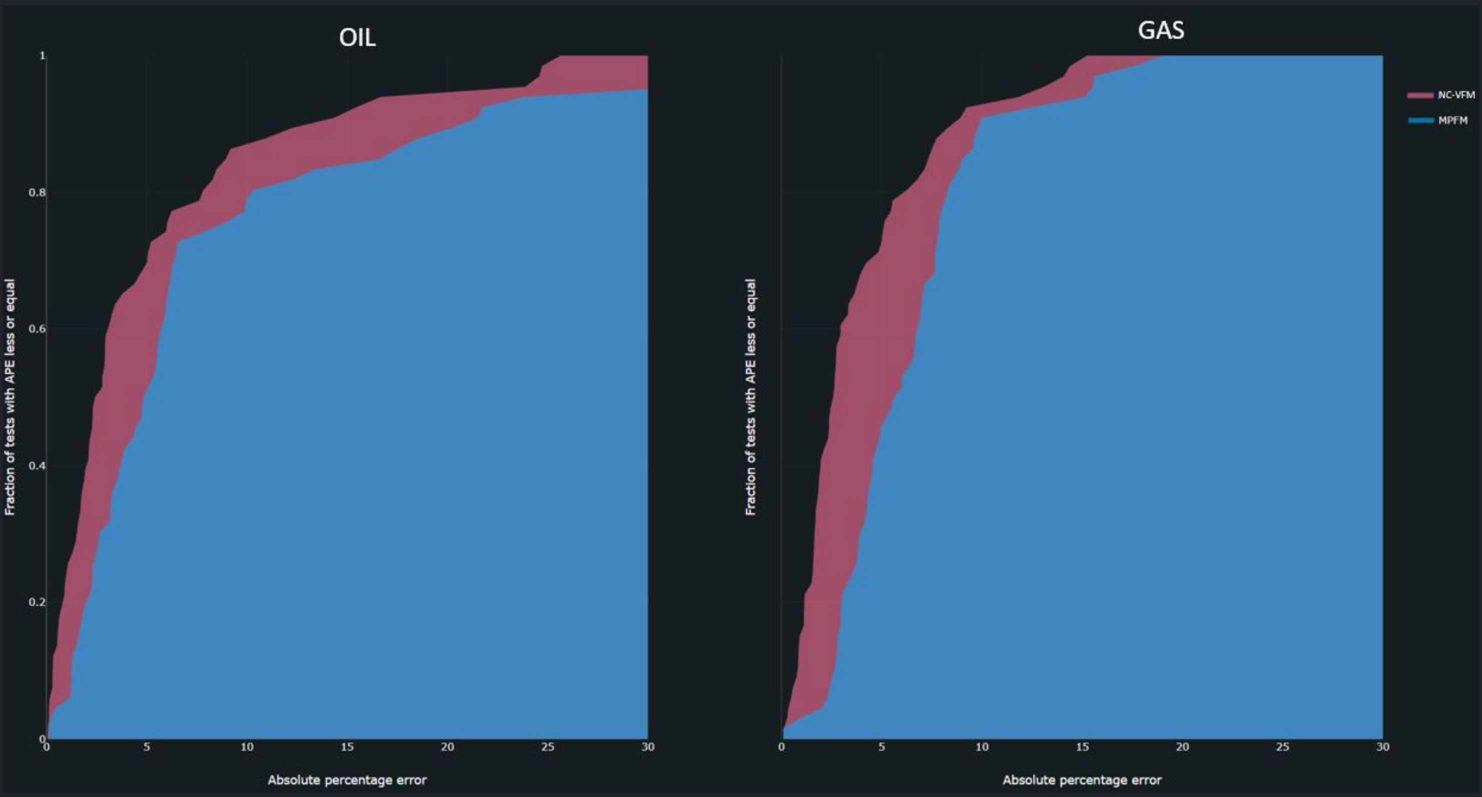
It is time to give you some intel regarding the current performance of NC-VFM for the wells that we deliver flow rates for. We have therefore calculated the test performance for the first quadrimester of 2023. We do highlight that this is the test performance, which means, as introduced earlier in this section, that the model has not seen the data before we use it for estimation. We have calculated the test performance in a biweekly fashion to obtain a realistic performance metric. This means, before every biweekly maintenance routine, we use the model to estimate the flow rate with the newest input data for our wells, and compare the estimation with the measured flow rate, such as well tests. Thereafter, we use the new data in model calibration. NC-VFM achieved an average performance across the wells for the first quadrimester of 2023 of 96.3%, 93.5%, and 93.4% for gas, oil, and water rate estimates, respectively.

These numbers correspond to an error of 3.7%, 6.5% and 6.6% respectively. It is worth mentioning that typical errors of MPFM reported in data sheets from manufacturers are reported to be between 5%-10% (Corneliussen 2005). Keep in mind that the performance is calculated by comparing our estimates to measurements that are themselves uncertain. Generally, well tests measured topside typically have a low uncertainty, some as low as 1%, but it is not uncommon that well tests can also have an uncertainty of say 5%, depending on the measurement technique, flow regime and more. Naturally, the higher the uncertainty of the tests, the more difficult it is to report NC-VFM performance as what we are comparing against cannot be considered certain.

Another way to illustrate the performance of NC-VFM is with a cumulative performance plot as shown below. This figure illustrates the fraction of tests that achieve an error of less than a specific threshold. For example, to find the median error across the wells we can read off the error on the x-axis for the fraction of tests equal to 0.5 (or 50% of the tests). We see from the figure that the median error is 1.3%, 2.2% and 2.3% for gas, oil, and water respectively, meaning a median performance accuracy of 98.7%, 97.8% and 97.7%. Further, we see that 94%, 86%, and 86% of all the tests had an error less than or equal to 10%.



If we compare the performance (against well tests) of the MPFM and NC-VFM for the wells that also have an MPFM available we get the cumulative performance shown on the next page. Reading the value on the x-axis for 50% of the tests, the median error for NC-VFM is 2.4% and 2.6% for oil and gas, respectively, while for the MPFM the median error is 4.8% and 5.6%.



There are several usages of NC-VFM that we also offer as services in Solution Seeker. The continuous flow rate estimates can be applied in well monitoring, providing situational awareness for engineers in between well tests. NC-VFM can be used as a stand-alone product or as a back-up system to an MPFM. The flow rates are also appropriate to apply in downstream applications such as allocation, production optimization, or reservoir modeling. In the following section we will deep-dive into our allocation service, FF. Although we recommend using NC-VFM into FF, the service is fully operable without NC-VFM using rate estimates that you already have access to such as MPFMs or other VFMs.

