geoLOGIC basinINTEL powered by BetaZi™

BetaZi SCIENCE

AN OVERVIEW OF PHYSIO-STATISTICS FOR PRODUCTION FORECASTING

EVOLVE, ALWAYS
That's geoLOGIC

BetaZi™ | geoLOGIC SYSTEMS

betazi.com

geoLOGIC.com

Introduction

Predictive Analytics is the science of using past facts to understand the present and predict the future. With inputs of monthly oil, gas and water production volumes, the BetaZi model automatically forecasts future hydrocarbon production from a well or completion by finding a million possible scenarios (called samples) which explain the production history of a well. The model then runs statistics over those wells to establish BetaZi's p-values.

Figure 1, below, illustrates BetaZi's basic product. The samples were generated from data (blue). They were then projected forward to return not just the single forecast which is typical in standard oil and gas production forecasting (for example, the exponential forecast illustrated by the dotted yellow line), but a range of possible outcomes which are presented as percentiles/p-values ranging from the p99 to the p1 (light green lines). The standard lower and upper bounds of the distribution are the p90 and p10 (darker green). P90 means that 90% of the BetaZi samples exceed this bound. P10 means that 10% of the samples exceed the upper bound. The central forecast (dark green) is the p50, or median of the samples. The width between the p90 and the p10 gives an estimate of the spread of risk of the well. The red line indicates data that was not used to make the forecast and was held out for testing.

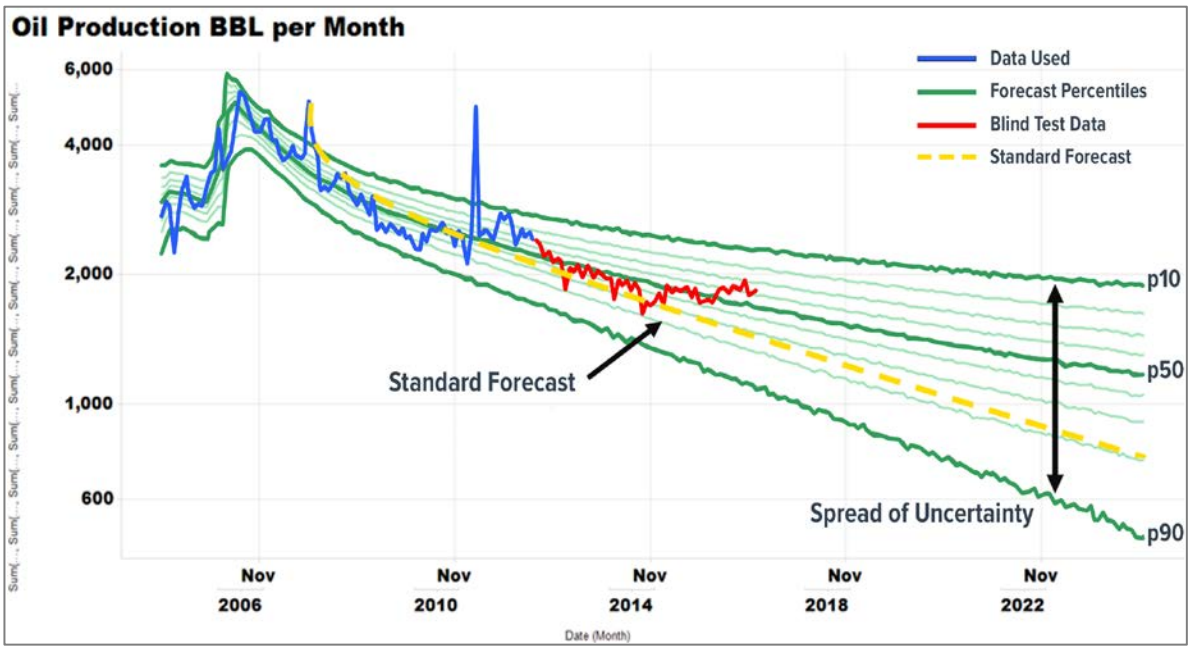

Figure 1: BetaZi's basic product is a forecast of the production from an oil or gas well which includes calibrated statistical bounds.

