Shale gas production forecasting is an ill-posed inverse problem and requires regularization

Justin B. Montgomery[a,b,1], **Samuel J. Raymond**[a,2], **Francis M. O'Sullivan**[b], **and John R. Williams**[a]

[a]Department of Civil and Environmental Engineering and; [b]MIT Energy Initiative, Massachusetts Institute of Technology, Cambridge, MA 02139

This manuscript was compiled on June 17, 2019

Extrapolation of a decline curve analysis (DCA) model fitted to a well's production rates remains the standard approach for forecasting shale gas production. A scaling curve was recently proposed as a way of connecting this approach with underlying physics but we show that it actually generates worse predictions than the traditional non-physical model. DCA is an ill-posed inverse problem and today's unconventional resource forecasts can be substantially improved by introducing Tikhonov regularization to address this. We compare forecasting accuracy with Barnett shale wells and find that this overlooked step is more critical than choice of model. Additionally, we show that deep neural networks—a purely data-driven approach with no physical interpretation—can outperform the other approaches considered here and serve as a future benchmark.

unconventional oil and gas | fracking | ridge regression | neural networks

The rapid growth of production from unconventional shale and tight resource systems—enabled by long horizontal wells combined with hydraulic fracturing—has turned the United States into the world's leading producer of both natural gas and oil. However, effectively planning for and managing the development of these resources requires accurate production forecasts which remain elusive due to a limited understanding of the complex physical processes behind production (1–3). Many wells under-perform compared to operators' projections and energy policy decisions are hampered by the immense uncertainty in the long-term productivity of new wells (4–7).

Rigorous physics-based models, such as numerical reservoir simulators, are rarely used to forecast unconventional oil and gas production due to the associated data and modeling costs as well as the challenge of adequately representing nano-scale flow and fracture properties (1, 2). Instead, it is standard to rely on decline curve analysis (DCA), in which a production model with a small number of parameters is matched to historical production rates using a least squares fit and the resulting curve is extrapolated into the future (1). The most widely used DCA model is the modified Arps curve, which builds on the hyperbolic decline relationship,

$$Q(t) = \frac{Q_i}{(1 + bD_i t)^{(\frac{1}{b})}},$$

originally introduced in 1945 (8), in which production rate $Q(t)$ at time t is determined by the parameters Q_i, b, and D_i. In order to avoid unrealistically high late-life projections for unconventional wells, this hyperbolic model is typically switched to a fixed exponential trend once production decline slows to some threshold rate, such as 10% annually (3, 9). Although widely used, DCA with modified Arps is a heuristic approach and lacks a physical basis with unconventional wells.

In order to better understand shale gas production decline behavior and put DCA on firmer theoretical ground, Patzek et al. developed a scaling curve model based on one-dimensional gas flow into planar fractures (10). Remarkably, this model could be reduced to two effective "scaling" parameters, $\mathcal{M}$ and τ, and a recovery factor curve $\mathrm{RF}(\cdot)$ to describe any Barnett shale well's cumulative production $m(t)$ at time t, as in

$$m(t) = \mathcal{M}\mathrm{RF}(t/\tau).$$

Because many of the underlying physical properties behind the parameters $\mathcal{M}$ and τ cannot be measured, Patzek et al. advocated a nonlinear least squares fit to production data, as is standard for DCA.

There has been a proliferation of production model formulas in recent years—ranging from purely empirical, like the modified Arps curve, to those that build upon physical theory, such as the scaling curve(1, 10). There is ongoing debate about which model is most appropriate for unconventional wells but it is beyond the scope of this brief report to review these, and we refer the reader to (3). Our goal is instead to shift attention toward a glaring shortcoming with the parameter estimation process of DCA more generally, which is evident in both these representative DCA models and actually worse with the scaling curve model.

DCA forecasting as currently implemented is unreliable due to potentially non-unique parameter estimates, particularly when the production history is short. This has also been described as a sensitivity in forecasts to initial parameter seeds in the nonlinear least squares algorithm (11). More saliently, this property of non-unique or non-identifiable parameters makes DCA a classic example of an ill-posed inverse problem which lacks a stable numerical solution (12–14). Consequently, DCA should incorporate regularization, as is typically used with this class of problems, but this critical step has surprisingly been overlooked in this context.

