# Paper 342-2012 Let Oil and Gas Talk to You: Predicting Production Performance

Keith R. Holdaway, SAS Institute Inc., Cary, NC, USA

## ABSTRACT

How do historical production data relate a story about the subsurface oil and gas reservoirs? Business analysts must perform accurate analysis of reservoir behavior using only rate and pressure data as a function of time.

This paper introduces methodologies to forecast oil and gas production by exploring implementations of the AUTOREG, ESM, and MODEL procedures in SAS/ETS<sup>®</sup>. The AUTOREG procedure estimates linear regression models when the errors are autocorrelated. The ESM procedure generates forecasts by using exponential smoothing models. Examples of the MODEL procedure arising in subsurface production data analysis are discussed. In addressing these examples, techniques for pattern recognition, implementing TREE, CLUSTER, and DISTANCE procedures in SAS/STAT<sup>®</sup> are highlighted to explicate the importance of oil- and gas-well profiling to characterize the reservoir.

# INTRODUCTION

The Reservoir Management Department (RMD) of Saudi ARAMCO National Oil Company (NOC) analyzes tens of thousands of producing wells across multiple oil and gas fields that cover hundreds of heterogeneous sandstone and carbonate reservoirs on the Arabian Peninsula. It is both complex and challenging to estimate reserves and forecast production rates from a short- and long-term perspective in the oil and gas industry. SAS proposes an advanced analytical methodology that encompasses a data preparation workflow to ensure integrity and reliability in the forecasts, and a workflow that ensures models adhere to Arp's<sup>1</sup> first principals that reservoir and production engineers traditionally implement. This suite of algorithms known as Arp's equations assume limited changes in operating strategies and make assumptions about the pressure and drive mechanisms in the reservoir. The oil and gas industry knows this process as Decline Curve Analysis (DCA)<sup>2</sup>. Engineers determine Estimated Ultimate Recovery (EUR) of reserves in existing reservoirs and forecast production trends implementing the equations in **Table 1**.

	Exponential	Hyperbolic	Harmonic
	b=0	0 <b<1< th=""><th>b=1</th></b<1<>	b=1
D1	ln(q <sup>1</sup> q <sup>2</sup> ) / t	D <sup>1</sup> (q <sup>1</sup> /q <sup>2</sup> )^1/b	D <sup>1</sup> (q <sup>1</sup> /q <sup>2</sup> )
q <sup>1</sup>	q <sup>1</sup> exp(-D <sup>1</sup> t)	q <sup>1</sup> /(1 + btD <sup>1</sup> )^1/b	q <sup>1</sup> /(1 + btD <sup>1</sup> )
Qp	q1 - q2/D1	q <sup>1</sup> /D <sup>1</sup> (1 – b)(1 - q <sup>2</sup> /q <sup>1</sup> )^1-b	$q^{1}/D^{1}ln(1 + D^{1}t)$
t	ln(q <sup>1</sup> q <sup>2</sup> )/D <sup>1</sup>	(q <sup>1</sup> /q <sup>2</sup> )^b – 1/bD <sup>1</sup>	q <sup>1</sup> - q <sup>2</sup> /D <sup>1</sup> q <sup>2</sup>

#### Table 1: Arp's Equations with Varying Constants, b Values

Below you can view the glossary of terms for the Arp's equations<sup>3</sup> to better appreciate the first principles that underpin current forecast model workflows. The input data are the historical production observations that engineers collect periodically at the producing wells across the oil and gas fields. When the constant 'b' is equal to '0', it is an exponential type curve; when equal to '1', it is a harmonic type curve and when the value of 'b' is greater than '0', and less than '1', it is a hyperbolic type curve. **Figure 1** depicts the traditional DCA methodology that embraces the necessary steps to enable production engineers to interpret the type curves and estimate performance in the shortand long-term.

