SPE-173444-MS

Big Data Every Day: Predictive Analytics Used to Improve Production Surveillance

Scott Raphael, Clara P. Fuge, and Shawn Gutierrez, P2 Energy Solutions; Heidi A. Kuzma, and Nimar S. Arora, BetaZi LLC

This paper was prepared for presentation at the SPE Digital Energy Conference and Exhibition held in The Woodlands, Texas, USA, 3–5 March 2015.

Abstract

Terabytes of data are being collected every day in the oilfield. Intuition suggests that the more data the better and that if a little bit of information is valuable, then a lot must be incredibly more so. And yet, is it? It is not the data itself, but the ability to put it toward some useful task which matters. One of the most concrete ways in which Big Data can be brought to bear on the everyday is through *predictive analytics*, the science of using facts from the past to analyze the present and predict the future. In the oil field, a simple and powerful way of making use of data and predictive analytics is for automatic *surveillance by exception*. By learning from the past it is possible to predict what should be happening today and compare the prediction to what is actually going on. If the two do not match then a problem has been identified and a solution found. Addressing problems early is one of the cheapest ways to boost production.

Introduction

Terabytes of data are being collected every day in the oilfield. Intuition suggests that the more data the better and that if a little bit of information is valuable, then a lot must be incredibly more so. And yet, is it? It is not the data itself, but the ability to put it toward some useful task which matters. One of the most concrete ways in which Big Data can be brought to bear on the everyday is through *predictive analytics*, the science of using facts from the past to analyze the present and predict the future. In the oil field, a simple and powerful way of making use of data and predictive analytics is for automatic *surveillance by exception*. By learning from the past it is possible to predict what should be happening today and compare the prediction to what is actually going on. If the two do not match then a problem has been identified and a solution found. Addressing problems early is one of the cheapest ways to boost production.

Wells that are big producers are typically watched by many eyes. Extensive instrumentation records pressures, temperatures, volumes and equipment performance parameters at regular intervals which can be as often as every minute. A large staff can quickly respond to unusual readings. However, engineers and instruments are expensive. In fields with many small wells, a single person may be responsible for the production of a hundred wells for which he or she is receiving information only once a day. Horror stories abound in which wells failed to produce at expected rates before an operator caught on and the situation was remedied. In this case, a simple, inexpensive system which requires no instrumentation other than the production monitoring, well testing and production allocations can potentially save millions of dollars in deferred or lost production if the data can be used to effectively provide a metric against which to measure current production.

Furthermore, many of the parties who are interested in production surveillance may have limited access to data other than production records. These include non-operating partners, royalty owners and lenders who also have an interest in making sure that production does not drop below reasonable levels. In this case, the data they have to consider might be as limited as monthly reported volumes.

Therefore, although most published advances in surveillance are concerned with new types of measurements , data collection or aggregation [Brunei et al,, 2003; Goh et al., 2007; Poulisse et al, 2006], there is a need for improved surveillance based on data which is already being collected and is generally available. The key to improving surveillance is improving prediction. The more accurately one can predict what production should be; the easier it is to spot deviation from the ideal.

Method

The Production Forecast Algorithm (PFA) is a commercially available, patented, automatic system for oil and gas production forecasting [Kuzma et al, 2013]. It uses monthly oil, gas and water allocated volumes and records of working days (if available) to accurately forecast future volumes. Because it is a statistical algorithm, it is capable of producing calibrated percentile bounds. It is based on a Generative Model (GM) which is capable of *generating* artificial production histories which have the same statistics as production histories have in real life where “real life” is defined as large datasets containing public records. In other words, if the GM is left to run on its own, it will produce a set of synthetic records called *samples* that have the same statistical properties as other large datasets such as the Drilling Info or IHS databases. It is able to do so by combining a physical solution with a probabilistic model learned from tens of thousands of wells. When it is presented with data from an individual well or completion, it uses a Markov Chain Monte Carlo algorithm to focus its samples in order to explain the data. The inputs into a forecast are past monthly volumes of oil, gas and water and a representation of down time or working days.

