Please fill in the name of the event you are preparing this manuscript for.	SPE Eastern Regional Meeting
Please fill in your 6-digit SPE manuscript number.	191793
Please fill in your manuscript title.	Combining Decline Curve Analysis and Geostatistics to Forecast Gas Production in the Marcellus Shale

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

Given Name	Surname	Company
Zhenke	Xi	The Pennsylvania State University
Eugene	Morgan	The Pennsylvania State University

Abstract

Traditionally, in order to estimate the production potential at a new, prospective field site via simulation or material balance, one needs to collect various forms of expensive field data and/or make assumptions about the nature of the formation at that site. Decline curve analysis would not be applicable in this scenario, as producing wells need to pre-exist in the target field. The objective of our work is to make first-order forecasts of production rates at prospective, undrilled sites using only production data from existing wells in the entire play. This is accomplished through co-kriging of decline curve parameter values, where the parameter values are obtained at each existing well by fitting an appropriate decline model to the production history. Co-kriging gives the best linear unbiased prediction of parameter values at undrilled locations, and also estimates uncertainty in those predictions. Thus, we can obtain production forecasts at P10, P50, and P90, as well as calculate EUR at those same levels, across the spatial domain of the play.

To demonstrate the proposed methodology, we used monthly gas flow rates and well locations from the Marcellus shale gas play in this research. Looking only at horizontal and directional wells, the gas production rates at each well were carefully filtered and screened. Also, we normalized the rates by perforation interval length. We kept only production histories of 24 months or longer in duration to ensure good decline curve fits. Ultimately, we were left with 5,637 production records. Here, we chose Duong's decline model to represent production decline in this shale gas play, and fitting of this decline curve was accomplished through ordinary least square regression.

Interpolation was done by universal co-kriging with consideration to correlate the four parameters in Duongs' model, which also showed a linear trend (the parameters show dependency on the x and y spatial coordinates). Kriging gave us the optimal decline curve coefficients at new locations (P50 curve), as well as the variance in these coefficient estimates (used to establish P10 and P90 curves). We were also able to map EUR across the study area. Finally, the co-kriging model was cross-validated with leave-one-out scheme, which showed significant but not unreasonable error in decline curve coefficient prediction.

We forecasted potential gas production in the study area using co-kriging. Heat maps of decline curve parameters as well as EUR were constructed to give operators a big picture of the production potential in the play. The methods proposed are easy to implement and do not require various expensive data like

permeability, bottom hole pressure, etc., giving operators a risk-based analysis of prospective sites. We also made this analysis available to the public in a user-friendly web app.

Introduction

The unconventional shale gas revolution has played an important role in supplying the increasing energy demand world-wide and has brought people many benefits such as national energy security and job creation. However, the development of unconventional gas can be substantially risky because of the difficulty characterizing the reservoir and understanding how the reservoir behaves. Unlike conventional reservoirs, unconventional shale always has complicated transport mechanism and ultra-low permeability, while the induced hydraulic fractures act as additional factors in evaluating the well performance (Clarkson et al., 2012). As a result, effective methods that can forecast the unconventional gas production are critical for the risk control.

Decline curve analysis (DCA) is one of the basic tools used in production rate forecasting and recovery estimation (Bhattacharya and Nikolaou, 2013). Arps' hyperbolic rate-time equation (Arps, 1945) which has been widely applied in conventional play can also capture the feature of shale gas rate decline with its hyperbolic decline constant "b" sometimes greater than unity. Besides, the numerical simulation is another approach to quantify the analysis but the unique features, such as gas desorption, of unconventional shale must be taken into consideration. Some authors (Cipolla et al., 2010) addressed issues in techniques of conventional production analysis when they are applied to unconventional reservoirs. Anderson and Liang (2011) developed a probabilistic approach based on "Tri-linear Flow" model for unconventional reservoir characterization and forecasting. More models dealing with the complex fracture networks have been developed (Diaz De Souza et al., 2012; Moinfar et al., 2013). Additionally, with the advancement of the computing power and storage capacity of computer hardwares, a variety of machine learning or data mining methods can make contributions to production prediction. For example, an expert system built upon multi-layer artificial neuron network (ANN) has been developed by Ozdemir et al., (2016) for production performance prediction of infill drilling locations. Support vector machine (SVM), as a sophisticated method in statistics, works well in pattern recognition such as determining the location of sweet spots (Hauge and Hermansen, 2017).

Although the previous methods would excel, given sufficient data and under appropriate assumptions, there are still limitations. For example, numerical history matching requires solvers with strong robustness and various types of field data (Oliver and Chen, 2011). These conditions cannot always be satisfied because of the complexity of partial differential equations or the scarcity of data in some newly developed plays. Artificial neural network may be viable with less refined data, but still, a large amount of data and different data types are needed. In order to find out a proper ANN architecture, a significant amount of time can be required for model training and cross validation. Additionally, many approaches will only give deterministic results without quantifying the uncertainty.

In this work, we design a simple and elegant approach to forecast production that does not rely on field data like permeability, bottom hole pressure, etc. and can estimate the uncertainty of the predicted results. First, we fit the production decline curve for existing wells using Duong's model. Then combined universal kriging and co-kriging are adopted to interpolate the decline curve parameters at undrilled sites. Finally, a heat map of estimated ultimate recovery (EUR) can be constructed with the interpolated parameters, and production forecasts up to an arbitrary number of months in the future can be generated for any site in the spatial domain of the play. While we demonstate the approach on the Marcellus shale gas play, the methodology can be adopted to any other play, conventional or unconvential, given that sufficient regional production data is available and an appropriate decline curve model exists for the reservoir type.