5 A DATA-DRIVEN APPROACH TO SMARTER ALLOCATION AND GROSS ERROR DETECTION

Due to increasingly complex production systems, often involving multiple stakeholders sharing the same pipelines and production facilities, it is also becoming progressively urgent to develop more accurate and defensible allocation methods (Kanshio, 2020). The bottomline is that the various stakeholders require their fair share of the sales revenue. It is also considered widely important to have a precise perception of the oil and gas ownership for the purpose of production optimization (Hanssen and Foss, 2015). Otherwise, optimal control settings can potentially drown among the large measurement uncertainties that are associated with MPFMs and VFMs that are infrequently updated. Furthermore, wrong well allocation leads to uncertain reservoir simulation which can negatively affect drawdown strategy and placement of new wells. In fact, given that production allocation is often based on flow rate estimates from such meters, many consider measurement errors, i.e. poor data quality, to be the fundamental issue of allocation (Pobitzer et al., 2016). As it is highly costly, and even physically impossible in some cases, to uphold a valid flow meter performance for every production system operating point of interest, there is a growing demand for intelligent, digital solutions (Badings and van Putten 2015). This is where FlowFusion comes into the picture - an optimization-based allocation framework that delivers industry-approved flow rate estimates and error detections for measurements that are performing off-spec. In the following sections, we provide a run-through of some practiced allocation methods in the industry along with their limitations, as well as a short introduction to the optimization method FF is based on, namely data validation and reconciliation (DVR). We then proceed to explain the different modules of FF and why you too should engage in the industry-applied framework. To emphasize, you are allowed to skip to the good part of this section if the introductory topics are already known or otherwise not your cup of tea for the time being.

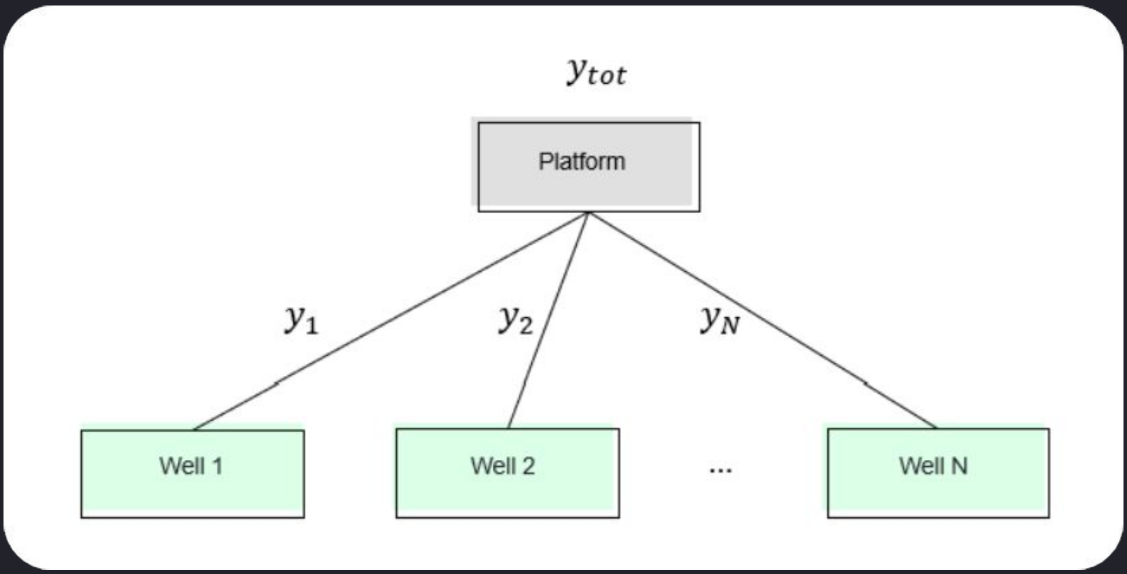
Conventional allocation methods

In the industry today, it is most common to use proportional allocation methods, such as by-difference allocation (BDA) and pro-rata allocation (PRA), to assign the oil and gas ownership throughout the production systems (Amin, 2016). In a simplified production system as that illustrated below, we define y_i^*, y_i and $\hat{y}_i$ to be the true flow rate, a single flow rate sensor reading, e.g. physical or virtual, and the allocated (or adjusted) estimate, respectively, at any given time for an arbitrary well *i* in a set of *N* wells. y_{tot}^* and y_{tot} represent the actual and measured total flow rates, respectively. For the time being these are assumed to be equal, both for simplicity and to account for industry practice. Hence, we have the following relationships:

$$y_{tot}^* = y_{tot} = \sum_{i=1}^{N} y_i + \epsilon_{tot}$$

where ϵ_{tot} is the total imbalance, or mismatch, that needs to be distributed to different origins. Slightly adjusting the above equation, ϵ_{tot} can be derived as follows:

$$\epsilon_{tot} = y_{tot} - \sum_{i=1}^{N} y_i$$



In BDA, the mismatch is assigned to a single sensor to satisfy the mass balance (Bjørk et al., 2016). Consequently, in our fictional case, the imbalance is attributed in full to a single well *j* and corresponding sensor:

$$\hat{y}_j = y_j + \epsilon_{tot}, \hat{y}_i = y_i \forall i \neq j$$

such that

$$y_{tot} = \sum_{i=1}^{N} \hat{y}_i$$

The fundamental assumption behind this strategy is that only one of the flow meters in the production system exhibits erroneous behavior. However, this is a gross simplification for any production system that involves more than one well, thereby limiting the performance and applicability of such an allocation method. An alternative is to use PRA, where the imbalance is distributed proportionally according to the magnitude of the flow rates (Amin, 2016). For instance, in the simplified example above, the following equations hold:

$$\hat{y}_i = y_i + \frac{y_i}{\sum_{i=1}^{N} y_i}\epsilon_{tot}$$

Although satisfying the mass balance of the production system in a slightly more justifiable fashion than BDA, PRA neglects that metering equipment across the production system almost always involves asymmetric uncertainties due to various reasons, e.g. individual set points for flow volume, malfunctioning instruments, differing multiphase flow dynamics, and asynchronous meter calibrations. Hence, this methodology also falls short in most cases.

