URTeC: 1928595

Generative Models for Production Forecasting in Unconventional Oil and Gas Plays

Heidi Anderson Kuzma* BetaZi LLC, Nimar S. Arora, BetaZi LLC, Kemal Farid, Merrick Systems Inc.

Copyright 2014, Unconventional Resources Technology Conference (URTeC) **DOI 10.15530/urtec-2014-1928595**

This paper was prepared for presentation at the Unconventional Resources Technology Conference held in Denver, Colorado, USA, 25-27 August 2014.

The URTeC Technical Program Committee accepted this presentation on the basis of information contained in an abstract submitted by the author(s). The contents of this paper have not been reviewed by URTeC and URTeC does not warrant the accuracy, reliability, or timeliness of any information herein. All information is the responsibility of, and, is subject to corrections by the author(s). Any person or entity that relies on any information obtained from this paper does so at their own risk. The information herein does not necessarily reflect any position of URTeC. Any reproduction, distribution, or storage of any part of this paper without the written consent of URTeC is prohibited.

Summary

Short and long term production forecasts are important for assessing the performance and value of unconventional oil and gas wells, setting production targets and predicting cash flows. Often the best way to try to estimate future production of a new completion is by analogy with production from other nearby completions with similar treatment histories. Forecasting production using type curves combines knowledge with experience using physics-based decline equations, type curve coefficients and an engineer's judgment. Modern algorithms for predictive analytics can be based on a generative model (GM) which does the same thing. A GM is a stochastic model which can be constructed to capture physics (solutions to differential fluid-flow equations) with experience (probability distributions that are calibrated using historic data). The result is a model which is similar to a type curve in that it can be used to give a reasonable forecast of the future of a completion based on only a small amount of actual information about that particular well; the rest it draws from what it has seen before. By combining a GM with an algorithm for statistical inference, it is possible to make accurate forecasts that include quantified uncertainty bounds. The combination of a forecast and uncertainty leads to powerful new ways of looking at future production including Forecast Certainty Maps.

Introduction

Trillions of dollars are invested annually based on forecasts of the future production from oil and gas wells. Reliable forecasts are needed to be able to set and meet realistic targets, compare the performance of different wells and completions, make short and long term financial forecasts, do due diligence when acquiring or divesting of production, and, in an environment where petroleum professionals are increasingly expensive and difficult to come by, maximize the efficiency of their engineering staff. Accurate production forecasting is particularly important in the unconventional space where stimulations are expensive and often have unforeseen consequences.

The variety of publications and forums which discuss the need for accurate production forecasts underscores how important the problem is, not only across all facets of the oil and gas industry, but to the public in general. Recent publications include [Sallh et al., 2014] who discuss the economic need for accurate forecasts in the North Sea, [Lake et al, 2013] who outline the financial consequences of unexpected steep declines in unconventional resources, [Hughes, 2013] who suggests in the journal *Nature* that scientists should call shale gas reserves into question, and [Hook et al, 2014] who outline to the Royal Society of London that "empirical findings have important implications for oil supply forecasting." On the technical side of the issue, articles have been published by [Harris and Lee, 2014] and [Shahamat et al., 2014]; a special series in the April issue of *Journal of Petroleum Technology* is devoted to forecasting techniques; and a series of SPE-sponsored Global Workshops on Production Forecasting are taking place in 2014 and 2015.

The production from a well which is produced at nearly constant pressure will decline roughly exponentially over its lifetime. The vast majority of production forecasts are made by fitting a curve to this decline. A theoretical smooth curve is usually chosen, parameters of the curve are found to make it match the well's production history, and the curve is projected forward in time to make a forecast. The theoretical curve might be an extremely simple exponential, a hyperbolic, an Arps curve or any one of a number of more complicated models including hybrid models (Shin et al, 2014), stretched Arps, or the Duong method. In the most sophisticated studies, iterative history matching uses the results of a numerical simulator to match the data, taking into account myriad variables such as porosity, permeability, formation thickness, reservoir geometry and detailed effects of gas desorption [Yu and Sepehrnoori, 2014].

