Solving the "Enigma" of Reserve Growth

Posted by Sam Foucher on July 16, 2008 - 10:00am
Topic: Supply/Production
Tags: reserves growth [list all tags]

14
diggs

digg it

This is a guest post from WebHubbleTelescope.

As a good rule of thumb, when you have a promising model describing some physical process, you might as well put it through its paces. Not only do you shake out some stubborn corner cases, but you often find something new and revealing. We did that in the last post, The Derivation of "Logistic-shaped" Discovery., grinding out the derivation of the classic Logistic/Sigmoid-shaped Hubbert curve based on the generalized Dispersive Discovery model. With this post, I use the same discovery model to derive the upward climb of the ***cumulative reserve growth curve*** which we empirically observe on many oil reservoirs and oil-bearing regions.

Update: *WebHubbleTelescope has posted an updated version of this post here.*

Many analysts have found this reserve growth behavior both curious and ultimately very important. I know that Khebab and Rembrandt have spent much time on TOD tracking this behavior as it plays an important role in how the peak will play out. Furthermore, I believe that the practice of "back-dating" discoveries based on reserve growth updates has muddied the waters and stalled progress in the basic understanding of the fundamental growth process. Unless we have a good model for the reserve growth dynamics we have to resort to using the heuristics supplied by USGS geologists, including the *modified Arrington* equation that Khebab has successfully used in the past. I find nothing wrong with using a heuristic and Khebab has really kick-started the "un-back-dating" approach with some excellent results. Still, a heuristic lacks some of the predictive power and room for insight that a fundamental model can provide.

The USGS crew have an interesting take on the reserve growth issue. In turns, the geologists working there have labeled fossil fuel (both oil and NG) reserve growth an "enigma" and a "puzzle".

> For that reason the United States Geological Survey (USGS) considers [this] analysis "arguably the most significant research problem in the field of hydrocarbon resources assessment."

I have to admit that it has puzzled me for a while, but then again most of us here don't work at the USGS for 40-hours a week. This post runs though a stochastic analysis that essentially

explains how reserve growth can happen.

One of the technical issues that the USGS have struggled with respect to their reserve growth analysis involves the use of "censored data". This essentially says that you should take special care of extrapolating data backwards considering you have only a truncated time-series data set of recent vintage. The "sweet spot" for good data basically doesn't exist, with very few values for old data and also very few values for the latest data. Working with such a limited data set, they massaged it the best they could and adequately normalized the fractional yearly growth, but they basically punted after this point and came up with only a heuristic to "explain" the trend. Pragmatically, I've decided to sweep the censoring problems under the rug, since in the long term, complaining about the possibility of statistical shortcomings still won't explain most of the trend, which rises steeply enough to make the cornucopians hopeful for great prospects ahead (listen to any right-wing radio program today for examples of this kind of unbridled optimism).

> *"We must accept finite disappointment, but we must never lose infinite hope."* -- Martin Luther King Jr.
>
> *"Reserving judgments is a matter of infinite hope."* -- F. Scott Fitzgerald in **The Great Gatsby**

Yet we still have left a riddle wrapped in a mystery inside an enigma**.** The actual problem with the half-baked reserve growth analysis has become obscured by the trickiness with using censored data. Wrapped inside, it really stems from a lack of a good value for the initial discovery estimate. Stating it bluntly, pick this number incorrectly and you can get numbers all over the map, with the possibility for some hugely absurd values.