Methodology

Data cleaning.

Processing raw data is a common step before extracting any interesting information from the data. In order to fit Duong's model into the flow rates that represent the true production feature of the reservoir, we suggest cleaning data following these steps in order:

1. Drop data from vertical wells (keep only directional and horizontal well flow rates);
2. Order flow rate by increasing date;
3. Remove values prior to completion date;
4. Remove any zero values at the beginning (before first positive value) OR remove first value if it is positive (may be partial observation);
5. Remove data before peak production, if any left;
6. Check for gaps, zero-values, or large changes (as created by shut-in, re-stimulation, etc.) in record and only keep data until first instance;
7. Throw out well if resulting record in less than 24 months long;
8. Normalized gas rate by perforation interval length, lateral length, or similar.

Decline curve analysis.

Duong (2011) proposed a decline curve model that works well for unconventional reservoirs which exhibit long periods of linear flow due to the pressure depletion within fracture networks. It models gas rate (q) as a function of time (t) and four parameters (a, m, q_i, q_∞). The four parameters are determined via two stages of sequential ordinary least squares (OLS) regression. First a and m are determined via OLS on:

$$ln\left(\frac{q}{G_p}\right)=m\ ln(t)+a \tag{1}$$

Then q_i and q_∞ are found by OLS regression of:

$$q=q_i t_{am}+q_\infty \tag{2}$$

Where

$$t_{am}=t^{-m}exp\left(\frac{a}{1-m}(t^{1-m}-1)\right) \tag{3}$$

Despite catching the physics of unconventional reservoirs, Duong's model is also advantageous in its numerical stability when fitting the curve. Since the four parameters can be obtained by implementing OLS on Equation (1) and Equation (2) respectively, they are guaranteed to be the solution of the global minimum of the squared error surface. This feature makes the fitting procedure faster than those using Arps' hyperbolic model which is done by nonlinear regression where the global minimum of squared error cannot be ensured even after many iterations.

Geostatistics method.

To calculate EUR at undrilled sites, we must first predict the associated parameters based on known data points. Although a variety of mathematical interpolation techniques are available, such as spline and polynomial, the spatial correlation among observations is ignored in them. Therefore, methods from kriging family that will handle the underlying spatial structures are applied in our work.

Intuitively, the four parameters (a, m, q_i, q_∞) for each well should be mutually related because they collectively quantify the same decline curve. Meanwhile a more reliable estimation can be obtained by utilizing the spatial correlation between different variables (Trangmar et al., 1986). Thus, co-kriging is employed in our work, which exploits the cross-correlation between several variables and calculate the unknown as a linear combination of them.

$$\hat{z}_2(x_0)=\sum_{i=1}^{n_1}\lambda_{1i}z_1(x_{1i})+\sum_{j=1}^{n_2}\lambda_{2j}z_2(x_{2j})+\ldots \quad (4)$$

λ_{1i} and λ_{2j} are the weights related to z_1 and z_2 respectively; n_1 and n_2 are the number of data points used to estimate $\hat{z}_2$ and x stands for the location. A complete system of co-kriging can be found in some other works (Journel and Huijbregts, 1978; McBratney and Webster, 1983; Vauclin et al., 1983).

While co-kriging takes advantage of multiple parameters, fitting the data, which involve obvious spatial trend (external drift), into a deterministic function of locations beforehand can also help estimation precision. This idea is conveyed from universal kriging that allows the kriging in the presence of unstationarity (Matheron, 1971). Therefore, we should utilize the features of the two kriging methods by "filtering out" the external drift from the data first and then co-kriging the associated residual, which leads to an indirect but comprehensive hybrid method, "universal co-kriging".

Besides, the scale of the shale play is usually large. So interpolating the parameters over single points that are very close to each other can hardly reveal the variation of parameter values. Instead, finding the mean and variance of the prediction over the area of a fixed-size block, namely "block kriging", can save people a lot of energy especially when constructing a heat map.

Since the outputs of kriging contain both mean and (co-)variance of each type of variable, we can determine the P10, P50 and P90 EUR by numerically integrating Duong's decline curve with its parameters sampled from a multivariate normal distribution. And thus, the uncertainty of the prediction is quantified.

To better explain this integrated approach, a case study on the Marcellus shale throughout Pennsylvania to West Virginia are presented in the following section.

Case Study

Data source.

Monthly production data were obtained for all wells in the Marcellus (11,904 wells out of 15,988 have production data, positions are shown in Figure 1), along with various header data for each well (geographical coordinates, completion date, etc.). The data have been cleaned by the procedure introduced previously. Furthermore, we only used flow rates from horizontal and directional wells (records from vertical wells were dropped) since they play a major role in massive unconventional shale gas extraction. Besides, normalizing the flow rates with the perforation interval length can lead to an equitable comparison between wells. Data from only 5,637 wells finally met all the conditions and thus were later used for curve fitting. The data size in each processing stage is summarized in Table 1.

