BetaZi Science

--An overview of physio-statistics for production forecasting--

Heidi Anderson Kuzma, PhD

BetaZi LLC
11209 Brockway Rd Ste 302
Truckee, CA 96161
530.587.3858

Business Development Contact:
Dan Gualtieri
dgualtieri@betazi.com
(713) 305-1175

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, BetaZi 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. We then run statistics over those wells to establish our p-values.

Figure 1, below, illustrates our 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 red line), but a range of possible outcomes which are presented as percentiles/p-values ranging from the p99 to the p1 (light red lines). The standard lower and upper bounds of the distribution are the p90 and p10 (dotted 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 yellow line indicates data that was not used to make the forecast and was held out for testing.

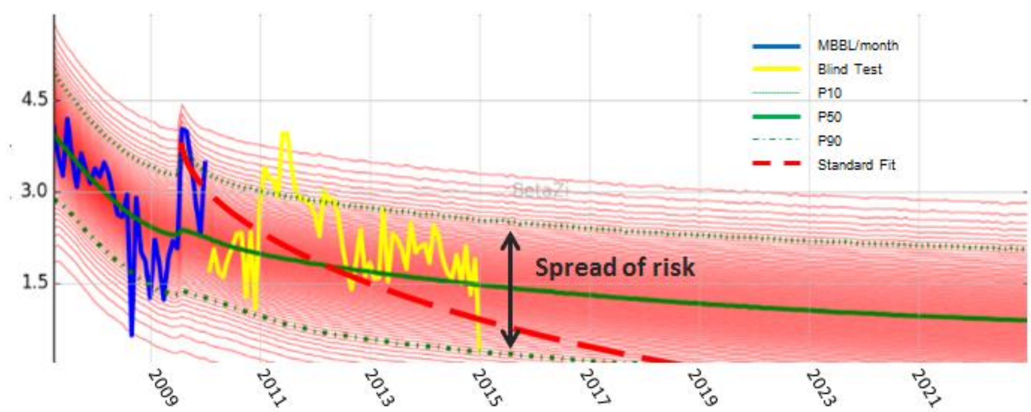

Figure 1: BetaZi's basic product is a forecast of the production from an oil or gas well which includes calibrated statistical bounds.

BetaZi runs statistical forecasts like this automatically on all of the individual wells in a lease, field, play or asset. Using correct statistical procedures, the forecasts from these individual wells can be aggregated (or "rolled up") to generate accurate statistical forecasts on groups of wells, leases, fields or plays. Calibrated type curves allow future production from undrilled wells to be added into the roll-ups. At every stage in the process, results are carefully tested using cross-validation.

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, it has used experience learned from big data 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 it requires no human intervention, no human bias is introduced.

BetaZi is big computation applied to the critical mission of predicting the production of energy from oil and gas wells. In this document, we will outline the science behind the technology and suggest how it can be applied to your critical mission, be it increasing production from better engineering, making better financial decisions, assigning values to potential deals of assets, or running quality control for bank financing.