To provide a path forward in unwrapping the riddle, I used the generalized Dispersive Discovery Model. In terms of modeling reserve growth, the dispersion generates a tail for accumulating further discoveries after the initial estimate occurs. For constant average growth, the model looks like this:

$$DD(t) = 1 / (1/L + 1/kt)$$

For the purposes of this analysis, I convert it to a reserve growth value `U(t)` and set `T=L/k` to make the math easier to handle later on:

$$U(t) = tT / (t+T)$$

Note that at time t=0, the discovered amount starts at zero and then the accumulation reaches some value proportional to L -- what one should consider as the characteristic depth or volume of the reservoir. It takes time=T for most of the search to reach this point. The basic premise of reserve growth and what USGS geologists such as Attanasi & Root[1] and Verma[2] frame their arguments on, has to do with the reserve growth considered as a *multiplicative factor of the initial estimate*. They see numbers that reach a value of nearly 10x after 90 years and claim (perhaps implicitly) that this has some real physical significance, almost offering up hope for still-to-come huge reserve benefits. Figure 1 exaggerates the claim for effect, as I want to make you aware that a finite asymptote certainly exists, but it gets obscured by the data trends commonly
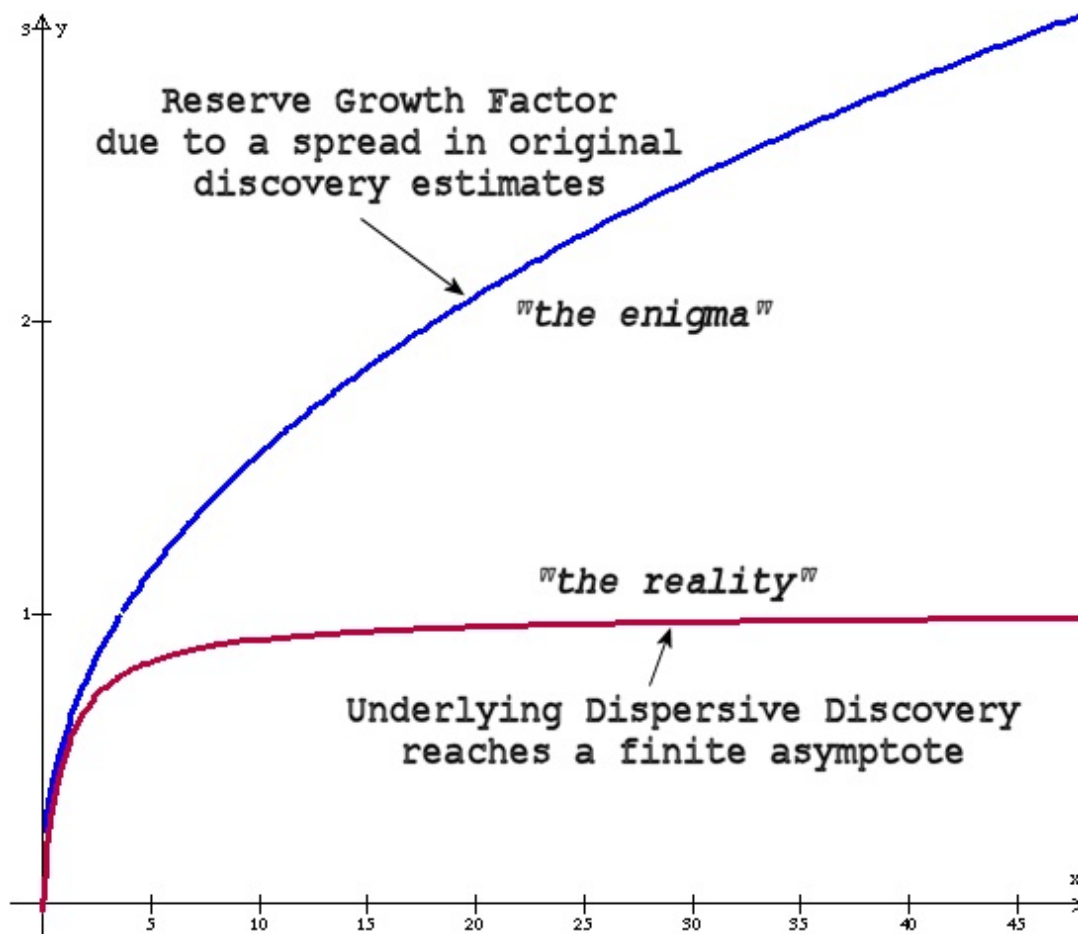
reported.

