SPE-187429-MS

Technical Potential: A Highly Effective Tool for National Oil Companies to Drive and Manage Business Plans – A Real Case Study at PETRONAS Malaysia

Vipin Prakash Gupta, PETRONAS; Satyajit Dwivedi, SAS INSTITUTE; Razeif B. Roslan, Rahim Masoudi, and W. Aziz Razali, PETRONAS

This paper was prepared for presentation at the SPE Annual Technical Conference and Exhibition held in San Antonio, Texas, USA, 9-11 October 2017.

Abstract

Sound business decisions require robust production forecasts whose uncertainties are quantified by an assessment of market expectations. While this applies to all E&P ventures, it is much more critical and relevant to National Oil Companies (NOCs) since they depend upon Production Sharing/Operating Contractors (PSCs/PACs) to deliver the business goals. It is critical to continuously improve processes in complex systems to enable a probabilistic range of performance forecasts. Data-driven methodologies support these business value propositions.

Technical Potential (TP) forms the basis for future expectation by defining what is achievable, and thus highlights the gap between possible performance and what is realized regarding hydrocarbon production. This knowledge transforms into initiatives that drive the processes for minimizing the gap. Assessment and forecast of TP workflows provide the appropriate tools for NOCs for driving the operator contractors towards better performance targets. PETRONAS in the past year has seen some clear, tangible benefits from the exercise.

However, the subject is tricky and the path to success, a potential minefield not only due to data frequency and metering issues that are inherent in production well testing but also the fact that TP forecast does not exactly follow the same set of rules as production. Soft-computing models based on data-driven analytics were used to resolve the issues, and a MAPE average of less than 20 was achieved for 80% of the producing assets. Data from a total of 74 producing fields in South East Asia was used to train and optimize data-driven algorithms.

The data-driven forecasts capture the trends, and temporal relationships surfaced through the patterns hidden in the well and reservoir historical data. Traditional type curve fitting and empirical DCA only provide a deterministic forecast, but a probabilistic data-driven methodology offers some key advantages: (1) A range instead of an absolute number and (2) Identification of time-series periods that focus on high and low performance periods optimizing business decision cycles.

We propose a hybrid approach where the traditional type curve fitting and DCA analysis results can act as input data points to a data-enriched probabilistic time-series analysis. The value of DCA cannot be

denied as a check. However, the current method will certainly enhance the value tremendously when used in conjunction with DCA.

Blind test workflows or what-if scenarios, a favorite of simulation engineers, to verify the accuracy of the forecast was implemented to verify the performance of the data-driven analytical models.

The paper introduces the relevance of TP methodologies and defines the key components of the production system, their relationships, measurement entities and methods of calculation. It then details the analytical methods for TP forecasts.

1.0 Introduction

It may be noted here that this paper is probably the first ever document being shared at the SPE or Petroleum Industry Forum. No references have been found on the subject during an extensive search of One Petro archives.

National oil companies operating in a multi-operator environment depend upon PA contractors to fulfill their commitments to the government. This situation over the years has manifested two clear trends. FIG1.1

1. Conservative forecasts that are almost always lower (substantially) compared to production performance
2. Many unnecessary investments by NOCs to fill up the gaps arising out of 'conservative' forecast.

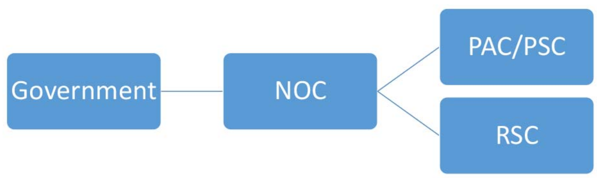


Figure 1.1—Relationship of NOC to Govt & Contractors

The first trend has indicated a gap between the actual production and potential to produce. To redeem the situation, it is imperative that NOCs be able to estimate this quantity. FIG 1.2 provides an example of this phenomenon and also emphasizes the need for NOCs to have a tool that can estimate the potential to produce.

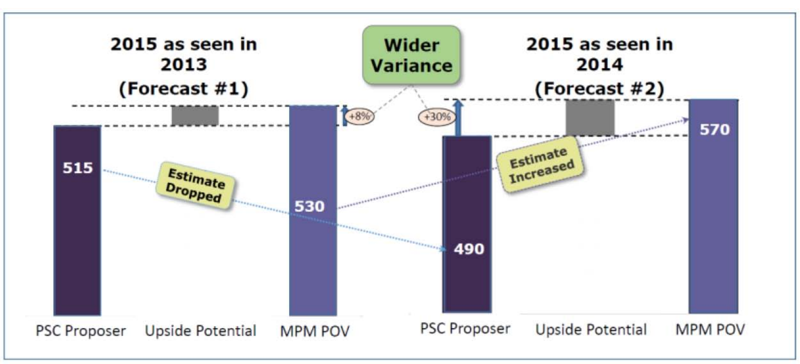


Figure 1.2—Conservative Forecasting

Technical Potential (TP) forms the basis for future expectation by defining what is achievable, and thus highlights the gap between potential performance and what is realized in terms of hydrocarbon production. This information can empower the NOCs to set realistic targets and minimize on unnecessary investments. This knowledge can be transformed into initiatives that drive the processes for minimizing the gap. Assessment and forecast of TP workflows provide the appropriate tools for NOCs to drive the operator contractors (PAC) towards better performance targets. PETRONAS in the past year has seen some clear and tangible benefits from the exercise.

Technical potential is calculated using production well test data. Since the well is isolated from the surface production system during the period of testing, well a test can be taken as a measure of the potential of the well. However, the well test data is often sporadic, and non-representative of the real potential even when the well is isolated due to the poor condition of the well itself. Such data cannot be used directly for either robust estimates even less so for generating forecasts.

The paper will discuss the steps that were taken to resolve the above issues through generating algorithms that were able to FIG 1.3

1. Scan through last 24 months of data to decide on the most representative value of TP.
2. Fill the gaps in the data using hard data points and DCA.

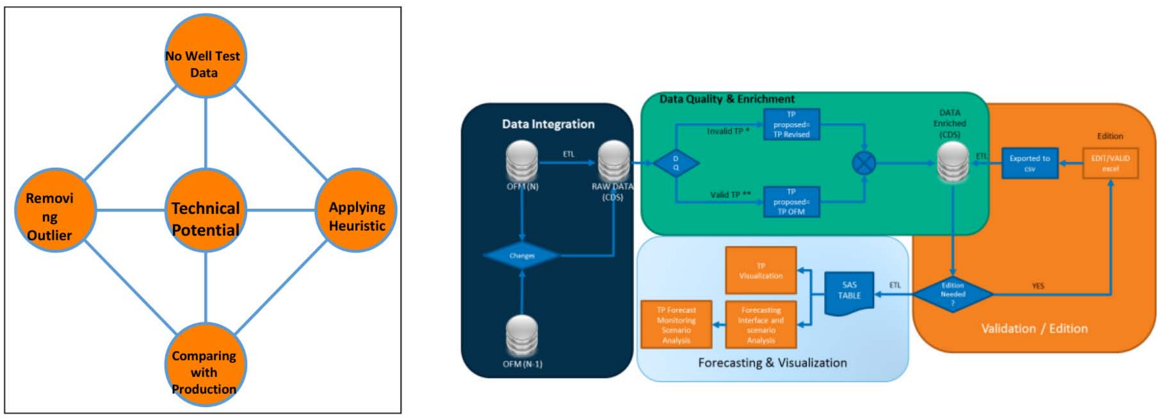


Figure 1.2—(1) Flow diagram depicting calculation of TP(2) Process Diagram for the Calculation of TP

The results were then used as input to create a TP-RECONSTRUCT (TPM3), virtually a continuous data stream that represents the technical potential or well-test at any point fulfilling the conditions for developing forecasts. TPM3 was checked using PDOIL

Traditionally TP forecasts have been done using pure DCA. However, the application not straightforward since the rules that apply to production and potential could be different. Additionally DCA provides an average forecast and is suitable for long term, however considering the major focus for TP being 12-18 months, DCA does not present itself as an optimum solution.

Considering the cyclic nature of patterns that were observed in the TPM3, data analytics based forecasts techniques were applied. The results threw up absorbing observations which will be discussed in the following pages. In all data from 74 Fields representing every part of the spectrum regarding diversity, be it depositional environment, drive mechanism, type of development.

The robustness of the forecast was tested using blind test techniques (Section 5), and accuracy was measured using the MAPE averages. The results are presented and discussed in section (Section 6)

2.0 Concept of Technical Potential

The journey of crude oil from the point of production in the reservoir to sales point can typically be described as a production system. A typical production system and the relationship between its various components is presented for reference at FIG 2.1

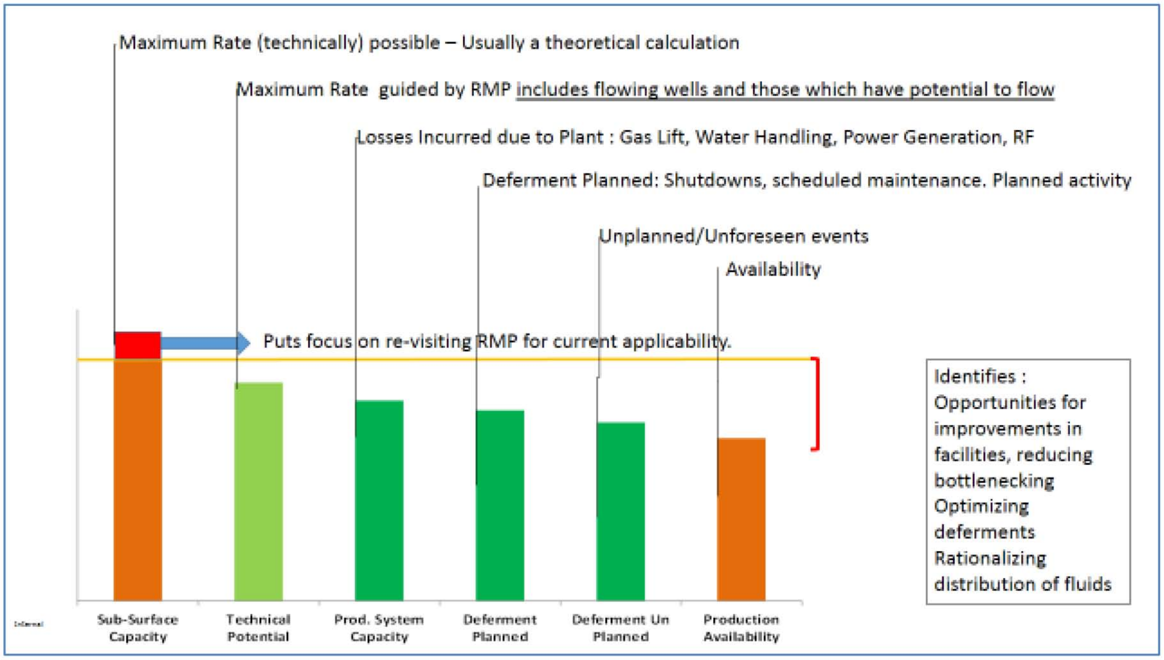


Figure 2.1—Relationship of components in a Production System

Detailed Description of terms in the above plot is provided below for reference.