Unfortunately it is very rare that actual declines are truly smooth, especially in unconventional fields where the data is complicated by numerous stimulations and transient production regimes. This means curve fitting almost always requires the intervention of a human hand; there are an infinite number of curves which match a given dataset reasonably well and it is up to a person to make the judgment of which is right. An engineer might constrain the forecast by knowledge of the parameters of curves used to fit the declines from similar wells, or by estimates of the total amount of hydrocarbons in a reservoir, or by his or her knowledge of a particular well or field. Very often, a *type curve* is introduced. A type curve is a normalized curve that is meant to define the standard or average behavior of a group of similar wells, such as a field, reservoir or play, or completions into a particular formation. When there is almost no information about a new completion except maybe an initial rate of production, the type curve gives a best estimate of what the well will do based on previous experience.

A skilled engineer, when asked to explain production, will often draw a theoretical decline on a piece of paper and then perturb it to show all of the different ways in which actual well behavior tends to deviate from the ideal. This kind of sketch illustrates what can and does happen in real life. When given data in the form of a paper plot of a production record, the same engineer can draw a theoretical curve and annotate the data with explanations of what might have happened that made artifacts in the data other than a simple decline. There might have been stimulation, the well was shut in or choked down, or the lift mechanism changed, for example.

The Generative Model (GM) outlined in this paper works in a similar way. It is a statistical model which captures both the *physics* of well decline (in terms of solutions to fluid flow equations) and *experience* (using statistics) in a manner that is not unlike the way that a person does it: by knowing how theoretical data should look and how various events tend to perturb it. A GM is similar to a type curve in that it encapsulates a priori knowledge; it gives a best guess for how a new well will behave based on analogy with existing wells.

A GM which includes events which might or might not happen in every data set (such as multiple shutins or stimulations) is called an Open Universe (OU) model [Milch and Russell, 2010]. The term Open is used to indicate that there is no fixed number of input parameters into the model. OU GMs such as this have been used to great effect in such problems as vehicle tracking, medical monitoring and, particularly, in earthquake location [Arora et al., 2013]. This is the first time that one has been applied to an important problem in oil and gas.

Theory

Generative Models and Statistical Inference

A Generative Model (GM) is a statistical model which, in the absence of any other data or input, can be used to generate realistic hypothetical outcomes of an experiment. Unlike theoretical models (including most so-called *forward models* in petroleum engineering and geophysics), it is not confined to the ideal. It doesn't stop with physically realistic curves; it also overlays synthetic artifacts and noise of the kind which tend to contaminate real data.

A GM is called *generative* because it can be used to generate *samples* which are realistic synthetic production histories that include noise and artifacts. If it is shown no data, it will generate these samples in the same proportion that they occur in real life – where "real life" to a computer means a large data set used for *training*. A GM is unlike a neural network, however, because a neural network uses training data to learn the full relationship between inputs and outputs. Once trained, it requires input in order to give an output. All a GM learns from training data is the parameters of internal probability distributions. The rest is developed by combining human expertise with precise statistical analysis. A GM will generate samples without any input other than "run." However, given a particular dataset, it can be used for *statistical inference*, that is, it can be used to infer the input parameters for models that do a reasonably good job of explaining the data. In the world of statistical inference, "reasonably" means "with high probability." The output of an inference algorithm is a suite of sample models usually found using a guided sampling algorithm such as Markov Monte Carlo Chain (MCMC).