Figure 1 shows an example of a GM forecast. The physical part of the GM ensures that the PFA is explaining data in terms that are physically plausible. The statistical part is used to explain normal deviations from physical behavior such as stimulations, shut-ins and bad data. Once the samples are drawn, they can be pushed forward to make a probabilistic forecast as is illustrated in Figure 1. Statstics are then computed over the samples. The typical output of the PFA is a set of percentile values (p-values) which reflect the expectation that actual production should exceed bounded values. For example, actual production should exceed the p90 90% of the time. It should exceed the p80 80% of the time and so on. The PFA is an example of practical, quantified *predictive analytics*. In the case of the PFA, past facts include both the statistics of the wells which were used to learn the probabilistic parts of the GM and the history of the individual well for which production is to be forecast.

Rigorous calibration with blind testing ensures that the statistics of the samples match real data. For example, when a p90 bound is computed, calibration studies show that in blind studies actual production exceeds the bound 90% of the time. Actual production exceeds the p80 bound 80% of the time and so on. Figure 2 shows an example of such a quantile-quantile (Q-Q) calibration study. Metrics such as the Q-Q are regularly used by statisticians and computer scientists but are relatively new to oil and gas.

Because of its statistical nature, the PFA can be used as an easily understood and controllable tool for surveillance, it can be used to set thresholds of a very specific nature: namely, the user can ask to be warned when production goes below a particular p-value such as the p90. In that case, the user can expect to receive warnings about 10% of the time from each well under surveillance. The number of alarms can be reduced by using filters that reflect other factors such as changes in water cut, known shut-ins for

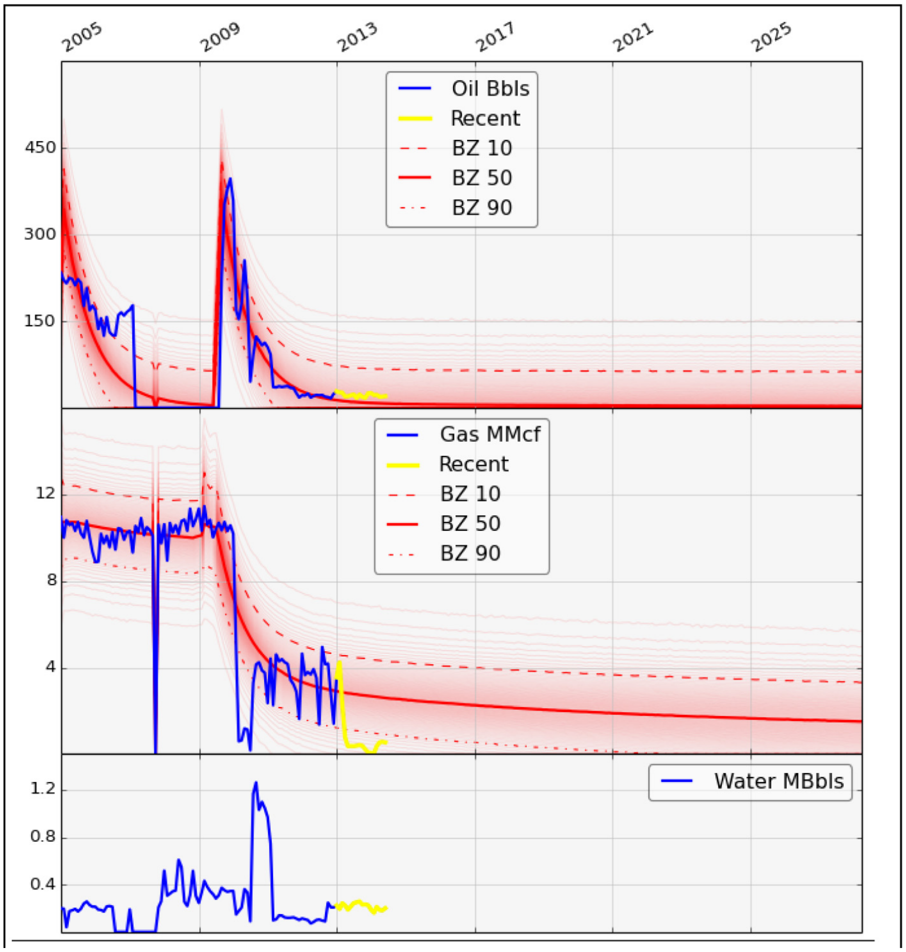

Figure 1—Forecast made using the Production Forecasting Algorithm (PFA) described in this paper. Blue indicates data that was used to make the forecast. Yellow indicates actuals that were held out for testing. Red lines are percentiles with heavy markings for p90 (lower bound), p50 (median) and p10 (upper bound).

