Comparative evaluation of multi-basin production performance and application of spatio-temporal models for unconventional oil and gas production prediction

M. E. Wigwe¹ · E. S. Bougre¹ · M. C. Watson¹ · A. Giussani¹

Received: 11 May 2020 / Accepted: 13 July 2020 / Published online: 24 July 2020
 © The Author(s) 2020

Abstract

Modern data analytic techniques, statistical and machine-learning algorithms have received widespread applications for solving oil and gas problems. As we face problems of parent–child well interactions, well spacing, and depletion concerns, it becomes necessary to model the effect of geology, completion design, and well parameters on production using models that can capture both spatial and temporal variability of the covariates on the response variable. We accomplish this using a well-formulated spatio-temporal (ST) model. In this paper, we present a multi-basin study of production performance evaluation and applications of ST models for oil and gas data. We analyzed dataset from 10,077 horizontal wells from 2008 to 2019 in five unconventional formations in the USA: Bakken, Marcellus, Eagleford, Wolfcamp, and Bone Spring formations. We evaluated well production performance and performance of new completions over time. Results show increased productivity of oil and gas since 2008. Also, the Bakken wells performed better for the counties evaluated. We present two methods for fitting spatio-temporal models: fixed rank kriging and ST generalized additive models using thin plate and cubic regression splines as basis functions in the spline-based smooths. Results show a significant effect on production by the smooth term, accounting for between 60 and 95% of the variability in the six-month production. Overall, we saw a better production response to completions for the gas formations compared to oil-rich plays. The results highlight the benefits of spatio-temporal models in production prediction as it implicitly accounts for geology and technological changes with time.

Keywords Spatio-temporal models · Data analytics · Unconventional reservoirs · Fixed rank kriging (FRK) · ST-GAMs

List of symbol

A_i	Row of matrix for any parametric model component	$f_j(x_{ji})$	Smooth function of covariates, x_{ji}
ST	Spatio-temporal	k	Basis dimension
s, x	Space (longitude and latitude)	t_p, c_r	Thin plate, cubic regression spline model
t, r	Time	AIC	Akaike information Criteria
$Y(s, t)$	True process of interest indexed in space and time	BAU	Basic areal unit
$y(s, t), Z(x; r)$	Potential or observed data indexed in space and time	GAM	Generalized additive models
$\mathcal{N}_s$	Neighbors of spatial location s	FRK	Fixed rank Kriging
$\theta_D, \theta_p, \theta_h$	Model parameter for data model, process model and hyper parameter	TOC	Total organic carbon
$g(\mu_i)$	Function representing the model	MSPE	Mean squared prediction error
θ	Parameter vector	RMSPE	Root mean squared prediction error
		PCV	Predictive cross-validation score
		SCV	Standardized cross-Validation score
		CRPS	Continuous rank probability score
		Tcf, Bcf, MMcf	Trillion, billion, million cubic feet
		Mbbl, bbl	Thousand barrel, barrel
		1bbl	42 gal = 5.615 ft ³ = 0.159 m ³
		1 ft	0.3048 m
		1 lbs	4.54 × 10 ⁻⁴ Metric ton

✉ M. E. Wigwe
 marshal.wigwe@ttu.edu

¹ Bob. L. Herd Department of Petroleum Engineering, Texas Tech University, Lubbock, TX, USA

Introduction

The data analytic lifecycle, in relation to big data problems and data science projects, is classified into six phases (EMC 2015). These consist of discovery, data preparation, model planning, model building, communication of results, and operationalization. Ideally, it is preferable to go in the feed-forward direction (from discovery to operationalization), but often, this is a back and forth process throughout the duration of the project. The goal of the discovery phase is to develop a problem statement and formulating an initial hypothesis that we test using data. The data preparation phase includes gathering the needed data from various sources, cleaning and performing necessary transformations, in addition to coming up with a plan for handling and storing the collected data and data generated during the project lifecycle. In the third phase, we assess methods for building the model in the next phase. The primary activity in this phase in exploratory data analysis (EDA) with the aim of examining relationships among variables and selecting those variables that are of interest in the project and those that show some promise during the EDA process for further consideration. This phase can also suggest appropriate models for consideration during the model building phase. The model planning phase involves developing a workflow for building the model. In the model-building phase, we construct the model using the information from previous phases, and the workflow developed in the model planning phase (Wigwe et al. 2020).

In the oil and gas industry, spatio-temporal and other machine learning models have received a range of applications in different projects. The general area of the application of data-driven analytics is called “*Petroleum Data Analytics*” (Mohaghegh 2016). Ettehadtavakkol and Jamali (2019) presented a spatio-temporal analysis of water production from Marcellus shale using kriging estimation. Siddiqui, et al. (2019) used machine learning modes to study fluid type variation and completion optimization in the Eagleford. Zhou et al. (2014) applied data mining techniques to evaluate gas production performance in the Marcellus. Wigwe et al. (2019a, b) presented both spatial and neural network techniques to analyze Bakken oil production while Zargari and Mohaghegh (2010) showed an application of machine learning models for the Bakken field development planning. Simha et al. (2019) integrated spatio-temporal unsupervised learning method with reservoir simulation to identify a unique scenario for assessing the impact of uncertainties on production. Although the modeling workflow presented in this paper is similar for other data analytic applications, the specific methods implemented, and their formulation is not native to the oil

and gas industry. As a result, we will present a brief mathematical modeling background and provide some details on the two techniques presented in this paper for spatio-temporal models before presenting results of application for oil and gas production prediction.

Spatio-temporal statistics

Almost every data collected is usually associated with space and time. In a “non-spatio-temporal” data, the ST components were either not recorded or discarded when space and time are not of interest to the observer or the research objective. In a spatio-temporal dataset, we include information about *where* and *when* the data was collected (Cressie and Wikle 2011). Hence, spatio-temporal statistics is the statistical analysis of such data. In ST data analysis, the analyst may be interested in one or more of the following goals (Wikle et al. 2019):

- 1 Gaining more understanding of the data.
- 2 Looking for relationship between two ST processes.
- 3 Making prediction in space and time.
- 4 Inference on model parameters.
- 5 Forecasting in time, etc.

Two traditional approaches to spatio-temporal modeling are a descriptive approach and a dynamic approach. In the descriptive approach, the spatio-temporal model uses the mean and covariance functions to characterize the ST process (Cressie and Wikle 2011; Wikle et al. 2019). The kriging method is based on this approach. Variability (or uncertainty) is captured through a marginal probability distribution. There are several reasons why we use this modeling approach in practice, one of them being a lack of understanding of the spatio-temporal process under study. The dynamic approach further incorporates how spatio-temporal processes evolve over time and are built from conditional probability distributions (Cressie and Wikle 2011). Stroud et al. (2001) proposed a modeling framework for space–time data that accounts for the spatial variability by modeling the mean function at each time as a locally weighted mixture of regression surfaces while they account for temporal variability by allowing component surfaces to evolve through time.