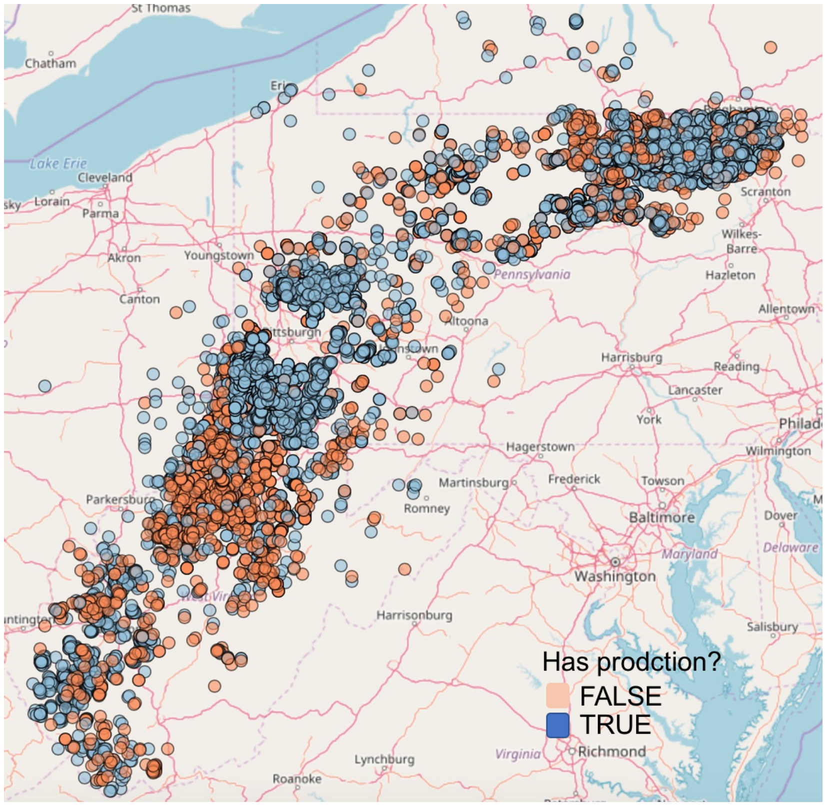

Figure 1-Well locations in study area

Table 1-Well numbers for each filtering stage

Criteria	Number of wells
Total	15990
Production data available	11904
Passed filtering conditions	8527
Horizontal & directional wells	5637

Parameter visualization.

The distributions of parameters are shown in Figure 2, we found parameters a and m are very likely to be subject to normal distributions while q_i and q_∞ appear to be log-normal type whose log-form is still a normal distribution. Since no multi-mode shape could be observed, we did not need to cluster the parameters in order to interpolate the unknowns accordingly.