- D<sup>1</sup> Initial rate of decline
- q<sup>1</sup> Initial rate of production
- q<sup>2</sup> Rate of production
- Qp Cumulative production
- t Time



#### Figure 1: Decline Curve Analysis Methodology

#### Challenge

The challenge for Saudi ARAMCO is to develop a consistent, robust, and effective methodology to analyze across disparate upstream silos the production data: oil, gas, and water. Such a solution enables the optimization of both production and field development in a user-friendly Web-based experience.

Reservoir engineers analyze thousands of wells to estimate oil reserves. But owing to high volumes of data on each well and the capabilities of existing systems, it is currently feasible to include only a limited number of wells in the analysis. To compensate, engineers use a deterministic sampling process that increases the uncertainty of forecasts. Engineers manually flag data errors and outliers in a time-intensive process. When Saudi ARAMCO mandate reviews of field redevelopment plans every three months, it is time for a new system to reduce resource constraints and to shorten the decision-making cycles.

#### Solution

The SAS<sup>®</sup> for Oilfield Production Forecasting solution offers a customized data management and advanced time series analytical approach that addresses the current inherent issues of a deterministic workflow by adopting a probabilistic perspective.

#### Results

It is now easier to identify productivity improvements in mature assets. The solution is a DCA workflow that rapidly flags problematic wells with robust data and reliable statistical accuracy indicators.

Adopting an advanced analysis of a decline curves methodology provides important forecasting results for future production from a single well or multiple wells across reservoirs and fields. SAS approaches the problem by delivering a product that helps you do the following:

- 1. Access well and reservoir data quickly and easily with an intuitive graphical user interface that integrates and profiles forecasting data, identifies outliers and missing values, and lets you select appropriate treatment strategies.
- Apply analytical functions consistently across the organization with a forecasting solution that automatically selects the best forecast model, but allows human intervention and benchmarking for repeatability. For gas reservoirs, multiple embedded water encroachment models are available.
- 3. Perform DCA quickly and flexibly with a robust analytical engine that provides what-if scenarios to model EUR based on a predefined range of easily adjustable, industry-standard default values.

4. Estimate unconventional sources of well production more accurately using best-fit prediction that uses smaller data sets when large volumes of historical data are not available.

## SAS FOR OILFIELD PRODUCTION FORECASTING

"We are not trying to create a program that replaces applications like OFM and ARIES, but one that will automate the DCA processes to a high enough accuracy that an engineer will only have to look into the wells (or reservoirs or fields) that are *problems*. The main objective is for production optimization and development planning, not reserves estimation."

Gary M. Williamson, Reservoir Engineer Saudi ARAMCO

Oilfield Production Forecasting introduces a refreshingly intuitive navigation of the disparate data sources and input parameters that are in alignment with a comprehensive suite of forecasting models. You can access these functionalities in a Web-based solution. The user is able to analyze an extensive array of "what-if" combinations of critical input parameters across multiple forecasting models. Whether working with oil or gas reservoirs, the user has access to a rich compendium of forecasting techniques that detail accuracy indicators and interactive graphs.

There is a need for reservoir engineers to establish a single information analysis integrated platform, and this requisite motivates the design and implementation of a Web-based DCA in an integrated portal. This platform integrates different data sources and applications that make up a reservoir engineers' toolbox for all fields and reservoirs within the scope of study, saving them the time to search and manipulate the data.

The Web-based DCA system enables engineers to dynamically perform analyses on field and reservoir data for individual wells or groups of wells by using different methods in an integrated portal. It provides several functions like best-fit forecasts for Cartesian, exponential, harmonic, and hyperbolic analysis, using linear and nonlinear regression with manual correction (both for curves and parameters). It also saves the production forecast along with parameters in a database to perform graphical comparison between multiple forecasts and manual input of empirical parameters (constant rate to percent reserve, reserves, decline rate, and decline type). This implementation allows engineers to perform dynamic production analysis with more advanced forecasting techniques, which is effective in business planning to generate the production forecast for the new wells and no drill case for the existing fields' wells.

#### **PRE-FORECASTING PREPARATION**

The SAS solution implements several preparatory steps in order to ensure more accuracy in the PROC MODEL execution.