BetaZi works because it uses advanced science to combine both physics and experience into its solutions. It uses solutions to differential equations to ensure that its samples always respect the fundamental physics of fluid flow. At the same time, the model uses experience learned from training on large datasets to understand how typical well behavior deviates from physics. This means that it does not get confused by messy, misallocated or aggregated production data complicated by multiple shut-ins, stimulations and missing records. And since BetaZi is fully automatic and requires no human intervention, no human bias is introduced.

Science

Traditional Forecasting vs. BetaZi Samples

There are many problems with traditional production forecasting, and almost all of them stem from the fact that it is done largely by hand. In the first place, it is tedious. It also uses up valuable time that could otherwise be applied to production engineering and optimization. Because it requires judgment, which varies from person to person, it is often difficult to reconcile forecasts made by different people. And often there just is no way to arrive at a correct answer. It is much better to outline a suite of possible outcomes than a single estimate. But this is difficult and expensive to accomplish.

The fundamental problem, illustrated in Figure 2, is that there are an infinite number of ways to fit the production history of a given well. Even a simple history is fraught with noisy, missing and bad data in addition to operational fluctuations such as equipment failures, shut-ins, pressure variations, new lift and multiple stages. Many different curves can be chosen which fit the data equally well. The engineer doing the forecast must apply his or her experience to try to find the right one. Forecasts made based on different choices of curves can vary enormously when they are pushed forward ten, fifteen or thirty years. It is common practice to be looking at a screen which computes the Expected Ultimate Recovery (EUR) from a well while the forecast is being made. Slight changes in the curve that is chosen can boost (or deflate) an EUR by as much as an order of magnitude.

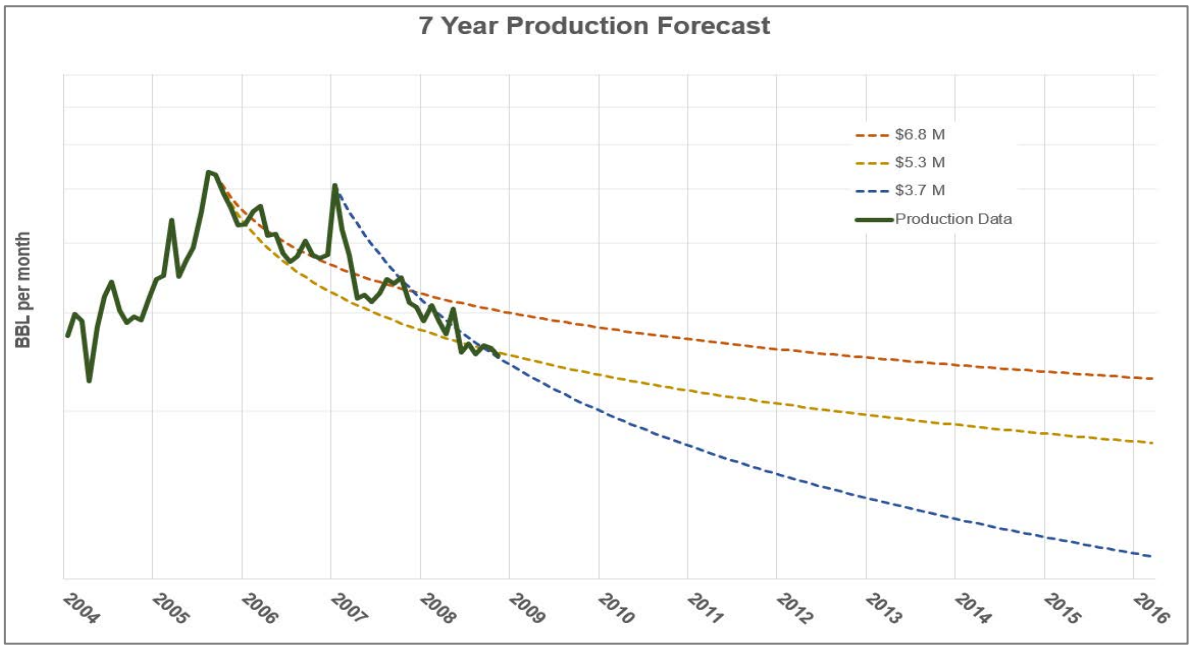

Figure 2: There are an almost infinite number of curves that can be used to fit a noisy dataset. In making a forecast, an engineer must pick just a few. In this case, even variations in a 7-year forecast can change the value of a well dramatically.