2.1 Technical Potential (TP)

TP for a well/field/reservoir is defined as the maximum possible rate of hydrocarbon that it can deliver without impacting the ultimate recovery (operating within RMP*) at any point in time. Represented by production well test for a well, the TP for more than one well can be calculated by aggregating the individual values. Identified, planned and budgeted events leading to change in TP need to be included in the calculation of future values of TP.

TP for an existing well is defined as the maximum possible rate that it can deliver without compromising resources' recovery at a particular point in time. The baseline assumptions of the TP estimation should be consistent with the approved Reservoir Management Plan (RMP) and the adopted Business Plan (WP&B). The estimation of TP is based on the results of the representative well test(s).

The TP for a facility or field or reservoir builds upon the summation of the estimated TP for all the results for individual wells. All active flowing wells (whether optimal flowing or not) plus wells that are not flowing but are not considered as idle wells (Wells that are shut-in but could be brought on stream by opening the choke e.g. swing wells).

The TP includes the forecasted gains associated with planned and approved production or system enhancement activities that are scheduled to occur within WPB duration.

If there are plans to drill a new well / work-over a well the TP should be estimated and included from the month that the well is expected to commence production. It can be calculated using analog wells and simulation techniques. FIG 3.1. Represents an average TP forecast. Note that changes in Technical potential are not following modifications in availability since lower availability due to deferments does not impact the potential of the system.

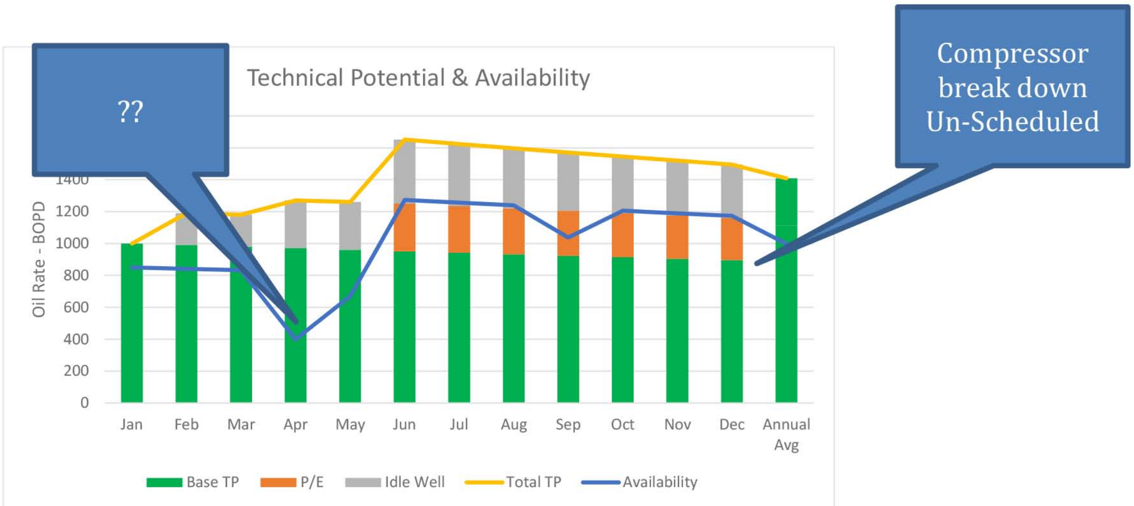


Figure 2.1—Technical Potential and Availability

Note: TP is not the 'technical limit' of a well or a field, it is a measure that allows NOCs/ Operators to understand relatively 'short-term' approved production levels and assessed the efficiency of the production system. It may be possible to achieve a higher rate of production than defined above if prudent reservoir management practices are not applied to the system.

Since the major focus of the paper is on estimation & forecast of TP & data-driven technologies that were used to generate the forecast. Description of other entities of the production system is placed for reference at Appendix I.

2.2 Subsurface Capacity

Subsurface Capacity for a well/field/reservoir is defined as the maximum possible rate of hydrocarbon with no RMP constraints at any point in time. All identified events leading to change in TP need to be included in the calculation of future values of TP.

Subsurface Capacity Is defined as the maximum production rate possible in a given production system with No RMP constraints and budget/resources constraints. It should include any gain associated with production/system enhancements that have been identified but may not yet be matured and approved. The base assumption will be a maximum drawdown, irrespective of RMP constraints.

The Subsurface Capacity for a facility or field or reservoir is based upon the summation of the estimated Subsurface Capacity for all the active and idle wells. It should also include gain associated with identified infill wells and workovers that are planned to occur within business plan period of the forecast.

The Subsurface Capacity can typically be estimated using nodal analysis, dynamic simulation or extrapolated from multi-bean well tests.

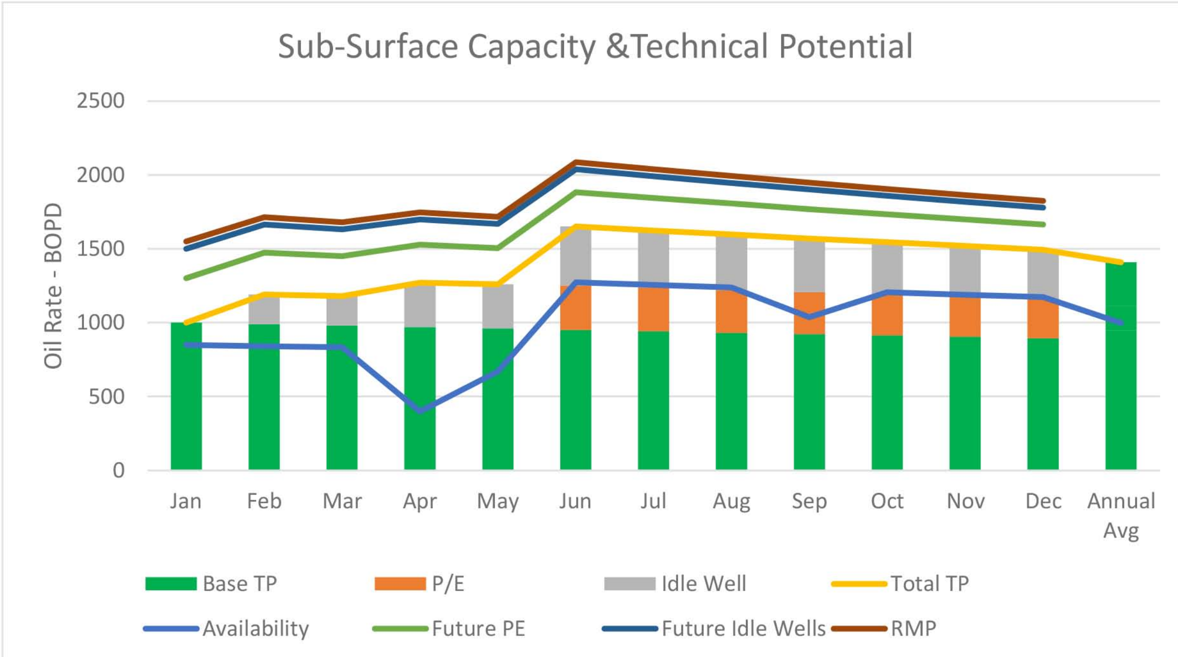


Figure 2.3—Technical Potential and Other components of Production System

Note: Subsurface Capacity is not the 'technical limit' of a well or a field, it is a measure that allows NOCs/ Operators to understand identified 'medium-term' production levels and assess the efficiency of the production system. The baseline assumptions are similar to that of Contingent Resources, i.e. there should be an identified project, but that the project does not need to be fully matured regarding definition.

2.3 Production System Capacity

Maximum hydrocarbon rate that a network of plants can deliver at the sale point with the output from the existing wells at any point in time.

The plant capacity may, therefore, be equal to the summation of the TP as defined above or lower. The hydrocarbon volume received at the Point of Sale, or Reference Point of the sales point is lower from the starting point due to losses due to shrinkage and other factors during transport. This loss is referred to as the Reconciliation Factor and assumes higher importance if the facilities are shared between assets. *In such a case the Production system capacity would be calculated after application of Reconciliation Factor.*

When more than one well flows into a system as would normally be the case in a producing asset, the production performance is likely to be affected by induced back-pressure or other factors such as described below, inherent to a multi-well flowing scenario.

- Distribution of Water Injection in a Water injection Reservoir
- Availability of Gas lift in a Gas Lift scenario
- Efficiency of water disposal systems
- Others…

Plant Capacity is defined as the maximum production rate that the network of wells plants can deliver after accounting for surface constraints.

The plant production capacity may, therefore, be equal to the potential technical summation of the TP as defined above or lower.

Plant capacity estimates are valid for the system extending from wellhead until the evacuation point i.e. the outlet from a platform. In certain circumstances, this may also be the point of sale.