Methodology

Spatio-temporal exploratory data analysis (EDA)

Visualization of spatio-temporal data poses specific challenges because of the number of dimensions needed to present the plots. At least three dimensions are suggested

as a minimum to be displayed at the same time (Cressie and Wikle 2011), representing two- or three-dimensional space and time. Some of these plots come in the form of maps, colors, and animations, and enable a simple presentation of important information that leads to the development of appropriate spatio-temporal models (Wikle et al. 2019). Static maps, multi-panel plots (Pebesma 2012), Hovmöller diagram (Cressie and Wikle 2011; Hovmöller 1949; Pebesma 2012; Wikle et al. 2019) and animations are all space–time plots. We generally use histograms and box-plots to show the distribution of a single continuous variable (EMC 2015; Navidi 2015; Westfall and Henning 2013; Wikle et al. 2019). For details and applications of these ST visualizations to oil and gas datasets see Wigwe and Watson (2021).

Spatio-temporal models

At a minimum, a dynamic spatio-temporal model, from a hierarchical modeling framework, requires the specification of a data model. This is a conditional model of the data, conditioned on the true process of interest and some model parameters. A process model captures how a spatio-temporal process evolves with time, along with some given parameters. Model for parameters is specified to yield the Bayesian hierarchical model (BHM) or estimates of the parameters are provided to yield the empirical hierarchical model (EHM). In general, we represent a spatio-temporal model by the stochastic process using Eq. (1) (Blangiardo and Cameletti 2015):

$$Y(s, t) \equiv \{y(s, t), (s, t) \in \mathcal{D} \subset \mathbb{R}^2 \times \mathbb{R}\} \quad (1)$$

where $\mathcal{D}$ is a subset of real numbers $\mathbb{R}$ in 2-D space and time. $y(s, t)$ implies that the process is indexed by space and time. In the framework presented by Cressie and Wikle (2011), at the top level of the hierarchical model, the data model is given by Eq. (2):

$$\{Z(x; r) : x \in D_s, r \in D_t\} | \{Y(s; t) : s \in \mathcal{N}_s, r \in \mathcal{N}_r\}, \theta_D \quad (2)$$

where $Z(x; r)$ is an observation (data) at spatial location x and time r , $Y(s; t)$ is the process of interest at spatial location s and time t and θ_D represents the model parameters for the data model, which could vary with space and/or time. At the second level, the process model is given by Eq. (3):

$$Y(s; t) | \{Y(w; t - \tau_1)w \in \mathcal{N}_s^{(1)}, \dots, Y(w; t - \tau_p)w \in \mathcal{N}_s^{(p)}\}, \theta_p \quad (3)$$

where $\mathcal{N}_s^{(1)}, \dots, \mathcal{N}_s^{(p)}$ are the neighbors of spatial locations s , which corresponds to time lags $0 < \tau_1 < \dots < \tau_p$ and θ_p is the model parameter for the process model, which could vary with space and/or time as well. At the third and final

level, the parameter model is given by: $[\theta_D, \theta_p | \theta_h]$, where θ_h represents hyper-parameters.

Fixed rank Kriging (FRK)

FRK facilitates optimal spatial prediction for large spatial and spatio-temporal (ST) datasets (Wikle et al. 2019; Zammit-Mangion and Cressie 2017). It constructs a spatial random effects model on a fine resolution discretized spatial domain known as a basic areal unit (BAU) whose primary use is to account for problems related to change of support (Wikle et al. 2019). The model decomposes spatial random processes using spatial (or ST) basis functions. Model parameters are estimated using the expectation–maximization algorithm. This prediction framework is computationally efficient because of the reduction in dimensionality using basis function. If covariates are included in the model, they must be specified in the BAUs.

Generalized additive models

This class of models is similar to generalized linear models whose linear predictors are smooth functions of the covariates (Hastie and Tibshirani 1986, 1990; Wood 2017). The model is given by Eq. (4):

$$g(\mu_i) = A_i \theta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + \dots \quad (4)$$

where $\mu_i \equiv E(Y_i)$ and $Y_i \sim EF(\mu_i, \phi)$. Y_i is a response variable, $EF(\mu_i, \phi)$ denotes an exponential family distribution with mean, μ_i , and scale parameter, ϕ , A_i is a row of matrix for any parametric model component, θ is the corresponding parameter vector, and the f_j are smooth functions of covariates, x_k . The model allows for a flexible specification of the dependence of the response on the covariates, but by specifying the model only in the smooth functions, rather than a parametric relationship. The smooth functions use penalized regression splines (or other splines) as basis functions with a specified number of dimensions, k , which controls model smoothness. *Thin plate regression splines*, *cubic regression splines*, and *Splines on the sphere* are some of the most popular basis used in GAM models. Basis functions (Wood 2000, 2003, 2017) enable efficient approximation in this set up by means of a smoothness parameter, which we choose by cross-validation–generalized cross-validation (GCV) score.

Model selection, validation, and diagnostics

Training-data validation, within-sample validation, forecast validation, hindcast validation, and cross-validation are the approaches used to compare model predictions with real-world observations (Hastie et al. 2009; Wikle et al. 2019). We compare the performance of several models on the

training and the test data using model diagnostics. These metrics are also used in selecting the best model. Graphically, for regression problems, the residual plots are useful for checking model assumptions. The conditional quantile plot (Wilks 2011) is also a useful diagram that plots the predicted values on the x-axis, and the quantiles from the *empirical predictive distribution of the observations associated with the predictions* on the y-axis. Any bias in the model predictions becomes apparent, depending on the position of the plot vis-a-vis the 45° diagonal line. The mean squared prediction error (MSPE) or its square root version (RMSPE) is the most used diagnostic/model validation statistic. It captures issues related to bias and variance. The predictive cross-validation score (PCV) and the standardized cross-validation score (SCV) are also measures used in evaluating model performance (Kang et al. 2009). Lower values of PCV and MSPE, and SCV closer to 1 indicate better model performance. Scoring rules for spatio-temporal predictions compare prediction distribution to a validation observation. The commonly used scoring rule for continuous variables is the continuous rank probability score (CRPS). Models with Lower CRPS are better. The Akaike information Criteria (AIC) is a model selection criterion that penalizes bias due to overfitting when evaluating models using training data, and the number of parameters used in fitting the model. Model parameters are estimated using maximum likelihood. When comparing several models, the model with the lowest AIC is the best. See these references for detailed treatment (Hastie et al. 2009; Hooten and Hobbs 2015; Westfall and Henning 2013; Wikle et al. 2019; Wilks 2011).

