Oilwell Production and Completion Regression Analysis

Tom Wilson

November 24, 2018

Literature Review

Introduction

For centuries, oilwells have been drilled vertically and completed by perforation only. Since the invention of horizontal drilling in the late 20th century, literature has shown that various completion parameters have a significant effect on oil production. Multiple recent sources agree that there is value in leveraging available data as input to statistical models steeped in machine learning.(Fulks et al. 2016), (Holdaway, Laing, and others 2015), (Pankaj et al. 2018), (Subrahmanya et al. 2014) While there is some disagreement about the limitations and applicability of such models, most agree that a general data anlysis workflow is applicable. (Fulks et al. 2016),(Temizel et al. 2015),(Groulx et al. 2017), (Pankaj et al. 2018)

Data Quality

One recurring theme within recent literature related to oilwell completion is that of data quality. Lopes and Jorge (2017) takes a statistical approach to dealing with gaps in data by breaking synthesysing gaps in otherwise complete data. Lopes and Jorge (2017) eventually shows that gaps in data did not have a significant effect on model selection or accuracy. On the other hand, Wang et al. (2016) claims that a input uncertainties model selection challenges were primarily caused by missing data and noise. Khodabakhsh, Ari, and Bakir (2017) took yet another approach via real–time models for detecting and classifying errors. Wang et al. (2016) suggests Basis Pursuit Denoising (BPDN) as a solution to data quality issues. Other sources did not comment on dealing with data quality or noise.

General Workflow

Regardless of the target variable or underlying dataset, there is wide-spread agreement that data analysis workflows are generally applicable and value-added. Quoting from Holdaway, Laing, and others (2015): "Data-driven workflows, models, and analysis can address a diverse array of business problems in the oil and gas industry." Groulx et al. (2017) found that their approach applied equally well to all basins(a geologic formation) and plays(oil barring zone) available. Groulx et al. (2017) showed that, generally, the number of performance measures directly correlate with the number of patterns identified. Groulx et al. (2017) indicates that Parallel coordinates approach makes identification of thresholds and correlation windows easy. Which can be valuable input for other regression efforts. Furthermore, the predictive proxy approach has a wide variety of applications to completion engineering and management. (Pankaj et al. 2018) Drilling and lifting in addition to completion also benefit from faster decision making according to Pankaj et al. (2018). Khodabakhsh, Ari, and Bakir (2017) agrees that the data analytic approach is applicable to the oil drilling industry in addition to oilwell completion practices. Subrahmanya et al. (2014) demonstrates that machine learning methods show considerable promise. Guevara et al. (2017) was able to outperform conventional techniques, such as kriging (a gaussian interpolation method). Future work in other domains was suggested by a variety of authors: vertical well logs (Guevara et al. 2017), geophysical (Luo 2018), fluid dynamics (Ezzatabadipour et al. 2017), exploration (Nikhalat-Jahromi and Jorge 2017), Remote sensing (Nikhalat-Jahromi and Jorge 2017), geobotany and geochemistry (Nikhalat-Jahromi and Jorge 2017). There is genuine excitement surrounding the application of machine learning and data analytics to various topics related to oil and gas production.

Feature Importance

Each author provides some insight into what features are most important for predicting oil production accurately. As a default position; High proppant (sand or other particulates used to prop open fractures in rock under high pressure), high-fluid (usually water, but sometimes gelling or cross-linking additives are included) completion designs described by Fulks et al. (2016) have shown success every basin. Fulks et al. (2016) goes on to say that degradable diversion (poly lactic acid diverts pressure but eventually degrades by bacteria) improves cluster efficiency and optimal lateral landing zone are important. In contrast, Temizel et al. (2015) lists in order of importance fracture half-length (which could be related to diversion), proppant amount, zone coverage (nearly equivalent to landing zone), and slurry volume (slurry is the combination of fluid with proppant). Ezzatabadipour et al. (2017) is more interested in two-phase flow patterns as a function of pipe condition. Luo (2018) purports that chemical ingredients are conducive to production and some, in fact, negatively impact gas production. (Luo 2018) Lopes and Jorge (2017) includes rock and fluid properties. While Guevara et al. (2017) expects to add well completion parameters to their model. Nikhalat-Jahromi and Jorge (2017) provides structural geology and reservoir properties (fault lines, water zones, etc.) as important features. Wang et al. (2016) focuses on trap and peel heights as well as gas flux, plans to add droplet size and gas-oil ratio to that list.