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

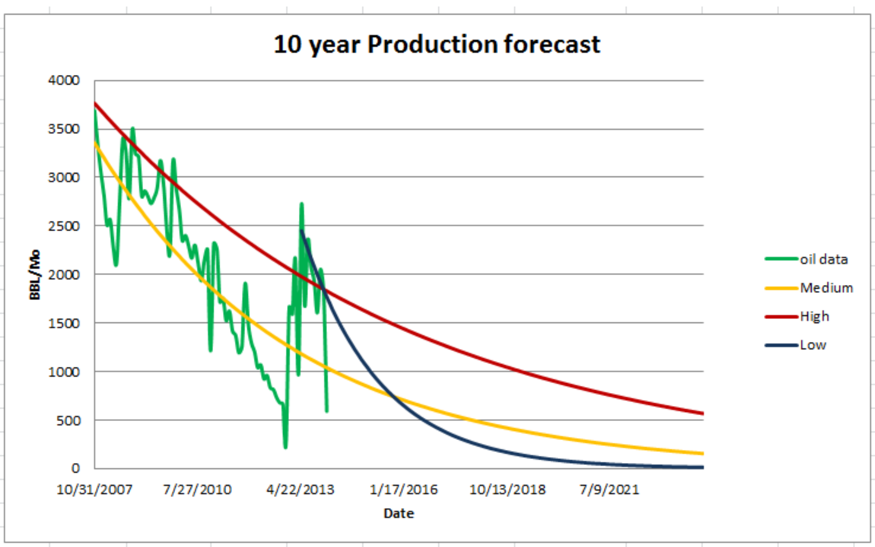

Figure 2: There are an infinite number of curves that can be used to fit a noisy dataset. In order to make a forecast, a person must pick just a few. In this case, 10 year forecasts based on the blue and red lines differ by a factor of seven.

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, shutins, 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.

Instead, BetaZi automatically finds *all* of the curves which 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 produces the equivalent of the forecasts that would be given if a million engineers were asked to give their opinions.

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 physical-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-stat model which, in the absence of data, will churn out synthetic well histories which have the same properties as production records do in real life, including operational artifacts such as shutins and artificial lift. When the model is given data, these samples are focused to explain that data.

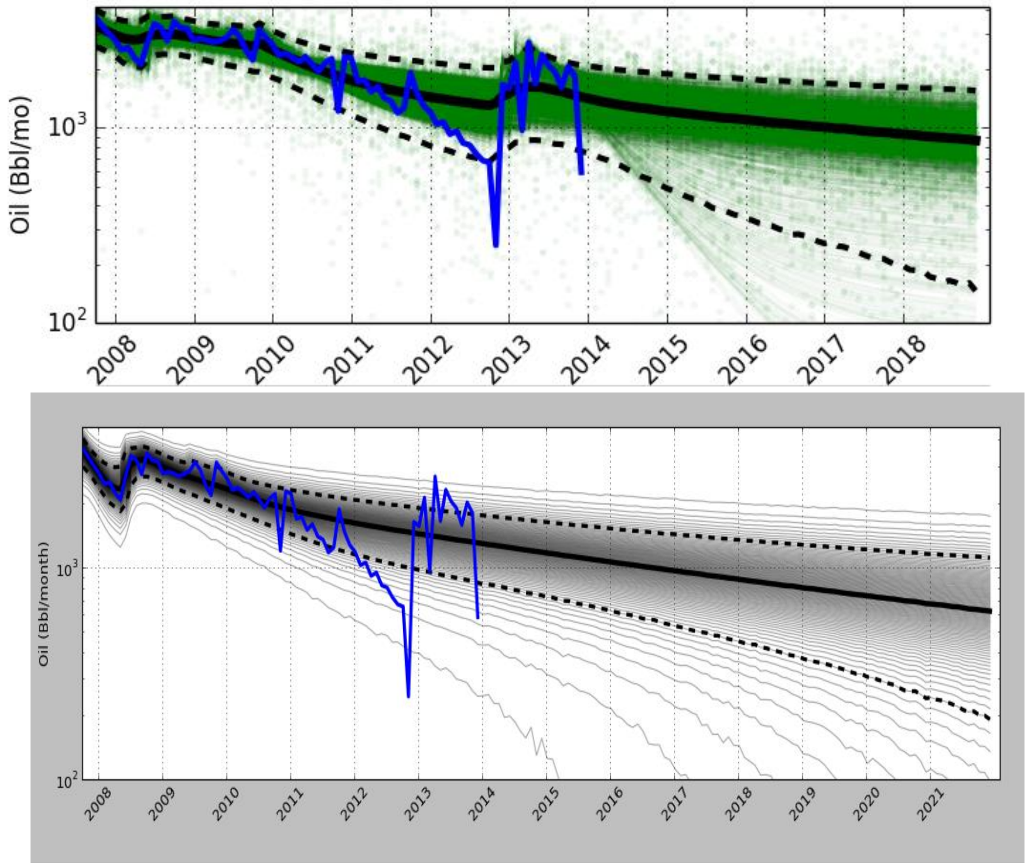

Figure 3: Top: Each green line represents a BetaZi sample. The black dotted lines are the p90 and p10. 90% of the samples fall above the p90. The heavy black line is the p50 or median of the samples. The fuzzy green dots represent noise. Bottom: The samples are abstracted into percentiles (light grey lines), again with the p90, p50 and p10 marked in bold.