Results and discussion of case studies

Application of spatio-temporal models for characterization of well performance is presented in this section. First, we present a geological overview of the five formations used as case studies. Using available data, we evaluate and compare well production performance across these formations on a yearly basis. We also carry out this comparative analysis using the first-year production data by completion year to capture any improvement in the performance of new wells over time. To build spatio-temporal models for each formation, we develop a workflow for each method that would be used across each formation. Using this workflow, we build the spatio-temporal models, evaluate the performance of the models using the model diagnostics presented under methodology and discuss the results. Model results plots presented, and discussions are for the major fluid phases in these formations. Models for the minor phases were also developed and the results presented in tabular form at the end.

Geology of formations and data preparation

The Bakken formation is in the Williston Basin and covers the western part of North Dakota, Montana, Manitoba and Saskatchewan in Canada and was first discovered in 1953. It has an areal coverage of about 200,000 sq. miles and the formation thickness ranges from 0 to 140 ft. The depth of the formation ranges from about 1000 ft in parts of Canada to about 15,000 ft in some areas of North Dakota (Kuhlman et al. 1992). The upper and lower formations are comprised mostly of shale while the middle formation is comprised mostly of sandstone, limestone, and siltstone. The upper and lower Bakken are organic-rich marine shales and are the petroleum source rock for the oil and gas produced from the Bakken petroleum system (Kumar et al. 2013; Li et al. 2015; Sonnenberg 2014; USGS 2008). The middle Bakken reservoir is the focus of most development activities in recent years (Kumar et al. 2013). The hydrocarbons that were generated in the formation resulted in over pressurizing the formation which then led to the creation of natural fractures. These natural fractures have been the main cause of increased permeability and productivity within the Bakken Formation (Jin et al. 2015; Tomomewo et al. 2019; Tran et al. 2011). The Bakken formation was estimated to contain between 3 and 4.3 billion barrels of recoverable oil and 1.85 trillion cubic feet of associated gas by the United States Geological Survey (USGS) in 2008 (USGS 2008). We retrieved production and completion data from North Dakota Industrial Commission (NDIC) and DrillingInfo. The Bakken analysis focuses on the McKenzie County, in which 2349 horizontal wells with completion data are available for study.

The Marcellus formation is an organic-rich shale that occurs in the subsurface of four states in the USA, these are Ohio, West Virginia, Pennsylvania, Maryland, and New York (Bartuska et al. 2012; Koesoemadinata et al. 2011; Yildirim et al. 2019; Zamirian et al. 2016). The formation is divided in two, the Upper Marcellus and the Lower Marcellus with the Lower Marcellus having a significantly higher concentration of organic matter as compared to the Upper Marcellus. It covers an area of more than 100,000 sq. miles. The EIA (2017) estimates oil reserves of 143 million barrels (MMbbls) and 410 trillion cubic feet (Tcf) of gas in place with recoverable gas at 50 Tcf. The Marcellus dataset contains 2020 horizontal wells completed since 2008 in three counties: Washington, Greene and Fayette in the south-west corner of Pennsylvania.

The Eagle Ford Shale formation is a hydrocarbon bearing formation found in South Texas. It is best known for producing variable amounts of dry gas, wet gas, Natural-gas liquids, Condensates, and more oil than other traditional shale plays. It is believed to be the source rock for many conventional oil and gas fields in the Texas Gulf Coast. The formations usually targeted are the Lower Eagle Ford, Upper Eagle Ford

and the Austin Chalk (Shelley et al. 2012). It extends across 26 counties from East Texas to the Mexican border with an acreage of about 20,000 sq. miles and the thickness ranges from 50 to 300 ft according to the Railroad Commission of Texas (RRC 2020). The formation is estimated to contain 66 trillion cubic feet of natural gas 8.5 billion barrels of oil and 1.9 billion barrels of natural gas liquids (USGS 2018). Most of the rock within the Eagle Ford formation is very brittle hence it is a good candidate for drilling horizontal wells and for hydraulic fracturing (Jaripatke and Pandya 2013; Lalehrokh and Bouma 2014; Nwabuoku 2011; Siddiqui et al. 2019). The Eagleford dataset contains 3413 horizontal wells located in LaSalle County, TX.

The Delaware Basin is one of the prolific basins in the USA comprising of stacked reservoirs which are the Wolfcamp shale, bone spring formation and Avalon shale. The Wolfcamp shale and bone spring formation are collectively called the “Wolfbone” play (Lohoefer et al. 2014a, b; Lalehrokh and Bouma 2014; Sharma et al. 2014; Yates et al. 2013). The Delaware Basin is in South East New Mexico (Eddy, Chaves, and Lea County) and West Texas (Culberson, Pecos, Loving, Terrel, Reeves, Ward and Winkler County). USGS estimates recoverable hydrocarbons to be in excess of 19 billion barrels of oil, 1.6 billion barrels of natural gas liquids and 16 trillion cubic feet of natural gas (EIA 2018, 2019). The Wolfcamp and Bonespring formation dataset contains 2295 horizontal wells drilled in Lea County, NM.

We gathered data of 10,077 horizontal wells from January 2008 to June 2019 from state agencies and drillinginfo. The data was cleaned, combined in a long format, and stored in a database for analysis (Table 1). The basic covariates needed for spatio-temporal modeling are space and time, where space is location (longitude and latitude) and time represents the number of months since January 2008. Hence, $t = 1$ for January 2008, $t = 12$ for December 2008 and so on. The space–time covariates are available for all formations and were used to characterize the spatio-temporal process for the formations. The Bakken formation has additional information. The wells have completion data (stages, perforated interval, pounds of proppants and volume of fluid), and geologic data (TOC content of upper, and lower Bakken along with isopach for upper, middle and lower Bakken (Source:

Nordeng and Helms 2010)). These additional covariates were included in the Bakken model. It was necessary to normalize the covariates before modeling as this expedited computation and removed any scaling effect. In this analysis, we used the min–max normalization to scale the variables to $[0, 1]$ using: $n_X_i = (X_i - \min(X)) / (\max(X) - \min(X))$. The six-month cumulative oil and gas production are the dependent variables. We apply log transformation (or log-link function for GAM model) to the dependent variables due to their lognormal distribution.

Comparative production analysis

Figure 1 shows a comparative plot of the production performance of wells in each formation. The Bakken, Eagleford, and Wolfbone formations primarily produce oil while the Marcellus is a gas-rich formation, with the Eagleford producing a good amount of gas, especially during its initial development years. Gas production from the Eagleford has been on a decline since 2010 with most of the new wells targeting oil-rich portions of the formation. The number of active wells has steadily increased in all formations while new wells are on the rise for the Permian basin formation. This plot is informative but does not fully display the performance of the new wells shown on the bottom right in Fig. 1.