Regularization is the process of introducing additional information to reduce the ill-posedness of an inverse problem. In machine learning, it also helps to avoid over-fitting by introducing some bias to reduce the overall variance of predictions (13, 15). Often, this takes the form of ℓ^2 Tikhonov regularization, where a squared penalty on the distance of parameter estimates θ from some expected parameter value θ_0* is added to the objective function. For least squares problem, the loss function $L(\mathbf{t}, F(\mathbf{t}, \theta))$ being minimized for observations $\mathbf{y}$ at times $\mathbf{t}$ with forward model predictions $F(\mathbf{t}, \theta)$, becomes

$$L(\mathbf{t}, F(\mathbf{t}, \theta)) = \|\mathbf{y} - F(\mathbf{t}, \theta)\|_2^2 + \lambda \|\theta - \theta_0\|_2^2 .$$

The ℓ^2-norm is denoted as $\|\cdot\|_2^2$ and λ is a weight hyperparameter controlling parameter shrinkage toward the mean. This can be tuned to optimize the bias-variance tradeoff and ensure model generalizability using cross-validation (13, 14).

*When θ_0 is zero this is often also called ridge regression.

When forecasting with limited well production data, it is currently common to avoid inversion altogether and rely on a type-well curve as a proxy. This is simply an average of production rates from older wells in the same area (2, 9). Ad hoc approaches may be used to normalize type-well curves to design metrics (e.g. lateral length) or peak production rate, but this approach ignores the impact of variations in geology and evolving development practices on temporal production dynamics. This is a particular concern as newer wells tend to be drilled much closer to neighboring wells and have more tightly spaced fracture stages in order to rapidly drain an area (4, 5). As we have previously shown, geology is exceptionally heterogeneous at even small scales in these basins and it is difficult to isolate the impact of constantly evolving technology on production, making the selection of suitable analogue wells for a type-well curve a highly fraught task (16).

To compare the accuracy of these forecasting approaches, we quantify the error in predictions of 10-year cumulative production for 4457 wells in the Barnett shale. These projections are made based on the first 6, 12, 24, and 48 months of production. The DCA models considered here—modified Arps and the scaling curve—are compared to a static county-based type-well curve, as established by the US Energy Information Administration (EIA) (9). We follow the same data processing and fitting procedure as (10) and for modified Arps use the methodology laid out in (9). These DCA models are then fitted with ℓ^2 Tikhonov regularization to understand the benefit this provides. A grid-search with 4-fold cross-validation is used to tune λ and θ_0, as described in (13, 14).

Given the many physical complexities of shale gas production, a DCA model with a small number of parameters considerably oversimplifies production dynamics resulting in model mis-specification error. To better understand the magnitude of this error we introduce an alternative data-driven extrapolation approach using a deep neural network (DNN) in the TensorFlow platform. This consists of 4 hidden layers of 15 fully-connected neurons (with ReLU activation functions), a regression layer outputting 10-year cumulative production, and an input layer with the number of input neurons determined by the number of production months used in the forecast. As with Tikhonov hyperparameter tuning, 4-fold cross-validation is used to separately train and test the DNN.

Results

Prediction accuracy, measured as root mean squared error (RMSE), is shown in Fig. 1. Despite its physical derivation, the scaling curve introduces substantial error compared to the modified Arps model. A standard eigenvalue analysis of the approximate Hessian (17) reveals an exceptional amount of *sloppiness* in the scaling curve model—meaning that parameters are highly correlated and can take on a wide range of values (casting doubt on their physical meaningfulness) with little change in goodness of fit. This is a common property of ill-posed problems and makes regularization even more critical with the scaling curve (14). By the time 48 months of production data are available for fitting, the modified Arps model with regularization nearly matches the reliability of the DNN forecast. In contrast, the continued underperformance of the regularized scaling curve at this point suggests that the physical explanation behind it lacks explanatory power.

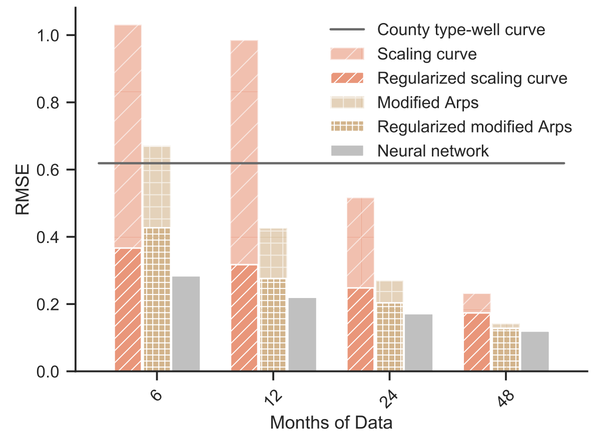

Fig. 1. A comparison of prediction accuracy for 10-year cumulative production with different approaches and months of production data.