Figure 3 illustrates the BetaZi samples used to explain the same data as in Figure 2. The top figure illustrates the samples themselves. Notice that there are samples which follow all three of the curves in Figure 2, including the blue curve which ignores data before the peak production in 2013. BetaZi has taken into account the possibility that the production regime might have changed in that year and the well is in for a steep decline. But most of the samples follow paths through the middle of the curve.

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 90^{th} percentile or p90 is defined as the line above which 90% of the samples fall. 80% of the samples fall about the p80 and so on.

Physio-statistics

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. 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 secret of its patented algorithm is to be able to figure out how many of these curves it needs to explain a given production history, searching through not just parameters but curves of increasing complexity. This means that a BetaZi forecast will include all of the standard models and beyond. 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 hardware, however, the generative model would not be able to simulate actual well histories without knowledge of how they tend to deviate from physics. This is the “statistics” part of physiostatistics. Using a probabilistic model learned from big data, 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 shutins, 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 *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 enormous amount of effort on calibration, in the form of back testing using blind data, to ensure that its forecasts are accurate.

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."

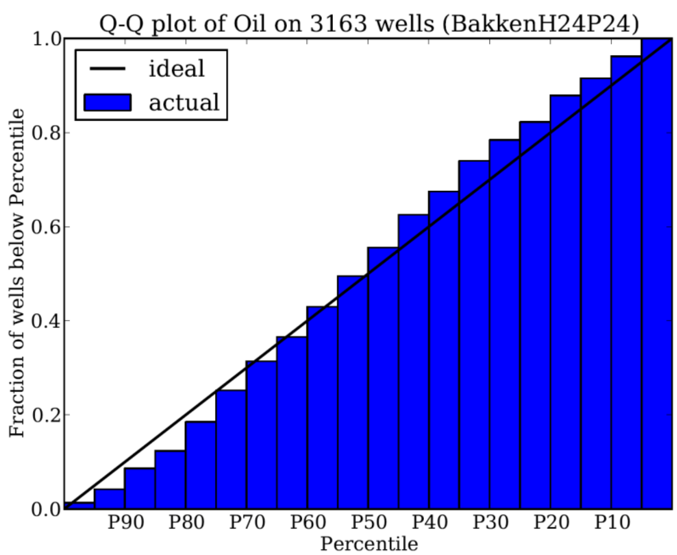

Figure 4: Quantile-Quantile plot.

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

Results

Roll-ups

Since the core BetaZi product is an automatic, calibrated forecast made on *every* well and completion in a dataset, the accuracy of results is established at a level of granularity that allows for uniquely rich roll-ups as well. As illustrated in Figure 5, statistics do not add. The p90 of a group of wells is not simply the sum of the p90 of individual wells, because multiple wells hedge each other: it is unlikely that they will all produce at their lower bound. Once BetaZi has computed individual forecasts, it can then statistically aggregate those numbers to correctly produce forecasts and bounds on groups of wells.

