



## **Extrapolating World Production**

Posted by Stuart Staniford on January 24, 2006 - 5:18am Topic: Supply/Production Tags: hubbert linearization, hubbert peak, oil prices, peak oil [list all tags]

2.0

This piece takes on how to model and extrapolate the world production curve. It is long and a bit complex. However, I think it's worth the effort, because there's some absolutely fascinating stuff going on in the world production curve.

We will try to build the graph to the right in a series of easy stages. It shows average annual oil production on a semilog plot with a variety of models fit to the data and extrapolated. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Data sources: API, ASPO, and BP.

Oh, and for you inpatient ones, the linearization stability analysis says the world

- URR is **2250 ± 260gb**
- K is 4.93 ± 0.32%
- the logistic peak is May 2007 ± 4.5 years

Those are intended as two-sigma error bars. However, see below the fold for more important caveats before taking

1.5 1.0 0.5 Production 0.0 1920 1860 1900 1940 1960 1980 2000 2020 Log (base 10) of Daily -0.5 -1.0 -1.5 -2.0 -2.5 -3.0 1860-1891 1891-1929 1929-1942 1942-1973 - 1973-1979 1979-1983 1983-2004 Gaussian Logistic Linear (1860-1891) Logistic-High Logistic-Low Linear (1891-1929) Linear (1929-1942) -Linear (1942-1973)

Linear (1979-1983)

this away - there is still some residual uncertainty about the validity of this extrapolation in my mind. Finally, the most pessimistic model doesn't reach sustained decline rates of 5% until 2038, so <u>peak is still a slow squeeze</u>.

Linear (1973-1979)

Firstly, the production data. I have finally managed to get pretty complete data for historical world production. The graph below has data from four sources. Jean Laharrere very kindly emailed me data from the American Petroleum Institute which for the global series runs from the start of production in Romania in 1857 through 1993. I believe this series is crude only. Then I have the data which appears to be what ASPO uses as historical data in their plots. I obtained this by the slightly unorthodox method of ripping it out of the data table for a chart in this <u>Powerpoint</u>

Linear (1983-2004)

presentation from Colin Campbell. That series runs 1930-2001. I don't know where ASPO ultimately got it from. Then I have more conventionally obtained <u>EIA Table 4.4</u> and <u>BP</u> data from 1970-2004, and 1965-2004 respectively. If we plot them all, we get:



Average annual oil production from various estimates. Click to enlarge. Believed to be all liquids, except API line is crude only. EIA line includes refinery gains, others do not. Sources: <u>ASPO</u>, <u>BP</u>, and <u>EIA</u>.

The discrepancies are likely due to different definitions of exactly what is being measured. The API data I believe does not include NGLs, which all the others do. The EIA includes refinery gains, which the others do not. The ASPO and BP data agree quite well in their overlap (though not exactly).

Now, we recently discussed the fact that the US production curve is <u>beautifully fitted by a</u> <u>Gaussian mathematical model</u>. That allowed for some <u>fairly stable extrapolations</u>. Also, the similarity of the logistic and Gaussian curve, except in the extreme tails, explained why the logistic also works ok in the US case. In the world case, even a cursory glance at the data indicate we are not going to be in quite such happy curve-fitting territory (though it's not nearly as sketchy as our <u>quick exploration of Kuwaiti production</u>). We will get to more modeling in just a second, but first let's look at the data a different way.

This graph shows the percentage change from one year to the next.



Percentage change in average annual oil production from one year to the next according to various estimates. Click to enlarge. Believed to be all liquids, except API line is crude only. EIA line includes refinery gains, others do not. Sources: <u>ASPO</u>, <u>BP</u>, and <u>EIA</u>.

You can see that, on the whole, growth rates have been steadily declining, though not smoothly so. Also, the noisiness of the curve is decreasing a lot: as <u>WebHubbleTelescope noted</u> in a slightly different context a little while back, this is significant. It's decreasing because we don't have lots of big new fields coming on to cause wild increases in production (and corresponding gluts, price collapses, and shut-ins). Of course, the point where the growth rates cross the x-axis and become negative is peak oil.

In order to have a single sequence to model, I proceeded to combine the production sequences as follows. Before 1930, I only have the API series, so I use that. From 1930-1964 I use the ASPO series. Then from 1965-2001, I use the average of the BP and ASPO values. Finally, from 2002-2004, I use the BP series. So this approximately models all liquids without refinery gains. The rest of this post all takes that combined series.

I was next moved to plot the data on a semilog plot. Partly this is because semilog plots make a gaussian curve into a quadratic (which <u>fits the US production curve beautifully</u>), and partly it was because of intriguing things I noticed in the growth graph. So anyway, here's just the data, before I start leading your eye with models of it. The x-axis is just the year from 1860 to 2010. The y-axis is logarithm to base 10 of daily production in millions of barrels. So "0.0" corresponds to 10mbpd, "1.0" corresponds to 10mbpd, "1.5" to 31.6mbpd, "2.0" to 100mbpd and so on.



Log (base 10) of average annual oil production from various estimates. Click to enlarge. Believed to be all liquids excluding refinery gains. Sources: API, <u>ASPO</u>, and <u>BP</u>.

So, is it just my eye, or does that look to you like a sequence of straight lines - with a little noise on top? Of course, a straight line on a semilog plot corresponds to true exponential growth in the production versus time graph (ie a constant growth rate).