Discussion

For the representative DCA models compared here, the step of introducing regularization matters more than model choice itself. This finding overshadows and somewhat trivializes the ongoing debate about which existing or new DCA model is best. Furthermore, the DNN sets a new benchmark in accuracy, although it lacks interpretability and requires abundant production data (making it less useful in new fields). The DNN exceeds the accuracy of DCA because it learns the best nonlinear mapping directly from the data, rather than presupposing an overly simplified model form. The layering of nonlinear activation functions in the DNN creates effectively infinite functional flexibility while the noise in the representative training sets serves to adequately regularize the model and prevent overfitting (18).

There is a Bayesian perspective to the approach we advocate which is informative. Type-well curves are undesirable because they are essentially a prior based on a group mean and provide a static forecast that cannot be updated based on observed production dynamics. On the other hand, DCA offers a maximum likelihood estimate. Tikhonov regularization is akin to a *max a posteriori* estimate balancing these sources of information (12). Other authors have suggested Bayesian modeling to quantify uncertainty in DCA forecasts (19). However, it is critical to note that the sloppiness and ill-posedness of the likelihood (DCA fit) means that the prior may dominate and should be chosen with care. This suggests a promising direction for future work—the use of hierarchical models to introduce other physical information into DCA to appropriately constrain and shape the prior.

Type-well curves and DCA are the foundation of production projections today, from the US EIA to companies' reporting of reserves. A lack of awareness of the ill-posed nature of DCA and the importance of regularization is unnecessarily introducing error and uncertainty into these forecasts. Ill-posed inverse problems are also prevalent in other domains, including biological and earth sciences (14, 18). It is important to remain vigilant about this property and, when data is abundant, identify data-driven approaches to mitigate it.

ACKNOWLEDGMENTS. This work was supported by the MIT Energy Initiative. We would like to thank Emre Gencer and Simone Cenci for their helpful suggestions to this research.

1. Lee J, Sidle R (2010) Gas-reserve estimation in resource plays. *Society of Petroleum Engineers Economics & Management* 2:86–91.
2. McGlade C, Speirs J, Sorrell S (2013) Methods of estimating shale gas resources – Comparison, evaluation and implications. *Energy* 59:116–125.
3. Tan L, Zuo L, Wang B (2018) Methods of decline curve analysis for shale gas reservoirs. *Energies* 11(3).
4. (2018) Peering inside the Permian. *The Economist*.
5. Olson B, Elliott R, Matthews C (2018) Fracking's secret problem–oil wells aren't producing as much as forecast.
6. (2018) Annual energy outlook 2018 with projections to 2050, (US EIA), Technical report.
7. Ikonnikova S, Smye K, Browning J, Domisse R (year?) Update and Enhancement of Shale Gas Outlooks, (Bureau of Economic Geology), Technical Report September 2018.
8. Arps JJ (1945) Analysis of decline curves. *Transactions of the American Institute of Mining Engineers* 160:228–247.
9. Eia (2018) Oil and Gas Supply Module of the National Energy Modeling System. (May).
10. Patzek TW, Male F, Marder M (2013) Gas production in the Barnett Shale obeys a simple scaling theory. *Proceedings of the National Academy of Science* 110(49):19731–19736.
11. Hong A, Bratvold RB, Lake LW, Ruiz Maraggi LM (2019) Integrating Model Uncertainty in Probabilistic Decline-Curve Analysis for Unconventional-Oil-Production Forecasting. *SPE Reservoir Evaluation & Engineering* (August 2018):23–25.
12. Stuart AM (2010) Inverse problems : A Bayesian perspective Inverse problems. *Acta Numerica* 19:451–559.
13. Johansen TA (1997) On Tikhonov Regularization, Bias and Variance in Nonlinear System Identification. *Automatica* 33(3).
14. Gábor A, Banga JR (2015) Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Systems Biology* 9(1).
15. Bishop CM (2011) *Pattern recognition and machine learning*. (Springer).
16. Montgomery J, O'Sullivan F (2017) Spatial variability of tight oil well productivity and the impact of technology. *Applied Energy* 195:344–355.
17. Transtrum MK, Machta BB, Sethna JP (2011) Geometry of nonlinear least squares with applications to sloppy models and optimization. 036701(October 2010):1–35.
18. Kim Y, Nakata N (2018) Geophysical inversion versus machine learning in inverse problems. *The Leading Edge* 37(12):894–901.
19. Gong X, Gonzalez R, McVay DA, Hart JD (2014) Bayesian Probabilistic Decline-Curve Analysis Reliably Quantifies Uncertainty in Shale-Well-Production Forecasts. *SPE Journal* 19(06):1047–1057.