Evidently, conventional allocation methods are often unable to fully capture the complexity of the production system at hand, thereby failing to distribute the imbalance aptly and justifiably. As Badings and van Putten (2015) puts it, "the methods are limited in terms of error correction performance, and can only be applied to [a] single-tier system" (p. 2). Not to mention that, in our experience, most applications of BDA and PRA are based on merely one source of flow rate information despite the fact that some assets have multiple flow meters at their disposal. Consequently, there is an industry driven demand for more sophisticated strategies for allocation, such as FF, that can both cope with the complexity and exploit redundant flow rate information.

Data validation and reconciliation

Although having been present in the process industry for a while (Narasimhan and Jordache, 2000; Oliveira and Aguiar, 2009), optimization-based methodologies, such as DVR, have not been applied for production allocation within the petroleum industry before recent times. In the following we describe a steady-state DVR, a model-based optimization method and data correction technique that exploits process data redundancy to mitigate the effects of measurement errors (Bikmukhametov and Jäschke, 2020). One important assumption that is taken in DVR is that the sensor signals have no systematic error, only random error (Oliveira and Aguiar, 2009). An example of a systematic error is if a sensor always measures say 3 too much. In context of the oil and gas sector, the optimization problem for phase *p* can be mathematically formulated as follows:

$$\hat{y}_p = argmin_y\{(y - Cy^*)^T\Sigma^{-1}(y - Cy^*)\}_p : Ay^* = 0$$

which is a constrained weighted least-square optimization problem. The notation $\{.\}_p$ suggests that the variables are given with respect to phase *p*. By assuming that the sensor signals are independent of each other, we can neglect the covariance resulting in a diagonal covariance matrix

$$\Sigma_p = \{diag(\sigma_1^2, \ldots, \sigma_N^2)\}_p$$

where N_p is the number of sensors in the production system and $\sigma_{p,i}$ is the absolute standard deviation of the sensor signal i for phase p (Jiang et al., 2014). A_p is the constraint matrix, which may encode the mass balances. Lastly, C_p is a measurement matrix that is often applied to sample merely the sensor signals that best represent the true flow rates. The steady-state DVR problem that can be solved analytically as shown on the next page:

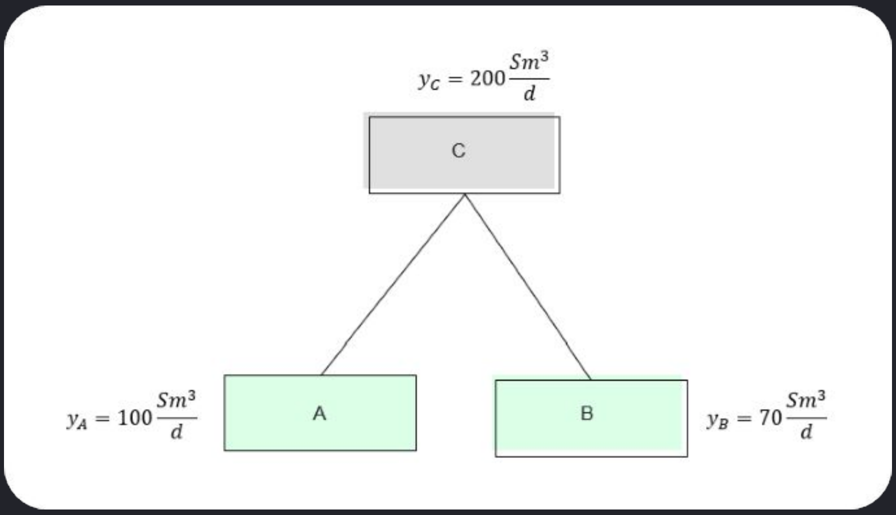
DVR solution:

$$\hat{y}_p = R_p y_p$$

$$R_p = \left[V(I - A^T(AVA^T)^{-1}AV)C^T\Sigma^{-1}\right]_p \qquad V = \left[(C^T\sigma^{-1}C)^{-1}\right]$$

But what does all these equations actually mean? In essence, the objective is to minimize the overall correction that is required for the constraints of the production system to be satisfied. This is however not exactly straightforward from the theory above, so let us delve into a simplified example to demonstrate.

Consider the production from two oil wells: A and B, that are commingled at the manifold C. The flow rates from A and B are measured with MPFMs but the flow rate at C is measured with fiscal metering such that the uncertainty for C can be neglected. The MPFM at A is recently calibrated and the uncertainty is quantified to 5%. The MPFM at B has not been calibrated for a while and the production engineers believe the error is approximately 15%. Let us further assume that at a given day, we are told that the MPFM at A showed that A produced 100Sm3 oil, the MPFM at B showed that B produced 70Sm3 oil, and that the MPFM at C showed 200Sm3 oil. The setup is illustrated in the below figure. From the numbers we see that there is a mismatch between the sum of A and B and the measured rate at C, and that we need to distribute an imbalance of 30Sm3 to the two wells as we believe the device at C is correct. If we considered a BDA approach, we would likely distribute all the imbalance to B as this device has the highest uncertainty. The results would thereby be as given in the table on page 28.



With the PRA approach, we would distribute the 30Sm3 /d according to the wells fraction of distribution to the total. This would lead to:

$$\hat{y}_A = y_A + \frac{y_A}{y_A + y_B}\epsilon_{tot} = 100 + \frac{100}{170} \times 30 = 117.6Sm^3/d$$

$$\hat{y}_B = y_B + \frac{y_B}{y_A + y_B}\epsilon_{tot} = 70 + \frac{70}{170} \times 30 = 82.4Sm^3/d$$

However, as we know that the measurement device of B is more uncertain than the measurement device at A, it does seem more fair that B is distributed more percentwise than A. The DVR approach does this, and we will go through how the problem is set up.

First of all, we need to find the variance matrix, the C matrix and the A matrix. With the uncertainties given, we will have that

$$y_A = 100 \pm 5Sm^3/d \qquad\qquad y_B = 70 \pm 10.5Sm^3/d$$

Assuming that the measurements are normally distributed and that the uncertainty cover a 95% confidence interval, the mean and variance can be approximated to:

$$(\mu_A, \sigma_A^2) = (100, 6.3) \qquad\qquad (\mu_B, \sigma_B^2) = (70, 27.6)$$

Note, in order to solve the problem, the variance of C must be set to a small, negligible value, for instance

$$(\mu_C, \sigma_C^2) = (200, 0.001)$$

The variance matrix, the C matrix, and the A matrix that incorporates the mass balance will become:

$$\Sigma = \begin{bmatrix} 6.3 & 0 & 0 \\ 0 & 27.6 & 0 \\ 0 & 0 & 0.001 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}$$

Using the matrices into the analytical equation of the problem, the solution becomes as given in the table on page 28. We see that compared to the PRA approach, A has gotten less production while B more. This seems fair as the uncertainty of B is much larger than that of A.