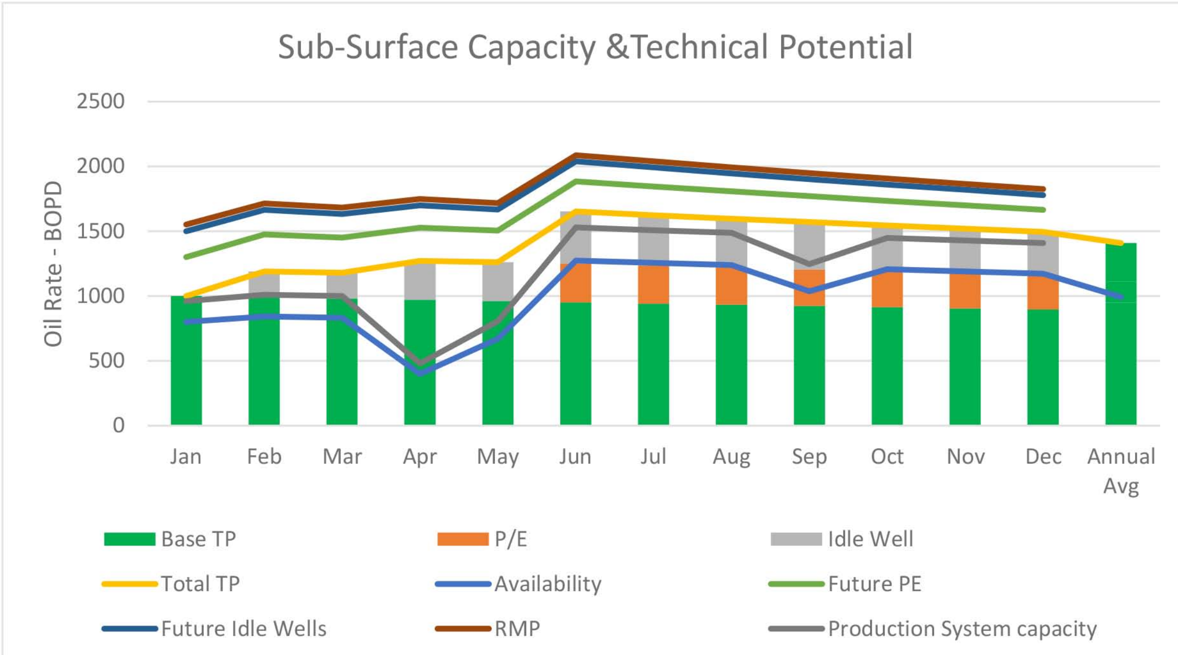


Figure 5—The relationship between Production system capacity and availability.

2.4 Deferment

TP, Plant Capacity, and Asset Capacity are all based on the assumption that there is no disruption to production. However this is rarely the case, and as such PAC Operators should estimate the extent to which the production system will not work at 100% operability. Such activities can be divided into two categories; Planned & Unplanned.

2.4.1 Planned Deferment: Amount of deferment that is planned to allow for the smooth functioning of operation facilities. Includes planned activities that are temporary in nature and result in an interruption to the flow of hydrocarbons. These activities are planned, and impact on production is included in projections. Examples of such activities include shutdowns associated with maintenance, drilling activities, well tests, well interventions, and simultaneous operations, etc. These deferments should not be captured by reducing the Plant Capacity because they are temporary in nature and are associated with specific activities.

2.4.2 Unplanned Deferment: Amount of deferment due to activities/events that are inevitable due to safety or reasonable causes that result in wells being shut-in for a temporary period until the interruption is rectified. Includes those unplanned activities that are temporary in nature that lead to disruption to the flow of hydrocarbons. Such interruptions could be due to environment (weather),

2.5 Availability

The volume of hydrocarbons remaining after application of all the above factors i.e. the volume that is available for sale and the generation of revenue. The forecasted availability is the basis of the business plan targets, commercial agreements and operational plans e.g. shipments.

2.6 Interdependencies

The Chart at Fig.2.1 explores the relationship between all the components defined above from Subsurface capacity to availability in a production system. TP of a system, therefore, remains unaffected by issues related to Plant, Asset or deferments except for cases where surface facilities are constrained to handle well output, and there are no plans to mitigate the situation (Table 2.1). Technical potential is largely dependent upon the number of wells contributing to production as well as factors that may impact this situation like drilling of new wells or restoration of idle wells. Production gains/losses associated with production enhancements, work-overs, new wells, rejuvenation of idle wells, or wells becoming idle also impacts the technical potential of a given asset.

Table 2.1—Factors affecting TP

No.	Item	Activity	Remarks
	Subsurface Capacity	• RMP • Well • Unapproved Optimisation, Workover, Enhancements	• Not restricted by sanding rates, voidage requirements. Often RMP restrictions limit the TP but are not considered in the Subsurface Capacity • Identified projects that are not yet approved
	Plant Capacity	• Gas Injection* • Water Injection* • Lifting • Oil handling • Water handling • Gas handling	• Limitations in oil, gas and water processing will result in a reduced Plant Capacity • Debottlenecking should increase Plant Capacity
	Technical Potential	• Optimization • Integrity • Mechanical • Lift • Drive Mechanism • Remedial	• Well optimization will often increase TP e.g. gas lift, lowering wellhead pressure • Integrity concerns may be reduced TP • TP will be limited by the lift curve and the pressure support • Remedial actions e.g. scale treatments, re-perforation, water shut-off are designed to improve TP • Secondary and tertiary recovery will increase TP once implemented

3.0 The challenge

Even after establishing the relevance and importance of TP to oil and Gas companies in general and NOCs in particular, the bigger challenge is to be able to estimate and more importantly to forecast TP. The greatest problem lies in the way TP is determined.

TP is based on production well test, and the sum of well tests for all the wells that are producing and those which can produce will determine the TP then for a field. A typical traditional way of calculation is presented in FIG 3.1

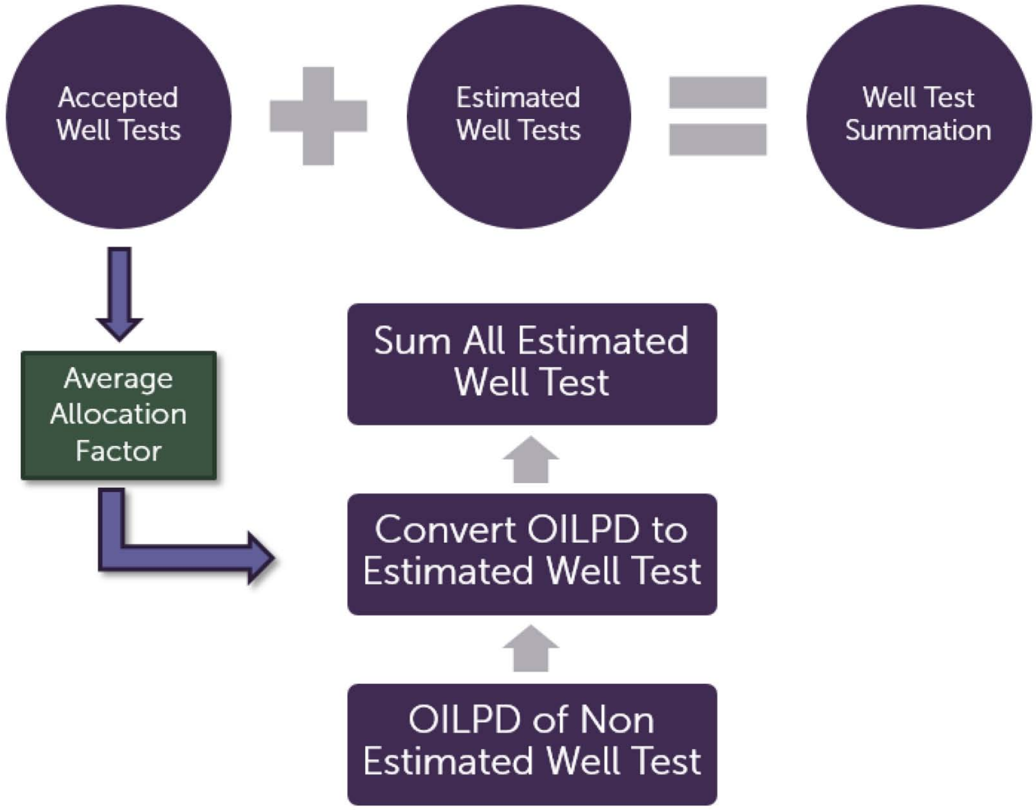


Figure 3.1—Manual Calculation of TP

However, the well tests data is usually collected sporadically and thus cannot be used directly for any estimation or forecast purpose. Moreover, every well test is not representative of the TP since the well condition may not be optimum. A two-step process was therefore introduced; (1) to select the representative wells and (2) to choose the well test that was not representative of TP. FIG 3.2