Figure 2 shows 12-months cumulative production per well for new wells only. Oil and gas production from new wells increase each year for all formations, except for gas production from the Eagleford for reasons already mentioned above. The increased productivity of new wells could be a direct result of technological improvements in drilling and completion design: longer laterals/extended reach wells, higher stages, more pounds of sand and injected volume of frac fluid, fracture complexity (zipper fracs) resulting in the opening of natural fractures, etc. The Bakken formation outperforms the other oil-producing formations consistently, with comparable production from the Wolfbone formation since 2016.

Model setup

Figure 3 shows the modeling workflow. For the FRK model, we construct the BAUs using the data. BAUs are the basic framework on which we build the model and carry out predictions. Figure 4 shows the constructed BAUs. In Fig. 5, we show the spatial basis function constructed for each formation using 2 resolution. Adding a second resolution helps capture finer details across space. With two resolutions, we generate a total of 77 basis functions across the spatial domain. There are 12 temporal basis functions with an aperture size of 6 months spanning the 11.5 years of production data (Fig. 6). Taking a tensor

Table 1 Number of horizontal wells in each formation

Formation	County	Well count
Bakken	McKenzie, ND	2349
Eagleford	La Salle, TX	3413
Marcellus	Washington, Greene and Fayette, PA	2020
“Wolfbone”	Lea, NM	2295
Total		10,077

Fig. 1 Comparison of production from all formations. Oil and gas yearly productions are reported on a per well basis to eliminate the effect of well count. Colors used are consistent across all four plots as shown in the legend

product of the spatial and temporal basis functions results in 924 spatio-temporal basis function that we supply to the FRK model.

For the GAM model, the type of basis and its dimension enable a reasonable approximation of the underlying data-generating process. We used the thin plate regression spline, “*tp*”, and cubic regression spline, “*cr*”, for the spatial and temporal dimensions, respectively, and their tensor product evaluated to obtain ST basis functions. Selection of the basis dimension, k , which is the number of basis function to construct, is an iterative process. It enables the model to capture the inherent variability in the data. After several iterations, we arrived at $k = (50, 20)$ as the appropriate basis dimension, leading to $k - 1 = 999$ spatio-temporal basis functions. This is equivalent to the FRK setup. Once these components are determined, we build the GAM model. If any of the covariates or smooth functions are unimportant, as reported by the p value at the 5% level, we drop the covariate and re-fit the model. Finally, we make predictions at test locations and visualize model results. Save the best model for deployment for an undrilled location.

Bakken formation

Figure 7 shows a multi-panel, faceted, spatial plot of the Bakken data. It shows the yearly oil production in barrels per day per stage, where the days represent the total number of days the well was online in the specified year. This is represented by the colors while the size of each bubble correlates with the number of stages. The facets are populated with some information like the year, the total number of wells and, how many of those wells were new completions. In 2013, for instance, the map contains 986 wells for which 406 are new completions. Although we see larger bubbles as we go across time, there does not seem to be any spatial correlation with the number of stages. Similar observations were made using the other completion variables available (the pounds of proppant, the volume of fluid, the perforated interval, and the per stage versions—plots not shown).

Another observation from Fig. 7 is the increase in activity around the north-east corner of the County. Wells drilled in this area have a higher initial rate than wells drilled in the western part, regardless of the number of stages used in completion. Geologically, this portion of the

Fig. 2 Comparison of 12-months cumulative production from new wells for all formations as a function of completion year. Oil and gas production are reported on a per well basis to eliminate the effect of

well count. Colors used are consistent across all plots. 2019 data contains at best four to six months of production data and hence could equal or surpass prior year's results

Fig. 3 Modeling workflow/flowchart

Middle Bakken has a higher thickness. Figure 8 shows the distribution of six-month cumulative oil production. The left figure shows actual production values that suggest, on average, increasing oil production with time as the distribution shifts toward the right. Each plot has a similar distribution with different characteristics (average, standard deviation, kurtosis, and level of skewness). Adding this temporal dimension gives us further insight into incremental production or the addition of reserve as new wells come online. What the plot does not show is that technology has “changed”. Operators have increased the length of laterals, the number of stages, volume of fluid, and pounds of proppants pumped in these completion-dependent unconventional wells. Consequently, generating a similar plot (on the right) that captures this dependence on the size of the completion becomes necessary. The figure on the right shows that the temporal dimension does not have a notable effect on the per stage normalized production, as the six-month cumulative oil production per stage has a distribution with parameters that are in a comparable range across time. If completion data is not available, then modeling oil production as dependent on time is advisable.

Fig. 4 Basic areal units (BAUs) generated for: **a** Bakken, **b** Wolfbone, **c** Eagleford, and **d** Marcellus. We can perform FRK predictions on these grids

The correlation plot matrix (Fig. 9) shows some problems with correlated variables. A solution will be to use principal components (PCs) as variables in the model (Everitt and Hothorn 2009; Hastie et al. 2009; Jolliffe 2002; Zhou et al. 2014). An additional advantage of using PC is for dimensionality reduction when the number of covariates is high. There are 9 covariates in the Bakken data, which would yield 9 PCs. This is a moderate number of covariates. Some authors have argued about using the total PCs for regression analysis rather than the first few that capture enough variability in the data. This is because of the possibility that lower PCs may actually have a significant influence on the model at the 5% level (see Jolliffe 2002). Following this suggestion would yield a similar result to the original covariates. Hence, we would carry out this analysis with the original scaled variables regardless of the correlation.

The Temporal histogram (Fig. 8) suggests an approximately lognormal distribution, but a gamma distribution will be more appropriate (Fig. 10). Gamma distribution resulted in the best fit.

Table 2 shows a summary of the full model for the six-month oil production. This result implies that only the completion variables are significant at the 5% level in this model (as shown by the probability column) and the completion year does not influence the model performance. The result also shows that including the ST component would significantly improve the model. Consequently, the final model contains the significant variables. We used the final model for prediction at test locations. Figure 11 shows the actual and predicted values for the GAM and FRK model. The FRK model outperforms the GAM model and would be the

preferred model for this dataset. We evaluated other model diagnostics and selection criteria and present the results in the “[Model summary](#)” section.

Eagleford formation

Figure 12 shows a map of the Eagleford with horizontal wells drilled since 2008. Wells with higher gas production are on the southern portion. In a previous study (Wigwe and Watson 2021), we observed an increased activity from 2008 to 2019 with most wells brought online in 2013 and 2014. We also found that there is a great deal of variability in oil and gas production both spatially and temporally across La Salle county. Hence, the goal of this study is to develop ST models to capture this variability in order to make predictions of oil and gas production for undrilled locations. We present and discuss the results of FRK and GAM spatio-temporal models of six-month gas production. A summary that includes model results for oil production is presented at the end.

Figure 13 shows the distribution of six-month oil and gas production using temporal histograms. A lognormal distribution or a gamma distribution with a log-link function would be appropriate to model this data. The left panel shows decreasing gas production with time as the distribution shrinks. This reduction in gas production is most likely due to the drilling of more wells in the oil-rich sections of the formation. The oil production (on the right figure) has not increased reasonably, but the minimum has “moved” with time toward the right (Wigwe and Watson 2021).