The DVR problem does not necessarily need to use only one sensor signal if there are more available. Imagine that the asset has a second sensor for the two wells, for instance a VFM, where both have an uncertainty quantified to 10%. Let's further assume that the two VFM measure 90 Sm3/d and 75Sm3 /d at A and B, respectively.

$$y_A, MPFM = 100 \pm 5Sm^3/d \qquad\qquad y_B, MPFM = 70 \pm 10.5Sm^3/d$$

$$y_A, VFM = 90 \pm 9Sm^3/d \qquad\qquad y_B, VFM = 75 \pm 7.5Sm^3/d$$

The matrices in the DVR problem would then become:

$$\Sigma = \begin{bmatrix} 6.3 & 0 & 0 & 0 & 0 \\ 0 & 20.3 & 0 & 0 & 0 \\ 0 & 0 & 27.6 & 0 & 0 \\ 0 & 0 & 0 & 14.1 & 0 \\ 0 & 0 & & & \\ 0 & 0 & 0 & 0 & 0.001 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad A = \begin{bmatrix} 1 & 1 & -1 \end{bmatrix}$$

and the resulting allocation results as given in the table on page 28. It does make sense to distribute less to B than with using DVR with one measurement as the difference between the two sensor readings are less than for A. Correspondingly, A must be given somewhat more of the imbalance.

Of course, the example shown here is very trivial and the production system is typically far more complex. Additionally, systematic error like bias in the sensor signals or correlations between meters is very common. Having systematic error on our sensor signals will result in the solution from DVR converging to a false solution even if the sensor uncertainties are known (Narasimhan and Jordache, 2000). We also never know exactly what the uncertainty of the sensors are at any given time, which means that the uncertainties must be estimated.

So what does FF bring to the table that is different from the general DVR problem described in this section? That is what we will demonstrate in the coming section.

Method	Allocation results [Sm3]		
	A	B	C
BDA	100	100	200
PRA	117.6	82.4	200
DVR (one measurement)	105.5	94.5	200
DVR (two measurements)	107.5	92.5	200

FlowFusion

FF is our industry-tested software service for reconciliation and allocation. As previously mentioned, it is a fully data-driven approach that exploits the information that lies in production data such as quantifiable uncertainties, well tests, redundant flow rate data, and detectable MPFM or VFM errors. The main methodology used for reconciliation is the data validation and reconciliation described in the previous section, yet, FF is also so much more than just the reconciliation method.

We generally divide FF into four modules as illustrated to the right: data processing, uncertainty estimation, the reconciliation problem, and gross error detection. We will go through the four modules and explain how they together give you high-quality rates for allocation.

Data processing

The first module is about data processing and is essential to tackle poor data quality and ensure realistic and accurate rate estimates. The module takes advantage of Squashy and Washy, introduced in Section 3, to extract high-quality data with reduced influence of noise from our databases. For instance, to lessen uncertainties related to process dynamics such that the mass balance assumption of the system upholds, it is important to use averages of the sensor signals over a certain time period.

With Squashy we can calculate the FF rates for any given time period, for instance every steady-state period or for fixed time periods of say 6 hours, 12 hours, or 24 hours. For daily allocation, the natural choice is 24 hours. Together with engineers we make sure to choose a time period that makes sense for the production system in question (for example with regards to its transportation delays) and the way the flow rates are supposed to be applied.

Furthermore, as previously mentioned, one of the assumptions of the DVR method is that the sensor signals used in reconciliation are unbiased. We know that this is rarely the case and that sensors can drift in time or even be defect for longer periods, which will cause the reconciliation problem to converge to faulty solutions. The data processing module therefore incorporates additional methodologies to tackle biased sensor signals. First, an additional set of logic is incorporated to automatically remove nonphysical behavior. For instance, if the flow rate data shows a positive flow when the well is supposed to be closed or if the topside measurement shows zero flow when some or all of the wells are *seemingly* producing, we can manipulate the data such that it makes physical sense. These types of errors, that traditionally demand human intervention and tedious hours - at least if accumulated - where you manually overwrite the data, are handled automatically in the FF framework. Secondly, we are able to exploit two algorithms that detect if the flow rate data has frozen or become “stray”, meaning that the signal has fallen out and shows an obviously faulty behavior for some time. The figure below illustrates a period where one of the sensor signals has frozen, in which case it is removed altogether from the reconciliation problem in that specific period. As can be guessed, these algorithms can only be applied if there are several flow rate sensors per well, such that it is possible to fall back on other sources of information. It is also acknowledged, and quite humbly at that, that different assets have different challenges with their flow rate measurements or estimates. Hence, the development of sound error detection algorithms is very much a topic to be discussed with experienced production engineers.

Lastly, if the original sensor signal is suspected to be biased, CalPal, described in Section 3, is applied to calibrate the rates. This virtual calibration has the potential to remove the bias from the measurements without having to completely remove the sensors from the reconciliation problem. As a consequence, this technique is especially beneficial for production systems where there is only one sensor for each well. As shown in Section 3, a study we did on calibrating MPFMs showed an average reduction in error (when compared to well tests) of 9.4% for oil. However, it also demonstrated a whopping 57% in a best case scenario. All in all, the complete set of tools in the data processing module ensures a more sound and realistic starting point for the reconciliation and greatly increases the accuracy of the calculated reconciled rates.

Uncertainty estimation

The second module involves estimating the uncertainty of the different flow rate signals that are used in the reconciliation problem. As mentioned earlier, the uncertainties are rarely known exactly and are usually be guesstimated together with production engineers that know the wells. In FF, the uncertainty estimation module automates the process by collecting information from well tests. These are extracted either from our Well Test Application or by connecting to your databases to automatically read the well test results.

Of course, the module also allows for qualified subjective assumptions made by the engineers if well tests do not suffice or are otherwise not available. In the end, the uncertainty *or variance* estimates we calculate are live and condition-based, and considers both the level of noise in the signal that we get from Squashy, see Section 3, the recent accuracy compared to high-quality well tests and subjective assessments if so desirable.