Figure 1: Potential for huge reserve growth

The contrived swindle of the USGS explanation has to do with exactly *when* the original estimate becomes available. Conceivably you can make estimates that occur very early in the lifespan of a reservoir, and you will get very low estimates for estimated discovery size. To take it to one extreme, you might find the initial estimate to fill a *sewing thimble*. Now, if that estimate grows at all, you can get huge apparent reserve growth factors, some fraction approaching infinite in fact. In contrast, you can wait a couple of years and then report the data. The later years' growth factor will proportionately account for much less of an increase. Now if you consider that in other parts of the world, countries report reserves less conservatively then the USA, then the reserve growth factors can vary even more wildly. I used USA oil data from an Attanasi & Root paper[1] which you can find a dump of here. Initially, I plotted the data as a fractional yearly growth curve, basically reproducing the trend that A&R report:

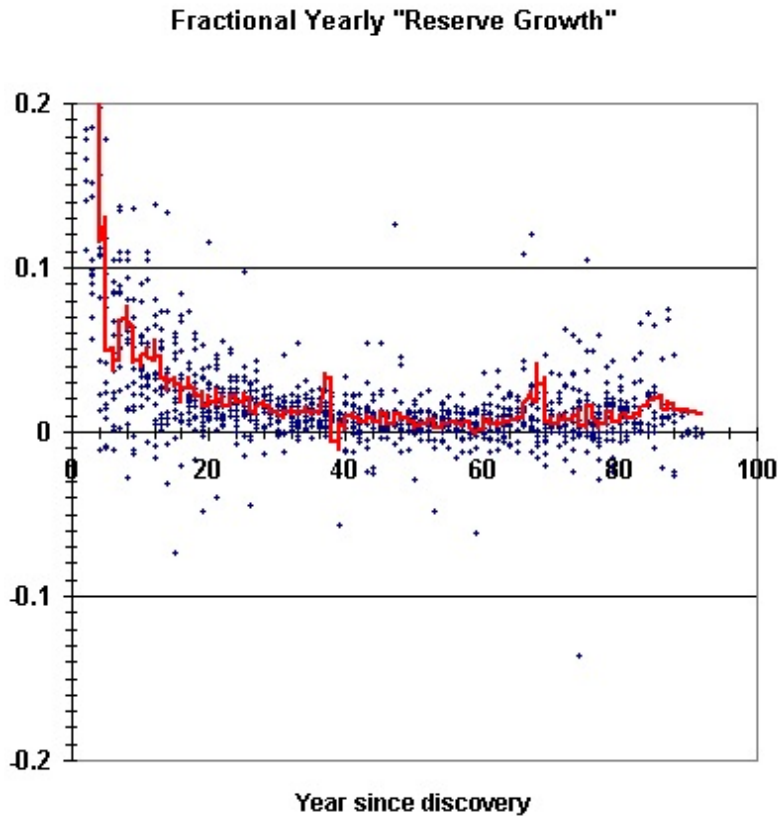**Fractional Yearly "Reserve Growth"**

Figure 2: Trend duplicated from the A&R data. The blue '+' symbols represent the data and the red curve represents a small-window moving average. To generate the curve correctly, you must normalize the yearly growth *per reservoir*, as the reservoir sizes vary widely.

The key insight to understand the growth factor in terms of the DD model has to do with averaging the initial discovery point over a relatively small window of time starting from **t=0**. This effectively samples the infinite values of growth against other finite values. The use of the sampling/integration window brings down the potentially infinite (or at least very large) growth factor to something more realistic. The math on this derives easily into an analytic form, and we end up with this function, where **A** indicates the time integration window:

$$U\_bar\ (t) = (A - T \ln((t+A+T)/(t+T)))\ /\ (A - T \ln((A+T)/T))$$

The term on the right of the main divisor assures that the average **U_bar** starts at 1 for time **t=0**. Alternatively we can set **A** to some arbitrary value and skip the integration, which assumes that all initial discovery estimates start at **t=A/2**. This results in the conceptually simpler:

$$U\_delta\ (t) = 2*(t+A/2)*(A/2+T)/A/(t+A/2+T)$$

The latter equation obviously starts at 1 and reaches an asymptotic value of **1+2T/A**. Both the values of **A** and **T** describe the ultimate asymptote, but the non-zero **A** serves to avoid generating a singularity at the origin. For **T= 24.6** and **A=6.6**, the figure to the right shows the negligible differences between using an integration

window versus assuming a delta shift in the first
estimate. The two curves essentially sit very close to one
another.

