| | |
|---|---|
| Please fill in the name of the event you are preparing this manuscript for. | SPE Eastern Regional Meeting, 2018 |
| Please fill in your 6-digit SPE manuscript number. | SPE-191788-MS |
| Please fill in your manuscript title. | Accounting for Serial Autocorrelation in Decline Curve Analysis of Marcellus Shale Gas Wells |


Please fill in your author name(s) and company affiliation.

| Given Name | Surname | Company |
|---|---|---|
| Eugene | Morgan | The Pennsylvania State University |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |





## Abstract

Current decline models fail to capture all of the behavior in shale gas production histories. That is, upon fitting one of these models, one often sees significant and sustained deviation of the flow rate data points from the decline trend. One way to measure this "lost signal" is to look at the autocorrelation in the residuals about the fitted decline model. Indeed, with many shale gas wells we see significant amounts of autocorrelation, especially when comparing the flow rate at one time to the next (lag one). Theoretically, this serially autocorrelated error can impact decline curve analysis in two ways: 1) inefficient estimation of decline curve parameters, and 2) lost signal in the data. Borrowing from time series statistics, there are two conventional ways of dealing with these potential problems: 1) estimate the decline curve parameters with generalized least squares or generalized nonlinear least squares, and 2) fitting an ARMA model to the residuals and adding it to the fitted decline curve.

This paper investigates the practical implications of these two procedures by exercising them over decline curves fit to 8,527 Marcellus shale gas wells (all wells from that play with viable data for the analysis). The study explores the effect that generalized regression methods and ARMA-modeled residuals have on six different decline curves, and performance is measured in terms of sum of squared residuals (a metric for goodness-of-fit, calculated on the training data (first 24 months of each record)) and mean absolute percent error (a standard metric for forecasting accuracy, calculated on the testing data (all production rates after 24 months)).

We find that inclusion of the ARMA-modeled residuals largely improves the goodness-of-fit for any decline curve, and improves the forecasting accuracy for the Hyperbolic decline curve and Duong's model. The use of generalized least squares or generalized nonlinear least squares has little benefit in fitting the decline curves, except for the Logistic Growth model, where it improves both fit and forecasting accuracy.


## Introduction

Decline curve analysis (DCA) serves as a popular method for determining EUR and forecasting production. By only requiring past production rate data, DCA has an advantage over other methods (Seidle et al. 2016). The initial decline curves were designed for conventional reservoirs by Arps (1945) in two forms: the Exponential model,

$$q_t = q_i \exp(-D_i t), \quad (1)$$

and the Hyperbolic/Harmonic model,

$$q_t = q_i (1 + D_i b t)^{-1/b}, \quad (2)$$

where $0 < b < 1$ is hyperbolic and $b = 1$ is harmonic. Many unconventional wells show $b > 1$. Some modern advances are designed for shale gas production, such as the Power Law Loss-ratio model (Ilk et al. 2008),

$$q_t = q_i \exp(-D_\infty t - D_i t^n), \quad (3)$$

the Stretched Exponential (Valko and Lee 2010),

$$q_t = q_i \exp(-(t/\tau)^n), \quad (4)$$

the Logistic Growth model (Clark et al. 2011),

$$q_t = K n t^{n-1}/(a + t^n)^2, \quad (5)$$

and Duong's model (Duong 2011),

$$q_t = q_i t^{-m} e^{\frac{a}{1-m}(t^{1-m}-1)} + q_\infty. \quad (6)$$

In all of the above equations, our independent variable is time $t = 1, 2, \ldots, T$ and our dependent variable is flow rate $q_t$. As one can see, these models have different coefficients, although some share common ones (e.g., initial flow rate $q_i$).

Despite the inherently temporal nature of production histories, there exists little published work that applies time series statistics to modeling and forecasting of well production rates. A few studies have applied time series statistics (ARIMA modeling) to production (Ayeni and Pilat 1992; Ediger et al. 2006; Yusof et al. 2010). Machine learning offers another source of improvement beyond decline curves. There has been some work exploring the application of nonlinear autoregressive neural networks with exogenous inputs (NARX networks) to production data (e.g., Sheremetov et al. 2014). In another example, Frausto-Solís et al. (2015) forecast oil production with a "Simulated Annealing based on Machine learning" approach. Cheng et al. (2010) examine the autocorrelation of residuals (after fitting Arps' hyperbolic model), but only in order to determine the appropriate block size for a block bootstrapping approach to probabilistic forecasting. No discussion is given to the effect of autocorrelated error on the estimation of the decline curve parameters, and no attempt is made to model this autocorrelation structure in the residuals.

Autocorrelation, $\rho(h)$, refers to the degree which an observation at time, $x_t$, depends on any previous observations, $x_{t-h}$, where $h$ is called the "lag". Autocorrelation can be estimated from a sample by

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad (7)$$

where $\gamma(h)$ is the autocovariance at lag $h$. It is estimated by

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \overline{x})(x_t - \overline{x}), \quad (8)$$

where $\overline{x}$ is the sample mean of $x$. Much of the dependency of production rates on time is described by any one of the decline curves above. However, this paper shows that significant autocorrelation can exist in the residuals after fitting a decline curve. This finding is important for two main reasons: 1) estimators of decline curve parameters that do not consider this autocorrelated error are inefficient; and 2) this autocorrelation in the residuals can be modeled and added to the decline curve fit, leading to a better performing model of production decline.