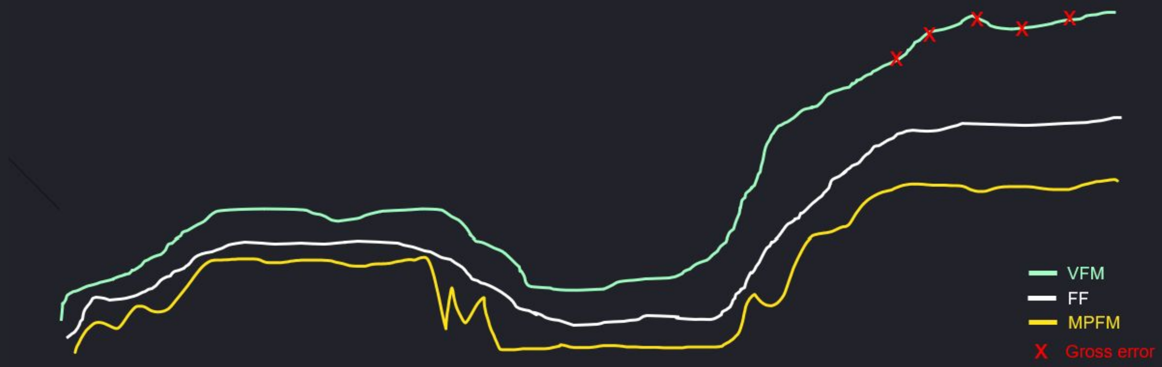
The figure below illustrates this with a made-up, yet representative, example. The white bars indicate the time periods in which a certain well was tested and the corresponding mean values of the flow rate signal at those times. The green line indicates the calculated uncertainty of the MPFM signal. For the first test, the MPFM is far away from the value of the test and the uncertainty will increase. In the next period we see that the MPFM signal has a higher fluctuation than previously, and the uncertainty increases somewhat more. However, during the second test we see that the MPFM is almost spot on and the uncertainty will therefore decrease. This automated uncertainty quantification is used for all the sensor signals in the reconciliation problem.

The third module is the reconciliation, which solves the DVR problem introduced above for whichever desired time period and production system setup given the validated data from the “data processing” step and the variances calculated in the “uncertainty estimation” phase. FF exploits all flow rate information available for the asset, whether it be one or five sensors. The FF framework also supports asymmetrical setups, meaning that a certain well can have one sensor while another can have five.

As a result you obtain the most likely (average) flow rate for each node in the production system, that, together, also fulfill any mass balances that are specified. However, in addition to the flow rates themselves, you also get their associated uncertainties, which can be further exploited to determine how trustworthy they are and where you perhaps want to place the flow rates yourself. Typically, if multiple sensor signals for a well are closer together, the uncertainty related to its FF rates are often lower and vice versa.

The last module, gross error detection (GED), can be used to warn about abnormalities related to any of the sensors, for instance, drifting sensors. Hence, it is a different means to identify biased sensors and possibly rule them out of the reconciliation problem (in a repeated run of the problem). By using statistical tests, the individual sensor measurements that are input to the reconciliation problem are compared to the calculated FF rates, resulting in GEDs should the deviations be too high.

This is illustrated in the simplified figure below. The white FF rate is the outcome of the reconciliation problem for a specific well, considering both the uncertainties of the input sensor signals and the mass balance it is supposed to satisfy. We see that the reconciled flow rates are in proximity to both of the input sensors in the beginning, but towards the end we observe that the green VFM is drifting away from the FF rates and that, consequently, it is detected as being out-of-spec. Given that the GEDs are correctly activated, it could then be beneficial to recalculate the FF rates in that specific time period without the drifting estimates of the VFM in the problem.



Naturally, GED can occur for several reasons and it is important to keep in mind that it may not always be the individual rate sensors that are faulty - it could also be the topside measurement we are trying to match against that is erroneous. In such a case, we often obtain out-of-spec detections for multiple sensors, thereby making it harder to know when to rule out sensor signals and not. In theoretical terms, FF applies a maximum power measurement test for the GEDs. We will spare you the mathematical details of this type of test, but for the interested reader we refer to Narasimhan and Jordache (2000). A disadvantage of GED is that sensor redundancy is required to fully utilize the methodology. For assets where a node have merely one MPFM or VFM, we cannot remove it from the reconciliation problem. For such cases, however, CalPal can be essential to remove the bias from the sensors such that they can still be used in the reconciliation problem.

As with NC-VFM, FF is offered as-a-service. We connect automatically to your databases and if set up, FF is smoothly integrated with the Well Test Application. FF therefore operates with low intrusiveness in your everyday work life and can, if you allow it, be your new colleague for smarter allocation.

FF as-a-service

INTELLIGENT
Delivers the most likely flow rate estimates for your wells applying data validation and reconciliation, mass balance enforcement and live, condition-based uncertainties

LOW INTRUSIVENESS
FF is operated through an automated workflow, fetching and writing back data in a way that causes little intrusiveness in your everyday routine

HANDY ERROR DETECTION
Using statistical methods, FF notifies you about faulty or deviating rate measurements

SINGLE SOURCE OF TRUTH
Fuses all available rate information together to give you robust flow rate estimates that reduce biases and variance

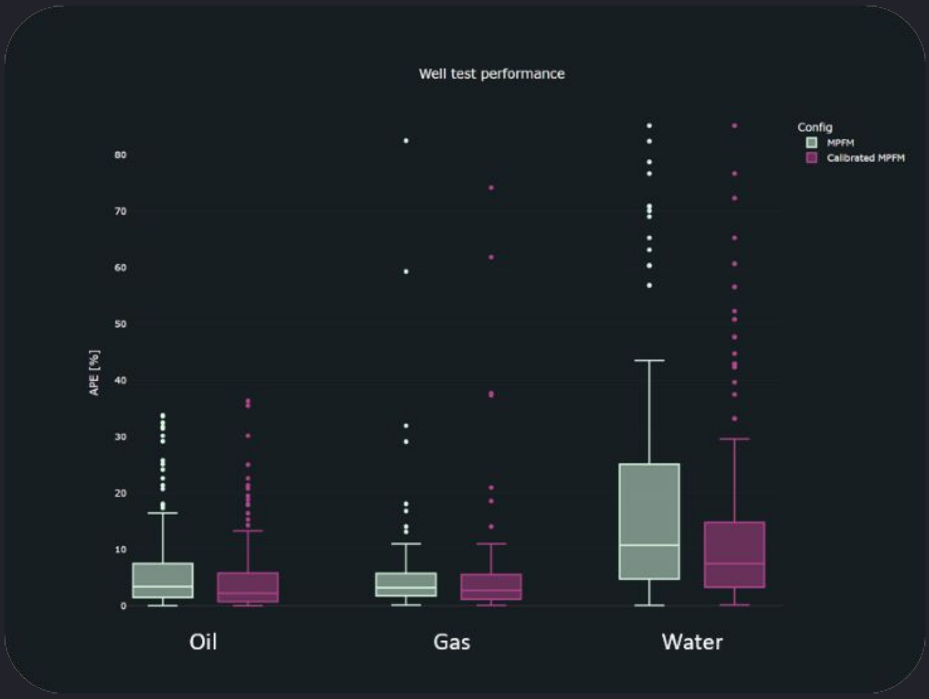
FLEXIBLE FRAMEWORK
A highly flexible framework that can handle a variable number of available measurements and be tuned to your liking