Fig. 5 FRK model Spatial Basis functions for: **a** Bakken, **b** Wolfbone, **c** Eagleford, and **d** Marcellus. Two resolutions produce 77 basis functions across each spatial domain

Fig. 6 Temporal Basis function for FRK models with 12 basis functions. The same specification was used across all four formations. The tensor product of the spatial and temporal basis function yields 924 ST basis function for the FRK model

Figures 14 and 15 show model results for FRK and GAM, respectively. Both models result in a good match with the FRK performance slightly better for this formation. The

right figures show the distribution of prediction error for each model. As we will show later, the model predictions tend to give a better match for gas production compared to

Fig. 7 Spatial map of wells in the Middle Bakken in McKenzie County. Colors represent yearly oil production in barrels per day per stage while the size is mapped to the number of stages. See Wigwe et al. (2020)

Fig. 8 Temporal histogram showing the distribution of bakken oil production data

oil production across all the formations. This is related to the better response of gas to completion than oil due to its smaller molecular size.

Marcellus formation

The space–time plot shown in Fig. 16 shows the development of the Marcellus in three counties in Pennsylvania. Fourteen of the wells drilled in 2008 clustered around the

center of Washington County. With time, development continued into the other two counties such that by June 2019, there were 1976 active wells producing from the Marcellus in these three counties. Of the 12,000 points plotted on the map, only 1000 cases were less than 100 MMcf annually, indicating how prolific these wells were.

Figure 17 shows the temporal boxplot of six-month gas production. The average production increased from 250 to 2000 MMcf and the distribution for each period is more

Fig. 9 Correlation plot of Bakken variables **Fig. 10** Suggested choice of distribution for fitting six-month oil. Legend is ordered from best to worse fit **Table 2** GAM full model summary for the Bakken

Variable	Estimate	SE	t Value	Pr(> t)
Intercept	11.26129	0.21653	52.008	< 2E-16
Number of stages	0.53478	0.08827	6.058	1.61E-09
Perforated interval (ft)	0.23373	0.07566	3.089	0.002031
Pounds of proppant (lbs)	0.38041	0.11912	3.194	0.001424
Total fluid volume (bbl)	0.31369	0.08748	3.586	0.000343
Completion year	-0.03706	0.27675	-0.134	0.893492
TOC lower Bakken	0.03092	0.09052	0.342	0.732647
TOC upper Bakken	-0.19148	0.11132	-1.72	0.085543
Height lower Bakken	-0.18101	0.10434	-1.735	0.082919
Height MiddleBakken	-0.30186	0.15607	-1.934	0.053227
Height upper Bakken	-0.29808	0.18417	-1.618	0.1057
ST component: <i>te(Lon, Lat, t)</i>				< 2E-16

The estimates are the variable coefficients in the model

right-skewed due to higher producing wells. This increase is due to technological changes related to improvements in drilling and completion design. In line with the previous analyses, a gamma or lognormal-type distribution is appropriate for modeling this data.

Figures 18 and 19 show the results of FRK and GAM models. While the FRK model tends to underestimate the high gas volumes, resulting in the higher prediction error, both models sufficiently capture the spatio-temporal variability in the data. Figure 20 shows the diagnostic conditional quantile plots of the results. The plot is useful for capturing any evidence of bias in the predicted values. As shown, the GAM model does not show bias in the predictions while the FRK model shows bias when predicting gas production above 3 Bcf. We select the GAM model for prediction of the expected performance of undrilled locations.

Fig. 11 Actual versus predicted six-month oil for GAM (left) and FRK (right) for the Bakken formation

Fig. 12 Map of Eagleford showing horizontal wells by major phases. We analyzed wells located in La Salle County in this study

Delaware basin: Wolfcamp and Bonespring – “Wolfbone”

We present data for oil production in the New Mexico side of the Delaware Basin (in Lea County). Average oil production increased gradually with time, with more drastic increases recorded from 2016 to 2019 (Fig. 21). Gamma distribution provided the best fit to the data for modeling of oil production. Figure 22 shows perspective and contour plots of the GAM model. The plots are 3-D visualization

of predicted oil production. The contour lines represent predicted six-month oil production. It shows the flexibility of the GAM model in characterizing the variability in the data across the spatio-temporal domain. The behavior of this model mimics reality. Figure 23 shows the predicted and actual oil production from the GAM and FRK models. Both models have reasonable performance and tend to overestimate production (biased high), although the FRK model has a superior performance for this formation in general.

Fig. 13 Temporal histogram for the Eagleford oil and gas data, each plot shows a similar distribution across time

Fig. 14 Actual versus predicted six-month gas (left) along with the prediction error (right) for the Eagleford formation from the FRK model

Model summary

There are variations in spatio-temporal development and the production response of the formations studied. The Bakken development has focused on the north-eastern part of the formation in McKenzie County. The Eagleford development covers La Salle County with active development focused on the SW-NE diagonal. For the Marcellus, initial development was in Washington County and spread to the other two counties over time. Table 3 presents a summary of all final models constructed for the unconventional plays along with the model diagnostics discussed

in the “[Model selection, validation, and diagnostics](#)” section. We constructed different models for oil and gas production. The mean represents the mean of the data while the predicted mean is the mean of the predicted values from each model. The GAM model performs poorly at predicting individual data points but performs well when estimating the mean of the data. Hence, if the goal of the study is capturing the mean, then select the GAM model, otherwise, the FRK model is better suited for the analysis.

The importance of drilling and completion technology, and geological consideration to the successful development of unconventional resources is well documented (Chong et al. 2010; Kolawole et al. 2019; Kolawole and

Fig. 15 Actual versus predicted six-month gas (left) along with the prediction error (right) for the Eagleford formation from the GAM model

Fig. 16 ST map of the marcellus wells in three counties shown, PA. Colors represent yearly gas production. Each panel shows the number of wells and new drills

Ispas 2019; Maity and Ciezobka 2019; Pope, et al. 2010; Sharma et al. 2014; Soliman, et al. 2014). For this reason, completion and geologic parameters were included in the Bakken models because these variables were available. However, the performance of the models did not show much improvement over a case (results not presented) which utilized only location and time parameters. This

could be because the ST parameters have implicitly accounted for geologic and technology since the data reflect these changes have improved well performance over time (Wigwe et al. 2019a, b). As a result, we expect the models for the other formations to reasonably describe geological and technological variations across each play over time in the absence of these variables in the model.