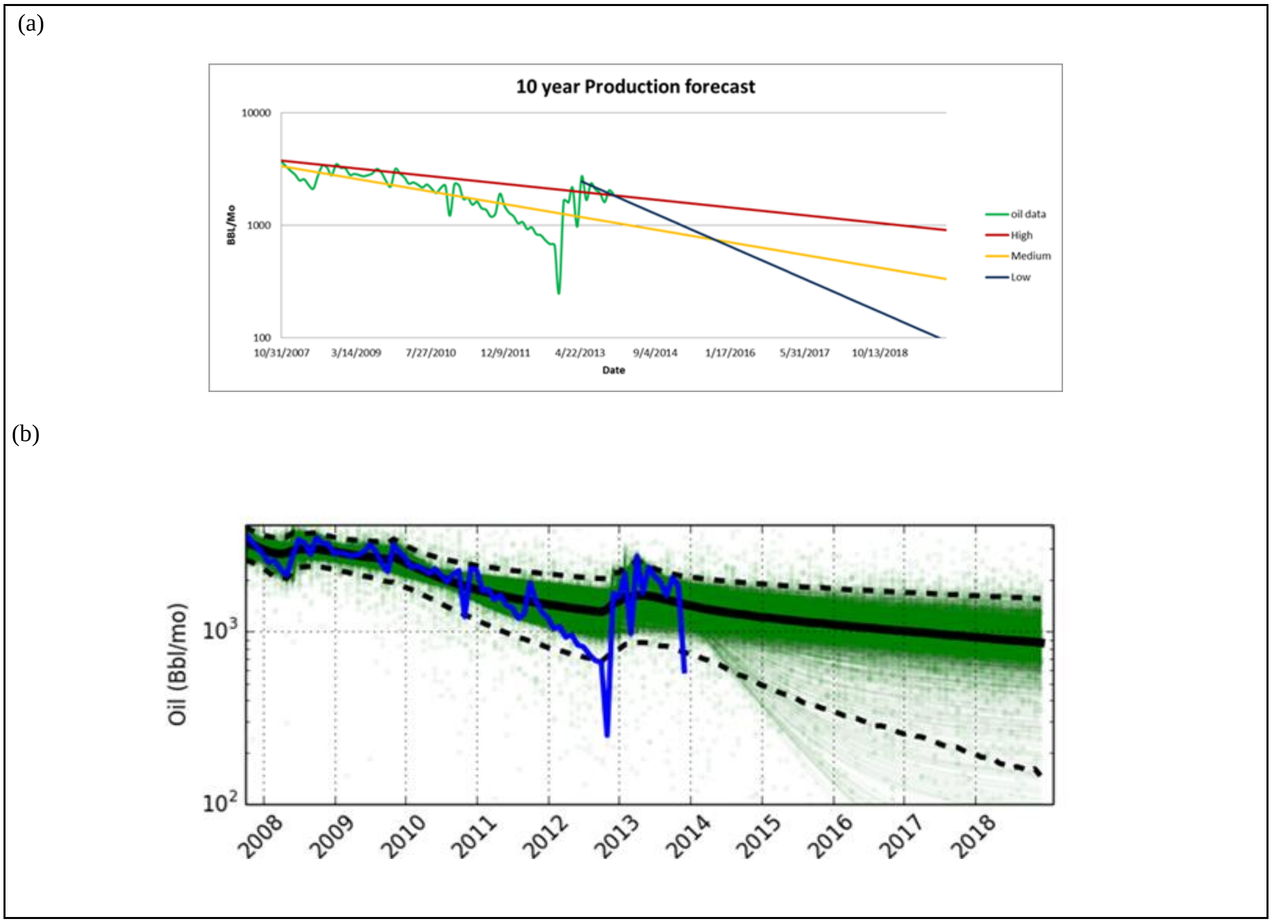


Figure 1: a) Three possible production forecasts made by fitting an exponential curve to the production history of an unconventional well. b) Statistical Inference uses a Generative model to explain the same data (blue). The green lines represent samples perturbed by noise (dotted green). The black lines represent percentiles p90, p50 and p10. The p90 is a lower bound (90% of the samples occur above this line). The p50 is the median. All of the exponential models shown in (a) are represented in these samples as well as a number of more complicated curves, including ones which explain the data after 2013 as being part of a new, rapidly declining production regime after a small stimulation

When these samples are projected forward in time, they give a forecast that is actually a *posterior probability distribution.* By doing statistics over these it is possible to assess the likelihood of any possible outcome.