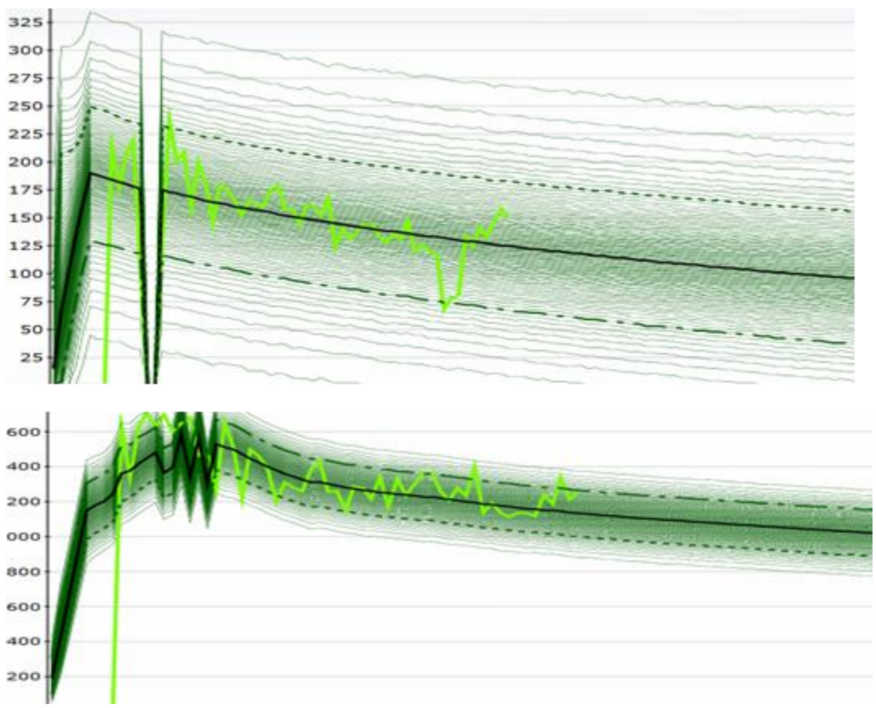

Figure 5: BetaZi rollups. Top: The ratio of the distance between the bounds of a forecast (the width) and the median forecast (p50) is about 1:1 for a single well. Bottom: for a group of 32 wells, the ratio is about 1:4.

A roll-up does not have to stop with the aggregation of producing wells. In most finance and acquisition applications, a client is not only interested in the future of Proved Developed Producing (PDP) assets, or wells that have current production. He or she is also interested in the future of Proved Developed Non-Producing (PDNP) assets (wells which were once producers and might be reopened) and Proved UnDeveloped (PUD) assets (offset wells which might be drilled or completed and are expected to have similar production to existing wells). BetaZi has rigorous methods for handling these cases and including PDNP and PUD production in the roll-up.

PDNP

Even if a well is obviously shut in with both working days and production at zero in the historical record, BetaZi will forecast production as if the well were going to be re-opened and produce at 30 (or 31) days a month[1]. Figure 6 shows a forecast on a California well which was shut in for 14 years and re-opened. This was a historical study, so the BetaZi forecast started in 2012. When a production schedule is provided with the data, PDNP will delay its forecast of the reopening to correspond with scheduled dates.

Figure 6: Forecast for a well which was shut in from 1998 through 2012. The blue indicates data that was used to make the forecast. The yellow represents actual re-opening volumes. BetaZi started forecasting new production starting in 2012.

Type Curves and PUDs

PUD production is more difficult to forecast than PDP or PDNP because it requires forecasting in the absence of any historical records. Nothing is known except how nearby wells with similar characteristics have behaved in the past. If a client provides information about which nearby wells that have similar geology and operational characteristics, BetaZi will automatically compute a statistical type curve to be used to estimate PUD production.

[1] BetaZi always assumes that future production will be for a full month every month. This is not fixed, however, and a client may request that production be forecast at a lesser rate of operating days.

The standard method of forecasting without data is to use a type curve. Typical type curves are made by normalizing the production histories from nearby wells, averaging them together and fitting a curve to the average. BetaZi takes this logic a leap forward. Instead of averaging, we find a statistical model for well production which is based on the physical part of our generative model. This results in not just a single curve, but a set of p-values which accurately represent the probability of future production for the PUD asset.