The STEPAR method consists of the following computational steps:

- 1. Fit the trend model as specified by the TREND= option using ordinary least squares regression.
- 2. This step de-trends the data. The default trend model for the STEPAR method is TREND=2, a linear trend model.
- 3. Take the residuals from step 1 and compute the autocovariances to the number of lags specified by the NLAGS= option.
- 4. Regress the current values against the lags, using the autocovariances from step 2 in a Yule-Walker framework. Do not bring in any autoregressive parameter that is not significant at the level that is specified by the SLENTRY= option. (The default is SLENTRY=0.20.) Do not bring in any autoregressive parameter that results in a non-positive-definite Toeplitz matrix.
- 5. Find the autoregressive parameter that is least significant. If the significance level is greater than the SLSTAY= value, remove the parameter from the model. (The default is SLSTAY=0.05.) Continue this process until only significant autoregressive parameters remain.
- 6. If the OUTEST= option is specified, write the estimates to the OUTEST= data set.
- 7. Generate the forecasts using the estimated model and output to the OUT= data set. Form the confidence limits by combining the trend variances with the autoregressive variances.

The solution tolerates missing values in the series and estimates autocorrelations from the available data. This method requires at least three passes through the data: two passes to fit the model and a third pass to initialize the autoregressive process and write to the output data set.

```
/* Run a Default Linear Model */
PROC FORECAST
       DATA = WORK.TMP0TempTableInput
       OUT=WORK.LinearForecast INPUT(LABEL="Forecast of WORK.FORECAST INPUT")
       OUTALL
       OUTEST=work.FCS1PARAMETERESTIMATESINPUT 0002
       METHOD=STEPAR
       INTERVAL=MONTH
/* Forecast Period default is 240 Months */
       LEAD=&FORECASTPERIOD.
       TREND=2
       ALPHA=0.05;
       ID Time;
       VAR MEASURE;
RUN;
/* Define Arps' Equations and Models Calculation */
PROC SQL;
CREATE TABLE WORK.PRED OUTPUT AS
        SELECT PRED Table.T ,
                PRED Table.Time,
                 PRED_Table.MEASURE,
(&Lin_Expo_crate. * EXP(&Lin_Expo_Di. * PRED_Table.T )) AS L_PRED_EXP,
(&NLin_Expo_crate. * EXP(&NLin_Expo_Di. * PRED_Table.T )) AS N_PRED_EXP,
(&Lin_Harm_crate. / (1- &Lin_Harm_Di. * PRED_Table.T )) AS L_PRED_HAR,
(&NLin_Harm_crate. / (1- &NLin_Harm_Di. * PRED_Table.T )) AS N_PRED_HAR,
(&Lin Hyp crate. / (1- &Lin Hyp n. * &Lin Hyp Di. * PRED Table.T) ** (1/ &Lin Hyp n.))
AS L PRED HYP,
(&NLin_Hyp_crate. / (1- &NLin_Hyp_n. * &NLin_Hyp_Di. * PRED_Table.T) **
(1/ &NLin_Hyp_n.)) AS N_PRED_HYP
        FROM Work.PRED Table AS PRED Table;
```

QUIT;



Figure 2: SAS for Oilfield Production Forecasting

In **Figure 2**, you see an example of the visualization that enables engineers to compare rates of oil production for a single well over a finite period of time. Engineers adopt both linear and non-linear models to fit the monthly oil production volumes input into the PROC MODEL code.