Statistical Inference vs. curve fitting

Figure 1 illustrates the difference between statistical inference with a GM and standard curve fitting. In Figure 1a, a set of exponential models were fit by hand to the production history of an unconventional well in the Bakken formation of North Dakota. (Exponential models were chosen for simplicity, but any other standard model could have been used). Note that three possible models are shown which fit the data and that one of them has a different starting point from the others.

The samples from the GM shown in Figure 1b include all of these fits as well as many others, some of which include the possibility that there was a small stimulation and steep decline in 2001. (Roughly 10,000 are shown out of a million computed using a proprietary GM and inference software.) This type of forecast is much richer than one which offers a single number. It makes it possible to set meaningful bounds. For example, a reasonable lower bound is the p90 (90% of the GM samples exceed this number). Similarly, the best forecast can be set to represent the median of the samples, the lower bound, and an upper bound set at the p10. These are displayed in the figure.

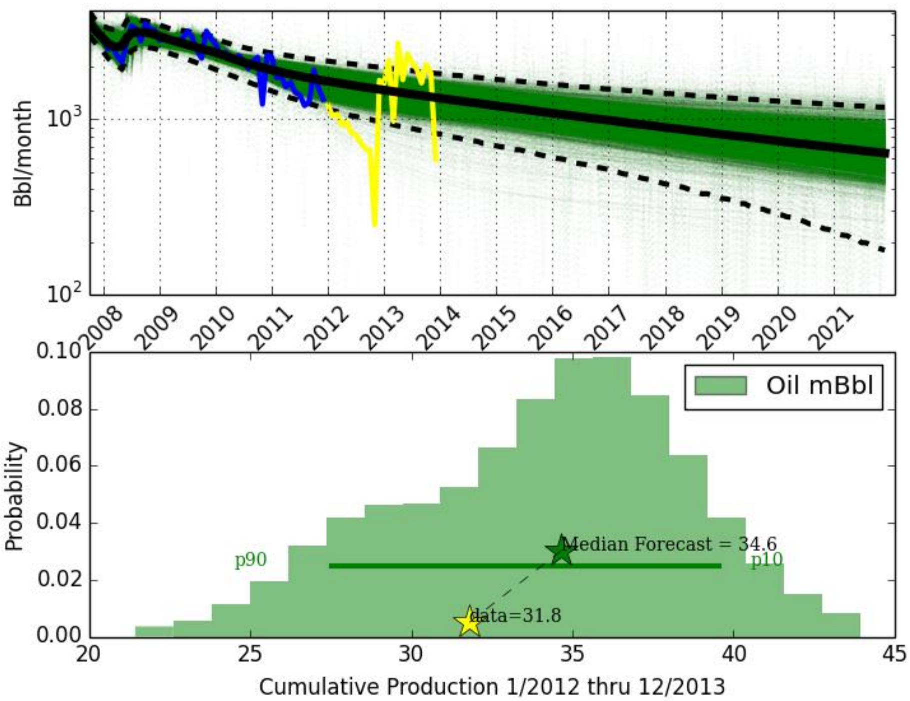


Figure 2: top) Two years of data (yellow) is held out for testing. Note that although the data is noisy, agreement with the bounds is still good. Bottom) A histogram of cumulative production between 2012 and 2014 show with actual cumulative production during the test period (yellow). Although the production history is noisy, cumulative production still matches well with the prediction.

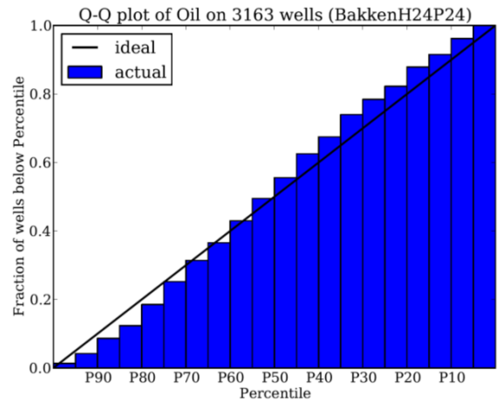
Figure 2 shows the outcome of an experiment in which two years of data were held out of the forecast for testing. The bottom of the figure shows a histogram of the expected cumulative production from the well during those two years compared to what actually happened. The histogram is heavy-tailed on the low end indicating that there is considerable downside to this well (at least for a prediction made in early 2012 – in fact, production did decline steeply in that year).