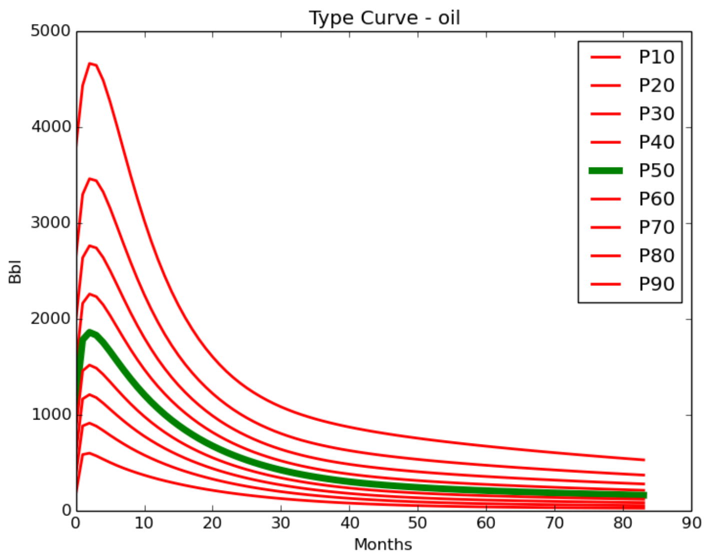

Figure 7: Example of a BetaZi type curve generated using 42 vertical wells producing from the same formation.

Figure 7 shows an example of a BetaZi type curve. Instead of a single, normalized decline curve, the BetaZi type curve gives all of the p-values of a normal forecast. In this case, it also shows that there is typically a period before a well reaches its peak production after opening.

The critical factor in being able to trust a type curve is back-testing. BetaZi type curves are tested using a process of 10-fold cross-validation. 10% of the data which is available to make the curve is withheld for testing via a Q-Q plot [see Figure 8]. The process is repeated until all of the data has been used for both making a type curve and for testing. Cross-validation is a standard technique used by computer scientists to establish the ability of an algorithm to *generalize*, that is, to predict data that it has not seen before.

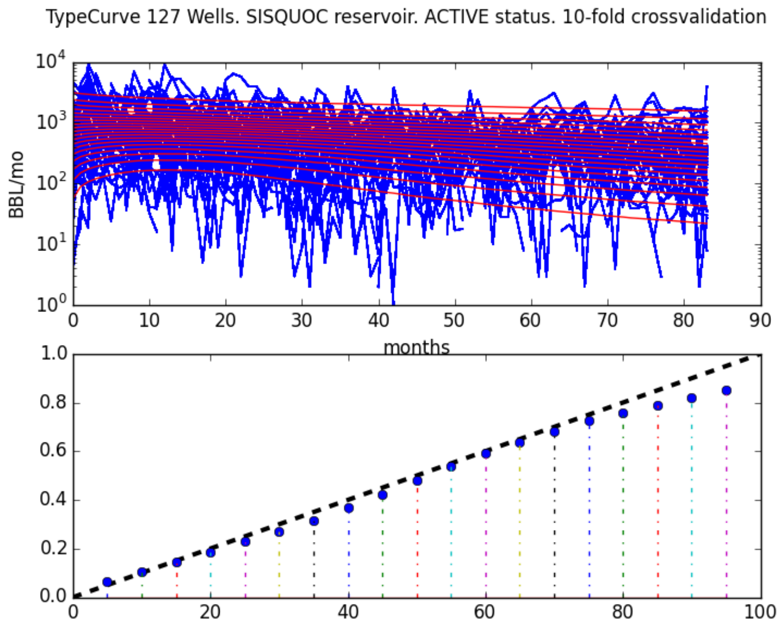

Figure 8: A different type curve is set against a background of production from wells which were not used to make it. A q-q plot (like this one) based on the data that was held out is used to validate it.

Future volumes and cumulatives

The final results of a BetaZi run include forecasts on every individual well in the run. If a client has provided suitable information, the results also include: a set of PDNP predictions; a set of type curves; and a set of rollups forecasting production on existing production, scheduled reopenings, and scheduled new completions.