To elaborate on the first reason, consider the linear regression equation:

$$y_t = \beta' x_t + \varepsilon_t. \quad (9)$$

An implicit assumption in fitting this equation (by least squares or maximum likelihood) is that the error at each time, $\varepsilon_t$, is independently and identically distributed (i.i.d.), which leads to the ordinary least squares (OLS) estimator for the coefficients $\beta$ being

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (10)$$

This assumption is broken when the residuals display autocorrelation, which implies dependency between times. While this doesn't theoretically introduce any bias in the estimation of the coefficients, it does make their estimation inefficient, which means that the variance of the estimated value is not minimal (Cochrane and Orcutt 1949). One can mitigate this inefficiency through "generalized least squares" regression, in which the correlation structure of the error is specified. The OLS estimator above then becomes

$$\hat{\beta} = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y}. \quad (11)$$

This is the generalised least squares (GLS) estimator of the linear regression coefficients, in which the covariance matrix $\boldsymbol{\Sigma}$ contains the serial autocorrelation in the off-diagonal elements (the diagonal elements contain the variance of $\varepsilon_t$). If $\boldsymbol{\Sigma}$ is unknown (which it usually is), one can estimate it by exercising Eq. 8 on the residuals from the OLS regression. This same premise holds true for nonlinear regression with nonlinear least squares (NLS) (Gallant and Goebel 1976), which is used in fitting all decline models presented above, except Arps' Exponential model, which can be linearized.

Generalized (nonlinear) least squares only pertains to the estimation of the regression coefficients. Even after estimating the coefficients in this manner, the residuals may still contain autocorrelation. Such structure can be modeled and incorporated in the decline curve, say in an additive way. Suppose we generically define the decline curve as $f(t;\theta)$, where $\theta$ is the vector of decline curve parameters we wish to estimate, then our regression equation is

$$q_t = f(t;\theta) + e_t. \quad (12)$$

If we know that $\varepsilon_t$ is at least partially composed of an autocorrelated signal, we can capture that signal with an autoregressive model of order $p$ (AR($p$)):

$$e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + \cdots + \phi_p e_{t-p} + w_t, \quad (13)$$

where now $w_t$ is i.i.d. Gaussian white noise with zero mean. Alternatively, the moving average model of order $q$ (MA($q$)), treats the signal as a linear combination of white noise terms:

$$e_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}. \quad (14)$$

The mixed autoregressive moving average (ARMA) model is then:

$$e_t = \phi_1 e_{t-1} + \phi_2 e_{t-2} + \cdots + \phi_p e_{t-p} + w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \cdots + \theta_q w_{t-q}, \quad (15)$$

where $\phi_p \neq 0$ and $\theta_q \neq 0$, and the time series $e_t$ is stationary (Shumway and Stoffer 2010).

While these theoretical points should be taken into consideration when fitting decline curves to production data, whether they afford any practical benefit remains to be tested. This is the primary goal of this paper: to perform a thorough comparative quantitative analysis between the traditional regression procedure for decline curves (OLS or NLS) and 1) regression via generalized (nonlinear) least squares (GLS or GNLS), 2) additive inclusion of an ARMA model with the decline curve, and 3) the combination of GLS/GNLS with ARMA-modeled residuals.

Another goal of this paper is to demonstrate the proposed improvements to DCA on a BIG dataset. Doing so gives a more convincing validation of the research than on a small dataset, which may appear to be preferentially selected to give a favorable validation. Working with BIG data also increases the potential for observing patterns that may not be apparent in smaller datasets, such as trends in behavior across a play, and exploring the limitations of the methodology at hand. To this end, I use all viable gas

production histories from the Marcellus shale, which totals 610,192 monthly rate observations from 8,527 wells.

## Methodology

The goal of this paper is to investigate the effect that the inclusion of autoregressive and/or moving average terms may have on decline curve performance, not only in terms of decline curve parameter estimation via generalized least squares (GLS) or generalized nonlinear least squares (GNLS), but also in terms of ARMA modeling of the residuals. The overall methodology to make these comparisons is as follows:

1. Fit decline curve by OLS (Arps' exponential) or NLS (hyperbolic, logistic growth, power law loss ratio, stretched exponential, and Duong's model) to training data, $q_{train}$
2. Calculate residuals, $e_{train} = q_{train} - \hat{q}_{train}$, where $\hat{q}_{train}$ are the predictions from the fitted decline curve
3. Iteratively fit ARMA models of varying order to the residuals to get optimal orders, $p$ and $q$
4. Re-fit decline curve by GLS (exponential model) or GNLS (all other models) with ARMA($p$,$q$) correlation structure
5. Repeat step 3 on residuals from GLS/GNLS fit

Thus, there are four distinct outputs whose performance needs to be evaluated: 1) the base model from step 1, 2) the base model combined with the ARMA-modeled residuals from step 3, 3) the GLS/GNLS fit model, and 4) the GLS/GNLS fit model with the ARMA-modeled residuals from step 5. The latter three are tested against the first (base case). To evaluate performance, all regressions in this paper are performed on the first 24 months of data from each well for the training data, and the fitted decline model forecasts are compared to the remainder of the record at each well. In practice, one would use all available data in $q_{train}$.

To elaborate more on these steps, in the first step, Arps' exponential model can be fit with OLS after taking the natural log transform of Eq. 1. However, all other decline curve models considered in this paper cannot be linearized and required NLS in order to estimate the decline parameters. Here, the Levenberg-Marquardt fitting algorithm gives relatively robust results, conditional on the convergence tolerance and maximum number of iterations allowed.

In steps 3 and 5 above, the orders $p$ and $q$ of the ARMA models are determined by iteratively fitting models of varying order and taking the one with the best AIC. Orders of $p = 0, 1, 2, \ldots, 5$ and $q = 0, 1, 2, \ldots, 5$, as well as all combinations, are explored. It is worth noting at this point that one could attempt to model these residuals with other explanatory variables, such as those related to the operation of the well (e.g., re-fracking events, choking of the well, etc.), but such data were not available for this analysis.

## Data and Results

### Marcellus Shale Dataset.

The West Virginia Geologic and Economic Survey (WVGES 2018) and DrillingInfo supply monthly production histories, along with various metadata, for these wells. Out of an initial population of 15,990 wells, only 8,527 have suitable data for decline curve analysis. This subset was determined after filtering and cleaning the production histories to remove data before the stated completion dates, removing probable partial observations (months where not all days exhibited production), removing zeros, and keeping only a continuous record of gas rates until the first (if any) gap in the record. Furthermore, after all these pre-processing steps, only records with more than 24 months of data were retained, in order to have a sufficient number of data points for robust curve fitting and to retain some testing data for assessing forecasting accuracy.