Calibrated Uncertainty and Quantile-Quantile plots

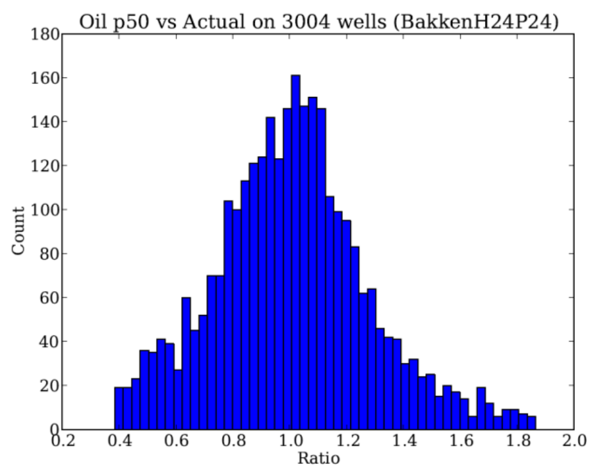
It is a simple matter to add error bars to any forecast, however it is much more difficult to ensure they are meaningful. Histograms and bounds are not useful unless there is some evidence that the samples used to make them actually do correspond to real statistics. *Calibrated Uncertainty* means that the probability distributions that are assigned to individual forecasts have been validated by computing how well on average they hold up against actual data. An algorithm can be calibrated by asking a few deceptively simple questions: *Does actual production exceed its p90 (lower bound) 90% of the time? Does it exceed the p10 (upper bound) 10% of the time? What about the p80, p70 and points in between?* If the answer to these questions is yes, uncertainty is calibrated, and bounds and histograms are correct.

To test the calibration of an algorithm, statisticians use Quantile-Quantile (Q-Q) plots which are largely unknown in oil and gas. A Q-Q plot is a statistical tool used to determine when two probability distributions are equivalent. In production forecasting, it used to compare the distribution of the difference between forecasts and actual historical production. To make a Q-Q plot, an algorithm must be run on a significant number of historical records (several thousand at least), leaving off some of the production for testing. The number of times that actual production exceeded the algorithm's calculated percentiles (p90, p80, p70 and so forth) is computed and presented on a plot

(a)



(b)



(c)

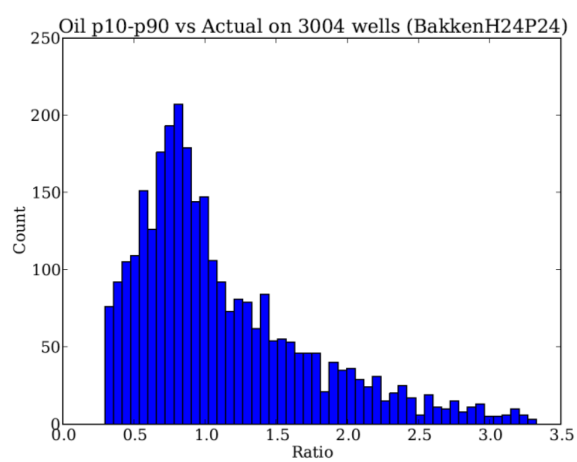


Figure 3: a) Q-Q plot comparing two years of cumulative oil production forecast to actual production percentiles. The curve deviates slightly from an ideal straight line because some of the wells were stimulated during the test period. b) The ratio of the forecast p50 to actual production. It is centered slightly high at about 1.1 meaning that forecast p50s tended to be 10% higher than actual production. c) The ratio between actual production and the uncertainty of a forecas showing that, although the spread in uncertainty is usually lower than the meadian value of the forecast, some forecasts can have spreads that are as much as three time greater than the p50.

such as the one in Figure 1. If this plot shows a straight line, then the probability distributions match and the uncertainty measures presented by the algorithm can be considered calibrated.