The graphical output in **Figure 2** illustrates a non-linear harmonic (b=1), a non-linear exponential (b=0) and a non-linear hyperbolic (0 < b < 1) as well as a linear harmonic suite of trends. You can select the optimum or best fit trend/curve based on deterministic experience. You can also endorse *a priori* interpretation by studying the statistical accuracy indicators that the SAS for Oilfield Production Forecasting solution determines post executing the following macro code:

```
----- */
/* MAE */
       E L PRED CART = ABS(Error);
       Sum E L PRED CART + E L PRED CART;
     _____
 2- Mean Absolute Percentage Error (MAPE)
     ..... */
/* MAPE */
       if ABS(MEASURE) NE 0 then do;
           E2 L PRED CART = E L PRED CART / ABS (MEASURE);
       end;
       else E2 L PRED CART = E L PRED CART;
       Sum E2 L PRED CART + E2 L PRED CART;
/* _____
 3- Percent Mean Absolute Deviation (PMAD)
 ----- */
/* PMAD */
       E3 L PRED CART = ABS (MEASURE);
       Sum E3 L PRED CART + E3 L PRED CART;
/* _____
 4- Root Mean squared Error (RMSE)
 ----- */
/* MSE & RMSE */
       E4 L PRED CART = (Error**2);
       Sum E4 L PRED CART + E4 L PRED CART;
   end;
```

#### **MODEL-BASED BOOTSTRAP**

Saudi ARAMCO is seeking an improved methodology for probabilistic quantification of reserves estimates using DCA. To evaluate the uncertainty in reserves estimates using a probabilistic approach, it is reasonable to provide a distribution of reserves estimates with three confidence levels (P10, P50, and P90) and a corresponding 80% confidence interval. You might ask: how reliable is this 80% confidence interval? Is the true value of reserves contained within this interval 80% of the time? Using traditional statistical analyses, it is apparent that true values of reserves estimates lie outside the 80% confidence interval much more than 20% of the time. Thus, you are underestimating uncertainty, often significantly. The challenge then is to determine not only how to appropriately characterize probabilistic properties of a complex production data set, but also how to improve the reliability of the uncertainty quantifications. In order to preserve the data structure, Figure 3 employs a rigorous model-based bootstrap algorithm. It uses the decline models (hyperbolic or exponential equations) to fit the production data and constructs residuals from the fitted model and observed data. You then generate a new series by incorporating random samples from the residuals into the fitted model. To consider correlation within the residuals and to preserve data structure, you use a block resampling approach to generate residual realizations. To determine the size of the blocks, you use the autocorrelation plot of residuals, which can help to detect the randomness or possible correlations within residual data and confidence band that can help to detect significantly nonzero points out of the band of a particular confidence level on the autocorrelation plot.



# Figure 3: Model-based Bootstrap Methodology /\* Estimating Linear Trend Model \*/

```
PROC AUTOREG data=WORK.Selection;
model y = t / nlag=2 method=ml
run;
/* Forecast each of the monthly time series */
PROC ESM data=WORK.Selection out=outesm print=forecasts;
             id DATE interval=month;
             forecast y;
             run;
PROC MODEL data=WORK.Selection(where=(rate > 0));
             %if &lin Fix QI eq Y %then %do;
                    parms a=0;
                    if &lin Fixed QI**2-2*production*a < 0 then
                           predicted rate = 0;
                    else
                           predicted rate = (&lin Fixed QI**2-2*production*a)**0.5;
                           outvars qi a;
                    %end; %else %if &lin_Fix_A eq Y %then %do;
                           parms qi=&qi;
                           bounds qi >= 0;
                           if (qi^*2-2^*production^*\&lin Fixed A < 0) then
                                  predicted rate = 0;
                           else
                                  predicted rate=(qi**2-
                           2*production*&lin Fixed A)**0.5;
                           outvars qi;
                    %end; %else %do;
                           parms a=0 qi=&qi;
                           bounds qi >= 0;
                           if qi^*2-2^* production*a < 0 then
                                  predicted rate = 0;
                           else
```

## SEGMENTATION OF AN OILFIELD

Geologists, reservoir, and petroleum engineers and geophysicists attempt to breakdown the geo-scientific silos and through collaborative methodologies aggregate the disparate data sets and determine ideal locations for field reengineering in mature or brown fields.