Simulation

Another common thread among recent literature is the idea of simulating new data from predictive models. Temizel et al. (2015) sees simulations as a tool to determine not only feature importance and significance, but also effect direction. Temizel et al. (2015) goes on to say that simulations can replace commercial fracture simulators. The approach described in Pankaj et al. (2018) creates a "parametric explosion of parameters"; enabling optimal well completion design to be determined much faster than with traditional methods, a matter of minutes instead of months.(Pankaj et al. 2018).Bozoev and Demidova (2016) agrees that simulation tends to be an appropriate approach to choosing the optimum completion of the wellbore. Like-wise, Ezzatabadipour et al. (2017) notes that investigations could be improved by exhaustive searches. Wang et al. (2016) also observes that resorting to simulated data from very high-resolution numerical simulations is a good compromise. Liu et al. (2018) uses simulation to study early hydration stages of a cement slurry which effect the transmission of hydrostatic pressure.

Model Limitations

Though there is much fanfare surrounding data analytic approaches applied to oilwell completions, many authors note that there are limitations to such endeavors. Fulks et al. (2016) says that establishing baseline performance is necessary. Temizel et al. (2015) takes the stance taht conclusions from data driven models are specific to the model used. Careful consideration is necessary for normalizing both performance measures and inputs. Before discernible conclusions, patterns detected must be reviewed to refine insights. (Groulx et al. 2017) Pankaj et al. (2018) notes on that a calibrated model is fundamental to a reliable prediction. Bozoev and Demidova (2016) has concern that analytical approaches do not take into account interference effects. Wang et al. (2016) adds that next stage numerical models must be able to handle mixtures of oil and gas in order to simulate realistically.

Conclusion

For centuries, oilwells have been drilled vertically with little care about completion. Today, horizontal drilling has changed that. literature shows completion parameters have a significant effect on oil production. And that these parameters can be used to make accurate predictions. While there is some disagreement about the limitations and applicability of such models, most agree that a general data analysis workflow is applicable.

Regression Analysis

Introduction

Welldatabase is an aggregator of publically available data related to oil and gas production. Many states have an entity responsible for collecting and enforcing reporting requirements which vary between juisdictions. In Texas, for example, Texas Railroad Commision (TXRRC) regulates the oil and gas industry. Location and depth of well casing is tightly regulated. TXRCC requires all hydocarbon production that leaves the well site to be reported monthly. In the state of texas, oil wells are tested for once per year. This analysis will focus on two counties in west texas, Midland County and Pecos County. Our goal is to model 12 month cumulative oil production as a function of publically available drilling and completion parameters summarized in the following table.

column name	description	units
api	14 digit unique well identifier	##-###-#####-####
surfacelatitude	surface location	degrees
surfacelongitude	surface location	degrees
county	texas county	either Midland or Pecos
producingformation	oil production geologic formation	e.g Wolfcamp, Spraberry
wellboreprofile	type of drilling	Verical, Directional, or Horizontal
trueverticaldepth	depth of production casing	feet
laterallength	legth of horizontal component of well	feet
totalbasewatervolume	volume of water used durring hydrolic fracturing	gallons
totalproppantmass	mass of sand used to prevent closure of fractures	lbs
fluidsystem	type of fluid additives	one of slick-water, linear, or cross-link
surfactantpresent	weather or not sufactant was used during completion	True or False
scaleinhibitorpresent	wether or not scale inhibtors were used during completion	True or False
claycontrolpresent	weather or not clay control was used during completion	True or False

column name	description	units
acidtreatmentpresent	weather or not acid was used during completion	True or False
chokesize	open proportion of surface gate valve during well test	0 is closed, 1 is open
oil	first 12 month cumulative oil production	bbl

Preprocessing

each table of welldatabase must be summarize and deduplicate by api number then joined to the header on api where wellboreprofile=HORIZONTAL. The python code for this processing is available on github

No auto-correlation

It is tempting to use the reported monthly oil production as is, but that would lead to a target variable that is highly auto-correlated, that is the previous month's production is a good esitmate of the next month. For this reason, we will focus our effort on modeling cumulative oil production.

Linear Relationship

oil is not very not very linear log(oil) chokesize

Little Multicollinearity

totalpropantmass and totalfracmass and totalbasewatervolume are highly correlated

Multivariate Normality

totalproppantmass has too many zeros set totalproppantconcetration=0 to null totalproppantconcetration = totalproppantmass/totalbasewatervolume

The regression model

It is expected that the relationship between these predictors and oil production is highly non-linear so it is appropriate to consider high degree polynomials with interations and quadradic terms.