maintenance and the like. In the following experiments, the PFA was used to make statistical forecasts of production from oil and gas wells using monthly volumes and working days during a test period. The forecasts were then compared to actual data that occurred after the test period. When the actuals fell below thresholds set by the PFA bounds, an experimental "alarm" would have been triggered.

The PFA is also used to make historical "what if" forecasts. Often wells undergo significant changes which are either physical such as the transition from one production regime to another or operations such as a choke change, shut-in or stimulation. When an operational change occurs, a what-if forecast made using data collected before the change gives a reliable way of measuring the effect of the change.

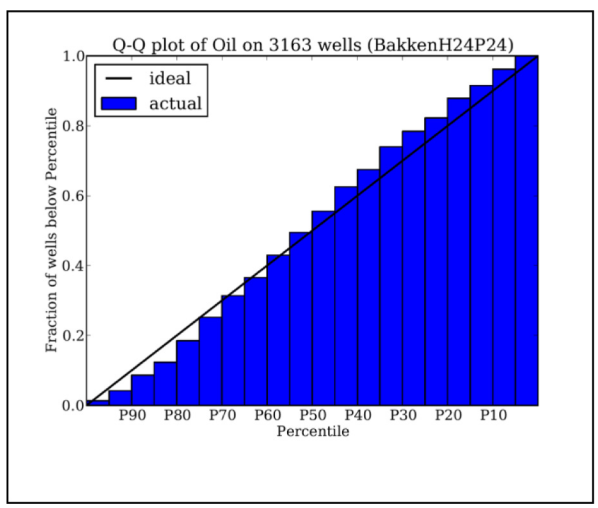

Figure 2—Example of a Q-Q plot.

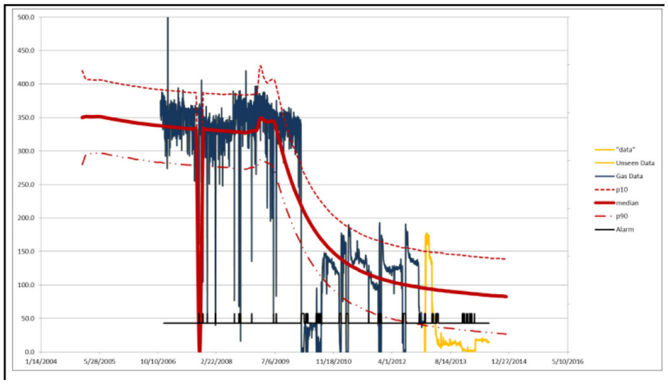

Figure 3—(Left) Production alarms (black) are set using a 30 day moving average of daily production data. Although the alarm is triggered on April 10 when production declines steeply, the 30 day average quickly resets to the lower level and the alarm is not triggered again until production declines even more a few months later. (Right) Production alarms are set by comparing daily production to a forecast made using the blue section of the dataset. The alarms from the moving average are also shown in black for comparison. In this case, the warning would have been given also on April 10, 2013. Continued warnings would have alerted interested parties that the situation was not remedied. (This is the same forecast that is illustrated in Figure 1).

Example 1 (single well)

If daily data is available, it is not difficult to recognize when a well is producing less than normal, even taking into account the fact that the production from a well declines over its lifetime. Big dips in production are obvious. Smaller dips can easily be found automatically by comparing production to a running average of previous production (over 7 or 30 days for example). Figure 3 shows a year's worth of production records for a small gas well in Texas. Alarms are marked as if they were set to trigger any time that production fell more than 50% below the average of the last 30 days production. The alarms do appear to trigger at reasonable times in the record.