So what do these numbers mean? Essentially, **A=6.6**
means that the first discovery estimate on occurs *on
average* **3.3** years after they first made the discovery.
This makes intuitive sense because if we make the
estimate too early, we end up with the equivalent of a thimble-full of oil. For **T=24.6**, this means
that it takes about 25 years for the majority (i.e. the fast part) of the dispersive search to take
place. The rest of the long tail results from the slower dispersion. The curve does eventually reach
the asymptote for a cumulative growth factor of **8.5**.

In terms of a spreadsheet, you can turn the CGF formula into a discrete generating function, with
the yearly estimates based on the growth factor of the preceding year. I plotted the curve directly
against the A&R data in Figure 3 and Figure 4. After the hairs on the back of my neck settled
down, I realized that this simple formula has some nice understandable properties. It essentially
generates growth factors based solely on the maximum entropy dispersion in the underlying
model. In other words, the "enigmatic" reserve growth has turned from a puzzle into a
mathematical result resulting solely from simple stochastic effects.
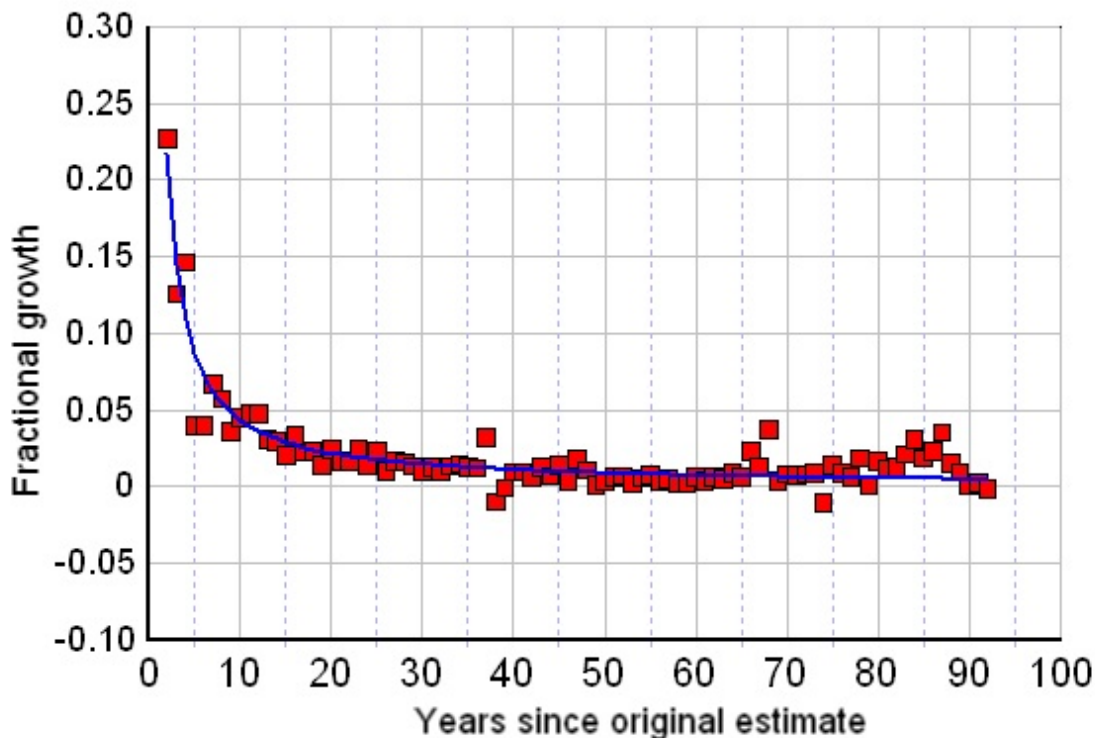
Figure 3: Fit to fractional yearly growth of the A&R data