Instead, BetaZi automatically finds *all* of the curves that can be used to fit a given production history. Its million samples occur with approximately the same frequency that different production outcomes would occur if it were possible to have the same well produce a million times in different experiments. Or, putting it a different way, BetaZi generates the equivalent of a million engineers' plausible estimates.

Generative Model

BetaZi is able to produce a million samples which simulate real outcomes because it is based on a very sophisticated probabilistic *Generative Model* which uses logic that is not unlike that inside a trained engineer's head. To do so, it uses calibrated *physio-statistical inference* (physio-stats), which is a way of combining physical laws with the statistical behavior of physical systems as learned from big data in order to do predictive analytics and estimate uncertainty with repeatable and testable results. The BetaZi generative model is a physio-statistical model which, in the absence of data, will churn out synthetic well histories having the same properties as production records do in real life, including operational artifacts such as shut-ins and artificial lift. When the model is given data, these samples are focused to explain that data.

Physio-statistics

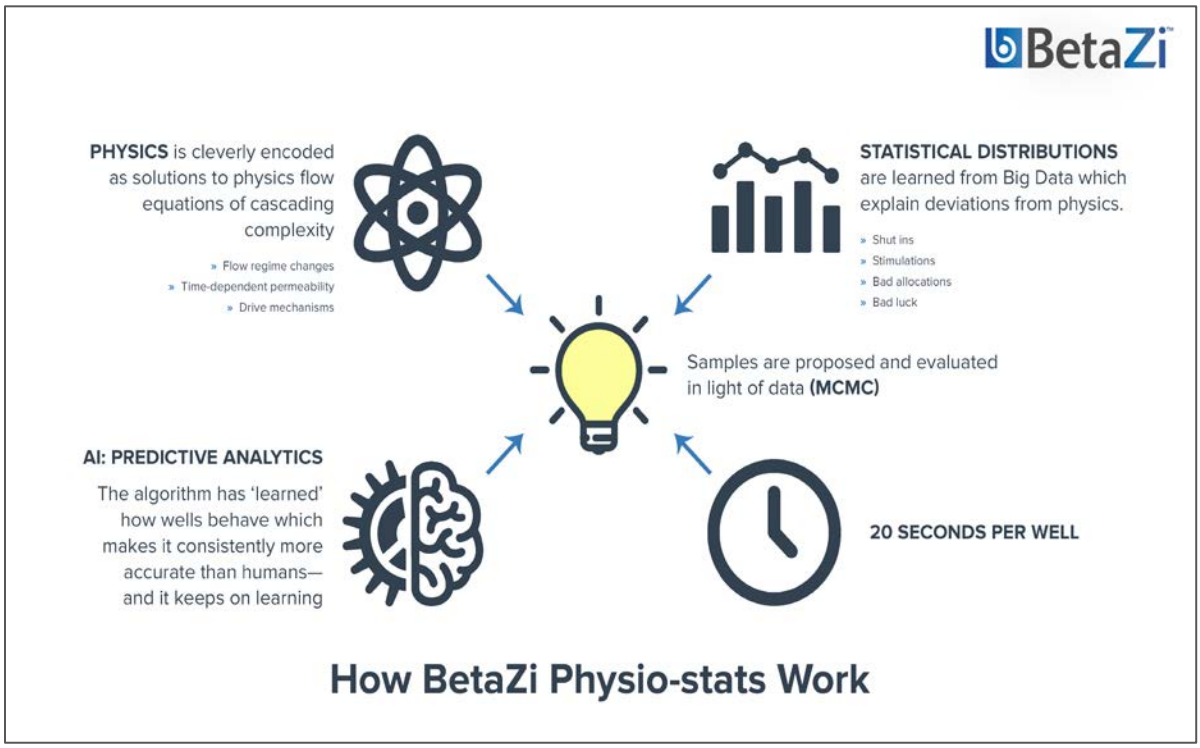

Figure 3: A simplified conceptual rendering of the BetaZi physio-statistical machinery.