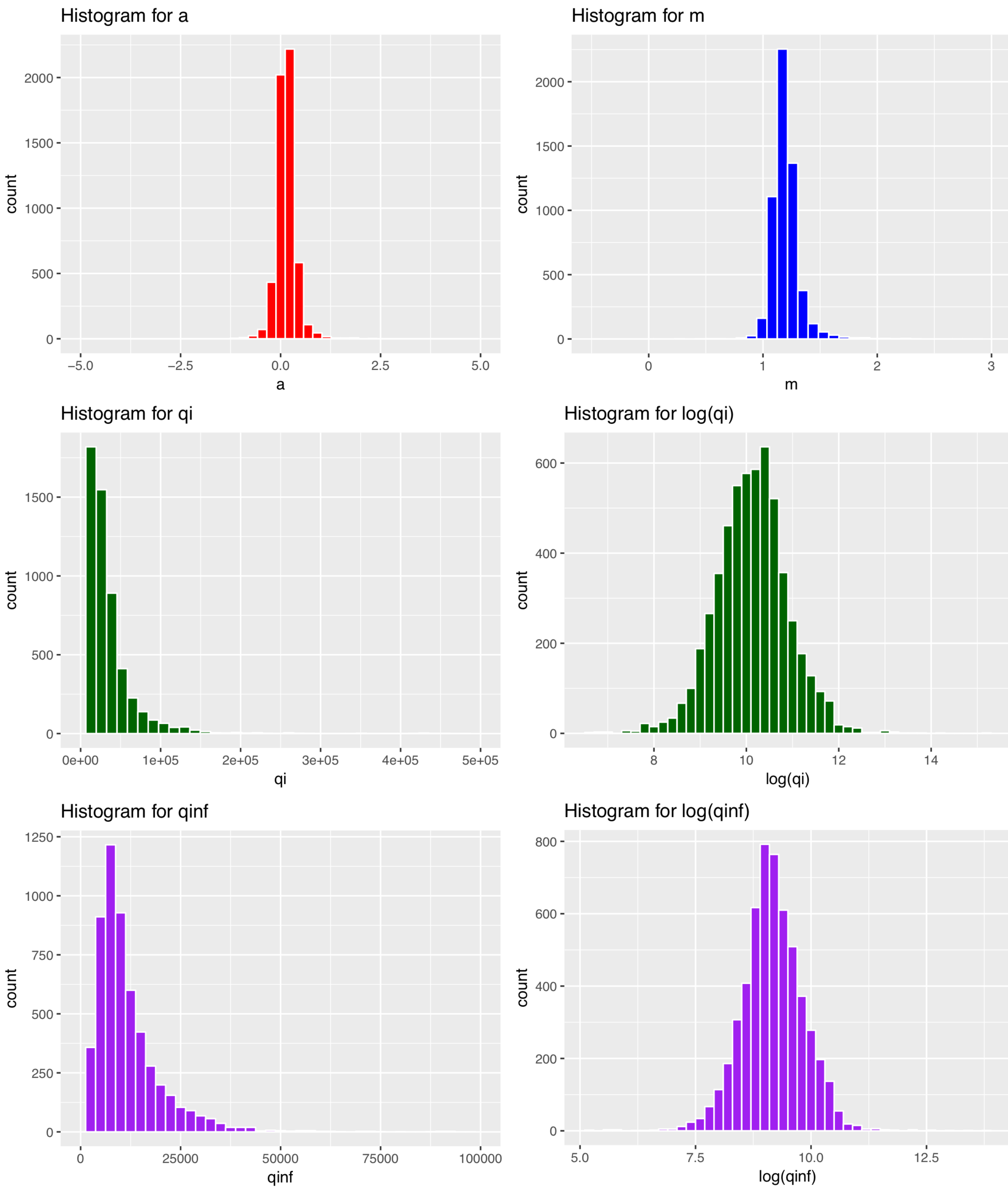

Figure 2-Duong's model parameter distributions

Additionally, we found the correlations between a & m and $\log(q_i)$ & $\log\left(q_{inf}\right)$ (or $\log(q_\infty)$) from the scatter-plot matrix in Figure 3. This is an indicator that co-kriging should be used in the further interpolation because it makes multiple parameters attribute to the prediciton.

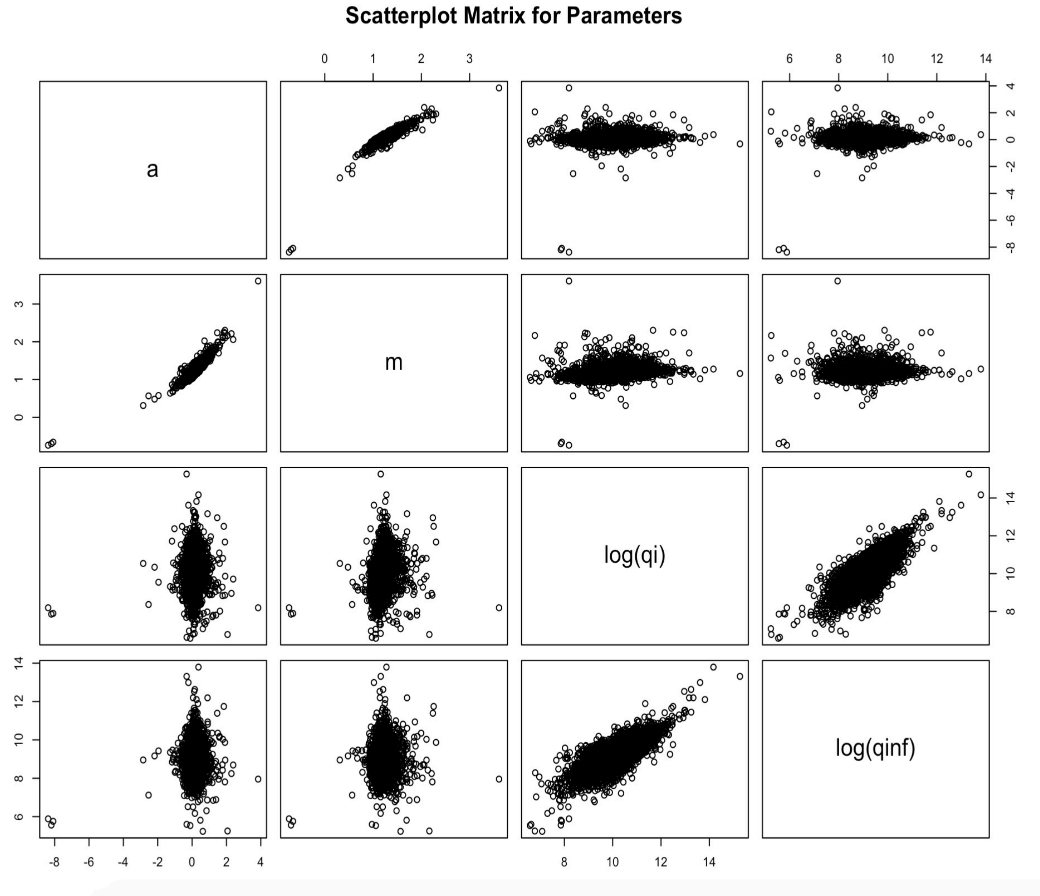

Figure 3-Scatter plot matrix for parameters

External drift.

To ensure the stationarity of our data before geostatistical modeling, we can fit a background trend (external drift) for universal kriging. Because we have m multiple response variables (in this case, the four decline curve parameters) and p multiple explanatory variables (in this case, spatial coordinates x and y, which are the UTM Eastings and Northings (both in km), respectively, plus an interaction term xy), we fit a multivariate multiple linear regression model for the background trend:

$$Y=Z\beta+\varepsilon \tag{5}$$

where Y is an n × m matrix (n being the number of observations, or wells in this application) containing each response variable as a column, Z is an n × (p+1) matrix containing the explanatory variables plus a series of ones (for fitting an intercept) as columns, β is a (p+1) × m matrix containing the regression coefficients, and ε is an n × m matrix of error terms. The regression coefficient estimates, their p-values, and the adjusted R^2 values are reported in Table 2.

Table 2. Coefficient estimates and their p-values (in parenthesis), as well as the adjusted R^2 values, for the multivariate multiple linear regression.