FlowFusion real-life results

Now that we have presented the fundamentals of our service for reconciliation and allocation, namely FF, we would like to demonstrate some real-life results. The results are anonymized. First, let's look at one example where the FF rate is calculated every 24 hours. Here we have three rates as input to the reconciliation problem. The white curve is the calculated FF rate and the shaded region around this curve is the calculated uncertainty of the FF, based on the uncertainties of the respective input sensors. We see that the FF rate is, for the most part, closest to VFMs 2 and 3. This is natural as the two sensors are very close to each other and the algorithm will believe these values to be more certain than that of VFM 1. From day 9, we see that the FF rate creeps closer to VFM 1. This can indicate one of two scenarios: either the uncertainty of VFM 1 is reduced, for instance if there has been a new well test indicating this, or there is a larger imbalance to make up for between the sum of the well-specific flow rates and the topside measurement such that the well in question gets attributed some of the extra flow. Which scenario we have here is difficult to tell without further investigations. However, as part of the service is a dashboard with visualizations of the flow rates and their uncertainties that can be used to deep-dive into the results, thereby accessing this kind of information.

The figure below illustrates an example where the data processing module comes in handy. Here we see that VFM 1 has suddenly dropped to a new and likely faulty value. With the stray and frozen detection algorithms, this behavior can be easily captured and, given the sensor redundancy, such sensors can be automatically filtered out of the reconciliation problem. Observe, for instance, that the white FF rate is not drawn towards the faulty estimates of VFM 1 - evidence that VFM 1 is, in fact, filtered out.

The figure below demonstrates the usage of GED. We see that VFM 1 is quite far from the FF rate and that GEDs have indeed occurred. This can be utilized to recalculate the FF rates without VFM 1 if desired. In this case, however, we see that VFM 1 has not influenced the FF rates significantly as they are not drawn towards its estimates, yet, there can be other cases where we see the opposite behavior where it would be beneficial to filter out the gross error detected sensor.



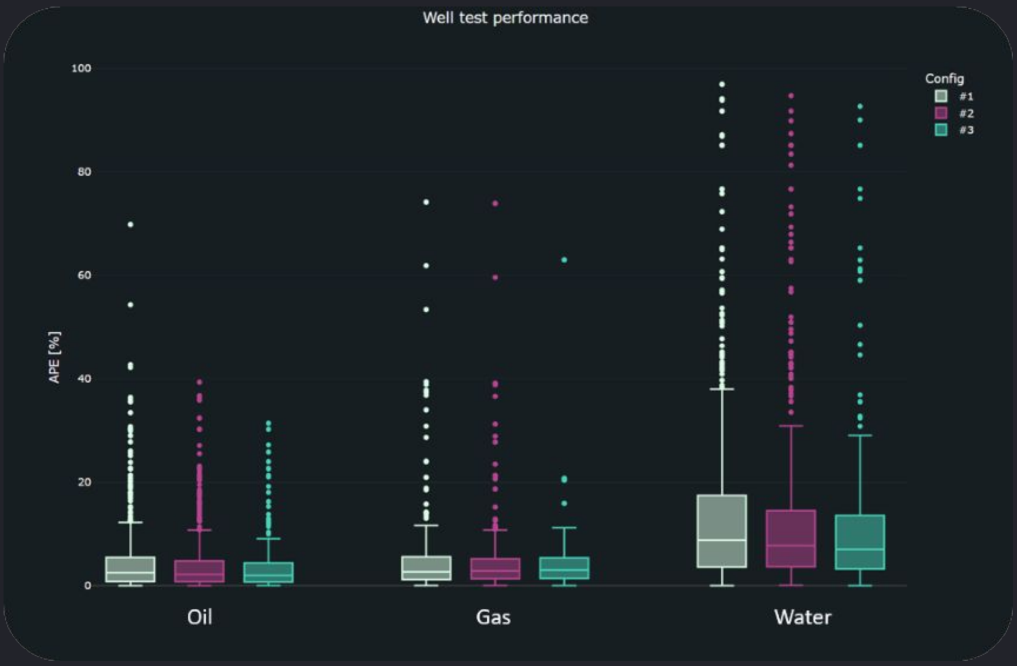
The figure to the left springs from a study we did in-house to demonstrate the benefit of calibrating biased MPFMs with CalPal. For this study, we calculated FF for steady-state periods instead of 24-hour intervals. In doing so, we are able to directly compare the reconciled rates to the results from *stable* well tests, that do not usually have time spans of 24 hours. There are two configurations shown in the figure on the left hand side, the first where the original MPFM is applied in the reconciliation problem and the second where the MPFM is calibrated with CalPal and used as a replacement for the original MPFM. No other rate sensors were used as input.

The absolute percentage errors (APEs) - retrieved by comparing the FF rates against well test results - for the wells in the study is grouped together in box plots differentiating the errors for the oil, gas, and water phases. The median error (the line in the middle of the boxes) is lower for the FF rates using the calibrated MPFMs than the original MPFMs. For oil, the median error is reduced from 3.4% to 2.2%, for gas from 3.2% to 2.8%, and for water from 10.8% to 7.5%. Notice that there are some errors that are very large - up to a 100% error. However, keep in mind that if the production from a well is very small, absolute percentage errors will quickly blow up. For instance, if your well produces 1 Sm3/d but the MPFM estimated 2 Sm3/d, your error will be 100%.

35%

Error reduction for the oil phase using calibrated sensors in FlowFusion

Another study we did in-house looked at the benefit of having and exploiting redundant measurements of the flow rates for the different wells. The results are shown in the figure to the right, for which the three configurations indicate the results when using all possible configurations of one, two, or three input sensors into the reconciliation problem. The boxes illustrate the distribution of APEs when comparing FF rates to well test results. Here, the median error decreases steadily going from one to two to three sensor signals for the oil and water phases.