After a million of these samples have been computed, they are abstracted in terms of the percentiles which are stored, displayed and used in further BetaZi computations. By convention, the 90th percentile (p90) is defined as the line above which 90% of the samples fall. 80% of the samples fall above the p80 and so on. These percentiles are also known as *p-values*. The p50 is the statistical mean.

The BetaZi Generative Model is unlike anything else used in the oil and gas industry. There are a number of products which will produce statistical forecasts using Markov Chain Monte Carlo (MCMC) sampling (which BetaZi uses to focus its samples). However, all they really do is draw samples to find the parameters of a curve equation which has been established a priori.

There are many such standard equations used to describe production declines, including exponential decline, hyperbolic decline, the Arps equation, stretched exponentials and more. They increase in complexity until they cannot be written in simple, closed form but are the solutions produced by high end numerical simulators. Each of these curve types can be thought of as the solution to differential equations which include increasing numbers of parameters. Unfortunately, if the wrong equation is chosen, it does not matter how sophisticated a method is used to solve it; the fit to data will not be good and it will not produce a meaningful forecast.

BetaZi bypasses this issue. It uses a set of smaller curves to describe the overall shape of each of its samples. The ingenuity of its patented algorithm is to be able to figure out how many of these curves it needs to explain a given production history, which it accomplishes by searching through not just parameters but curves of increasing complexity. This means that a BetaZi forecast will include all of the standard models and then some. Because all of these functions add up to solutions to differential fluid flow equations, BetaZi ensures that all of its samples will lie within the realm of physical plausibility.

Even with all of this capability, however, the generative model could not simulate actual well histories without the additional knowledge of how they tend to *deviate* from physics. This is the “statistics” part of physio-statistics. Using a probabilistic model trained on large datasets, the generative model is able to explain many of the features of production histories which have more to do with operations than physics, such as shut-ins, stimulations and the addition of artificial lift. It is also how the generative model, when left to run in the absence of data, is able to produce synthetic well histories that are very close to actual ones and not simple, idealized curves. The standard BetaZi prior, or collection of distributions describing the statistical part of the model, was derived using about 150,000 well histories from the United States and abroad.

Calibrated Uncertainty

Every BetaZi forecast includes meaningful and precise p-values: bounds which indicate both the upside potential and the risk of an asset. It is easy to assign statistics and error bars to any plot. The trick is proving that they are realistic. To do this, BetaZi LLC spends an unprecedented amount of effort on calibration, in the form of back testing using blind data, to ensure that its forecasts are accurate. (To date, no other forecasting program does this or can do it.)

Quantile-Quantile comparisons (also known as Q-Q plots), such as the one shown in Figure 4, are used by statisticians to compare probability distributions and the performance of automatic algorithms. A straight line across the plot diagonally ensures that BetaZi bounds and percentiles match reality. In blind tests, 90% of the time, actual production does in fact exceed the BetaZi p90. 80% of the time it exceeds the p80, and so on, up to the p1. Statistics that pass this test can be called “calibrated.”