### Example with Duong's Model.

We start with a detailed illustration focusing on Duong's model exercised over all available Marcellus shale gas wells. First, a comparison is made between two methods of fitting Duong's decline curve model: the prescribed step-wise ordinary least squares (OLS) procedure (as outlined in Duong (2011)) and a non-linear least squares (NLS) procedure. Figure 1 compares the sum of squared residuals (SSR) from these two approaches, as fit to the 8,527 Marcellus records. For any one record, the better fitting model will give the smaller SSR, defined as:

$$SSR = \sum_t^T (q_t - \hat{q}_t)^2. \quad \ldots\ldots(16)$$

The NLS approach generally gives lower SSR than the OLS approach, with some exceptions (the vast majority of points fall below the 1:1 line in Figure 1). This indicates that NLS fits Duong's decline model to the data better than the standard OLS approach.

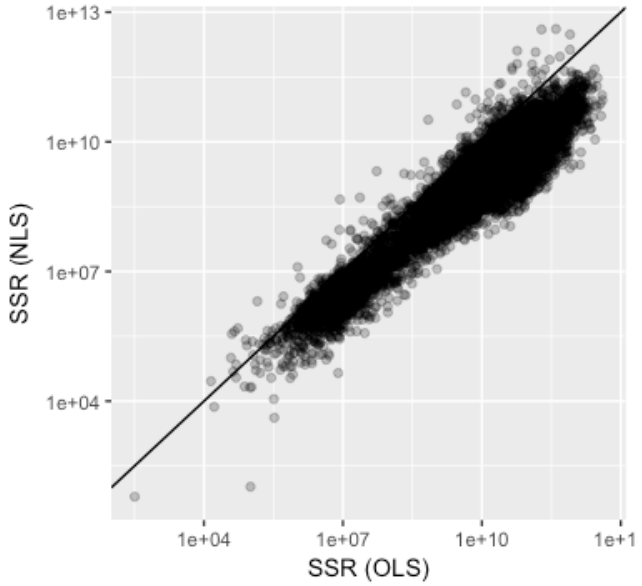


**Figure 1. Sum of squared residuals from the non-linear least squares fitting approach versus sum of squared residuals from the traditional step-wise OLS approach. The diagonal line is the 1:1 ($x = y$) line.**

Furthermore, because it fits the data better, the NLS fitting method yields less serial autocorrelation in the residuals, $e$. The Durbin-Watson test statistic serves as a metric for the significance of sample autocorrelation values (in this case, we only look at lag-1 autocorrelation, or the correlation between consecutive observations in time):

$$DW = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}. \quad \ldots\ldots(17)$$

The value for $DW$ is between 0 and 4, with smaller values indicating increasing positive autocorrelation and larger values indicating increasing negative autocorrelation ($DW = 2$ indicates no autocorrelation). Figure 2 shows that the NLS fitting approach generally gives larger $DW$ values than the OLS approach, again with some exceptions (majority of points fall above the 1:1 line, with few below). Thus, the bias introduced through the comparatively poor fits of the OLS approach gives greater positive autocorrelation in the residuals.

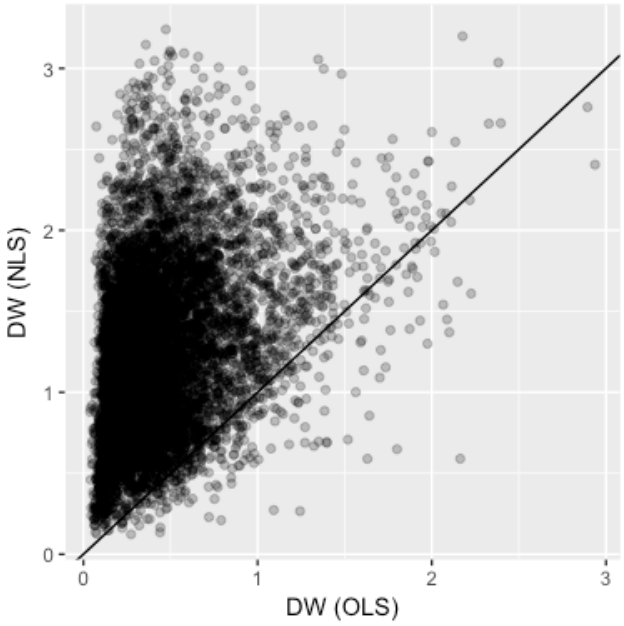


**Figure 2. Durbin-Watson test statistics for the lag-1 autocorrelation value from the non-linear least squares fitting approach versus the traditional step-wise OLS approach. The diagonal line is the 1:1 ($x = y$) line.**

Although the NLS fitting approach reduces the magnitude of serial autocorrelation in the decline curve residuals, there is still considerable autocorrelation in a significant proportion of the well production histories. In Figure 2, 2946 points fall below $DW = 1$, which is generally accepted as critically low, for the NLS approach (or approximately 35% of the 8,527 wells).

With the NLS regression established as the superior fitting procedure in the baseline case (over OLS), we can proceed to investigate the improvements afforded by GNLS regression and ARMA modeling of the residuals. For the rest of the analysis, the first 24 months of every well production history is used for model fitting, while the remainder of the record is reserved for validation (calculation of mean absolute percent error below). The next step is to fit an ARMA model to the calculated residuals from the NLS regression of Duong's model. Figure 3 shows the distributions of autoregressive order $p$ and moving average order $q$ from fitting ARMA models to the Duong model residuals. While the case $p = 0$ and $q = 0$ are predominant (as would be expected from the DW results above), there are significant numbers of $p = 1$ and $q = 1$ instances. There are a very small number of higher-order AR and MA components.