The Mature Fields Evaluation consists of these three steps:

- Implement graphical user interfaces in the preliminary stages (control and data validation) to aid the global field analysis and diagnose the producing wells.
- Transform the data (modeling, smoothing, extraction of production indicators) to make data conducive to the multivariate analysis (MVA) techniques.
- Use statistical clustering procedure to segment the field into different zones based on well profiling.

The goal of the analytical process is to understand the distribution of the Water Cut (Sw) values in relationship to the cumulative liquid production across the field.

By profiling the individual wells according to certain criteria such as Water Cut, minimum Free Water Level distance (FWL), cumulative liquid, well bore type within different time phases and incremental geographic regions, it is possible to classify those wells and appreciate through the analytical indicators of similarity/dissimilarity a potential segmentation of the field.

Geoscientists map this appreciation of the statistical results to identify the production mechanisms such as best producers, depletion, pressure maintenance, and thus locate poorly drained zones and potentially lucrative field reengineering tactics and strategies based on a more complete understanding of the distribution of water throughout the field.

After normalizing the Water Cut curves with the cumulative liquid, the problem of finding similarities between wells basically becomes finding a similarity between the water cut and cumulative production scatter plots.

There is a wide variety of distance and similarity measures available in the cluster analysis technique, but as the well data are now in a coordinate form, it is appropriate to use a non-Euclidean distance for clustering, computing a distance matrix by using the DISTANCE procedure.

You can then convert similarity measures to dissimilarities prior to running the PROC CLUSTER. You implement the DISTANCE procedure to compute the Jaccard coefficient between each pair of wells. The Jaccard coefficient is defined as the number of variables that are coded as 1 for both wells divided by the number of variables that are coded as 1 for either or both wells. The CLUSTER procedure requires dissimilarity measures, and you then select the DJACCARD coefficient.

Figure 4 illustrates the dissimilarity matrix. A small value means that two wells are similar with respect to their water cut mark.

```
proc distance data=ARAMCO.M_ClusteringInput
    (where= (TimeWindow='Window1'))
    method=DJaccard absent=0 out=ARAMCO.M_DistanceMatW1;
    var anominal(G1-G100) anominal(WellType) interval(DistFWL) ;
    id WellBoreName;
run;
proc cluster data=ARAMCO.M_DistanceMatW1 method=Ward outtree=ARAMCO.M_treeW1;
    id WellBoreName;
run;
```



Figure 4: Dissimilarity Matrix Using Jaccard Metric

The hierarchical cluster analysis successively joins data points to form groups of similar behavior until all records are joined in one group.

You can see in **Figure 5** the default resulting graph of the hierarchical cluster analysis. It is similar to a tree structure starting with the single records at the bottom and finishing with all records joined into one single group at the top.

This scales the graph based on the distances of the groups and makes dissimilarities more visible. The higher the distance is to the next join (represented by the lines in the graphs that join previous groups), the larger the distance between the groups and the more dissimilar the groups. You can draw conclusions from the dendrogram such as the presence of two distinct groups in the data around the water cut and cumulative liquid behavior of the 43 wells that constitute the cluster in time-window 1, thus generating two clusters. It is then possible to profile the two clusters to identify the main characteristics of each group. You obtain the clusters by implementing the Ward algorithm for minimum variance.





#### Type Wells

Engineers often use type well methodologies to predict production so as to determine the performance of future wells. Although predicting production based on analog can be considerably uncertain, it might be the only method available when specific information such as porosity, reservoir thickness, and saturations about a location is unknown. One of the objectives is to provide a workflow that develops type wells based on unique statistically valid selection methods. You can then predict the production performance in development planning using the type wells.

This application allows the automated grouping of wells based on multiple selection criteria (rates, water cut, etc.) into Type Well classes. You then obtain a statistically based selection of the valid production for each type well grouping and then develop a production forecast using DCA.

An important step in the process of DCA is the selection of production data used for the forecast. Besides the decline of the reservoir, there can be technical or economic reasons to cut back production for a well or a group of wells. The engineer must decide based on his or her knowledge which data to use for the forecast. An automated filtering and selection of data can support the decision process.