Response	Intercept	*x* coefficient	*y* coefficient	*xy* coefficient	Adjusted R^2
a	-4.4e+00 (8.9e-03)	7.0e-03 (2.0e-02)	-1.4e-06 (2.6e-02)	9.4e-04 (1.1e-02)	0.04

m	8.7e-02 (9.0e-01)	3.5e-03 (4.9e-03)	-6.9e-07 (9.7e-03)	1.9e-04 (2.0e-01)	0.04
ln(q_i)	3.9e+01 (3.5e-25)	3.8e-03 (5.7e-01)	4.4e-07 (7.6e-01)	-7.2e-03 (5.3e-19)	0.19
ln(q_∞)	3.9e+01 (4.0e-33)	-1.6e-02 (4.8e-03)	4.5e-06 (3.5e-04)	-7.3e-03 (1.5e-24)	0.16

Table 2 shows that while the adjusted R^2 values are fairly low for all four models, the coefficients are generally significant (low p-values), albeit also generally low in magnitude. Nevertheless, these models have better goodness-of-fit (higher R^2) than simpler models with fewer terms, including models of just an intercept (mean; as would be used for simple kriging). This criteria of higher R^2 is the reason for including the interaction term, xy.

Fitting linear model of co-regionalization.

Predictions can be strengthened by using the strong correlation between a and m and between q_i and q_∞. A linear model of co-regionalization not only models the spatial autocorrelation of individual variables, but also the between-variable spatial dependence. Here, we fit a Matern variogram model with nugget to all direct and cross variograms shown in Figure 4.

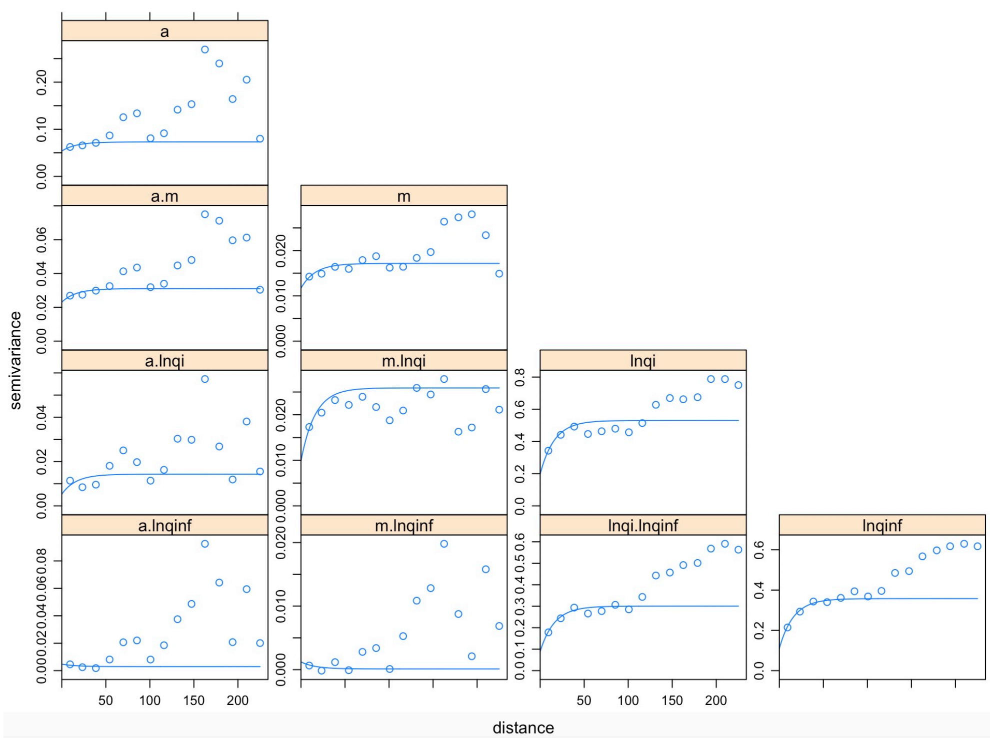

Figure 4-Fitted direct and cross variograms

Results and Discussion

Block co-kriging with external drift.

Figure 5 and 6 show the mean and variance respectively of each decline curve parameter. They are estimated within 5 km × 5km blocks spanning the domain of the Marcellus.