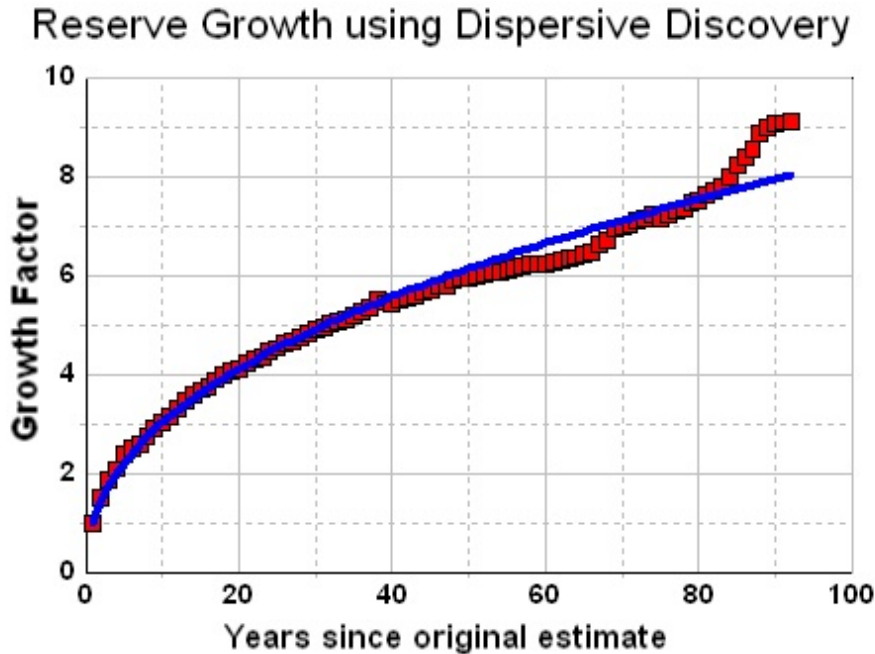## Reserve Growth using Dispersive Discovery

Figure 4: Fit to cumulative reserve growth of the A&R data. This was calculated from the generating function in Figure 3. This correlates well with the direct use of the reserve growth equation.

Plotting the same data against an Attanasi & Root chart in Figure 5, it lays cleanly on top of it, showing discrepancies only on some very old outlier data. A&R went through the rationale of discounting the outliers, but I consider all the data valuable, *if it gets used in the context of a decent model.*
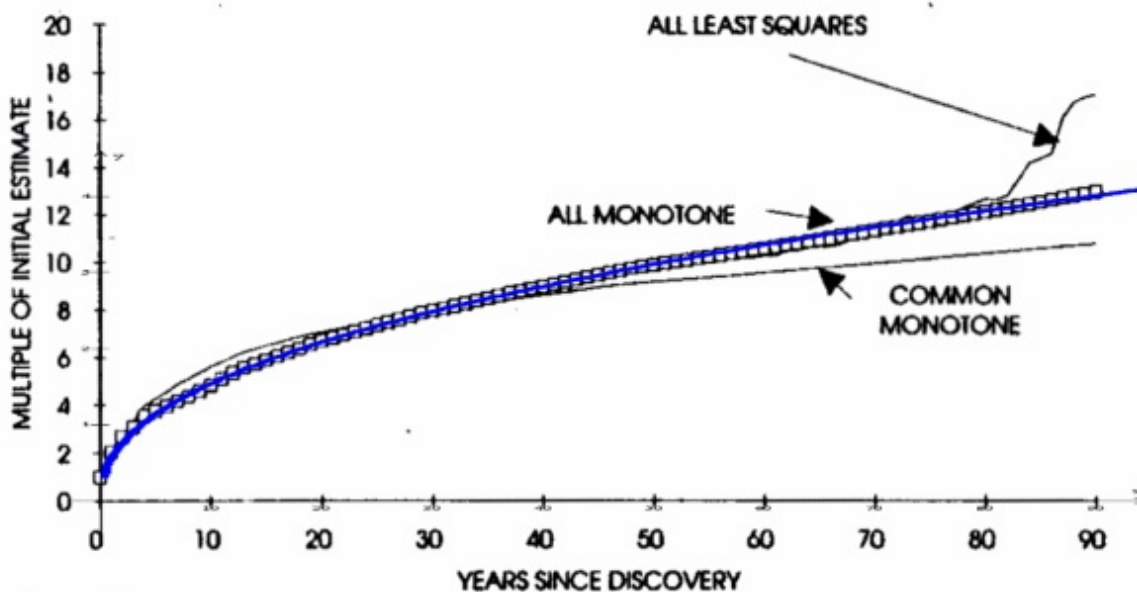