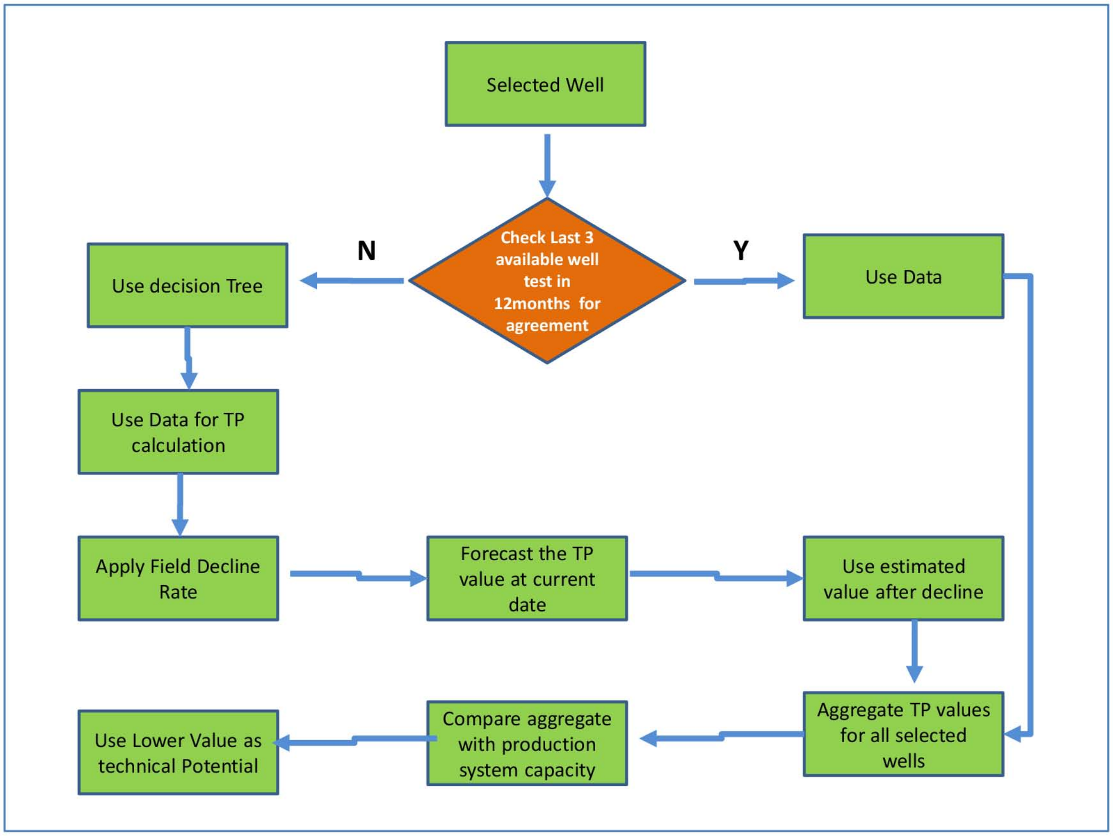


Figure 3.2—Flow Diagram for Well Selection to be Used for TP calculation

Once all the wells had been run through the decision mentioned above analysis loop, the next task was to define an algorithm that could reconstruct TP as continuous data.

3.1 TP Reconstruct: Algorithm Development.

Some of the challenges faced by the team in development of algorithm were

1. How much should historical data be used?
2. What should be the basis of comparison?

Statistical analysis of data related to the time of intervention that could be responsible for the change in the source of TP indicated an optimum period of 18 months. For item 2, rigorous workflows were set up (Section 5.0) and resulted compared with trends observed for PDOIL. (FIG 5.1) TPM3 (copyright protected), based on conditions mentioned above gave the best results (FIG 3.2) and was selected to be applied for generating TP Reconstruct data for all wells. This subject is dealt with in greater detail in (Section 5.0)

4.0 Forecasting

Unlike production, the challenge in the forecast of technical potential lies in the fact that temporary conditions related to decline in production vz., artificial lift performance, skin development, influx related issues, lower efficiency of surface production facilities, etc. will not impact technical potential. However, these conditions may be used by the practicing engineer to forecast a production declining trend for short term forecasts especially. It is in fact, this aspect of TP that pushes the focus on mitigating and rectification of the circumstances that lead to decline rather than accepting a decline.

One important observation while doing trend analysis on the results obtained after application of TPM3 is a manifestation of patterns that (FIG 5.2) that were seen to be cyclic in nature. The frequency of reproduction of cycles was found to be variable. However, the signature of the cycle was unmistakable and was found to be independent of the phase of development indicating the shape of the trend to be similar to an inherent characteristic or Signature. These repeating patterns were the basis of initiating a search to seek alternative methods for forecasting TP as opposed to DCA which is traditionally used forecasting future production profiles in the petroleum industry.

4.1 Data Analytics Driven Forecasting

Historically forecasting for production estimates has been based on standard algorithms for type curves which are deterministic in nature and is driven by an understanding of the surface or sub-surface updates, development plans for well. Most DCA based forecasts need manual intervention for selection of the trends that represent the basis of projected behavior. This makes the process Slow, Expensive and Subject to bias. Additionally, the forecast essentially is a smooth line based unable to capture the undulations depicted in the production patterns of the field.

Also, a critical point that was missing from the traditional forecasting/estimation process was that the accuracy of forecasts was never an organizational level metric that was monitored. Therefore there is not much of motivation for looking into a new approach of forecasting that is based on forecast accuracy with statistics of fit which could integrate structured data and unstructured data from a daily operating report or unplanned downtime records within the organization.

Data Analytics-driven forecasting differs from the traditional modeling techniques fundamentally on four accounts: -

a. *The models are based on patterns observed in the time-series and its relationship with previous values and other variables unlike a fixed mathematical formula based forecasting*
b. *The models are tested for forecast accuracy using statistical measures and not left individual discretion*
c. *Providing the forecasts with a confidence interval (e.g. 90%, 95%, 99%) band and not an absolute number for future periods and*
d. *Ability to automate the hierarchical forecasting process for a large number of time-series taking into consideration changing patterns in incremental field data on a periodic basis.*

The strength of analytics-driven forecasting lies in the detection of trends; cyclicity (FIG 5.2), randomness and the relationship with variables that have an impact on various production estimates. The data-driven approach can detect patterns both fundamental and technical in nature.

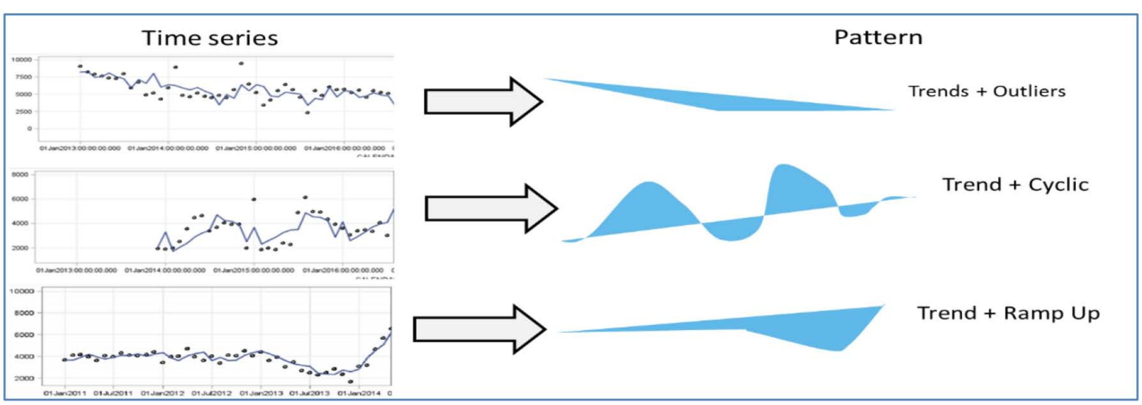


Figure 4.1—Cyclicity or Manifestation of Patterns

Fundamental in the sense of observed patterns over months & years and technical in a sense to capture the impact of the unplanned shutdown, commissioning of new wells, etc. To leverage the capabilities of data analytics and to make it an enterprise practice, a sound analytics process needs to be adopted. The following section approach describes a framework that has proven to be robust, scalable and generates accurate long term and short term technical potential forecasts.

5.0 Methodology/ Workflow

The 7 step data analytics process for generating potential technical forecast across the three regions in Malaysia is given below:-

- Step 1: Reconstruct of the Technical Potential based on the well test data
- Step 2: Data Quality Pre-Treatment in the Time Series before the Classification and baseline forecasting process.
- Step 3: Classification of Fields for understanding categorical patterns and homogenous groupings
- Step 4: Generating baseline time series forecast and validate the accuracy with stability test using Out-of-Sample Data.
- Step 5: Integrating with Decline Curve Analysis for shorter time-series
- Step 6: Schedule Monthly / Daily Rolling forecast with incremental data
- Step 7: Monitor the Accuracy and re-diagnose the time-series upon degradation of the forecast accuracy

The detailed process flow for the data analytics driven forecasting is given in (fig), and the description of each step is provided below:-

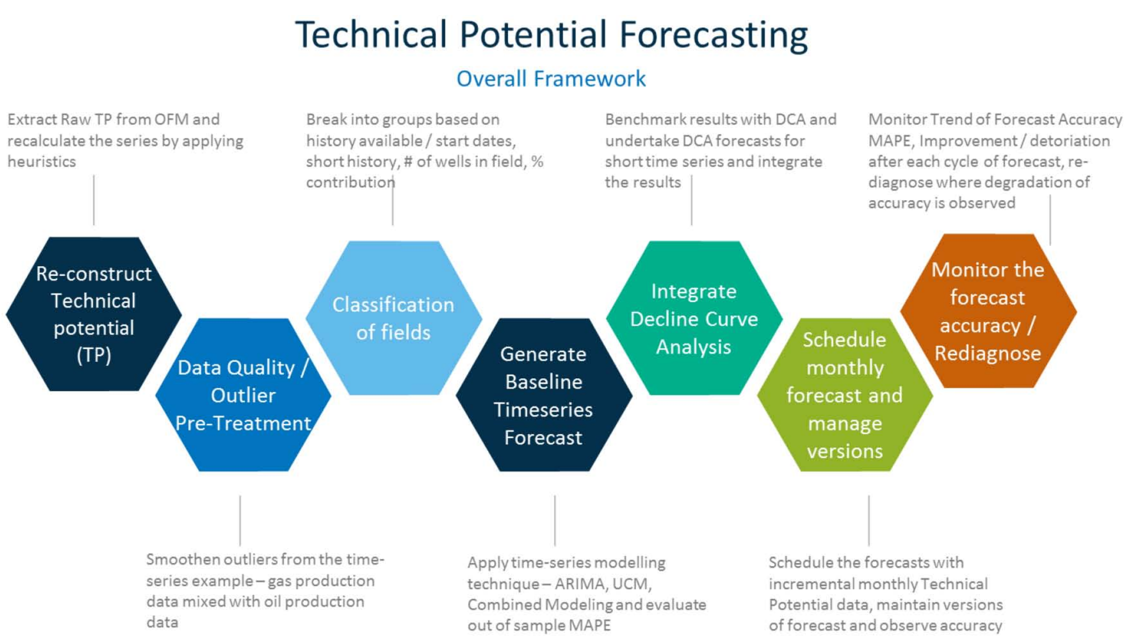


Figure 5.1—Flow Diagram for Technical Potential Forecasting

5.1 Step 1: Reconstruct of the Technical Potential based on the well test data

This step is one of the foundation elements of the data analytics forecasting process. As described in the sections above, the technical potential time-series is a calculated timestamped variable that is generated from well test data from active wells and the well tests that represent the technical potential of the well. The reconstruction process started at the string level; aggregated up to the well and then to the field level. Several analytical and iterative heuristic techniques were applied as described under:-

1. Identification of active and non-active wells for a given period of calculation.
2. Aggregation set creation for a given time-period considering the transition from active to inactive stage
3. Adjustment of well test data to reflect the true technical potential of the string with decline rate factors. Consistency checks and balances with the field, well and string level production data.
4. Technical potential history field level consistency with particular focus to removing any outliers thru the application of normalization using statistical analysis techniques. For Detailed description, refer to Appendix I

Three different methods with variations in the above steps were used to generate three potential technical estimates. The TP estimate from the third method was found functionally fit and statistically consistent to be adopted for the next steps in the data analytics forecasting process.