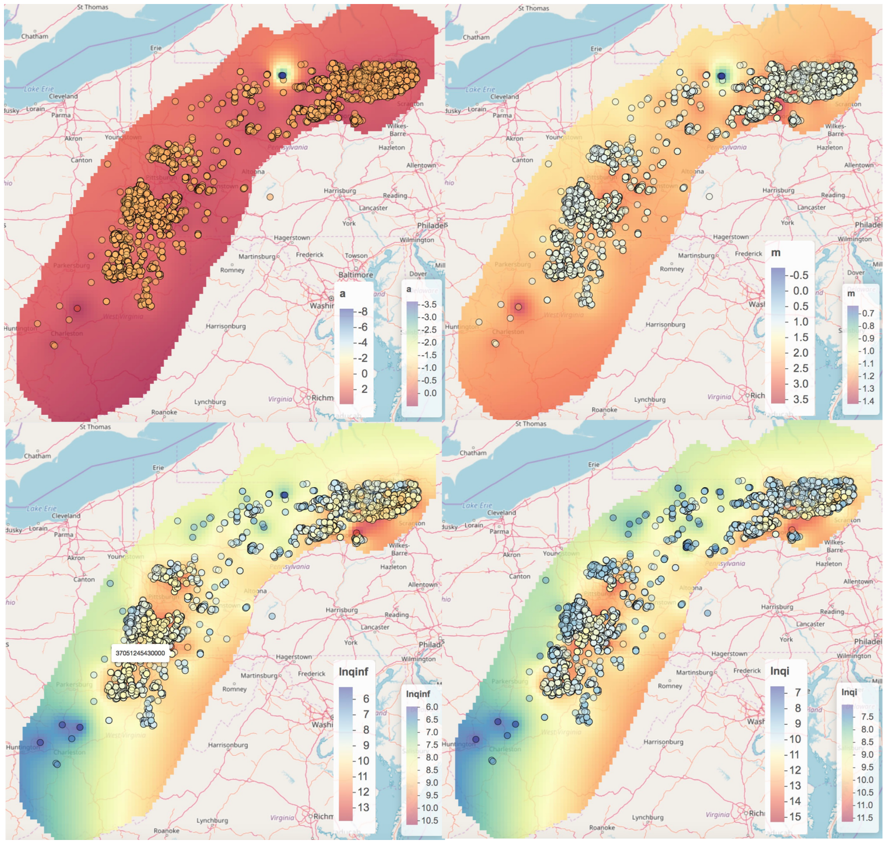

Figure 5-Predicted mean values for parameters (Legends in the right corner: left one for point value, right one for grid value)

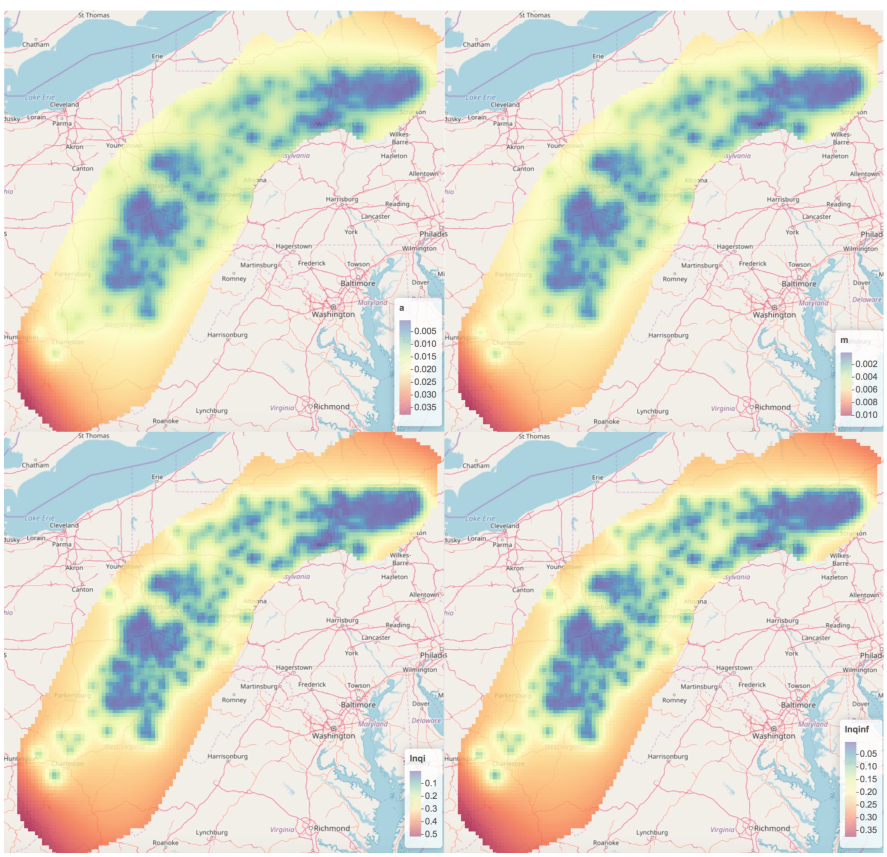

Figure 6-Variance for each parameter

Estimated ultimate recovery.

Estimated Ultimate Recovery EUR (at P10, P50, and P90 levels) is estimated by drawing a random sample of size 1,000 from a multivariate normal distribution for a, m, $\ln(q_i)$, and $\ln(q_\infty)$. The mean values and covariance matrix are constructed from the co-kriging predicted values in each block. The samples of the log-transformed variables are back transformed via:

$$q_i = exp\left(ln\left(q_i\right) + 0.5\left[var\left(ln\left(q_i\right)\right)\right]\right) \tag{6}$$

$$q_\infty = exp\left(ln\left(q_\infty\right) + 0.5\left[var\left(ln\left(q_\infty\right)\right)\right]\right) \tag{7}$$

In order to correct for bias, each sample is then used to calculate a 25-year EUR value:

$$EUR[Mscf] = \int_{1}^{25\times 12} q(t; a, m, q_i, q_\infty)dt \tag{8}$$

where q (t; a, m, q_i, q_∞) represents Duong's model. Taking the P10th, P50th, and P90th percentiles of the resulting EUR vector gives P90, P50, and P10, respectively. These are mapped below:

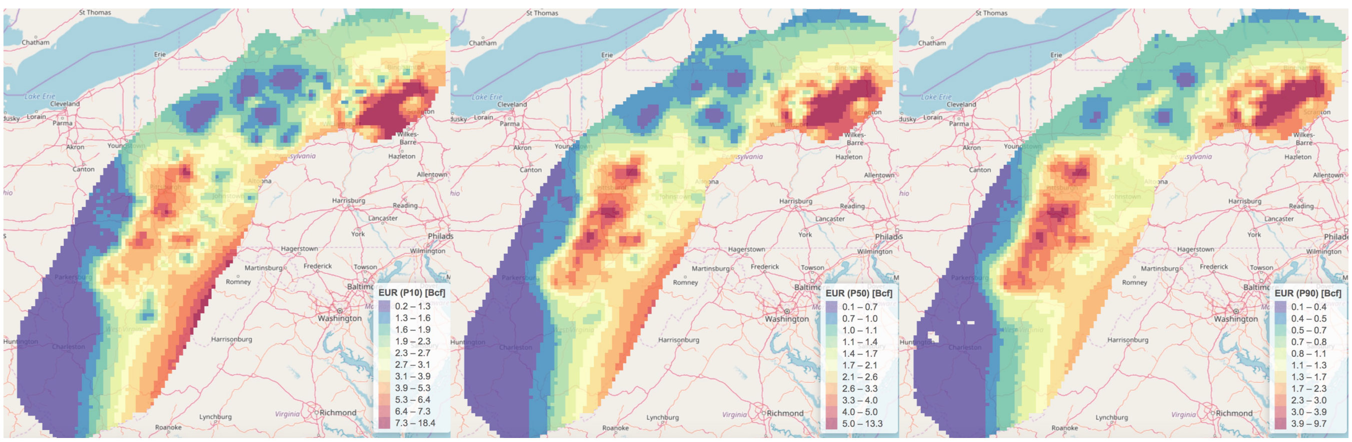

Figure 7 P10 P50 P90 EUR heat maps

Validation.

Leave-one-out cross-validation offers a rational way to test the quality of the universal co-kriging used above. In this validation scheme, a single well (with known decline curve parameters) is removed from the population of wells, and the universal co-kriging model is re-fit to the remaining data. Prediction using this new model is made at the single well site, and the predicted decline curve parameters are compared against their known values at the site. Additionally, the predictions of the decline parameters from the background trend (Equation (5)) are also recorded for comparison. We repeat this process for 100 randomly selected wells. This sample size of 100 was chosen as a balance between increasing statistical robustness of the analysis of errors and decreasing computational time (this process is computationally costly, with these 100 samples taking approximately 32 processor hours). Histograms of the resulting percent error of each parameter (difference between the true value and the predicted value divided by the true value) are presented in Figure 8.

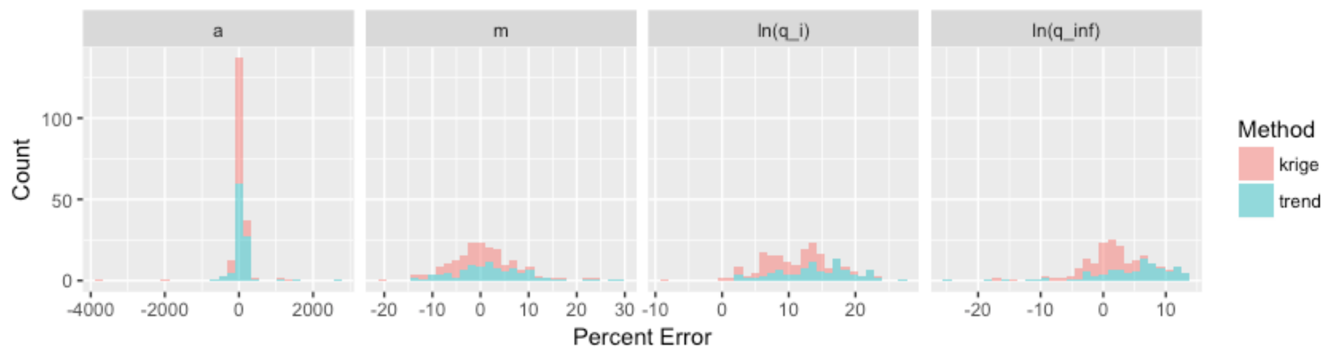

Figure 8. Histograms of percent error for each decline curve parameter and each method: universal co-kriging and backgroud trend.

Difference of mean hypothesis tests indicate whether the kriged predictions are statistically different from the true values. Here, we run one-sample t-tests on the percent error of each parameter to see if the mean percent error is significantly different from zero. So the null hypothesis is that the mean percent error equals zero and the alternative hypothesis is that it is not equal to zero. The results (estimate of mean percent error and associated p-value of the hypothesis test) are reported in Table 3. As the hypothesis tests are conducted at the 95% confidence level, any p-values less than 0.05 indicate statistically significant difference from zero mean (one can reject the null hypothesis in favor of the alternative). We see in Table 3 that most p-values for kriging are well above this 0.05 threshold, with $\ln(q_i)$ being the only exception, whereas all p-values for the background trend are below this threshold. In other words, on average the universal co-kriging gives predictions that are not significantly different from the true values (except in the case of q_i), and has much greater predictive accuracy than the background linear trend model alone, which gives predictions that differ strongly from the true values.

Table 3. One-sample t-test results for each decline curve parameter and for both the universal co-kriging predictions and background trend predictions.

	Krige		Trend	
Parameter	Mean	p-value	Mean	p-value
a	-3.62	0.94	107.37	0.003
m	-0.44	0.53	2.32	0.006
$\ln(q_i)$	9.79	7.56E-36	13.97	5.63E-45
$\ln(q_\infty)$	-0.10	0.83	4.44	6.35E-09

Besides these hypothesis tests which relate the central tendency and standard deviation of each histogram in Figure 8, it is also worth discussing the extremes of the histograms as well. For most cases, the percent error rarely exceeds +/- 20%. However, for parameter a we see some outlier cases (four of them, for kriging) where percent error can exceed +/- 1000% and get as high as -3800%. These outliers can be explained by the very small magnitude of the true a values for these four cases: they are the closest to zero.