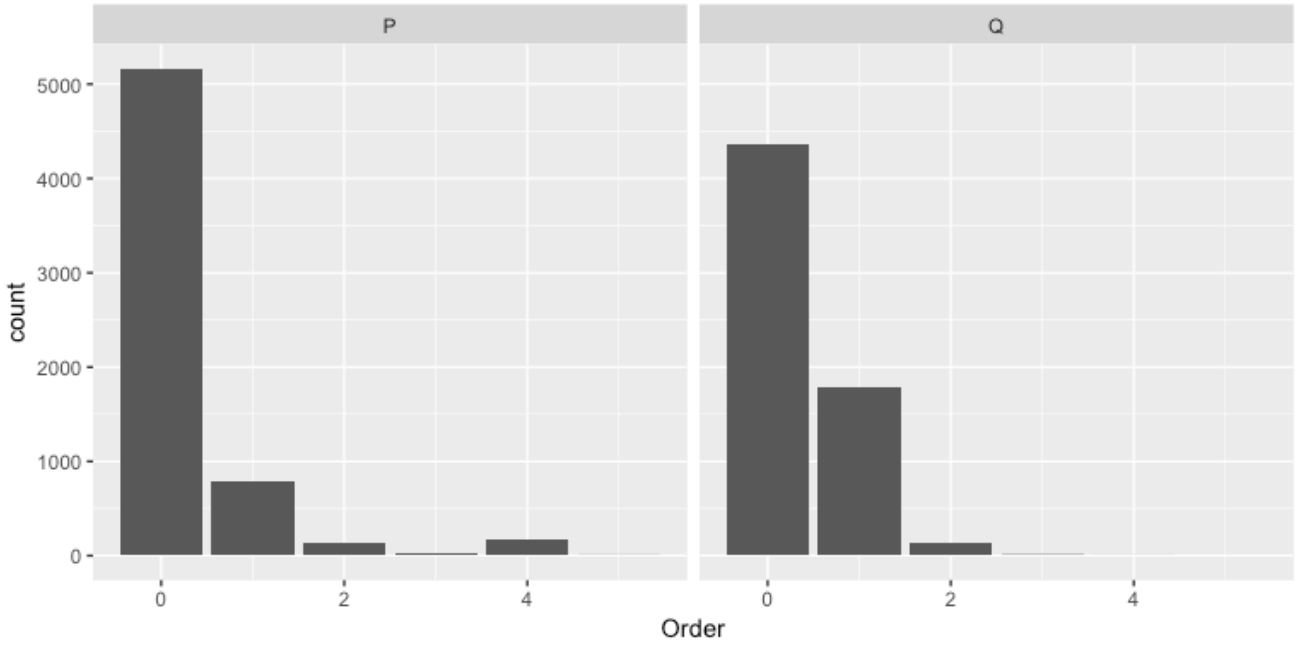


**Figure 3. Bar charts of values for $p$ and $q$ after fitting ARMA models to the residuals from Duong's model.**

GNLS regression of Duong's model is run on all production records with any non-zero $p$ or $q$ value (if both $p = 0$ and $q = 0$, then there is no serial autocorrelation structure to specify and GNLS regression reduces to a weighted least squares regression). Separately, the fitted ARMA model from the OLS/NLS residuals is added to the OLS/NLS decline curve fit (as in Eqs. 12-15). Figure 4 shows an example of these different fitting procedures for Duong's decline curve model, where those methods incorporating ARMA-modeled residuals fit the training data (before and including 24 months) better and, in this case, the use of GNLS gives better agreement with the testing data (better forecasts). The effect of these two procedures is measured by two performance metrics: 1) SSR, which indicates the goodness-of-fit to the training data (first 24 months), and 2) mean absolute percent error (MAPE), which assess the predictive accuracy of the model forecasts on the reserved testing data (after 24 months). SSR is defined in Eq. 16 (except now the summation is only over the first 24 months), and MAPE is defined as

$$MAPE = \frac{1}{T-25}\sum_{t=25}^{T} 100\,\% \left|\frac{q_t - \hat{q}_t}{q_t}\right| \quad \text{..... (18)}$$

for this application. Figure 5 shows the SSR and MAPE values for all the wells fit with Duong's model plotted by whether they incorporate ARMA-modeled residuals or not ("w/ ARMA" versus "w/o ARMA") and also by whether they use GNLS or not ("GNLS" versus "NLS"). Here we see that including the ARMA-model generally gives a better fit (lower SSR) and a better forecast (lower MAPE), whereas the use of GNLS generally gives a worse fit (higher SSR) and worse predictive accuracy (higher MAPE).