Fig. 17 Temporal boxplot shows bi-annual Marcellus gas production. The dashed lines represent the group average

Fig. 18 Actual versus predicted six-month gas for GAM (left) and FRK (right) for the Marcellus formation

Fig. 19 Distribution of prediction error using GAM (left) and FRK (right) for the Marcellus formation

Conclusion

We presented an application of spatio-temporal models for production evaluation in a multi-basin study of four

unconventional formations in the USA: the Bakken, Marcellus, Eagleford, and Wolfcamp and Bonespring formations. For each formation, we presented ST plots, including the multi-panel space–time plots, temporal histogram, and temporal boxplot. These visualizations enabled us to select

Fig. 20 Conditional quantile plot of six-month gas in Bcf for GAM (left) and FRK (right) showing that both models produced predictions with no bias for the Marcellus dataset

Fig. 21 Temporal boxplot shows bi-annual Wolfbone oil production, with the most increase seen from 2014–2015 to 2016–2017 period

a suitable distribution and suggest appropriate covariates to include in the model. Using the workflow presented in Fig. 3, we fit two ST models (Fixed rank kriging, FRK, and spatio-temporal generalized additive model, GAM), and compared their predictions. The FRK model produced better results compared to the GAM model. Both models overestimated predicted production values, with more bias in the GAM model predictions than the FRK model results. The GAM model performed better at predicting the mean of the data. We select the GAM model if this is the goal of the study. In summary:

- 1 From the production evaluation, we observed incremental average oil and gas production for new wells in each formation from 2008 to 2019. The Bakken formation consistently outperformed the other oil-producing formations during this period for new wells (Fig. 2).
- 2 Spatio-temporal models have applications in the oil and gas industry. With properly formulated models, we can perform spatio-temporal production evaluation successfully, regardless of the specific basin studied as shown in this paper. These techniques highlight the importance of

space and time in production prediction as it takes geology and technological changes with time into account.

- 3 The tensor product of space and time showed a strong influence on the GAM model.
- 4 Overall, the models account for between 60 and 95% of the variability in the six-month production in the oil-producing and gas-producing formations.
- 5 The FRK model performs better than the GAM model across all formations and production streams evaluated, excepting the Marcellus model that favors the GAM model as shown by the computed model diagnostics in Table 3. This could be because the FRK model is a specialty model designed specifically for spatial and spatio-temporal datasets, and as a result is able to capture the covariance structure of the dataset. Hence, the FRK model would be the preferred model for prediction of the performance of undrilled locations.
- 6 Both models tend to overestimate oil and gas production, with the GAM model showing a much higher bias compared to the FRK model.
- 7 Overall, the gas models show a better response to capturing the variability in gas production than the oil models

Fig. 22 GAM model predictions capture spatio-temporal variability in six-month oil production (in barrels) across the formation. The un-contoured regions indicate no activity or absence of data

Fig. 23 Actual versus predicted six-month oil for GAM (left) and FRK (right) for the Wolfcamp and Bonespring formations

do on oil production. This behavior of the models is consistent across all four formations. We posit that this observation is related to the response of the fluid phase to completion, with the molecular size of the gas phase playing a key role in this regard.

- 8 There are several variables that could be included in a model formulation as apparent when carrying out a reservoir simulation study. The results presented in this paper suggests that space and time have strong correlation with oil and gas production, and this is based on

sound scientific principles that are in line with Tobler's First Law. Hence, we recommend these methods to provide a first pass result when studying a given field before commissioning a full-scale reservoir (simulation) study. With the availability of more covariates that would be influential on the dependent variable, we expect model performance to improve.

Table 3 Summary of model diagnostics and selection criteria

	Diagnostic	Formation	BK-Oil	BK-Gas	EF-Gas	EF-Oil	MC-Gas	WC-Gas	WC-Oil
R^2		FRK	0.915	0.929	0.9457	0.8457	0.9294	0.8611	0.7274
		ST-GAM	0.614	0.6332	0.9131	0.7369	0.9394	0.6725	0.6066
AIC		FRK	51,651	54,743	80,540	71,580	54,549	50,666	48,368
		ST-GAM	55,239	58,636	82,077	73,326	54,243	52,427	49,120
Bias		FRK	51,668	54,760	80,558	71,598	54,566	50,683	48,384
		ST-GAM	55,256	58,653	82,096	73,344	54,260	52,444	49,137
PCV		FRK	−0.0169	−0.0221	−0.0816	−0.0346	−0.016	−0.0265	−0.009
		ST-GAM	−0.1044	−0.1665	−0.1695	−0.0786	−0.0213	−0.1698	−0.0864
SCV		FRK	0.1134	0.1294	0.3987	0.179	0.0526	0.1849	0.1325
		ST-GAM	0.5286	0.946	0.6397	0.2852	0.0466	0.7403	0.2082
RMSPE		FRK	1.288	0.9149	1.6007	1.4586	1.0027	0.8383	1.5997
		ST-GAM	145.16	162.35	12.42	11.82	3.57	16.37	10.98
CRPS		FRK	0.3368	0.3597	0.6314	0.4231	0.2293	0.43	0.364
		ST-GAM	0.727	0.9726	0.7998	0.5341	0.2158	0.8604	0.4563
Mean			0.1188	0.1437	0.2698	0.1649	0.1196	0.1876	0.1971
Predicted mean		FRK	0.281	0.378	0.3533	0.2192	0.117	0.3747	0.2698
		ST-GAM	11.06	11.35	11.14	10.29	13.37	11.51	11.15