Figure 3a shows a Q-Q plot computed with the same GM and inference as above, using historical oil and gas production from 3163 wells in two counties of North Dakota. Again, two years of data was held out to test the predictions. Ideally, the plot would be a straight line; however, actual production is skewed high by a few percent in the upper percentiles. This is because some of the wells in the test dataset were stimulated during the test period. A skew upward like this is to be expected because engineers are always seeking and finding ways to boost production outside of its predicted ranges.

The median (p50) curve is the closest the GM returns to a single answer. Figure 3a shows the ratio between the predicted p50 and actual production for 3004 of those wells (159 wells were excluded because they had missing data which was not accounted for in working days). The mean ratio of the prediction to actuals was slightly under 1.1, meaning that on average the p50 predicted less than 10% higher than actual production. The shape of the distribution of errors is not Gaussian. More than 70% of the time, the p50 was within 25% of actual production. Figure 3b shows the ration of the spread between the p10 and p90 to actual production, giving a notion of the tightness of the prediction. This spread is centered at about 0.8, indicating that the error bounds are usually about 80% of production. The distribution is heavy tailed on the high side, however, meaning that sometimes the uncertainty in the prediction can be two or three times higher than the prediction itself.

GM interpreted as a type curve

A type curve is a representation of the average behavior of a group of completions or wells. It gives an engineer a best guess as to what production will be if little or no other data is available. For example, if an infill well has not yet been drilled, a type curve can be used to estimate possible production. When a completion is newly opened, initial production and pressures can be combined with the type curve to refine the estimate.

A GM is essentially the same thing with more information. In the absence of data, its samples are snapshots of possible well behaviors. A GM can be calibrated using only a subset of wells, much the same as a type curve can be made on a single field, reservoir or play. When given only a small amount of data, such as an initial volume, the bounds of the forecast will be reasonable, albeit large. Figure 4 shows what happens when only two months of data are used for inference on the same well in Figure 1. Note that, although the bounds are much wider when little data is available, the p50 retains its same general character.

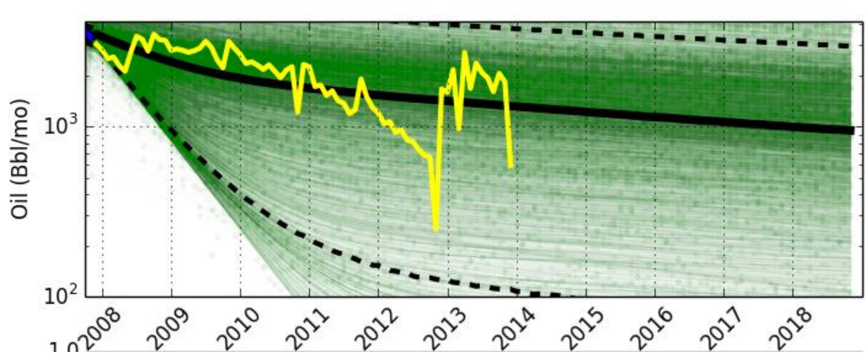


Figure 4: The same well using only two months of data to make a forecast. The p90 and p10 bounds are much wider than they are in the forecast made using all of the data, but the median forecast is nearly the same.