Although the well was shut-in and repressurized several times during its history, it clearly underwent a major reduction in production in April 2013. A user of the moving average system would have been warned when production started to fall on May 10 and then again when it fell more steeply on April 10. But, from the perspective of a moving average, even persistently low production quickly becomes normal.

The other side of the figure, however, gives more information. To create it, it was assumed that the PFA used oil, gas and water records up to April 1, 2013 to make a forecast. (This is the same forecast shown in Figure 1). In this figure, if an alarm had been set to trigger when production falls below the p90 lower bound, then it would also have given warning on April 10, and every day thereafter. The warning would have been much more specific. The message from the moving average alarm was that production was falling. The prediction alarm tells the user that production has fallen to a below-normal level that should only be reached 10% of the time, and furthermore it is staying low. In this case, the reason for the drop was listed as "junked equipment in hole," a situation that was not recorded until June 1.

Example 2 (group of wells)

Figure 4 shows a different representation of a forecast made using the PFA a year later at the end of June 2014. In this example, monthly instead of daily volumes are shown and all of the percentiles (from the p1 to p99) are plotted. Notice that although the rise in production in April, 2014 does not reach the p90 bound in the previous forecast, in this forecast, the fact that production has been down for a year is enough to make the algorithm believe that the well is now in a new regime and bringing the forecast levels down. The April rise is taken into account in the p50.

Statistics do not add. The p90 of a group of wells is not the sum of the p90s of the individual wells. That is because it is unlikely that all of the wells in a field will be low producers at the same time. Instead, proper statistics on a group of wells need to be computed by sampling from the full probability

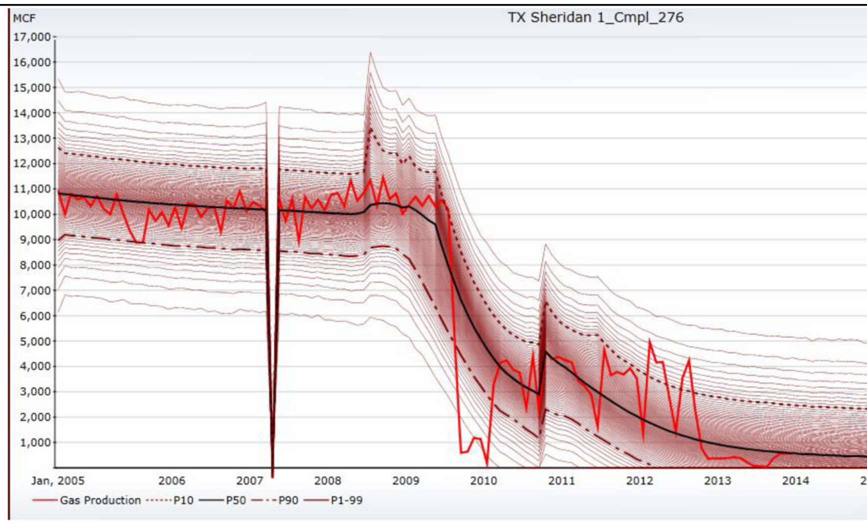

Figure 4—A forecast on the same well as Figures 1 and 3, this time made using data through June, 2014.