Acknowledgements The authors wish to acknowledge the North Dakota Industrial Commission's Oil and Gas Division and Drilling-Info for providing datasets used for this research. We acknowledge use of the public license for R programming language, RStudio IDE for R, and contributed libraries.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bartuska JE, Pechiney JJ, Leonard RS, Woodroof RA (2012) Optimizing completion designs for horizontal shale gas wells using completion diagnostics. In: SPE Americas unconventional resources conference. Society of Petroleum Engineers, Pittsburgh, Pennsylvania USA, <https://doi.org/10.2118/155759-MS>
- Blangiardo M, Cameletti M (2015) Spatial and spatio-temporal bayesian models with r-INLA. Wiley, New York. <https://doi.org/10.1002/9781118950203>
- Chong KK, Grieser WV, Passman A, Tamayo HC, Modeland N, Burke BE (2010) A completions guide book to shale-play development: a review of successful approaches toward shale-play stimulation in the last two decades. In: Canadian unconventional resources and international petroleum conference. Calgary, Alberta, Canada: Society of Petroleum Engineers. <https://doi.org/10.2118/133874-MS>
- Cressie N, Wikle CK (2011) Statistics for spatio-temporal data. Wiley, Hoboken
- EIA (2017) Marcellus shale play, Geology review. Retrieved from https://www.eia.gov/maps/pdf/MarcellusPlayUpdate_Jan2017.pdf
- EIA (2018) The Wolfcamp play has been key to Permian Basin Oil and natural gas production growth. Retrieved from <https://www.eia.gov/todayinenergy/detail.php?id=37532>
- EIA (2019) Permian basin, Wolfcamp and Bone Spring Shale Plays Geology review. Retrieved from https://www.eia.gov/maps/pdf/Wolfcamp_BoneSpring_EIA_Report_July2019.pdf
- EMC Education Services (2015) Data science and big data analytics. Wiley, Indianapolis. <https://doi.org/10.1002/9781119183686>
- Ettehadtavakkol A, Jamali A (2019) A data analytic workflow to forecast produced water from Marcellus shale. J Nat Gas Sci Eng 61:293–302. <https://doi.org/10.1016/j.jngse.2018.11.021>
- Everitt BS, Hothorn T (2009) A handbook of statistical analyses using R, 2nd edn. Chapman & Hall/CRC, Boca Raton
- Hastie T, Tibshirani R (1986) Generalized additive models (with discussion). Stat Sci 1:297–318
- Hastie T, Tibshirani R (1990) Generalized additive models. Chapman & Hall, New York
- Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd ed. Springer, New York, <https://doi.org/10.1007/978-0-387-84858-7>
- Hooten MB, Hobbs NT (2015) A guide to Bayesian model selection for ecologists. Ecol Monogr 85(1):3–28. <https://doi.org/10.1890/14-0661.1>
- Hovmöller E (1949) The trough-and-ridge diagram. Tellus 1(2):62–66. <https://doi.org/10.1111/j.2153-3490.1949.tb01260.x>
- Jaripatke O, Pandya N (2013) Eagle ford completions optimization - an operator's approach. In: SPE/AAPG/SEG unconventional resources technology conference. Unconventional Resources Technology Conference, Denver, Colorado, USA, <https://doi.org/10.1190/urtec2013-072>
- Jin H, Sonnenberg SA, Sarg JF (2015) Source rock potential and sequence stratigraphy of bakken shales in the williston basin. In: Unconventional resources technology conference. Unconventional

- Resources Technology Conference, San Antonio, Texas, USA, <https://doi.org/10.15530/URTEC-2015-2169797>
- Jolliffe IT (2002) Principal component analysis, 2nd edn. Springer, New York
- Kang EL, Liu D, Cressie N (2009) Statistical analysis of small-area data based on independence, spatial, non-hierarchical, and hierarchical models. *Comput Stat Data Anal* 53(8):3016–3032. <https://doi.org/10.1016/j.csda.2008.07.033>
- Koesoemadinata A, El-Kaseh G, Banik N, Dai J, Egan M, Gonzalez A, Tamulonis K (2011) Seismic reservoir characterization in marcellus shale. In: 2011 SEG annual meeting. Texas: Society of Exploration Geophysicists, San Antonio
- Kolawole O, Ispas I (2019) Interaction between hydraulic fractures and natural fractures: current status and prospective directions. *J Pet Explor Prod Technol*. <https://doi.org/10.1007/s13202-019-00778-3>
- Kolawole O, Esmaeilpour S, Hunky R, Saleh L, Ali-Alhaj HK, Marghani M (2019) Optimization of hydraulic fracturing design in unconventional formations: impact of treatment parameters. *SPE Kuwait Oil Gas Show Conf Soc Pet Eng*. <https://doi.org/10.2118/198031-MS>
- Kuhlman RD, Perez JJ, Claiborne EB (1992) Microfracture stress tests, anelastic strain recovery, and differential strain analysis assist in bakken shale horizontal drilling program. In: SPE rocky mountain regional meeting. Society of Petroleum Engineers. <https://doi.org/10.2118/24379-MS>
- Kumar S, Hoffman T, Prasad M (2013) Upper and lower bakken shale production contribution to the middle bakken reservoir. In: SPE/AAPG/SEG unconventional resources technology conference. Unconventional Resources Technology Conference, Denver, Colorado, USA, <https://doi.org/10.1190/urtec2013-001>
- Lalehrokh F, Bouma J (2014) Well spacing optimization in eagle ford. In: SPE/CSUR unconventional resources conference – Canada. Society of Petroleum Engineers, Calgary, Alberta, Canada, <https://doi.org/10.2118/171640-MS>
- Li H, Hart B, Dawson M, Radjef E (2015) Characterizing the middle bakken: laboratory measurement and rock typing of the middle bakken formation. In: Unconventional resources technology conference. San Antonio, Texas, USA, <https://doi.org/10.15530/URTEC-2015-2172485>
- Lohoefer D, Keener B, Snyder DJ, Ezeldin S (2014) Development of the wolfbone formation using open hole multistage vertical completion technology. In: SPE hydraulic fracturing technology conference. Society of Petroleum Engineers, The Woodlands, Texas, USA, <https://doi.org/10.2118/168643-MS>
- Lohoefer DS, Keener B, Ezeldin S, Snyder D (2014) A one-year production study between cemented multistage and openhole completion technologies for vertical wells in the permian basin. In: SPE annual technical conference and exhibition. Society of Petroleum Engineers, Amsterdam, The Netherlands, <https://doi.org/10.2118/170928-MS>
- Maity D, Ciezobka J (2019) An interpretation of proppant transport within the stimulated rock volume at the hydraulic-fracturing test site in the permian basin. *SPE Reserv Eval Eng* 22(02):477–491. <https://doi.org/10.2118/194496-PA>
- Mohaghegh SD (2016) Shale analytics : data-driven analytics in unconventional resources. Springer, Cham. Retrieved from <https://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=4803562>
- Navidi WC (2015) Statistics for engineers and scientists, 4th edn. McGraw-Hill, New York
- Nordeng SH, Helms LD (2010) Bakken Source System–Three Forks Formation Assessment. North Dakota Department of Mineral Resources (ND DMR). <https://www.dmr.nd.gov/ndgs/bakken/bakkenthree.asp>
- Nwabuoku KC (2011) Increasing lateral coverage in eagle ford horizontal shale completion. In: SPE annual technical conference and exhibition, Society of Petroleum Engineers, Denver, Colorado, USA, <https://doi.org/10.2118/147549-MS>
- Pebesma, E. (2012). spacetime : Spatio-Temporal Data in R. *Journal of Statistical Software*, 51(7). <https://doi.org/10.18637/jss.v051.i07>
- Pope CD, Palisch TT, Lolon E, Dzubin B, Chapman MA (2010) Improving stimulation effectiveness: field results in the haynesville shale. In: SPE annual technical conference and exhibition. Society of Petroleum Engineers, Florence, Italy, <https://doi.org/10.2118/134165-MS>
- RRC (2020) Eagle ford shale information. Retrieved April 2, 2020, from <https://www.rrc.state.tx.us/oil-gas/major-oil-and-gas-formations/eagle-ford-shale-information/>
- Sharma A, Yates ME, Pope T, Fisher K, Brown R, Honeyman L, Bates B (2014) Horizontal well development in unconventional resource play using an integrated completion and production workflow: delaware basin case study. In: SPE/EAGE European unconventional resources conference and exhibition. Society of Petroleum Engineers, Vienna, Austria, <https://doi.org/10.2118/167708-MS>
- Shelley RF, Saugier LD, Al-Tailji W, Guliyev N, Shah K (2012) Understanding hydraulic fracture stimulated horizontal eagle ford completions. In: SPE/EAGE European unconventional resources conference and exhibition. Society of Petroleum Engineers, Vienna, Austria, <https://doi.org/10.2118/152533-MS>
- Siddiqui F, Rezaei A, Dindoruk B, Soliman MY (2019) Eagle ford fluid type variation and completion optimization: a case for data analytics. In: SPE/AAPG/SEG unconventional resources technology conference. Unconventional Resources Technology Conference, Denver, Colorado, USA, <https://doi.org/10.15530/urtec-2019-598>
- Simha S, Tummala P, Kumar V, Singhal M, Viswanathan M, Kawar R, Dijk H (2019) Integrated reservoir modelling using spatio-temporal unsupervised learning and integrated visualization. In: Abu Dhabi international petroleum exhibition & conference, Society of Petroleum Engineers, Abu Dhabi, UAE, <https://doi.org/10.2118/197218-MS>
- Soliman MY, Wigwe M, Alzahabi A, Pirayesh E, Stegent N (2014) Analysis of fracturing pressure data in heterogeneous shale formations. *Hydraul Fract J* 1(2):8–12
- Sonnenberg SA (2014) The upper bakken shale resource play, williston basin. In: SPE/AAPG/SEG unconventional resources technology conference, Unconventional Resources Technology Conference, Denver, Colorado, USA, p. 12. <https://doi.org/10.15530/URTEC-2014-1918895>
- Stroud JR, Muller P, Sanso B (2001) Dynamic models for spatiotemporal data. *J R Stat Soc Ser B (Stat Methodol)* 63(4):673–689. <https://doi.org/10.1111/1467-9868.00305>
- Tomomewo OS, Jabbari H, Badrouchi N, Onwumelu C, Mann M (2019) Characterization of the Bakken formation using NMR and SEM techniques. In: 53rd U.S. rock mechanics/geomechanics symposium. American Rock Mechanics Association, New York City
- Tran T, Sinurat PD, Wattenbarger BA (2011) Production characteristics of the bakken shale oil. In: SPE annual technical conference and exhibition. Society of Petroleum Engineers, Denver, Colorado, USA. <https://doi.org/10.2118/145684-MS>
- USGS (2008) Assessment of undiscovered oil resources in the Devonian-Mississippian Bakken Formation, Williston Basin Province, Montana and North Dakota, 2008. Retrieved from https://pubs.usgs.gov/fs/2008/3021/pdf/FS08-3021_508.pdf
- USGS (2018) USGS estimates oil and gas in texas' eagle ford group. Retrieved from <https://www.usgs.gov/news/usgs-estimates-oil-and-gas-texas-eagle-ford-group>
- Westfall P, Henning KSS (2013) Understanding advanced statistical methods. CRC Press, Boca Raon