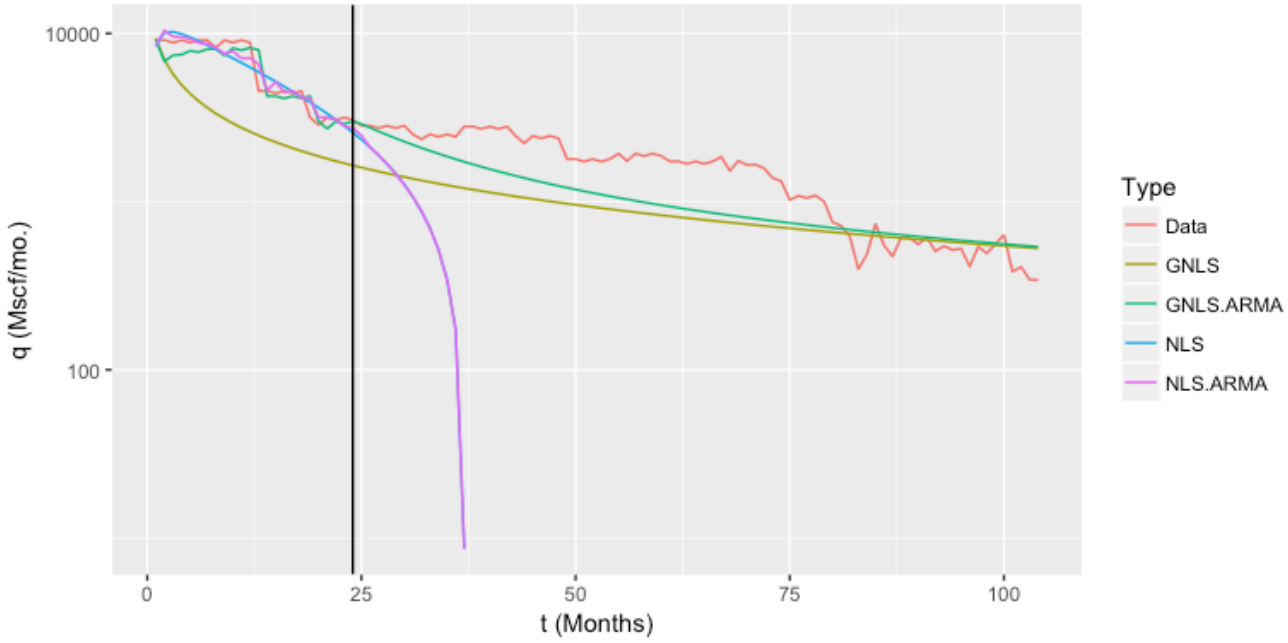


**Figure 4. An example from well API 37-051-24154. The vertical black line is at $t = 24$ months, representing the cut-off between training and testing data points.**

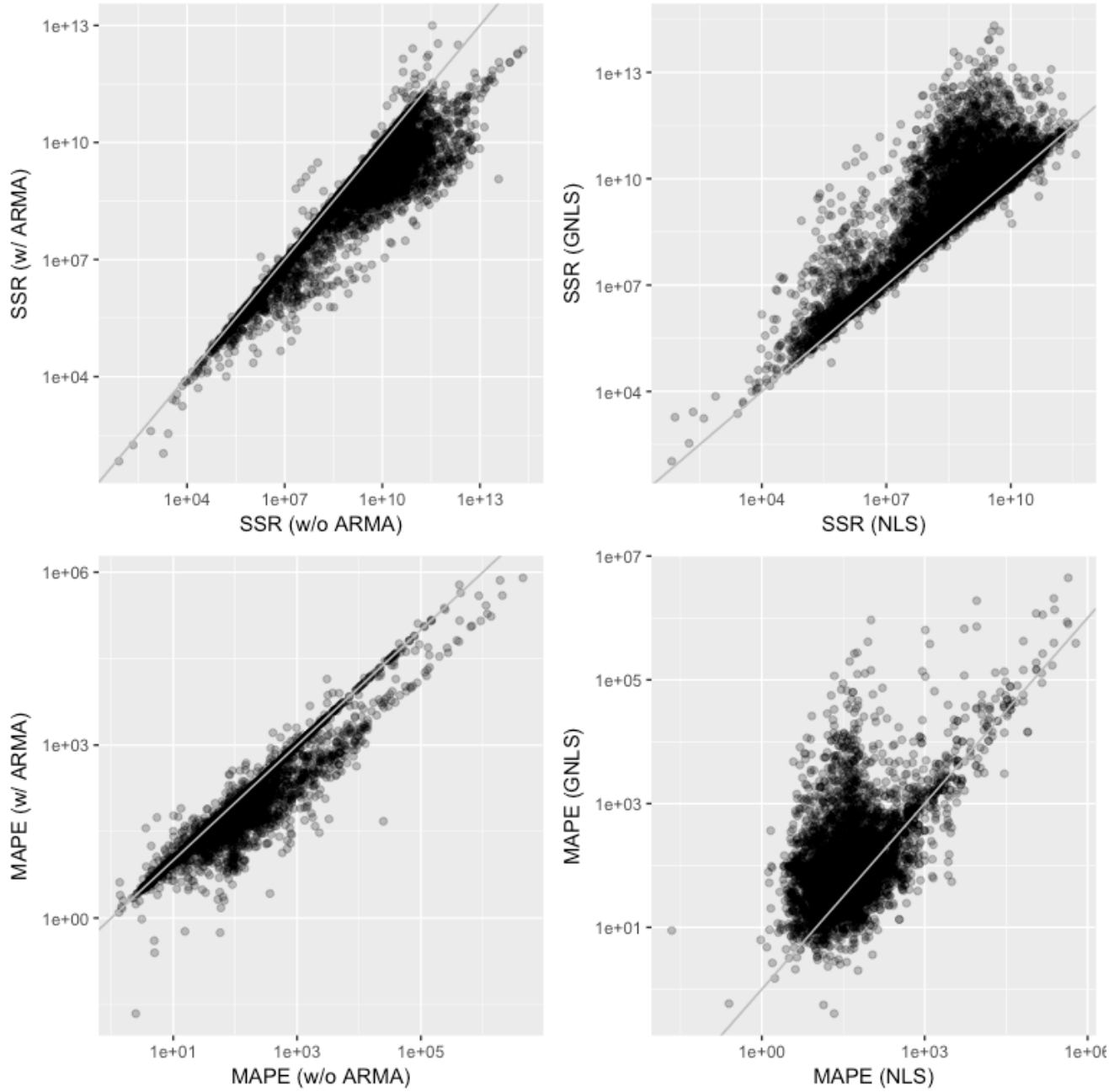


**Figure 5. Plots showing performance metrics (SSR and MAPE) from all wells fit with Duong's model and with $p > 0$ or $q > 0$. The performance metrics are plotted by the proposed fitting improvements (with ARMA and using GNLS for the regression) against the baseline case (no ARMA modeling of residuals and using NLS for the regression). The grey line is the 1:1 ($x = y$) line.**