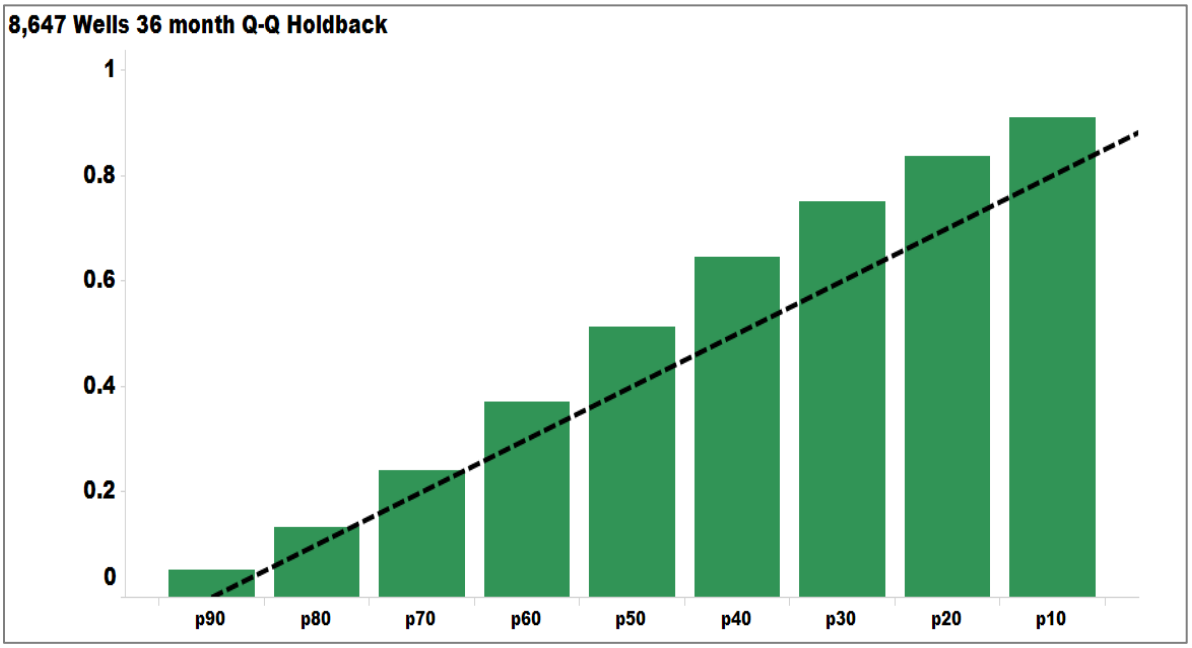

Figure 3: A passable Quantile-Quantile plot.

In practice, actual production exceeds the BetaZi upper bound (p10) more than 10% of the time because engineers are often able to stimulate, treat and otherwise coax extra production out of wells. This is, arguably, what engineers should be doing, not manually setting forecast curves.

Written by:

Heidi Anderson Kuzma, PhD– Founder and Chief Innovation Officer, BetaZi LLC

Dr. Kuzma has a PhD in Applied Geophysics with a focus on computer learning for physical applications from the University of California at Berkeley, School of Engineering; an MS from UC Berkeley in Applied Geophysics; and a Bachelor of Science in Geophysics and Planetary Science from the California Institute of Technology. Dr. Kuzma has more than twenty years' experience in mineral exploration for oil, gas and gold. She has contributed to and published on a large number of projects combining computer learning with physical systems, including: building a Self-Teaching Expert System (SeTES) for shale gas production and discovery, inventing a system for tunnel detection using the passive seismic signal from motor vehicles, and developing metrics for the evaluation of computer learning algorithms for seismic data analysis. Dr. Kuzma has also worked in atmospheric fluid flow, frame semantics, programming Graphics Processing Units, and Uncertainty Quantification via polynomial chaos.

geoLOGIC basinINTEL powered by BetaZi™

Delivered online. Updated monthly.

About basinINTEL

basinINTEL is a revolutionary software that provides data-driven, accurate, repeatable and unbiased production forecasts for every well in Western Canada. It combines BetaZi's Physics-based Predictive Analytics model with geoLOGIC's premium data to generate forecasts, in seconds. It can incorporate proprietary data to enhance business intel. You can map or chart your results for better illustration. It is delivered in a series of Spotfire projects and updated monthly.

About geoLOGIC

geoLOGIC systems ltd. is an industry leading oil & gas information provider based out of Alberta, Canada. We deliver premium data, innovative software and integrated analytics. We're continuously evolving to deliver an exceptional customer experience. For more information visit www.geoLOGIC.com.

About BetaZi

BetaZi LLC creates state-of-the art production forecasting solutions for the oil and gas industry using physics-based predictive analytics. BetaZi has been vetting its new science and providing meaningful intelligence to producers and financiers since 2011. Their research office is located in Truckee, California.