Figure 5: Fit on top of the original A&R chart duplicated from Reference 1. The blue line comes from the Dispersive Discovery reserve growth model.

Interestingly, with the limited data available, the reserve growth looks like it will continue on and eventually reach infinite values, but this becomes a mathematical impossibility if we integrate out the "thimble"-sized initial estimates. As a bottom-line, if we continue to make poor initial estimates for discoveries, we will continue to pay the price for acting surprised at the "huge"

reserve growth we have. In other words, the USGS has pawned off a contrivance on us by making us believe that their trend lines went beyond mere empiricism. If you look back at the previous "model" that the USGS's Verma postulated[2] based on the "modified Arrington" approach, you will realize that the trend line comes about purely from heuristic considerations. Their equation shows a fractional order power-law growth:

$$CGF = 1.7378(YSD)^{0.3152}$$

To top it off, this heuristic shows **unlimited growth**! Do you suppose that the USGS geologists figured out that by reserving judgment on the possibility of an asymptote, that they could pawn off a fast one on the public, and defer the reality for years to come?

You have to ask yourself how these professionals get away with publishing stuff based on hacking and punting, that really has such a simple statistical and mathematical underlying foundation. I really find nothing complicated about the mathematics (even though it has taken me some time to arrive at my current state of understanding). I call this combination of using simple models and using straightforward calculus and probabilities a form of *pragmathematics* -- just something you do to understand the physical foundation for the data we observe.

Referring again to the literature, you find hints that support the dispersive effects of accumulated reservoir estimates (note that they use the term dispersion).

> A graphic illustration of the very broad URA data **dispersion** that occurs when grouping fields across geologic types and geographic areas was provided by the National Petroleum Council (NPC) and is reproduced with minor modification in Figure FE5.

Figure FE5. Observed Growth Factors and URA Model Projections for the NPC Sample Fields
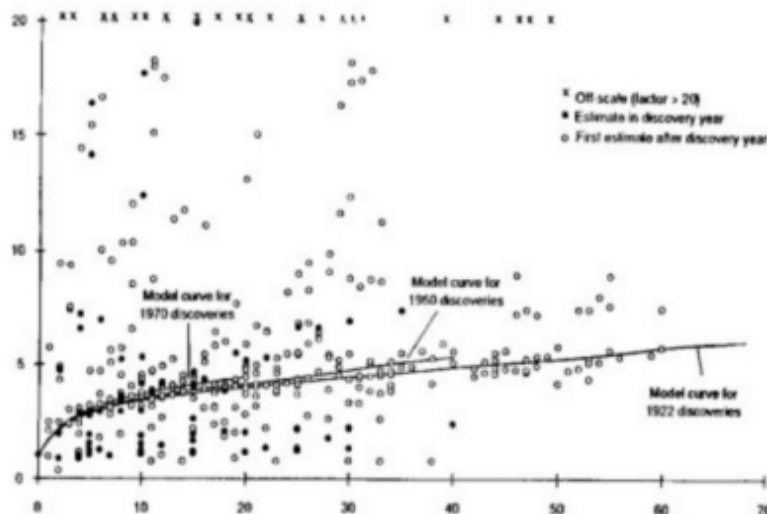
Source: Energy Information Administration, Office of Oil and Gas. Derived from National Petroleum Council, "Report of the Reserves Appreciation Subgroup fo the Source and Supply Task Group, 1992 National Petroleum Council Natural Gas Study" (Washington, DC, August 1992), unpublished open file text, Figure 14.