Paired t-tests assess the statistical significance of these patterns. Since each well gives performance metric values in all categories (w/ ARMA, w/o ARMA, GNLS, NLS), pairing by well and taking the difference of metric values at this level gives a more powerful hypothesis test. Furthermore, since we want to test whether the model fitting procedures are making an improvement, one-sided tests are appropriate. Specifically, with our null hypotheses being that there is no change in MAPE or SSR with the inclusion of ARMA-modeled residuals or with using GNLS, our alternative hypotheses are that MAPE and SSR are lower with these fitting methods than without ARMA and without GNLS. All tests are conducted at a 95% confidence level. Table 1 presents the estimates of the mean of differences and their associated p-values from these tests. These values agree with the patterns seen in Figure 5. SSR is significantly greater without ARMA-modeled residuals but there is no evidence that this is the case with NLS (in fact the negative

mean of differences here suggests GNLS gives worse fits). Similarly, MAPE is significantly greater without ARMA-modeled residuals, by 2605% on average, but again GNLS appears to give worse forecasts (this mean percent may seem excessively large, but it is likely influenced by extreme positive outliers in the difference). Additionally, Table 1 contains the one-sided paired t-test results for GNLS with ARMA-modeled residuals versus the base fitting procedure: NLS without ARMA-modeled residuals (in other words, looking at the combined effect of GNLS and ARMA). For both MAPE and SSR, Duong's decline model fit with GNLS and including the ARMA-modeled residuals does not give statistically significant better performance over just Duong's model fit with NLS. This is likely due to the worse fits from the GNLS procedure outweighing any benefit gained from the ARMA modeling of the residuals, which comes after and therefore depends on the GNLS fit (refer to step 5 in Methodology).

| | Decline Models | | | | | |
|---|---|---|---|---|---|---|
| **Alternative** | **Exp** | **Hyp** | **Pow** | **Str** | **Log** | **Duo** |
| **MAPE** | | | | | | |
| NLS > GNLS | 1.6e+01 (1.7e-01) | -1.8e+02 (1.0e+00) | -7.3e+10 (9.2e-01) | -1.9e+02 (1.0e+00) | 2.0e+02 (4.9e-04) | -3.5e+03 (1.0e+00) |
| w/o ARMA > w/ ARMA | -2.1e+01 (8.8e-01) | 7.4e+01 (7.3e-03) | -1.3e+10 (8.4e-01) | -2.3e+02 (1.0e+00) | -1.7e+02 (1.0e+00) | 2.6e+03 (5.1e-04) |
| NLS w/o ARMA > GNLS w/ ARMA | -4.6e+00 (6.7e-01) | -1.1e+02 (9.9e-01) | -8.6e+10 (8.4e-01) | -4.2e+02 (1.0e+00) | 3.3e+01 (1.7e-01) | -9.0e+02 (1.0e+00) |
| **SSR** | | | | | | |
| NLS > GNLS | -4.5e+10 (9.7e-01) | -8.6e+09 (1.0e+00) | -2.0e+08 (1.0e+00) | -3.5e+10 (1.0e+00) | 3.3e+10 (1.9e-12) | -2.6e+11 (1.0e+00) |
| w/o ARMA > w/ ARMA | 4.7e+10 (2.7e-02) | 9.3e+09 (1.5e-113) | 1.8e+09 (6.6e-74) | 3.5e+10 (4.8e-135) | 5.3e+10 (3.5e-34) | 2.4e+11 (4.8e-06) |
| NLS w/o ARMA > GNLS w/ ARMA | 2.1e+09 (4.9e-35) | 7.2e+08 (7.5e-12) | 1.6e+09 (7.9e-32) | -5.3e+08 (1.0e+00) | 8.6e+10 (9.5e-21) | -1.7e+10 (1.0e+00) |

**Table 1. One-sided paired t-test results for MAPE and SSE calculated over all decline models and at all viable well records. In each cell, number on top is the estimate of the mean of differences and the number in parenthesis is the p-value. Significant values (p-value < 0.05) are colored red.**

## Extension to Other Decline Curves

The same comparative procedure made on Duong's decline curve model above is applied to five other decline curves: Arps' Exponential (Exp), Hyperbolic (Hyp), Power Law Loss Ratio (Pow), Stretched Exponential (Str), and Logistic Growth (Log). Again, the Exponential model uses OLS and GLS, whereas all other models use the nonlinear variants. The same steps outlined in the Methodology are applied to all 8,527 Marcellus wells for all decline curves. The frequency of orders $p$ and $q$ of ARMA models for the six different decline curves are shown in Figure 6. Here, all other models show a similar distribution of $p$ and $q$ to Duong's model, with zero-order values being predominant, but $p = 1$ and $q = 1$ also having a significant occurrence.