Homoscedasticity

Residual Analysis

Summary

References

Bozoev, A M, and E A Demidova. 2016. "Selection of the Optimal Completion of Horizontal Wells with Multi-Stage Hydraulic Fracturing of the Low-Permeable Formation, Field c." *IOP Conference Series: Earth and Environmental Science* 33 (1): 012035. http://stacks.iop.org/1755-1315/33/i=1/a=012035.

Ezzatabadipour, Mohammadmehdi, Parth Singh, Melvin Deloyd Robinson, Pablo Guillen-Rondon, and Carlos Torres. 2017. "Deep Learning as a Tool to Predict Flow Patterns in Two-Phase Flow." *CoRR* abs/1705.07117. http://arxiv.org/abs/1705.07117.

Fulks, Robert, Simon Hughes, Ingo Geldmacher, and others. 2016. "Optimizing Unconventional Completions-an Integrated Approach." In *Abu Dhabi International Petroleum Exhibition & Conference*. Society of Petroleum Engineers.

Groulx, Bertrand, Jim Gouveia, Don Chenery, and others. 2017. "Multivariate Analysis Using Advanced Probabilistic Techniques for Completion Design Optimization." In *SPE Unconventional Resources Conference*. Society of Petroleum Engineers.

Guevara, Jorge, Matthias Kormaksson, Bianca Zadrozny, Ligang Lu, John Tolle, Tyler Croft, Mingqi Wu, Jan Limbeck, and Detlef Hohl. 2017. "A Data-Driven Workflow for Predicting Horizontal Well Production Using Vertical Well Logs." *CoRR* abs/1705.06556. http://arxiv.org/abs/1705.06556.

Holdaway, Keith R, Moray L Laing, and others. 2015. "Drilling and Completion Optimization in Unconventional Reservoirs with Data-Driven Models." In *SPE Asia Pacific Unconventional Resources Conference and Exhibition*. Society of Petroleum Engineers.

Khodabakhsh, Athar, Ismail Ari, and Mustafa Bakir. 2017. "Cloud-Based Fault Detection and Classification for Oil & Gas Industry." *CoRR* abs/1705.04583. http://arxiv.org/abs/1705.04583.

Liu, Kaiqiang1, chengxw@swpu.edu.cnzj-656@163.com Cheng Xiaowei1, Xingguo2 Zhang, Zaoyuan2 Li, Jia1 Zhuang, and Xiaoyang2 Guo. 2018. "Relationship Between the Microstructure/Pore Structure of Oil-Well Cement and Hydrostatic Pressure." *Transport in Porous Media* 124 (2): 463–78.

Lopes, Rui L., and Alípio Jorge. 2017. "Mind the Gap: A Well Log Data Analysis." *CoRR* abs/1705.03669. http://arxiv.org/abs/1705.03669.

Luo, & Zhang, H. 2018. "Mining Fracfocus and Production Data for Efficacy of Fracturing Fluid Formulations." In *Society of Petroleum Engineers*. Society of Petroleum Engineers.

Nikhalat-Jahromi, Hamed, and Alípio M. Jorge. 2017. "An Overview of Data Mining Applications in Oil and Gas Exploration: Structural Geology and Reservoir Property-Issues." *CoRR* abs/1705.06345. http://arxiv.org/abs/1705.06345.

Pankaj, Piyush, Steve Geetan, Richard MacDonald, Priyavrat Shukla, Abhishek Sharma, Samir Menasria, Tobias Judd, and others. 2018. "Need for Speed: Data Analytics Coupled to Reservoir Characterization Fast Tracks Well Completion Optimization." In *SPE Canada Unconventional Resources Conference*. Society of Petroleum Engineers.

Subrahmanya, Niranjan, Peng Xu, Amr El-Bakry, Carmon Reynolds, and others. 2014. "Advanced Machine Learning Methods for Production Data Pattern Recognition." In *SPE Intelligent Energy Conference & Exhibition*. Society of Petroleum Engineers.

Temizel, C, S Purwar, A Abdullayev, K Urrutia, Aditya Tiwari, and others. 2015. "Efficient Use of Data

Analytics in Optimization of Hydraulic Fracturing in Unconventional Reservoirs." In *Abu Dhabi International Petroleum Exhibition and Conference*. Society of Petroleum Engineers.

Wang, Shitao, Mohamed Iskandarani, Ashwanth Srinivasan, W. Carlisle Thacker, Justin Winokur, and Omar M. Knio. 2016. "Propagation of Uncertainty and Sensitivity Analysis in an Integral Oil-Gas Plume Model." *Journal of Geophysical Research: Oceans* 121 (5): 3488–3501.