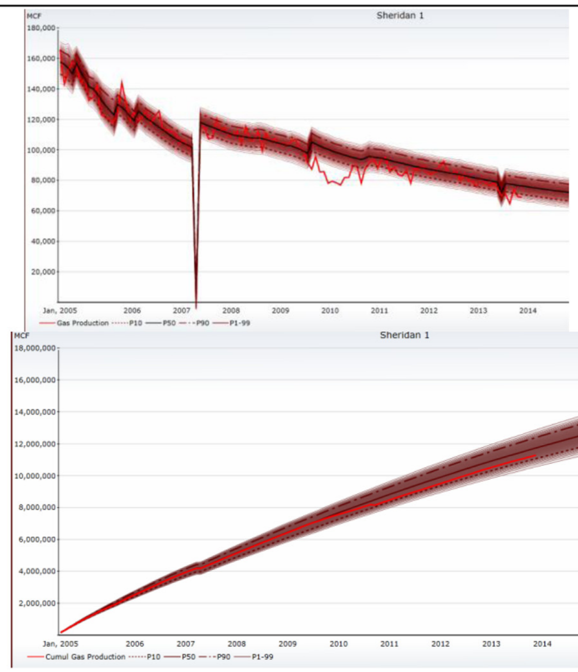

Figure 5—(Top) Statistical rollup of 19 wells in the same field as the well in previous figures. Note the drop in production in 2010. (Bottom) The same rollup shown in a cumulative view.

distributions of the individuals. Figure 5 shows a statistical rollup over 19 wells in the same Texas field as the well in Example 1 (including that well). Note that the width between the p90 and p10 in the rollup is considerably narrower than on the individual wells.

There is a significant decline to the field starting in 2014. This is because out of the 19 wells, 5 have anomalous decline. The drop below the p90 in 2010 is because a different subset of five of the wells had an unusual number of shut-in days. This message is reinforced in the bottom figure which shows the rollup statistics in a cumulative view. There is a 400,000 MCF discrepancy between actual and p50 production, most of which is due to the loss in 2010. Without the p50 and forecast statistics, there is no fixed point of reference for determining this loss.

Example 3 (geographical surveillance)

The map in Figure 6 represents the future of production from 3,544 Bakken wells. (This data is the same data that was used for the calibration study in Figure 2). The markers are sized proportional to a p50 forecast for cumulative production in 2013 that would have been made at the end of 2012. They are color saturated as a function of uncertainty; that is the distance between the p10 and p90 normalized by the p50. The oil forecast is coded in green, gas is in red. A yellow star marks every well for which production fell below its p90 during 2013, giving a visual display of wells which had shutins or otherwise reduced production for some period during the year. If either the oil or gas production from a well went below its p90, a theoretical alarm was rung. The size of the yellow stars indicates the number of alarms for a particular well. The filled blue contours indicate Hydrogen Index which is a proxy for the maturity of Bakken shales. The deepest blue is the most mature part of the formation, where production tends to be gas rich.

There were 938 alarms during the year experienced by a total of 267 wells (8% of the total). The most alarms experienced by a particular well was 12 although about half of the (100) wells which had alarms only had one alarm. In 2012, with 3586 wells in the dataset, there were 551 alarms, 6% of the wells experienced one alarm with 124 wells only experiencing one alarm. It is possible that the increased alarm rate in 2013 was due to more horizontal wells, or it could just be natural statistical variation.

It is interesting that if the markers for oil are removed from the map, the size and location of the yellow stars corresponds well to the location of big gas producing wells and is not necessarily correlated to the density of wells. Although this correlation has not yet been extensively studied, one suggested explanation is that most of the high producing wells with multiple alarms are likely relatively new horizontal wells which are technically more difficult to produce than vertical wells and may be subject to more downtime. It is also interesting to consider that the permeability, porosity and natural fracturing of the Bakken formation (particularly the middle Bakken) are more variable in the deep, mature regions of the shale than they are on the edges and production may be subject to more uncertainty [Pitman, 2001].

An investor might be interested in comparing the alarm rate of various operators as a proxy for the amount of production which might be lost due to downtime. 265 separate operators were listed in the 2013 dataset (although some of the separate listings were for different divisions of the same company). Only 50 companies operated wells which had alarms. Figure 7 shows the top 20 companies by well count with their alarm rate superimposed. There is no correlation between well count and alarm rate, although some operators fared much better than others.