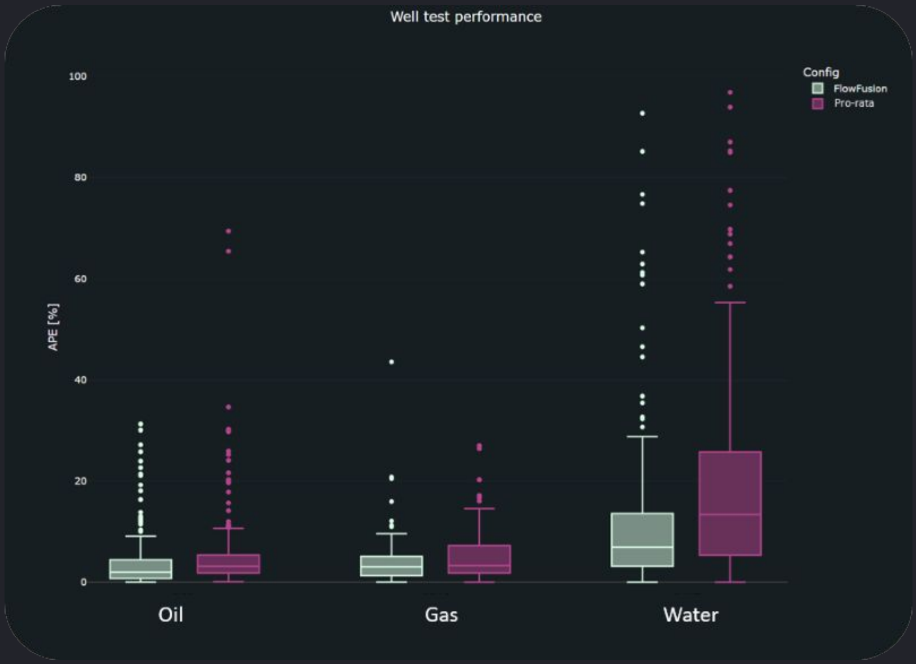


Nevertheless, for the gas phase, the median error increases slightly. Yet, the variance is smaller with fewer outliers, indicating all the while a more robust behavior. One reason for the increased median error can be that the flow rate sensors used for gas have quite different behaviors. Furthermore, the sum of the well-specific flow rates might also have a larger mismatch from the topside measurement than that for oil and water. In such cases, it could be more difficult to estimate the uncertainties of the different sensor and the reconciliation problem can, accordingly, have a harder time distributing the imbalance appropriately. Of course, we cannot only look at the well test performance when analyzing the benefit of redundant sensory information. In addition to general robustness, we also have to acknowledge the possibility to filter out faulty sensors. In the results shown above, no sensors were actively filtered out via GED. However, if GED had been applied to filter out faulty sensors in configurations #2 and #3, we strongly believe that we would have seen a decrease in the median error for all phases.

oil **38%** gas **6%** water **49%**

Error reduction using FlowFusion instead of pro-rata

Last but not least, we also compared the more sophisticated FF technology with the conventional allocation method PRA as previously described in this section. The results are shown in the figure below. All three sensors from the above study were applied to obtain the FF rates.



Observe how for all phases, the median and variance of the error distributions are smaller for FF, indicating a more robust and intelligent approach. In fact, comparing FF to PRA, the median error is reduced by 38%, 6%, and 49% for oil, gas, and water, respectively by applying the former. Again, with a bit more love gone into the gas phase, the reduction in error would likely be larger. Nevertheless, these are still very promising results - at least in our eyes.

6 WHAT WAS THE FUZZ ALL ABOUT?

We have now delved into the secrets of our industry-tested AI Flow Solutions. The two main technologies are NeuralCompass Virtual Flow Meter and FlowFusion, which together offer you AI-powered flow rates for both monitoring and allocation in real-time. NC-VFM is a commercially available and fully data-driven VFM that offers high-quality flow rates with high interoperability. It is based on more than 10 years of heavy research on the topic and experience in the field. NC-VFM has methods to exploit the massive amounts of underutilized production data that exists in the oil and gas industry and to tackle the typical data challenges we see in relation to such data.

The second technology, FlowFusion, is our service for smarter allocation. It consists of four modules: data processing, uncertainty estimation, reconciliation, and gross error detection, all essential to offer you the most likely flow rates for your wells. The service is mathematically based on a model-based optimization methodology called data validation and reconciliation, which smartly distributes the imbalance between well-specific flow rates and topside measurements. With the gross error detection module, available flow rate sensors can be monitored and faulty behavior like drift can be detected.

Both solutions are offered as-a-service. This means that the product you buy will be in constant development with the latest and greatest from Solution Seeker. It is not a one-time purchase that will be left alone after investment. Furthermore, maintenance of the products is included in the price and there will be no surprise bill if some wells require more love and care than others. The solutions are also of high interoperability as we use our platform ProductionCompass AI to seamlessly integrate with your databases to read the required data and write the flow rates back to your systems.

Not fully hooked yet? Do not hesitate to contact us for more information or perhaps a demo or two of our solutions. We look forward to hearing from you.