This approach has been applied for copyright, and therefore the detailed description of the process is out of the scope of the current paper.

Consistency check was carried out to observe the relationship of the Technical Potential reconstruct time series with the production data of the field.

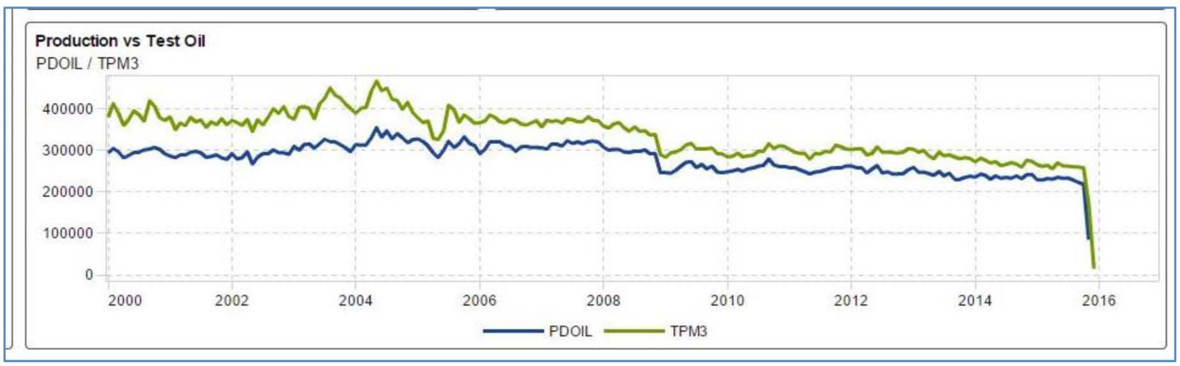


Figure 5.1—PDOIL & TPM3 Comparison

Based on the selected methodology 3, various patterns of Technical Potential line graphs were observed for different fields with different patterns as given below.

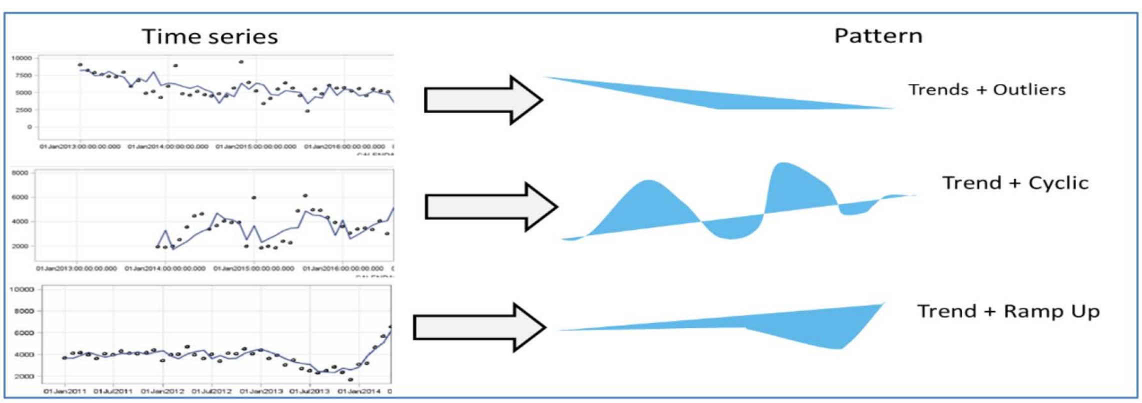


Figure 5.2—Cyclicity OR Manifestation of Patterns

It was observed that fields have different technical potential trends, cyclicity, jump, periods of high technical potential and sometimes a ramp up and a steady decline.

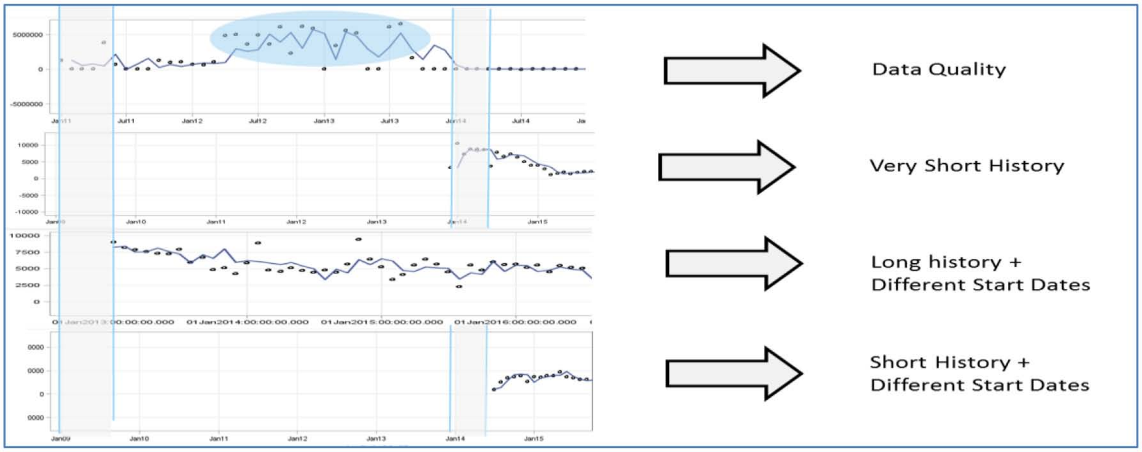


Figure 5.3—Interpretations related to Pattern types

Initial exploration of the technical potential time-series post reconstruction is critical. We observed that there are data quality issues in some of the fields that need to be fixed before forecasting. Example being gas production data was showing up in the crude technical potential. This was due to the data issues at Oilfield Management System tool that was a source to the Forecasting system. Also, it was observed that within a span of 3 years many fields do not have a potential technical history for the same duration and therefore different start dates of forecasting were needed for different fields.

5.2 Step 2: Data Quality Pre-Treatment in the Time Series before the Classification and baseline forecasting process.

Good forecasts depend on the high quality and reliable data. With a lot of variables being captured by OFM database about crude production, gas productions, well test data, water cut, etc., there are high chances that the data quality gets disturbed and goes unnoticed. Although it can be detected by the outlier treatment of the forecasting process, it was observed that due to the huge difference in the scale of crude production and gas production values the graphical output of TP forecast gets distorted. Therefore it was important to focus on data quality issues in the beginning before the time-series is used for forecasting. It was observed that in some fields gas production data is mixed up with crude production data during some parts of the years, thereby impacting the technical potential reconstruct.

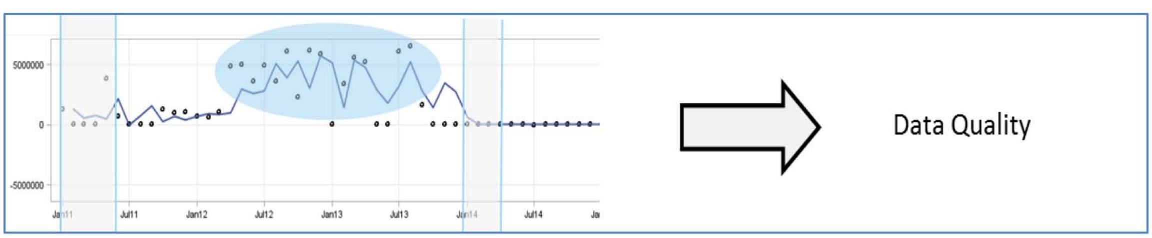


Figure 5.4—Gas Production Data visible in the crude production

Some fields had big spikes due to gas production data getting mixed up with raw data. Although the model picked up as an outlier, the huge difference in the unit of measure of crude and gas production skews the complete visualization graph.