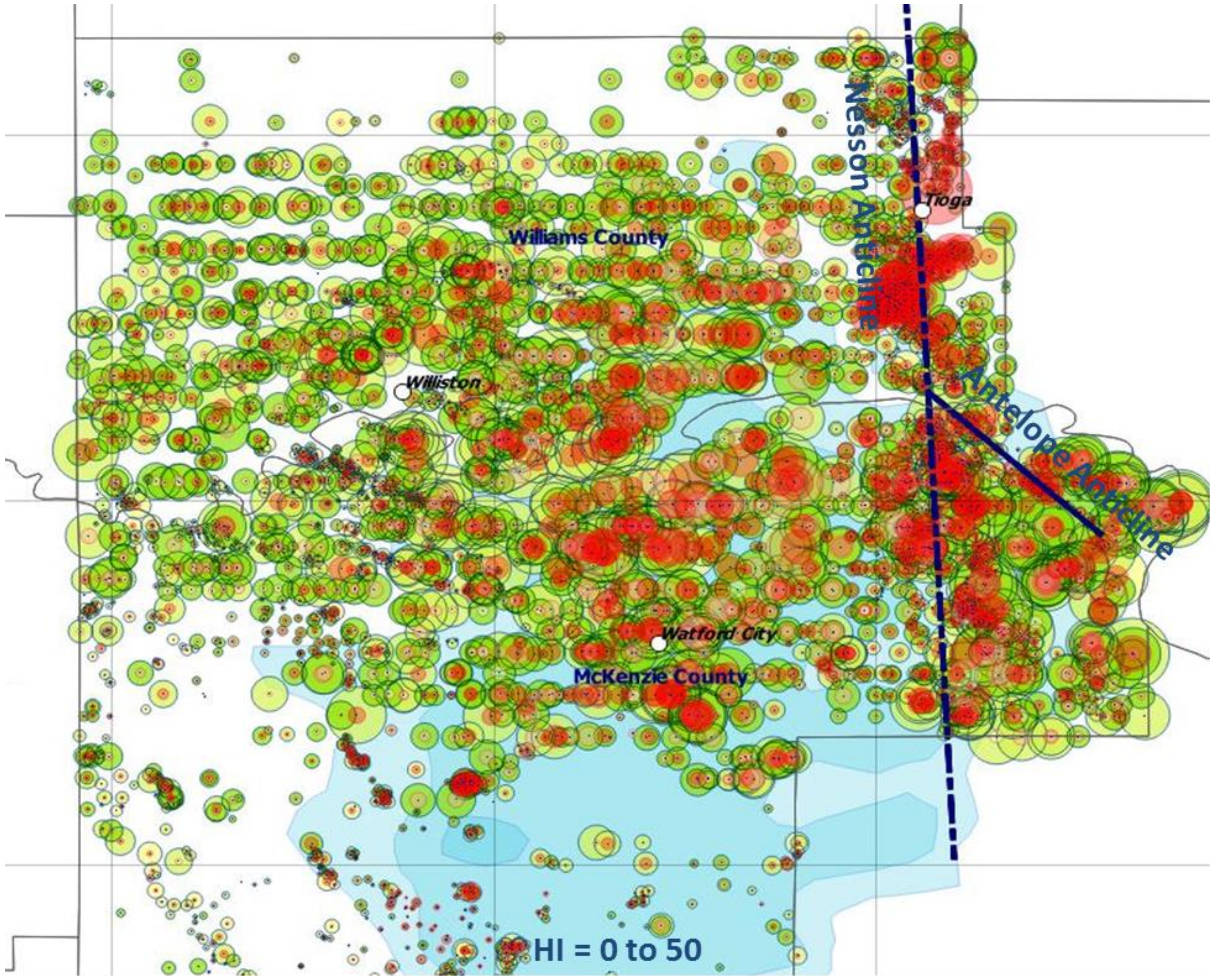


Figure 5: Forecast Certainty Map for Williams and McKenzie Counties, North Dakota for a five year forecast (2014- 2019). Oil (green) and Gas (red, normalized by BFE factor). Dot size is scaled by p50 forecast and color saturation shows uncertainty. The big, light red dot near Tioga (upper right) is a gas well that has high potential and high uncertainty.