Summary and Conclusion

The approach we are proposing in this paper is a combination of decline curve analysis and geostatistic method. First, the parameters that describe the natural gas production variation are obtained by curve fitting. According to their statistics, we can choose a proper geostatistical method (universal co-kriging) to interpolate those parameters for undrilled sites. Finally, we can estimate the EUR and associated decline curve with predicted parameters while quantifying the uncertainty at the same time. Without asking for extensive types of data, our approach is handy to implement and can provide operators with a good reference for early-stage explorations.

The heat maps generated in our case study have shown that the promising sites with high EUR are mainly located in the middle and northeast of the study area, which is consistent with the actual distribution of existing wells in the Marcellus. Although some areas along the southeast boundary of in the EUR heat map also appear to be sweet spots, the reliability of their EUR may suffer from the relatively high variance of the associated decline curve parameters. Furthermore, the validation not only indicates the superiority of the universal co-kriging over the background trend prediction but also shows reasonable percent errors in predicted parameters despite the "magnitude" issue for parameter a.

Lastly, some extensions and modifications may be applied in our future work such as adopting modified Duong's model (Joshi and Lee, 2013) to avoid potentially unrealistic forecasts or employing pattern recognition method (unsupervised learning) to further narrow down the range of promising areas based upon our current analysis.

References

Anderson, D. M., and P. Liang, 2011, Quantifying Uncertainty in Rate Transient Analysis for Unconventional Gas Reservoirs: the SPE North American Unconventional Gas Conference and Exhibition, p. 14–16 June. SPE 145088–MS, doi:10.2118/145088-MS.

Arps, J. J., 1945, Analysis of Decline Curves: Transactions of the AIME, doi:10.2118/945228-G.

Bhattacharya, S., and M. Nikolaou, 2013, Analysis of Production History for Unconventional Gas Reservoirs With Statistical Methods: Spe, doi:10.2118/147658-PA.

Cipolla, C. L., E. P. Lolon, J. C. Erdle, and B. Rubin, 2010, Reservoir Modeling in Shale-Gas Reservoirs: SPE Reservoir Evaluation & Engineering, doi:10.2118/125530-PA.

Clarkson, C. R., J. L. Jensen, and S. Chipperfield, 2012, Unconventional gas reservoir evaluation: What do we have to consider? Journal of Natural Gas Science and Engineering, doi:10.1016/j.jngse.2012.01.001.

Diaz De Souza, O. C., A. Sharp, R. C. Martinez, R. A. Foster, E. Piekenbrock, M. Reeves Simpson, and I. Abou-Sayed, 2012, Integrated Unconventional Shale Gas Reservoir Modeling: A Worked Example From the Haynesville Shale, De Soto Parish, North Louisiana, in SPE Americas Unconventional Resources Conference: doi:10.2118/154692-MS.

Duong, A. N., 2011, Rate-decline analysis for fracture-dominated shale reservoirs: SPE Reservoir Evaluation & Engineering, v. 14, no. 03, p. 377–387.

Hauge, V. L., and G. H. Hermansen, 2017, Machine Learning Methods for Sweet Spot Detection: A Case Study BT - Geostatistics Valencia 2016, in J. J. Gómez-Hernández, J. Rodrigo-Ilarri, M. E. Rodrigo-Clavero, E. Cassiraga, and J. A. Vargas-Guzmán, eds.: Cham, Springer International Publishing, p. 573–588, doi:10.1007/978-3-319-46819-8_38.

Joshi, K., and J. Lee, 2013, Comparison of Various Deterministic Forecasting Techniques in Shale Gas Reservoirs: SPE Conference, p. 12, doi:10.2118/163870-MS.

Journel, A. G., and C. J. Huijbregts, 1978, Mining geostatistics: Academic press.

Matheron, G., 1971, The theory of regionalised variables and its applications: Les Cahiers du Centre de Morphologie Mathématique, v. 5, p. 212.

McBRATNEY, A. B., and R. WEBSTER, 1983, Optimal interpolation and isarithmic mapping of soil properties: V. Co-regionalization and multiple sampling strategy: Journal of Soil Science, doi:10.1111/j.1365-2389.1983.tb00820.x.

Moinfar, A., A. Varavei, K. Sepehrnoori, and R. T. Johns, 2013, Development of a Coupled Dual Continuum and Discrete Fracture Model for the Simulation of Unconventional Reservoirs, in SPE Reservoir Simulation Symposium: doi:10.2118/163647-MS.

Oliver, D. S., and Y. Chen, 2011, Recent progress on reservoir history matching: A review: doi:10.1007/s10596-010-9194-2.

Ozdemir, I., Q. Sun, and T. Ertekin, 2016, Structuring an Integrated Reservoir Characterization and Field Development Protocol Utilizing Artificial Intelligence: Proceedings of the 26th ITU Petroleum and Natural Gas Symposium and Exhibition.

Trangmar, B. B., R. S. Yost, and G. Uehara, 1986, Application of Geostatistics to Spatial Studies of Soil Properties: Advances in Agronomy, doi:10.1016/S0065-2113(08)60673-2.

Vauclin, M., S. R. Vieira, G. Vachaud, and D. R. Nielsen, 1983, The Use of Cokriging with Limited Field Soil Observations 1: Soil Science Society of America Journal, v. 47, no. 2, p. 175–184.