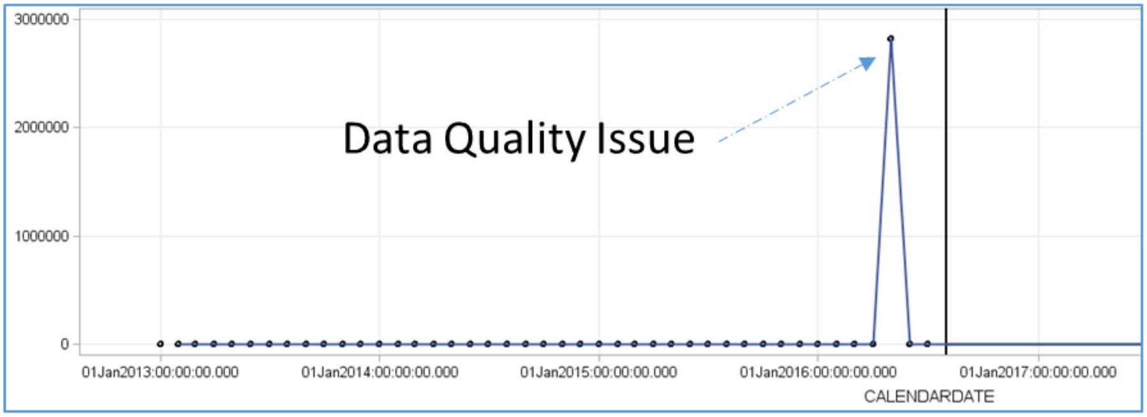


Figure 5.5—Gas Production Data visible in the crude production

There were some fields with very short history and intermittent patterns as shown below:-

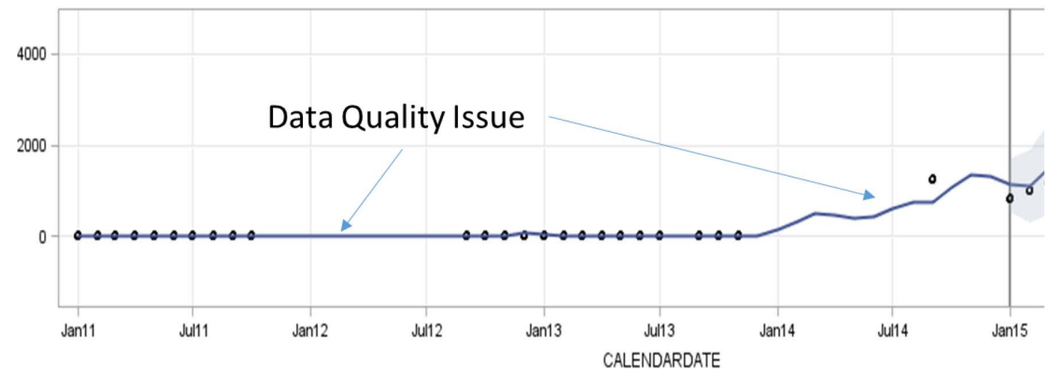


Figure 5.6—Short History Data and impact

A data quality process needed to be to remove the bad data and replace with imputed value of the Technical potential.

5.3 Step 3 – Classification of fields for understanding categorical patterns and homogenous groups

Classification of fields is an important step to understanding the characteristics of the field such as # of wells, # of years of data available, % contribution to the field to the overall production, type of structural changes to the production pattern of the field in the past, observed cyclicity from time-series statistics. This enables to create homogenous groups of fields from data analytics forecasting standpoint to which standard settings of the analytical forecasting models can be applied. It also allows us to understand which fields are to be considered for Type Curve based forecasting (in our case decline curve).

In the current project we found the following:-

Field Distribution	**Observation**
Top 15% of the fields	Add 60% of production
Top 30% of the fields	Add 80% of production
Top 50% of Fields	Add 90% of production
69 Fields	Had history more than 60 months
5 Fields	History less than 24 months
4 Fields	DCA candidates with shorter history than 12 months

Considering that 50% of the top fields contribute 90% of production, extensive exploration to understand the patterns of these fields were carried with specific emphasis on trend, ramp & cyclicity of technical potential time-series. This analysis brought about the fact that many fields operate under specific inherent characteristics that repeat itself in a cycle.

The separate path was created for modeling highly cyclic fields, less cyclic fields, fields with lesser history and appropriate configuration were built for each of this groups.

Last but not the least, based on the % contribution to the total production made us look at Forecast Value Add (FVA) (Section 7.0) against each of the fields on the efforts required for model fine tuning.

5.4 Step 4 – Generate baseline forecast and validating the robustness with stability analysis

The baseline models were created based on the groupings arrived at Step 3. The model family that was used in the forecasting process was ARIMA (Auto Regressive Integrated Moving Average) and UCM (Unobserved Component Model). In addition to ARIMA & UCM, a combinational model approach was also included in the overall process that would combine two or more than two individual models to come up more accurate models.

5.4.1 ARIMA model analyzes and forecasts equally spaced univariate time series data, transfer function data and intervention data. It predicts a value in a response time series as a linear combination of its past values, past errors (shocks or innovations) and current and past values of the other time series (ARIMAX Model). The model building is divided into three stages – identification, estimation & diagnostic checking and forecasting stage. 1) <u>Identification</u> is made using graphs, statistics, studying Auto Correlation Factors, and Partial Correlation Factors, transformations, etc. to achieve stationary and tentatively identify patterns and model components. 2) <u>Estimation</u> is the process of determination of coefficients and estimate through the software application of least squares and maximum likelihood methods. 3) <u>Diagnostics</u> is done using graphs, statistics, Auto Correlation Factor's and Partial Auto Correlation Factors of residuals to verify whether the model is valid. If found valid then the model is selected, otherwise repeat the steps of Identification, Estimation, and Diagnostics for other models.

Regarding y, the general forecasting equation is:

$$\hat{y}_t = \overbrace{\mu}^{\text{constant}} + \overbrace{\phi_1 y_{t-1} + \ldots + \phi_p y_{t-p}}^{\text{AR terms (lagged values of } y)} \underbrace{- \theta_1 e_{t-1} \ldots - \theta_q e_{t-q}}_{\text{MA terms (lagged errors)}}$$

By convention, the AR terms are + and the MA terms are –

AR – Auto Regressive | MA – Moving Average

5.4.2 UCM Model can be considered to be multiple regression models with time-varying coefficients. It decomposes time-series into trends, cycles, and the regression effects of explanatory variables. It provides a variety of diagnostic tools to assess the fitted model and to suggest possible extensions or modifications. Components of UCM provide the succinct description of the underlying mechanism governing the time-series. Similar to the Dynamic Models, popular in Bayesian time series, captures the versatility of ARIMA and interpretability of Smoothing Models.

The fully specified Unobserved Components Model is written as

$$y_t = \mu_t + \gamma_t + \psi_t + r_t + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{m} \beta_j x_{jt} + \varepsilon_t \, .$$

In the above equation, y represents the time series to be modeled and forecast, μ t the trend component, t γ the seasonal component, ψ t the cyclical component, t r the autoregressive component, and t ε the irregular component. All of these components are assumed to be unobserved and must be estimated given the time series data on t y_t and x_{jt}, hence the title unobserved components model.

5.4.3 Combined Model Forecast combination, also called ensemble forecasting, is a technique for improving forecast accuracy and reducing the variability of the resulting forecasts. It is an automatic process of combining automatic models that the system formulates from diagnosis of input series using an average of individual model forecasts. Models respond differently to end of the time span. One forecast model might overestimate, and the other forecast model might underestimate. Therefore a combination of forecasts using multiple models gives a better result. The combined forecast uses combination weights w_i to combine the fitted model forecasts as a weighted average of the predicted values with possible restrictions on the combination weights. The combination weights can correspond to a simple average, once can specify or they can be estimated using a variety of methods from the fitted forecast results.

$$\hat{y}_{c,t} = \sum_{i=1}^{M} w_i \hat{y}_{i,t} \text{ denotes the combined forecast at time } t$$

In our current setting, we had used simple average serves to correct for these different dynamics in the respective candidate models and OLS encompassing test to access forecast quality with an estimation region missing percentage of 25% and a missing horizon percentage of 50%. In two fields out of the top 30 fields, the combined model performed better than individual models

5.4.4 The Exponential Smoothing Model (ESM) models fit a time trend model using a smoothing scheme in which the weights decline geometrically as you go backward in time. The forecasts from exponential smoothing are a time trend, but the trend is based mostly on the recent observations instead of on all the observations equally.

Supposeis X_t actual observations Y_t and is smoothed series, $t = 1, 2, \ldots n$.

$$Y_t = \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + \alpha(1-\alpha)^3 X_{t-3} + \ldots \text{ where } 0 < \alpha < 1.$$

Exponential smoothing was not recommended due to two reasons a) the absence of seasonality in potential technical forecasts and b) two year ahead forecast at monthly level would not performance well and the forecasts will become a straight line since most recent one with maximum weight and then successive less on previous observations which do not capture the patterns of the field level as compared to ARIMA & UCM.

The ARIMA & UCM models can be trained to provide more weight to recent history. This enables the selection of the best fit model out of the array of candidate models. We also observed that the Exponential Smoothing Model (ESM) sometimes did well during the training period but failed in the validation period, and the observed Out-of-sample MAPE* at field levels were very high.

5.4.5 Blind Testing

Two years of technical potential data 2015 and 2017 was taken out of the sample, and the prior history was made to train the model in Forecast Server. SAS Forecast Server was configured to run ARIMA, UCM and Combined Model across all the 74 fields consisting of 1684 Wells with a new hierarchy of Regions & Operators. Based on the patterns observed and the result of statistical test configurations was done for a group of fields or individual field's level. The different types of configuration that were carried out are defined as:

- the start point of field level technical potential time-series
- cyclicity based on the field technical potential historical characteristics

- automatic outlier detection
- mode families such as ARIMA, UCM and additional default models in SAS repository.
- a model combination approach
- out-of-sample period of 2 years / one years
- hold out period for selecting the best fit model

5.4.6 Quality Check

Mean absolute percent error is calculated based on the following formula: - Average of Absolute of the (Actual Value – Forecast Value) / Actual Value * 100 across the period chosen for evaluation. The process of evaluating candidate model, best fit model and validate the model is based on the MAPE statistics given in the figure below (fig). The training data is used to arrive at the candidate models of ARIMA, UCM and Combined models with the parameters. Based on these models the forecast is done over two years, and the MAPE is observed across the training period and validation period. In case looking at the patterns of the time-series weight to the recent history is required then a hold-out period is carved out from the training period and the candidate models are tested on the holdout period. Depending on the patterns the top candidate model may not remain the top model in the holdout period.

Figure 5.7—– showing the model selection, validation and forecasting process

The robustness of forecast was qualitatively evaluated using the concept of blind testing. Operationally, this means that the portion of the potential technical data for the most recent years or months are not used for model building and the forecasts generated by the models for the same years or months is compared. The statistics of fit that measures the error for this particular period is called the out-of-sample MAPE. The lower the error in the blind tests determines the robustness of forecasts.