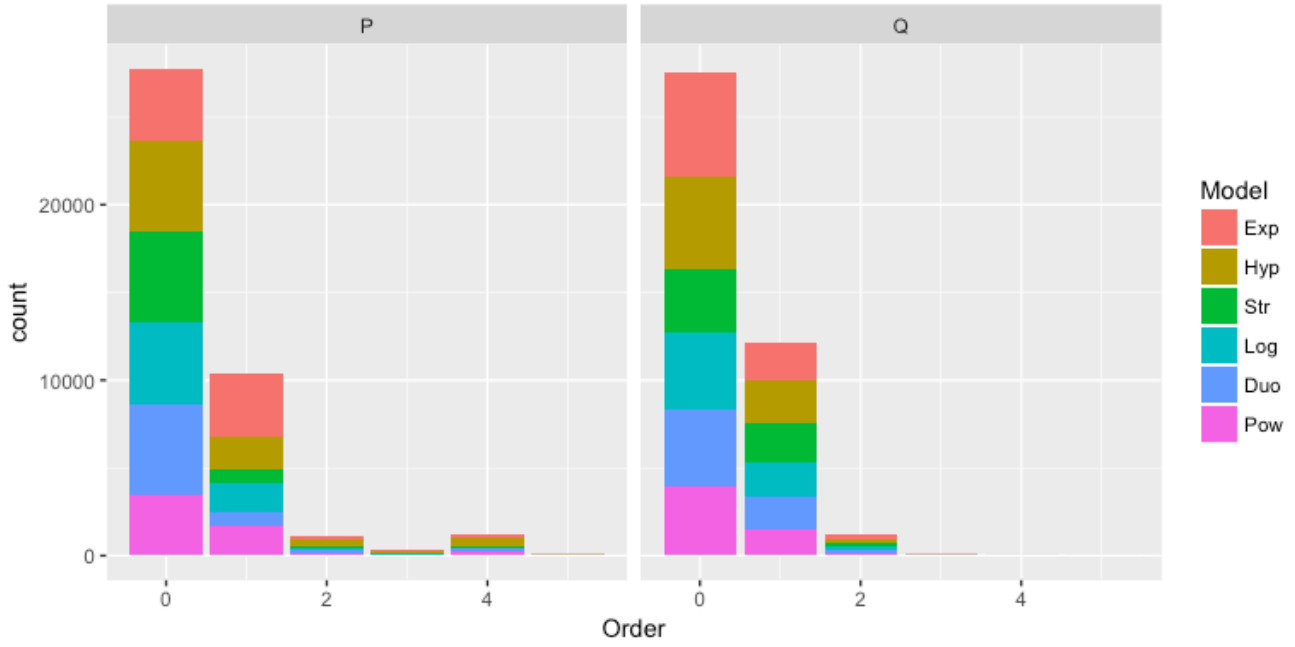


**Figure 6. Stacked bar chart of ARMA order values, $p$ and $q$.**

Furthermore, the same one-sided paired t-tests as used for Duong's model above are run on the results from the other decline curves, and the results are presented in Table 1. In support of Table 1, and in lieu of the sort of scatterplots in Figure 5, Figure 7 shows boxplots of the raw SSR and MAPE values, organized by decline curve type and fitting procedure. However, the main conclusions should be drawn from the t-tests (Table 1), because these boxplots do not pair the data by well. In Table 1, we see that including ARMA-modeled residuals always improves the fit to the training data (all models have positive estimates of mean of differences for SSR for the alternative hypothesis that "w/o ARMA > w/ ARMA", and these estimates are all statistically significant, some drastically so). The decline curves only gain benefits to their forecasting ability (in terms of MAPE) sporadically; in addition to Duong's model having a lower MAPE for the "w/ ARMA" case, the Hyperbolic model also benefits from including ARMA-modeled residuals, whereas the Logistic Growth model has lower MAPE when fitting with GNLS. The Logistic Growth model seems to benefit the most out of all six decline curve models, where GNLS, ARMA, and GNLS plus ARMA all improve the goodness-of-fit and forecasting accuracy, except for the "w/ ARMA" case under MAPE.

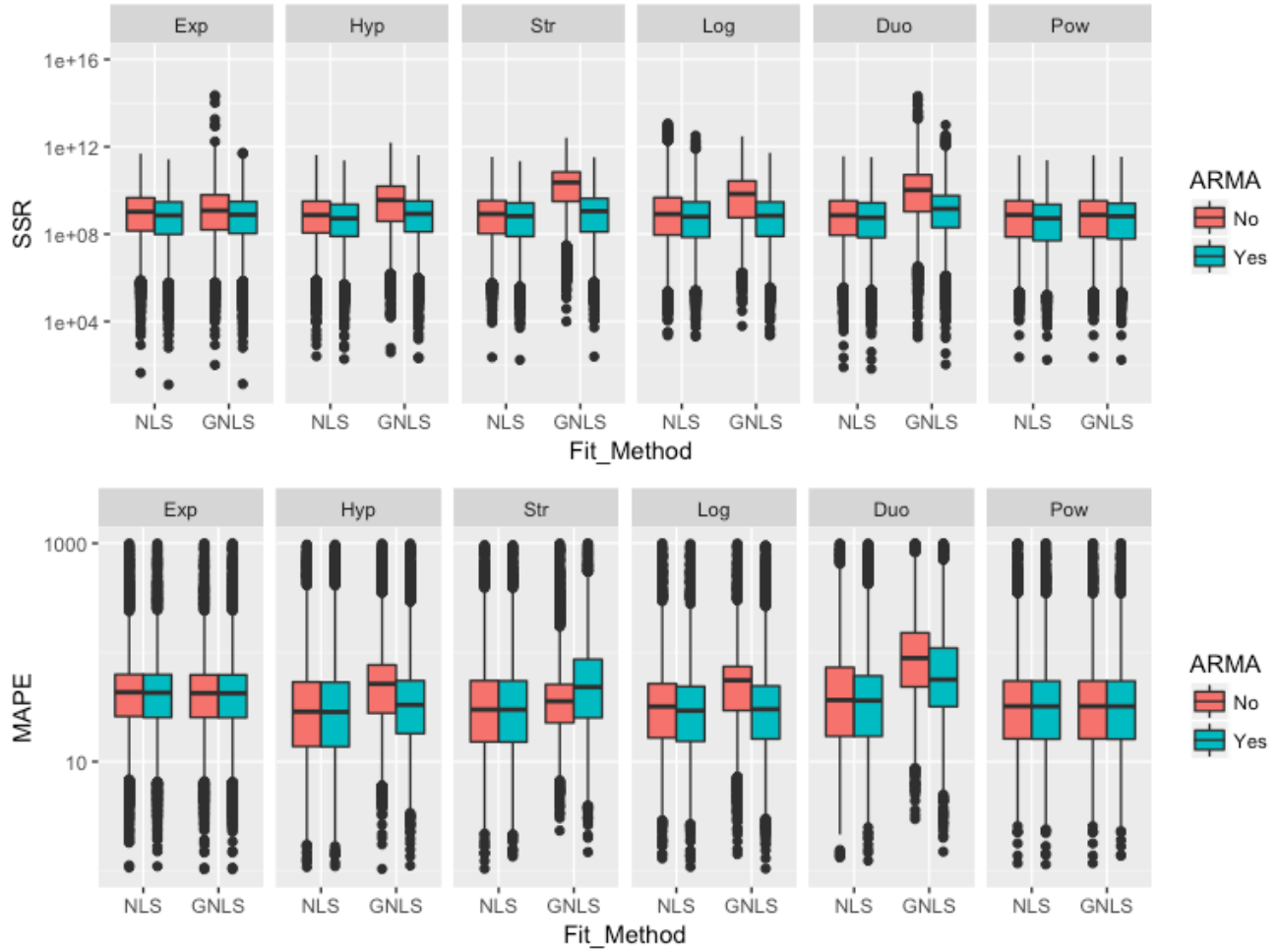


**Figure 7. Box plots of SSR and MAPE values for all viable Marcellus production histories after fitting each decline curve model under the various combinations of NLS versus GNLS and with ARMA-modeled residuals and without. The $y$ axes have been clipped from 1e1 to 1e16 for SSR and from 1 to 1000 for MAPE in order to exclude outliers and get a better view of the majority of the data.**