- Wigwe ME, Watson MC (2021) Presentation of oil and gas spatiotemporal big data visualization techniques as tools to aid in dynamic spatio-temporal models - SPE-200864-MS. In: SPE western regional meeting. Bakersfield, CA: Society of Petroleum Engineers
- Wigwe ME, Westfall PH, Watson MC, Giussani A, Nasir EA (2019a) Evaluation of the effect of well parameters on oil production. In: Proceedings JSM (ed) Statistical computing section. American Statistical Association, Alexandria, pp 1781–1803
- Wigwe ME, Watson MC, Giussani A, Nasir E, Dambani S (2019) Application of geographically weighted regression to model the effect of completion parameters on oil production—case study on unconventional wells. In: SPE Nigeria annual international conference and exhibition. Lagos, Nigeria: Society of Petroleum Engineers. <https://doi.org/10.2118/198847-MS>
- Wigwe ME, Bougre ES, Watson MC, Giussani A (2020) Spatio-temporal models for big data and applications on unconventional production evaluation. In: SPE/AAPG/SEG unconventional resources technology conference. Unconventional Resources Technology Conference (URTeC 2855), Austin, Texas, USA. <https://doi.org/10.15530/urtec-2020-2855>
- Wikle CK, Zammit-Mangion A, Cressie N (2019) Spatio-temporal statistics with R. CRC Press, Boca Raton, Florida. <https://doi.org/10.1201/9781351769723>
- Wilks, D. S. (2011). *Statistical Methods in the Atmospheric Sciences*. Elsevier Science. Retrieved from <https://books.google.com/books?id=IJuCVtQ0ySIC>
- Wood SN (2000) Modelling and smoothing parameter estimation with multiple quadratic penalties. *J R Stat Soc Ser B (Stat Methodol)* 62(2):413–428
- Wood Simon N (2003) Thin plate regression splines. *J R Stat Soc Ser B (Stat Methodol)* 65(1):95–114. <https://doi.org/10.1111/1467-9868.00374>
- Wood SN (2017) Generalized additive models, 2nd ed. CRC Press, Boca Raton. <https://doi.org/10.1201/9781315370279>
- Yates ME, Sharma A, Itibrou T, Smith L, Fisher K, Brown R, Bates B (2013) An integrated approach for optimizing vertical wolfbone wells in the delaware basin. In: SPE/AAPG/SEG unconventional resources technology conference. Unconventional Resources Technology Conference, Denver, Colorado, USA, <https://doi.org/10.1190/urtec2013-071>
- Yildirim LTO, Wang JY, Elsworth D (2019) Petrophysical evaluation of shale gas reservoirs: a field case study of marcellus shale. In: Abu Dhabi international petroleum exhibition & conference. Society of Petroleum Engineers, Abu Dhabi, UAE, <https://doi.org/10.2118/197838-MS>
- Zamirian M, Aminian K, Ameri S (2016) Measuring marcellus shale petrophysical properties. In: SPE western regional meeting. Anchorage, Alaska, USA: Society of Petroleum Engineers. <https://doi.org/10.2118/180366-MS>
- Zammit-Mangion A, Cressie N (2017) FRK: an R package for spatial and spatio-temporal prediction with large datasets. Retrieved from <http://arxiv.org/abs/1705.08105>
- Zargari S, Mohaghegh SD (2010) Field development strategies for bakken shale formation. In: SPE eastern regional meeting. Morgantown, West Virginia, USA: Society of Petroleum Engineers. <https://doi.org/10.2118/139032-MS>
- Zhou Q, Dilmore R, Kleit A, Wang JY (2014) Evaluating gas production performances in marcellus using data mining technologies. *J Natl Gas Sci Eng* 20:109–120. <https://doi.org/10.1016/j.jngse.2014.06.014>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.