Fig details of the model families that were selected as best models for the fields are given below:-

Total	ARIMA Model	UCM Model	Combination Model	DCA Candidate
74	53	8	4	6

The MAPE distribution of best fit models for the 69 fields are given below:-

Basic Statistical Measures			
Location		Variability	
Mean	13.09328	Std Deviation	8.98541
Median	10.69000	Variance	80.73767
Mode	16.67000	Range	40.24000
		Interquartile Range	10.38000

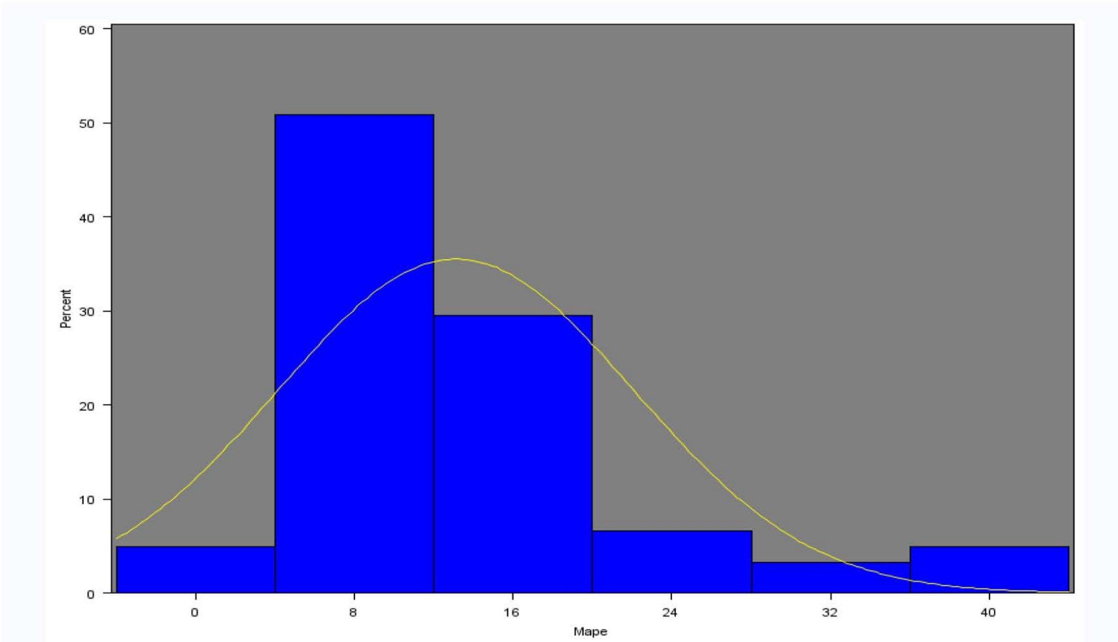


Figure 5.8—– Statistical Measures & Distribution analysis of the MAPE across fields

5.5 Step 5 – Integrate with Decline Curve Analysis for shorter time series/benchmark with data analytics based forecasting

Data Analytics based forecasting requires at least 2-3 years of history of fields. Fields which have a history of fewer than 18 months of history were taken as candidates for using type curve (decline curve analysis) based forecasting techniques. However, as a robust framework, the importance to integrate type curve analysis in the overall process and measure the accuracy of it as well has been realized, and work is undergoing to integrate at the forecast accuracy level. This will provide an essential element of "lift" in forecast accuracy that is being achieved of the data analytics based model over and above the traditional methods of forecasting.

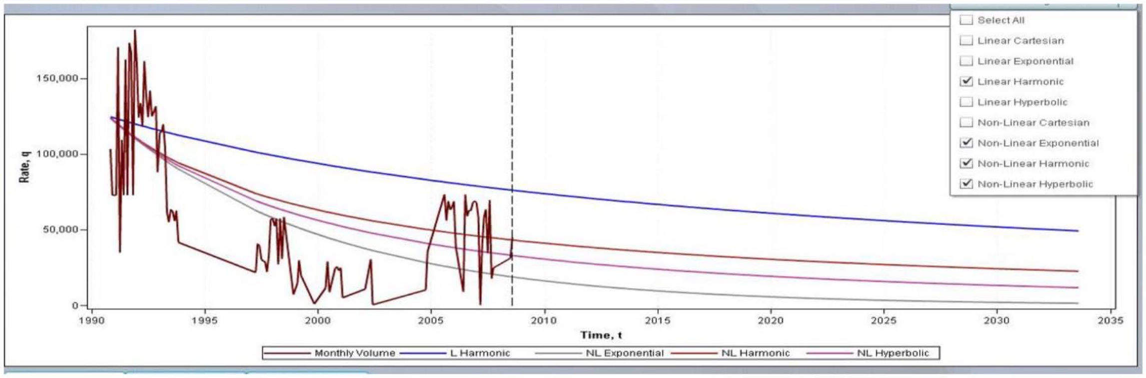


Figure 5.9— Decline Curve Analysis using SAS

Moreover, the integration of Type curve based analysis in the data analytics forecasting process provides a comprehensive and complete system of production parameter forecasting that gives the users the flexibility and confidence of using one common platform.

In the result section, we have provided the comparative analysis of data analytics driven forecasts and type curve forecasts to identify the improvement in forecast accuracy.

5.6 Step 6 - Schedule Monthly / Daily Rolling Forecast with incremental data

Continuous Rolling forecasts is a good practice. Each additional periodic piece of information is a part of a larger pattern in the time-series. Therefore it was important to consider the new data for each period and generate rolling forecasts for two years. This enabled us to visualize the patterns that were coming up in the short term, and the automated modeling process views it and provides weight to the recent trends, patterns that are developing in the field.

5.7 Step 7 - Monitoring the Accuracy of Forecast and Re-diagnose

To manage better one needs to monitor the outcome continuously. The forecast accuracy is measured periodically for every time-series with incremental actual TP data and the forecasted TP data. Trend lines of the error % - MAPE (fig) over a period of each field is plotted, and degradation (if any) observed. An alert is configured if the error % - MAPE value goes above a certain level. Every six months a thorough model performance audit is planned on each of the fields for the accuracy and need for re-diagnosis would be done for the time-series to detect new patterns in the time-series that would enable achieving higher accuracy.

6.0 Results

Total 74 fields across three regions in Malaysia was used for forecasting Technical potential at the monthly level. The forecast error MAPE was observed for the entire history and validation was done with 1/2 years ahead Out of Sample forecasts.

Analysis of the forecast accuracy

- *34 fields contributing to 90% of the production had total forecast accuracy greater than 90%. The 1-year blind test out-of-sample forecast accuracy was higher than 80%.*
- *43 fields contributing to 95% of the production had total forecast accuracy greater than 85%*

This indicates that the models were able to forecast quite accurately the technical potential in advance of 1 year. Rolling forecast is made with monthly data, and we see that the models were able to capture the structural changes and forecast accuracy improved with each additional data given to the model

The one year ahead blind test is better than the two years ahead of the forecast which is natural as forecasting accuracy decays with the length of the forecast horizon. Therefore from Technical Potential estimate standpoint, we could use the one-year ahead forecast, as confirmed Technical Potential to do the monthly planning and to improve the understanding of the reserves pool and the 2nd year forecast as "tentative," can be used for high-level planning for field development activities and discussed with the operators.

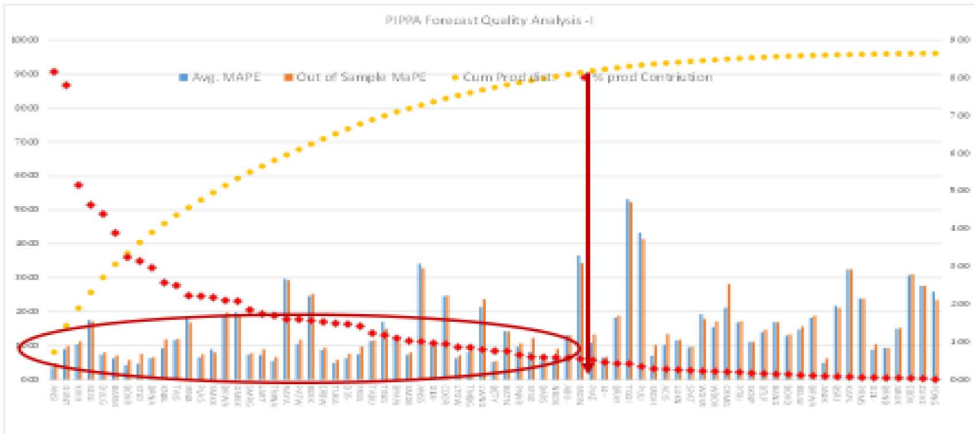
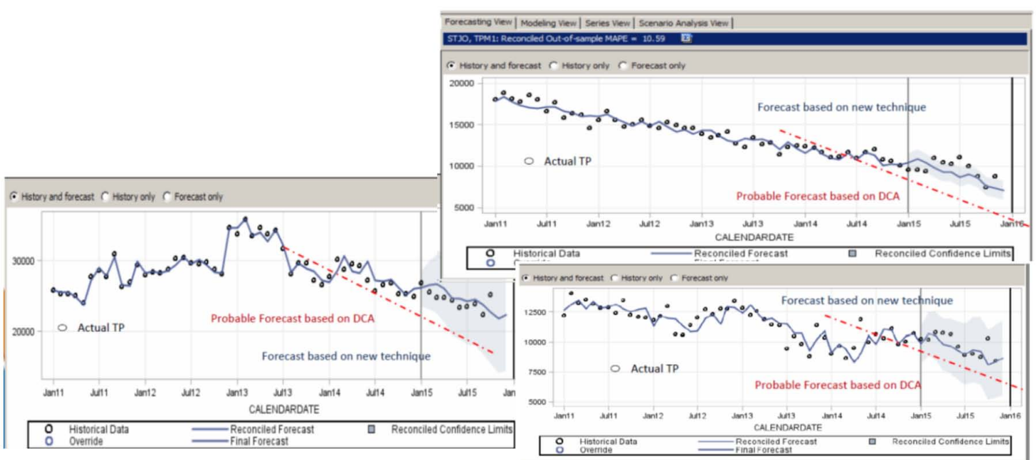