Normally the reservoir engineer selects a time period where a constant production rate and decline is visible. Sometimes the minimum number of valid data points of month-on-month production rates for a time series analysis cannot be represented with the data available. An automated data selection supports the engineer to solve the following two basic issues:

- 1. replacing or excluding zero production rates
- 2. smoothing extreme variance in the data (for example, production cuts due to economic reasons, well tests)

While excluding zero production rates is straightforward, mathematics helps in the detection of outliers. In probability theory and statistics, the standard deviation of a statistical population, a data set, or a probability distribution is the square root of its variance. A data point is statistically considered an outlier if the value is more than two times the standard deviation from the mean. These are possible automatic treatments of detected outliers in a given data population:

- filtering out of all records with outliers
- filtering out records with zero production rates only
- replacing all outliers with the mean or median production rate

The value of mean and median of the data selected has been calculated excluding zero production rates. That ensures that the result of the mean and median calculation creates equal results with or without removing the zero production rates first.

For time series data used for forecasting, the identity and order of the observations are crucial. A time series is a set of observations made at a succession of equally spaced points in time. If records have been filtered out, this rule of thumb has been violated. The simple solution to the problem is to keep the order of the observations and to adjust the time ID of the records in the data population.

The system provides interpolation methods in its analytical software (PROC EXPAND) that could be used to interpolate missing values or to generate lower frequency output from higher frequency data (for example, generating quarterly estimates from monthly production data).

## CONCLUSION

The Web-based DCA system's current broad range of time series and forecasting capabilities enables users to model, forecast, and simulate business processes for improved field redevelopment strategies and well management. Users can model complex business scenarios and analyze the dynamic impact-specific events might have on the lifespan and production of an asset, be that a field or an individual well.

The system provides an intelligent application environment in which data from unrelated systems can be gathered, stored, analyzed, and distributed in a simple and timely manner. These technologies allow disparate systems to contribute data and information to an integrated, enterprise-wide business intelligence strategy. This ensures that the production data and all requisite reservoir data can be aggregated to provide a robust forecasting experience. In the future, the addition of grid computing will offer a cost-effective solution for customers who want to accelerate the forecasting process or increase the scale or scope (number of users, size of data sets, and the frequency of analysis) of the DCA.

The flexibility of the DCA solution enables future access to data mining processes to create highly accurate predictive and descriptive models based on vast amounts of data gathered from across an enterprise. It will also open up work flows for interactive data analysis and visualization capabilities, broad data preparation features, and deep analytics, thus enhancing a multi-faceted experience to field management and robust short- and long-term forecasts.

The system provides statistical tools that include the ability to automate the DCA process by working backwards through the data. It can measure the compatibility of fit of the data and then select only the data that fits into one production profile and do the decline on that data. This will allow us to compute declines and production forecasts for every producing well in the database with marginal impact on manpower.

### REFERENCES

- 1. J.J. Arps, "Analysis of Decline Curves," Trans., AIME 160 (1944): 228-247.
- L. Kewen and N.H. Roland. 2003. "A Decline Curve Analysis Model Based on Fluid Flow Mechanisms." 83470-MS. *Proceeding of the SPE Western Regional/AAPG Pacific Section Joint Meeting*. Society of Petroleum Engineers.
- 3. SPEE. 2000. "SPEE Recommended Evaluation Practice #6 Definition of Decline Curve Parameters." Available at http://www.spee.org/images/PDFs/ReferencesResources/REP06-DeclineCurves.pdf.

# ACKNOWLEDGMENTS

The author wishes to acknowledge the research and contributions of Gary M. Williamson, Dennis Seemann, and Husameddin Madani, Saudi ARAMCO.

## **CONTACT INFORMATION**

Your comments and questions are valued and encouraged. Contact the author:

Keith R. Holdaway SAS Institute Inc. C5188 SAS Campus Drive Cary, NC, 27513 Work Phone: +1 919 531 2034 Fax: +1 919 531 2034 E-mail: keith.holdaway@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.