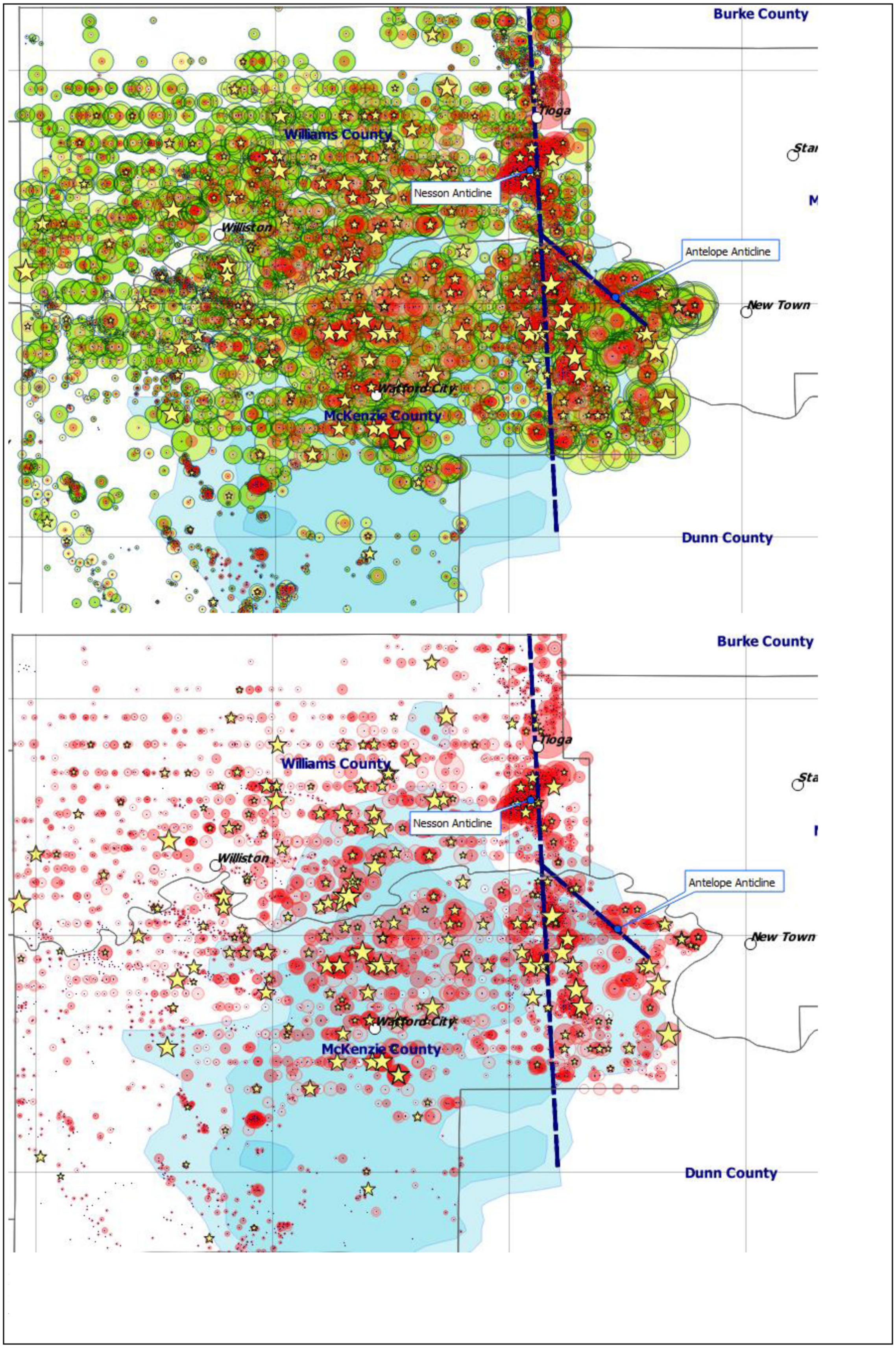

Figure 6—(Top) Map of approximately 3,400 Bakken wells with markers sized to reflect forecast production and color saturated to mark uncertainty. Oil is green, gas is red. Yellow stars mark wells which had production in 2013 that was less than the p90 and for which a theoretical alarm was triggered. Blue marks Hydrogen Index contours. (Bottom) The same map without the oil markers.