## Conclusion

The main conclusion to be drawn from the analysis in the paper is that, after fitting a decline curve, its goodness-of-fit can be greatly improved most of the time if the residuals are fit with an ARMA model and this ARMA model is added to the fitted decline curve. However, fitting past data better does not have much practical utility, at least in terms of production forecasting and estimation of EUR. This same treatment of including ARMA-modeled residuals only improves the forecasting accuracy of Duong's model and the Hyperbolic decline curve. Furthermore, estimating decline curve parameters via generalized (nonlinear) least squares as opposed to ordinary or nonlinear least squares only helps the Logistic Growth model achieve better goodness-of-fit and forecasting accuracy. A tangential conclusion made from the analysis in this paper is that Duong's model is generally fit better using nonlinear least squares than using the prescribed step-wise ordinary least squares procedure from Duong (2011).

It is possible that the results obtained in this study could dramatically change with larger training data, as may be used in practice. That is, the choice of 24 months worth of training data may have a large influence on the results, and setting this cutoff to a later time would sample more production rates under boundary-dominated flow conditions and the resulting fitted decline curve would represent that behavior better (and project it into the future). It would be worth repeating the analysis in this study over a variety of different training/testing partitions, as is done in Cheng et al. (2010).

Furthermore, in order to achieve a high level of statistical power, this study uses a BIG dataset of the entire population of Marcellus shale wells. Consequently, the fitting procedures in this paper were carried out in an automated fashion, where the algorithms used for the nonlinear least squares and generalized nonlinear least squares regressions are somewhat sensitive to the chosen starting values, convergence tolerances, and number of maximum iterations. While great effort was made in this analysis to find good heuristics for the starting values of each decline curve model, these could be suboptimal for some wells. Similarly, the global values used for convergence tolerance and maximum number of iterations may not have been appropriate for all wells. It is possible that different results may be achieved, and conclusions drawn, if one were to manually tailor these settings to each individual well. Nevertheless, the algorithms worked sufficiently for the majority of cases and I believe the conclusions are robust to the cases where they may have performed suboptimally.

## Acknowledgements

This work was made possible by the Deike Research Grant, College of Earth and Mineral Sciences, Penn State. The author thanks DrillingInfo for the donated academic license.

## References

Arps, J. J. 1945. Analysis of Decline Curves. *Transactions of the American Institute of Mining and Metallurgical Engineers* **160**: 228–47.

Ayeni, B. J. and Pilat, R. 1992. Crude oil reserve estimation: An application of the autoregressive integrated moving average (ARIMA) model. *Journal of Petroleum Science and Engineering* **8** (1): 13–28.

Cheng, Y., Wang, Y., McVay, D., et al. 2010. Practical Application of a Probabilistic Approach to Estimate Reserves Using Production Decline Data. *SPE Econ & Mgmt* **2** (01): 19–31.

Clark, A. J., Lake, L. W., and Patzek, T. W. 2011. Production Forecasting with Logistic Growth Models. Presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, 30 October-2 November. SPE-144790-MS. doi:10.2118/144790-MS

Cochrane, D. and Orcutt, G.H. 1949. Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms. *Journal of the American Statistical Association* **44** (245): 32–61.

Duong, A. N. 2011. Rate-Decline Analysis for Fracture-Dominated Shale Reservoirs. *SPE Res Eval & Eng* **14** (3): 377–87.

Ediger, V. S., Akar, S., and Ugurlu, B. 2006. Forecasting production of fossil fuel sources in Turkey using a comparative regression and ARIMA model. *Energy Policy* **34** (18): 3836–46.

Frausto-Solís, J., Chi-Chim, M., and Sheremetov, L. 2015. Forecasting Oil Production Time Series with a Population-Based Simulated Annealing Method. *Arabian Journal for Science and Engineering* **40** (4): 1081–96.

Gallant, A. R. and Goebel, J. J. 1976. Nonlinear regression with autocorrelated errors. *Journal of the American Statistical Association* **71** (356): 961–67.

Ilk, D., Rushing, J. A., Perego, A. D., and Blasingame, T.A. 2008. Exponential vs. Hyperbolic Decline in Tight Gas Sands: Understanding the Origin and Implications for Reserve Estimates Using Arps' Decline Curves. Presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, USA, 21–24 September. SPE-116731-MS. https://doi.org/10.2118/116731-MS

Seidle, J., Erdle, J., Hale, B., et al. 2016. *Estimating ultimate recovery of developed wells in low permeability reservoirs*, SPEE Monograph 4.

Sheremetov, L., Cosultchi, A., Martinez-Munoz, J., et al. 2014. Data-driven forecasting of naturally fractured reservoirs based on nonlinear autoregressive neural networks with exogenous input. *Journal of Petroleum Science and Engineering* **123** (November): 106–19.

Shumway, R. H. and Stoffer, D. S. 2010. *Time Series Analysis and Its Applications: with R Examples*. 3rd Edition, Springer. DOI 10.1007/978-1-4419-7865-3

Valko, P. P., and Lee, W. J.. 2010. A Better Way to Forecast Production from Unconventional Gas Wells. Presented at the SPE Annual Technical Conference and Exhibition, Florence, Italy, 19-22 September. SPE-134231-MS. https://doi.org/10.2118/134231-MS

West Virginia Geological and Economic Survey. 2018. Selected Shale-Gas Resources: Marcellus Data Files. http://www.wvgs.wvnet.edu/www/datastat/Marcellus/Downloads/WVGES%20Marcellus%20Wells.xlsx (accessed 7 May 2018).

Yusof, N. M., Sulong, R., Rashid, A., et al. 2010. Malaysia crude oil production estimation: an application of ARIMA model. In 2010 International Conference on Science and Social Research (Cssr), 1255–9. IEEE.