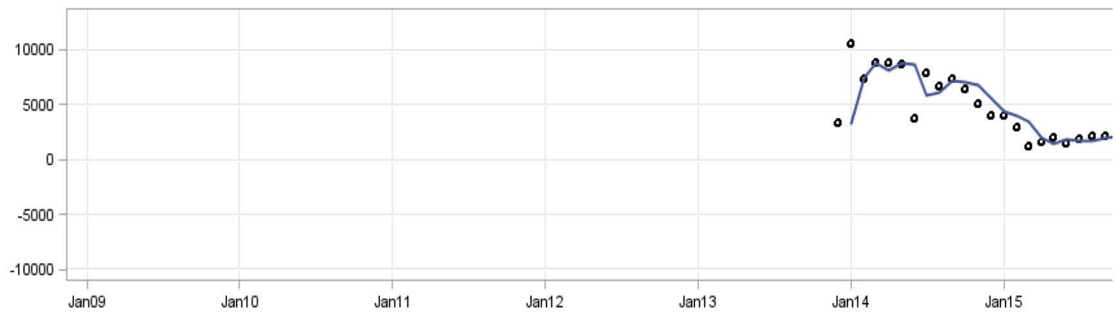
Figure 6.1—Over View of MAPE for all fields. Clear Indication of Forecast Robustness

The forecast estimates over a period of 2 years were compared with the possible decline curves for the fields. It was observed that the decline curves accuracy was significantly lower as compared to the data analytics driven forecasting.



Comparative benchmark analysis of forecast accuracy of data analytics based forecasting vs. Type curve based forecasting was undertaken as illustrated above. It shows that the data-driven analytics based forecasting can detect the patterns of the technical potential much better than the decline curve analysis and demonstrates significant improvement in forecast accuracy. This improved ability to predict the future technical potential would enable the planners to utilize the resources efficiently and effectively.

Due to shorter history, 6 active fields were not considered for the data analytics based forecasting and were segregated to be analyzed using Decline Curve Analysis. The process was integrated into the primary system with measurement of accuracy for the decline curve baseline forecast. The transition path from decline curve to data analytics forecasting method was built once sufficient history of Technical Potential is available for the fields.



Figure—Example of Field with Shorter History for Decline curve analysis

7.0 Accuracy Enhancements

To determine whether additional efforts for forecast improvements are making things better, it is important to look at "Forecast Value Add." Forecasting performance metrics mean absolute percent error (MAPE), shows the size of the forecast error. Considering FVA, the following parameters were used to filter the fields which would qualify for the enhancement efforts:-

1. % Contribution of the field to the overall production
2. % Error observed
3. Available history of time-series.
4. Forecast accuracy achieved using type curve (DCA) based forecasting

Considering the above, our focus would be mostly on the top producing fields. The current data analytics driven Technical Potential forecasting model used the univariate method and dummy regressors for outliers. Using this approach, the current levels of accuracy achieved are extremely encouraging and consistently performing. However, there were two aspects on accuracy enhancements have been planned in subsequent phases of the project. *Firstly*, the transition of the shorter time-series which are using Decline Curve Analysis into the data analytics is driven forecasting technique as soon as sufficient period of data is build up in the time-series. *Secondly,* the use of explanatory variables such as # of active wells planned field development activities and historical actual field developmental activities. To be able to utilize these variables in the forecasting process, necessary steps needs to be taken by the asset & operations for a systemic gathering of data and providing into the forecasting system.

8.0 Possible Impact on Industry

Major Oil & Gas Operators, especially NOCs face a decision that involves billions of dollars and thousands of lives essentially based on the production profiles. When it comes to short-term (1-2 years) the accuracy of the forecast is paramount since it could mean saving on a massive unnecessary investment. Higher accuracy of technical forecasts in advance of 6 months to 24 months provides the capability of upstream companies to plan the downtimes; either preempt it or delayed based on the overall production requirement leading to optimized investment.

In the case of PETRONAS, 9% incremental gain was achieved through negotiations based on the NOC assessments of TP using the algorithms described in the paper.

There is the certain significant qualitative impact that this exercise of data analytics driven forecasting brings to the entire organization:-

Making forecasting an enterprise process and not a case of multiple individual's silo effort facilitating collaboration across different units within the organization.

Setting a culture of measuring the forecasting accuracy as individual KPI in the Petroleum Management and Asset Management structure.

Establishing the importance of data and its relevance, surfacing the data quality issues and the need for standardization, need of developing a standard methodology for additional data collection and integration with the forecasting system.

Encouraged with the forecast accuracy achieved the analytics process is being extended from potential technical forecasting to the short-term (30-60-90 days) production forecasting. Initial the results obtained in the blind tests of the short-term production forecasting is, even more, encouraging with forecasting accuracy of 30-60 days be greater than 95% at field level for fields contributing to 90% of the production of the country. This will enable to optimize the planned and unplanned downtime and provide significant production optimization.

9.0 Way Forward

There is certain significant qualitative impact that this exercise of data analytics driven forecasting brings to the entire organization purely from the technique standpoint and the ability to leverage these insights into the planning process. Therefore Capacity building within the organization was one of the key aspects that was considered soon after this project and necessary knowledge transfer is being done to specific teams who would utilize these actionable insights while engaging with the operators for resource planning.

With promising results in the Technical Potential forecasting, the data-driven analytics based forecasting approach is currently being extended to the short-term production forecasting utilizing the planned and unplanned downtime data. The initial outcome of the project with measurement of the 30 days / 60-day blind test production forecast accuracy is showing excellent results.

Once the short term production forecasting is complete, the authors plan to write a paper on the challenges of short-term production forecasting and how data-driven analytics forecasting is integrating multiple PD / UPD data could bring in substantial accuracy improvement thereby bringing in operational advantage to the company.

10.0 Bibliograph

No	Term	Meaning
1	TP	Technical Potential
2	NOC	National Oil company
3	PSC	Production Sharing Contractors
4	PAC	Petroleum Arrangement Contractors
5	RSC	RISC Service Contractors
6	PDOIL	Producing Days Oil Rate
7	RMP	Reservoir Management Plan
8	WPB	Business plan (commonly Five-year duration)
9	ARIMA	Auto Regressive Integrated Moving Average
10	UCM	Unobserved Component Model
11	ESM	Exponential Smoothing Model
12	MAPE	Mean Absolute Percent Error
13	FVA	Forecast Value Added

References

Technical Potential, A Petronas Guideline, Vipin Prakash Gupta, 2016

SAS/ETS® 13.2 User's Guide, PROC ARIMA User Guide https://support.sas.com/documentation/onlinedoc/ets/132/arima.pdf

Combined Forecasts: What to Do When One Model Isn't Good Enough, Ed Blair, Michael Leonard, Bruce Elsheimer, SAS Institute Inc. http://support.sas.com/resources/papers/proceedings12/341-2012.pdf

Introduction to ARIMA Models by Robert Nau, Fuqua School of Business, Duke University

SAS/ETS® 13.2 User's Guide, PROC ARIMA User Guide, PROC UCM User Guide https://support.sas.com/documentation/onlinedoc/ets/132/ucm.pdf

Unconventional Data-Driven Methodologies Forecast Performance In Unconventional Oil and Gas Reservoirs, Keith R. Holdaway, Louis Fabbi, and Dan Lozie, SAS Institute Inc

Reserve Technical Potential Management System: Key in Rehabilitation of Mature Field M. Ruslan | F. Caretta | R. G. Artola (Schlumberger) | J. M. Gandolfo (YPF SA)

Production Forecasting with Logistic Growth Models A. J. Clark, L. W. Lake, T. W. Patzek, University of Texas at Austin

Malaysia crude oil production estimation: an application of ARIMA model, published on IEEE Science and Social Research (CSSR), 2010 International Conference by Nazuha Muda Yusof; Ruzaidah Sulong A. Rashid; Zamzulani Mohamed

Crude oil reserve estimation: An application of the autoregressive integrated moving average (ARIMA) model Journal of Petroleum Science and Engineering https://www.researchgate.net/journal/0920-4105_journal_of_petroleum_science_and_engineering **8** (1): 13–28 · July 1992

Forecasting Oil Production Time Series with a Population-Based Simulated Annealing Method, Article in ARABIAN JOURNAL FOR SCIENCE AND ENGINEERING **40** (4) · April 2015, Juan Frausto-Solis, Manuel Chi-Chim, Leonid Sheremetov

Appendix I

Method 1 and 3 – Part 1

1. Global ratio – SUM(PA) / SUM(WT)

   A. GR1 – Global ratio for the same strings overpass period (24 months)
   B. GR2 – Global ratio for strings within the same field in the same month
   C. Individual ratio – MEDIAN(PA/WT)
   D. IR1 – Individual ratio for the same strings overpass period (24 months)
   E. IR2 – Individual ratio for strings within the same field in the same month
   F. The final ratio of method one will be taking the median of GR1 and IR1, and the final ratio of method two will be taking from the median of GR2 and IR2.
      For example:
      GR1 = 0.94
      GR2 = 0.85
      IR1 = 0.97
      IR2 = 0.92
      Ratio for Method 1 is **MEDIAN(GR1,IR1)** = 0.955
      Ratio for Method 3 is **MEDIAN(GR2,IR2)** = 0.885
2. TP will be calculated by PA/<Final Ratio> when WT **NOT EXISTS,** or WT is **LOWER** than PA

By any chance, if the final ratio is **NOT** presented or **HIGHER** than **Method 1 and 3 – Part 2**
Requirements

1. Both PA and WT must exist on the individual basis (Applies to all)
2. WT must be higher than PA on the individual basis (Only applies to GR1 and IR1)
3. PA/WT must be higher than 0.7 on the individual basis (Only applies to GR1 and IR1)

TPM3 : The algorithms is under copy right to PETROANS, so the details can not be shared here.