Forecast Certainty Maps

A *Forecast Certainty* map, such as the one in Figure 5, is a visual representation of how production is likely to play out if no new investment (drilling or stimulation) is made in a field. Dots are sized according to their p50 (gas is scaled using a factor of 1 BOE per 6000 scf). Color saturation is a function of uncertainty (the ratio of the p10 to p90 distance normalized by the p50). This makes it possible to identify at a glance wells that have particular characteristics. Big, dark wells are likely to be predictably important producers. Big light wells, such as the big gas well near Tioga, are risky but might have a potentially high upside. Also shown are Hydrogen Index contours which are used as a stand-in for shale maturity in the Bakken. It is clear that there is a correlation between maturity and reliable gas production. (The correlation between liquids and maturity is less clear on this map.)

Although this map shows unnormalized cumulative production of oil and gas over five years, a similar map could be made with Estimated Ultimate Recovery, or production as a function of completion length and perforations, number of stimulations, or well age or initial volumes. To facilitate evaluating investment opportunities, it could be shown in terms of Net Value or Net Value divided by price in order to identify attractive investment opportunities. Without a reliable forecast, none of these maps would be possible. Without calibrated uncertainty, they would not be meaningful.

Conclusions

Production forecasts are a vital tool for the oil and gas industry. Currently, they are largely made using curve-fitting. Often, forecasts are adjusted by hand. A Generative Model works in a manner that is similar to that of a human engineer, taking not only idealized production curves but production artifacts into account. Combined with an algorithm for statistical inference, such a GM makes accurate forecasts that include uncertainty. If the uncertainty is calibrated by rigorous testing, then it can be trusted. This opens up a whole realm of interesting opportunities to explore production data either by itself or with other types of data.

References

Global Integrated Workshop Series http://www.spe.org/events/13abn2/

Arora, N., Russell, S., and Sudderth, E., 2013. "NET-VISA: Network Processing Vertically Integrated Seismic Analysis." Bulletin of the Seismological Society of America, vol. 103. No 2A. pp 709-729. DOI:10.1785/0120120107.

Harris, S. and Lee, W.J., 2014. "A Study of Decline Curve Analysis in the Elm Coulee Field", SPE conference paper SPE-169018-MS. DOI: 10.2118/169018-MS.

Hook, M., Savidsson, S., Johansson, S., and Tang, X., 2013. "Decline and depletion rates of oil production: a comprehensive investigation." Philosophical Transactions of the Royal Society A. 13, January 2014 bol.372 no. 2006 20120448. DOI: 10.1098/rsta.2012.0448.

Hughes, J.D., 2013, "Energy: A reality check on the shale revolution," Nature 494, 307-308. Doi:10.1038/494307a

Lake, L.W., Martin, J., Ramsey, J.D., Titman, S., 2013, A Primer on the Economics of Shale Gas Production. Just How Cheap is Shale Gas? Journal of Applied Corporate Finance, Vol 25. Pp 87-96. DOI: 10.1111/jacf.12045

Milch, B. and S. Russell. 2010. Extending Bayesian Net-works to the Open-Universe Case. In Rina Dechter, Hector Gener, and Joseph Y Halpern, Eds., Heuristics, Probability and Causality. A Tribute to Judea Pearl. College Publications.

Sallh, D., Hook, Mikael, Grandell, L., Davidsson, S., 2014 Evaluation and update of Norwegian and Danish oil production forecasts and implications for Swedish oil import, Energy, Vol 65, pp 333-345.

Shahamat, M.S., Mattar, L., Aguilera, R., 2014, Analysis of Decline Curves Based on Beta Derivative." SPE Conference Paper. SPE-169570-MS. DOI:10.2118/169570-MS.

Shin, H.J., Lim, J.S., Shin, S.H., 2014. Estimated ultimate recovery prediction using oil and gas production decline curve analysis and cash flow analysis for resource plays. Geosystem Engineering, Vol 17, Issue 1 pp 78 – 87. DOI: 10.1080/12269328.2014.886970

Yu, W. and Sepehrnoori, K., 2013. "Simulation of gas desorption and geomechanics effects for unconventional gas reservoirs". Fuel. Vol 116. Pp 455-464. DOI:10.1016/j/fuel/2013.08.032.