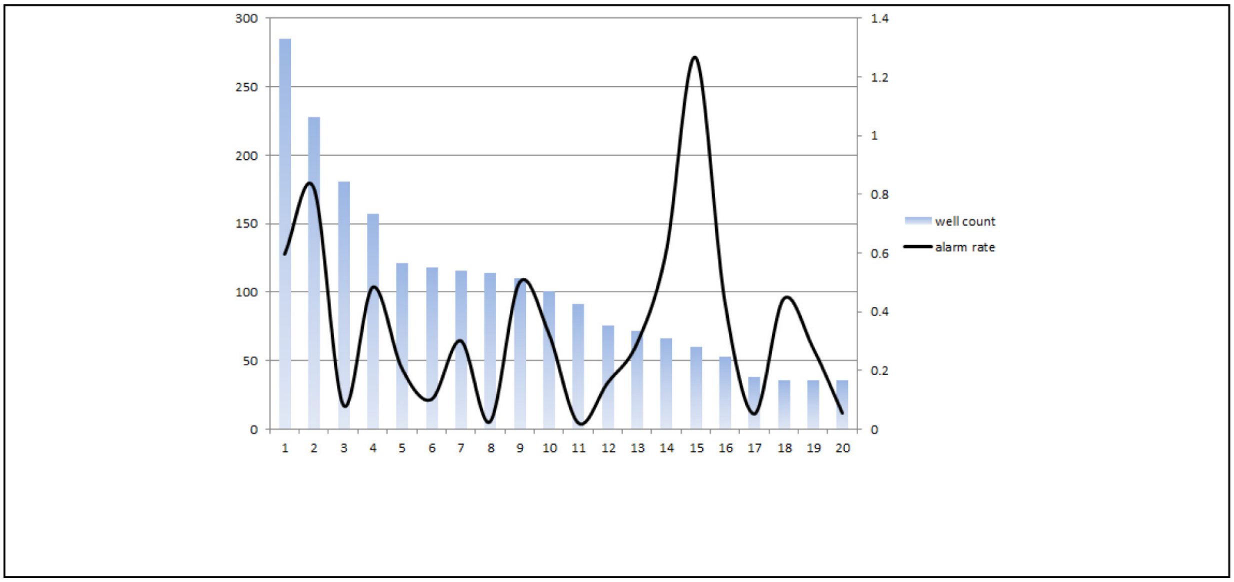

Figure 7—Study showing the 20 operators who had the highest well counts and also had alarms in 2012.

Conclusions

The key to surveillance is knowing what to expect. Using sophisticated predictive analytics, the PFA in this paper gives accurate, calibrated statistical forecasts against which actual production can be measured. The three given examples show that having a grounded metric for production is an effective means of computing automatic warning alarms, computing lost production and correlating production alarms to geographical location, geology and to particular operators.

References:

Bruni, Tl, A. Lentini, S. Ventura, R. Gheller, C.A. Maybee, J.E. Pinedo, 2003, A Technically Rigorous and Fully Automated System for Performance Monitorning and Production Test Validation., SPE, International Improved Recovery Conference in Asia Pacific 20-21 October, Malaysia. DOI: **http://dx.doi.org/10.2118/84881-MS**

Goh, K-C., C.E. Moncur, P. Van Overschee, J. Briera, 2007, Production Surveilland and Optimization with Data Driven Models. International Petroleum Technology Conference, 4-6 December, U.A.E. DOI: **http://dx.doi.org/10.2523/11647-MS**

Kuzma H.A., H.S. Arora, K. Farid. 2014. Generative Models for Production Forecasting in Unconventional Oil and Gas Plays. Unconventional Resources Technology Conference. DOI: **10.15530/urtec-2014-1928595**

Pitman, J.K., L.C. Price, J.A. LeFever, 2001. Diagensis and Fracture Development in the Bakken Formation, Williston Basin. *Implications for Reservoir Quality*. U.S.G.S. Professional Paper 1653.

Poulisse, H., P. van Overschee, J. Briera, C. Moncur, K.C., Gobi, 2006, Continuous Well Production Flow Monitoring and Surveillance, SPE, Intelligent Energy Conference and Exhibition, 11-13 April, The Netherlands, DOI: **http://dx.doi.org/10.2118/99963-MS**