Contact us
solutionseeker.no

Meet the authors

Christine Foss Sjulstad
Data scientist

Mathilde Hotvedt Wekre
Data scientist

7 SOURCES

1. AL-Qutami, T. A. and Ibrahim, R. and Ismail, I. and Ishak, M. A. (2018). 'Virtual multiphase flow metering using diverse neural network ensemble and adaptive simulated annealing'. *Expert System With Applications,* 93.
2. Amin, A. (2016). 'Using measurement uncertainty in production allocation'. In: *Upstream Production Measurement Forum.* Houston, Texas.
3. Antonelo, E. A. and Camponogara E. and Foss, B. (2009). 'Data-driven Soft Sensors in the process industry'. *Computers and Chemical Engineering,* 33(4), pp. 795–814.
4. Badings, T. S. and van Putten, D. S. (2020). 'Data validation and reconciliation for error correction and gross error detection in multiphase allocation systems'. *Journal of Petroleum Science and Engineering,* 195.
5. Balaji, K. and Rabiei, M, and Suicmez, V. and Canbaz, C. H. and Agharzeyva, Z. and Tek, S. and Bulut, U. (2018). 'Status of Data-Driven Methods and their Applications in Oil and Gas Industry Introduction to Data Driven Methods'. *Society of Petroleum Engineers*.
6. Bikmukhametov, T. and Jäschke, J. (2020). 'First principles and machine learning virtual flow metering: A literature review'. *Journal of Petroleum Science and Engineering,* 184.
7. Bjørk, R.N. and Skålvik, A.M. and Pobitzer, A. and Mosland, A. and Sætre, E. and Frøysa, K. (2016). 'Analysis of field and ownership allocation uncertainty in complex multi-field configurations'. In: *North Sea Flow Measurement Workshop*.
8. Clark, J.T. and Klavenes, J. and Stewart, J. (2022). 'How Process Industries Can Catch Up in AI'. BCG 02 March 2022. Available at: https://www.bcg.com/publications/2022/how-process-industries-can-catch-up-in-ai . (Accessed: 23.05.2023)
9. Corneliussen, S. and Couput, J. and Dahl, E. and Dykesteen, E. and Frøysa, K. and Malde, E. and Moestue, H. and Moksnes, P. and Scheers, L. and Tunheim, H. (2005). 'Handbook of multiphase flow metering'. *The Norwegian Society for Oil and Gas Measurements*.
10. Davenport, T. H. and Ronanki, R. (2018). 'Artificial intelligence for the real world'. *Harvard Business Review*, 96(1), 108–116
11. Enholm, I.M. and Papagiannidis, E. and Mikalef, P. and Krogstie, J. (2021). 'Artificial Intelligence and Business Value: a Literature Review'. *Information Systems Frontiers,* 24.
12. Fountaine, T. and McCarthy, B., and Saleh, T. (2019). 'Building the AI-powered organization'. *Harvard Business Review*, 97(4), 62–73.
13. Gabernet, A. R. and Limburn, J. (2017). 'Breaking the 80/20 rule: How data catalogs transform data scientists' productivity'. IBM 23 August 2017. Available at: https://www.ibm.com/cloud/blog/ibm-data-catalog-data-scientists-productivity#:~:text=Yet%20in%20most%20companies%2C%20the,percent%20to%20actually%20perform%20analysis. (Accessed: 01.05.2023)
14. Goodfellow, I. and Bengio. Y. and Courville, A. (2016). *Deep Learning.* Cambridge Massachusetts. London, England. The MIT Press.
15. Grimstad, B. and Hotvedt, M. and Sandnes, A.T. and Kolbjørnsen, O. and Imsland, L. (2021). 'Bayesian Neural Networks for Virtual Flow Metering: An Empirical Study'. *Applied Soft Computing,* 112 (1).
16. Gunnerud, V. and Sandes, A. and Uglane, V. T. and Ursin-Holm, S. (2016). *Assessment of flow networks*. European Patent Office no. EP3371414A. Available at: https://worldwide.espacenet.com/patent/search/family/055132414/publication/EP3371414A1?q=16791594.1 (Accessed: 31.05.2023).
17. Gunnerud, V. and Sandes, A. and Uglane, V. T. and Ursin-Holm, S. (2018). *Recording data from flow networks*. European Patent Office no. EP3619396A1. Available at: https://worldwide.espacenet.com/patent/search/family/059065626/publication/EP3619396A1?q=18725770.4 (Accessed: 31.05.2023).
18. Hanssen, K.G. and Foss, B. (2015). 'Production optimization under uncertainty-applied to petroleum production'. *IFAC-PapersOnLine,* 48(8).
19. Jansen, J.-D. (2015). 'Nodal Analysis of Oil and Gas wells. Theory and Numerical Implementation'. TU Delft.
20. Jiang, X. and Liu, P. and Li, Z.(2014). 'Data reconciliation and gross error detection for operational data in power plants'. *Energy,* 75, 14–23.
21. Kanshio, S. (2020). 'A review of hydrocarbon allocation methods in the upstream oil and gas industry'. *Journal of Petroleum Science and Engineering,* 184.
22. Kneight, A. (2019). 'A hybrid approach'. Oilfield Technology 30 April 2019. Available at: https://www.oilfieldtechnology.com/special-reports/30042019/a-hybrid-approach/. (Accessed: 21.04.2023).
23. Monteiro, D.D. and Duque, M. M, and Chaves, G. S. and Filho, V. M. F. and Baioco, J. S. (2020). 'Using Data analytics to quantify the impact of production test uncertainty on oil flow rate forecast'. *IFP Energies Nouvelles,* 75 (7), pp. 1–15. doi: 10.2516/ogst/2019065.
24. Narasimhan, S. and Jordache, C. (2000). 'Data Reconciliation and Gross Error Detection'. *Gulf Professional Publishing*, ISBN 9780884152552.
25. Oliveira, E.C. and Aguiar, P.F. (2009). 'Data reconciliation in the natural gas industry: Analytical applications'. *Energy Fuels,* 23, 3658–3664.
26. Pobitzer, A. and Skålvik, A.M. and Bjørk, R.N. (2016). 'Allocation system setup optimization in a cost-benefit perspective'. *Journal of Petroleum Science and Engineering,* 147.
27. Sadri, M. and Shariatipour, S. M. and Ahmadinia, M. (2019). 'The Impact of Regular Well Testing on the Accuracy of Allocation Calculations'. *The 37th international North Sea Flow Measurement Workshop.*
28. Sandnes, A. T, and Grimstad, B. and Kolbjørnsen, O. (2021). 'Multi-task learning for virtual flow metering'. *Knowledge-Based Systems,* 232, 107458, ISSN 0950-7051
29. Shapovalov, V. and Shapovalov, Y. B. and Bilyk, Z., and Megalinska, A.P. and Muzyka, I. O. (2020). 'The Google Lens analyzing quality: an analysis of the possibility to use in the educational process'.
30. Silver, D. and Schrittwieser, J. and Simonyan, K. and Antonoglou, I. and Huang, A. and Guez, A. and Hubert, T. and Baker, L. and Mattheew, L. and Bolton, A. and Chen, Y. and Lillicrap, T. and Hui, F. and Sifre, L. and van der Driessche, G. and Graepel, T. and Hassabis, D. (2017). 'Mastering the game of Go without human knowledge'. *Nature,* 550, doi: https://doi.org/10.1038/nature24270.
31. Thorn, R. and Johansen, G. A. and Hjertaker, B. T. (2012). 'Three-phase flow measurement in the petroleum industry'. *Measurement Science and Technology,* 24(1).