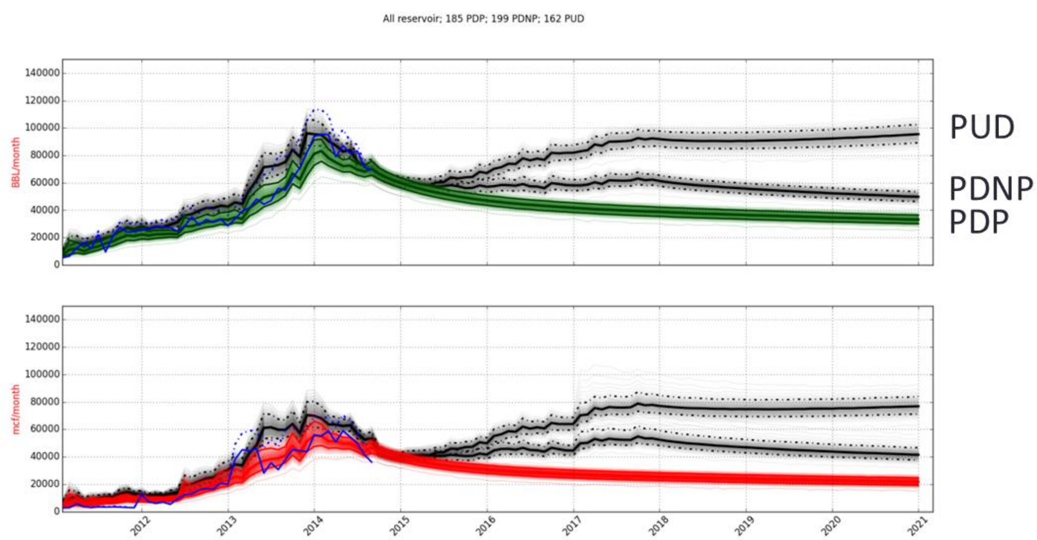

Figure 9: Example of a full BetaZi rollup including PDP, PDNP and PUD production.

It is only a short step between the volume forecasts of previous figures to computing future cumulative, future discounted cumulative volumes, Estimated Ultimate Recovery (EUR) or future discounted cash flows, as is shown in Figure 10.

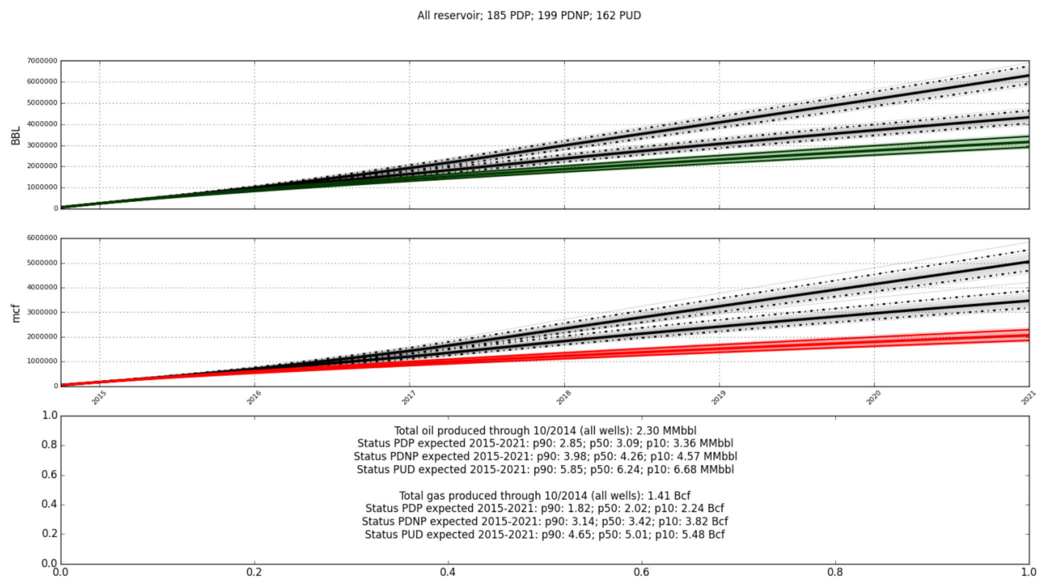

Figure 10: The volume forecasts in Figure 7 can be uses as the basis for computing future cumulatives.

To facilitate economic modeling, BetaZi forecasts and rollups can be exported into standard modeling packages such as PEEP and AIRES to form starting point for rigorous financial analysis.