I claim that this new dispersive discovery reserve growth model has the potential for filling in the back-dated discovery curves and providing better estimates of future production levels. I would recommend starting with the USA data and using Khebab's approach for regenerating a profile of un-back-dated discoveries and then applying the Oil Shock Model to estimate extrapolated production levels. It may take some time to sort out the mess, but we likely have all the pieces

necessary to formulate a complete model-based projection.

The actual enigma of reserve growth I think has to do with the cluelessness of the USGS and the secrecy and inscrutability of the oil industry. You would think they would have figured out the reserve growth puzzle long ago. I guess they thought that deferring the reality would surely provide us with infinite hope.

The "How Did We Get Ourselves Into This Mess?" Epilogue

A fraction of the population thinks that oil production remains a simple matter of turning the spigot clockwise to get more oil, and reserve growth a non-issue. And that raising the current reserve inventory becomes a conservative decision made entirely under the control of humans.

Overestimating the size of a reservoir has significant financial advantages. You can attract huge initial investment capital if you exaggerate a claim. But this can also attract charlatans. So I agree with the SEC's decision to put the stops on that practice, with big fines for fraudulent claims. And thus the oil companies tried to make their reserve estimates as high as they could without overdoing it. But they still had no good ideas on how best to characterize their estimates (not smart? who knows?). Some people suggest that the industry just makes "conservative" decisions. Yet, I don't think that ambiguously-framed "conservative" decisions matches with well with what you can accomplish with a good reserve growth model. I would describe the industry's approach not "conservative" but "non-predictive" and "safe". I can point to lots of other engineering areas where through the routine process of characterization one can make equally conservative estimates, but they do an excellent job of predicting the asymptotic behavior with safety margins for any errors or noise that creeps in.

As an example Andy Grove was the first to characterize silicon dioxide growth by evaluating simple models and thus began Intel. Getting oxide thickness correct could follow a process where someone checks the thickness every two minutes like they were watching a loaf of bread rise, but the technology and modeling has gone way beyond that now in Silicon Valley and China. Not so with the oil industry however; to the untrained eye it looks like they use the primitive trick of plunging a thermometer into the turkey roasting in the oven. This is safe but, jeez, how primitive!

Bottom line, I believe the current industry estimates for reserve growth remain just as safe and non-predictive as a turkey thermometer, but that the underlying reality (having to do with the dispersion of searches through the volume of a reservoir) can vastly improve our educated guesses. This makes it potentially predictive if the industry had the balls to use a good validated model. The simple model I use for reserve growth is complementary to what Andy Grove did in the 60's. Unfortunately, I also don't think that petroleum engineers have any control over marketing decisions. It's possible that oil industry engineers knew about the characterization outlined here all along but were superceded by management's decisions. They used it in their own internal books and presented a "safe" facade for the outside world.

Unfortunately, we turned the clock to a new century, and some nobodies on TOD have finally figured out the obvious that had stared at them and the USGS in the face for years.

Journal References:

[1] Attanasi, E.D., and Root, D.H., 1994, The enigma of oil and gas field growth: American Association of Petroleum Geologists Bulletin, v. 78, no. 3, p. 321-332.

[2] Verma, M.K., 2003, Modified Arrington Method for Calculating Reserve Growth — A New Model for United States Oil and Gas Fields, U.S. Department of the Interior, U.S. Geological Survey

Web References:

1. http://mobjectivist.blogspot.com
2. Finding Needles in a Haystack
3. Application of the Dispersive Discovery Model
4. The Shock Model (A Review) : Part I
5. The Shock Model : Part II
6. The Derivation of "Logistic-shaped" Discovery