Well, now I will lead your eye. Here's how it breaks down visually to me:



Average annual oil production on a semilog plot with piecewise exponential fit. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Data sources: API, ASPO, and BP.

I've color coded each time range that strikes me as a straight line, and added a linear fit (which would be an exponential in the underlying production graph). Seems to do quite well at describing the data, yes?

Not to say thats it's completely uniquely specified, or perfect. It seems to get better over time, and in particular, there's more than one way to handle the first world-war and early twenties. However, I still think it's capturing some interesting features of the data. Here's a table that give the average growth rates during each of those intervals (as computed from the slopes of the linear fits).

1860-1891	13.9%
1891-1929	7.9%
1929-1942	3.9%
1942-1973	7.4%

1973-1979	2.1%
1979-1983	-4.0%
1983-2004	1.5%

Many of these dates have some economic or political significance (though I didn't pick them that way - I was just looking at the data). I didn't know of any significance to the early 1890s, but it turns out there were massive droughts in the US in that timeframe, and then monetary problems, culminating in a financial panic in 1893. 1929 is of course the end of the roaring twenties and beginning of the great depression. That seems to have led to lower rate of growth in world oil production, which ends in 1942 as the US, much the world's largest producer at the time, enters the second world war. From then on, world growth in production is higher until the 1973 oil shock. Between the shocks, growth in oil production is lower but still positive (except for 1974). However, after the shocks, there is a brief period of declining oil usage until 1983, when we enter the region of slow growth in oil usage that lasted until very recently. I suspect that we are now on the threshold of a new era, one way or another.

While this model is very interesting descriptively, it obviously has very limited predictive power by itself, since it only asserts "oil production is piecewise exponential", but not how to predict the date or slope of the next link in that chain of exponential pieces.

So, let's add in the quadratic associated with a Gaussian peak in the production versus time graph. This next graph fits a quadratic to the entire data series from 1860-2004 (the black curve). Voila:



Average annual oil production on a semilog plot with quadratic (ie Gaussian) fit in addition to piecewise exponentials. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

I have to say this graph nearly blew my mind. Firstly, as in the US, the Gaussian does a surprisingly good job of fitting an enormous range of data (through four orders of magnitude change in the volume of production). But more significant is the nature of the noise. I don't think we have an adequate understanding yet of why the US curve is so Gaussian, still less this one. However, my speculation, following an idea of Khebab's, would have been that it's Gaussian because if you convolute a messy discovery curve with enough different and long lived development and production and decline processes, you end up with a Gaussian. If that was true, you'd expect the leading source of noise about the Gaussian to be lumpiness due to individual large fields being discovered and developed (eg Prudhoe Bay). I didn't check this in the US case, but I was vaguely assuming that to be likely.

However, that clearly can't be what's going on here. Here the leading source of noise about the Gaussians is the different trends in the growth rate (the colored straight lines that cross back and forth over the black line). They have "economics" written all over them, not developments of

The Oil Drum | Extrapolating World Production

http://www.theoildrum.com/story/2006/1/22/04219/1102

particular fields. Eg, where's the Ghawar bump (production started in 1951)? Or even the Middle East bump? They're just not there; they are buried in that beautiful straight line from 1942 to 1973.

In short, there is no clear visible evidence of the discovery curve showing through into production any time recently. And yet, there's that quadratic fitting the thing from end to end. Why?

The picture that emerges - and I stress that this is a highly speculative story at this point - is that there is some kind of random-exploration-through-oil-space reason for why production curves in large regions tend to be Gaussian. However, the economy tends to come to relatively long-lived cultural agreements about how fast oil usage should grow, which control the approximate rate of growth (with considerable yearly noise). These cultural agreements last until they get too out of whack with what is possible based on geological/exploration considerations (eg perhaps in 1973), or the economy goes through some kind of trauma which changes the expectation (eg the 1979 oil shock, or the 1942 need to mobilize production for the war effort).

Again, don't bank on that last paragraph - further work is required to substantiate that, or to suggest a better way of looking at the situation. But it's certainly an interesting working hypothesis. (I should also mention that it's worth looking at WebHubbleTelescope's <u>oilshock</u> <u>model</u> for additional backround.)

However, in extrapolating forward to do prediction, we found in the US case that the Gaussian does a bad job before the peak - it's <u>predictions are unstable</u>. The problem is not that the Gaussian does not fit the data well. Instead the problem appears to be that there are too many different Gaussians that fit the data about equally well, and they have a broad range of implications going forward. Until there is post peak data, therefore, the Gaussian projection is not well-constrained and it jerks around violently depending on exactly what data range is used for the fit.

For this reason, we return to our old workhorse of Hubbert linearization, which was the most stable method pre-peak on the US data (I described the basic rationale for this method, such as it is, <u>several months ago</u>.)

What I have done in the following graph is to keep the same color codings of regions as I used in the semilog plot, so we can see what we are linearizing.



Hubbert linearization of global oil production. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Color coding of data ranges matches the previous plots. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

In his book *Beyond Oil*, Deffeyes picks 1983 to start his linearization, and it's obvious why. If we do the same, we get the fit shown above, with a URR of 2250gb, and a K of 4.93%. If we return to the production domain and insist on a peak date that means the logistic cumulative production by the end of 2004 matched the **actual** production to the end of 2004 (1059gb in my composites series), we end up with a smooth peak in May 2007.

Professor Deffeyes was obviously fitting to a different series -- mostly likely crude alone, without NGLs. He doesn't cite his data source, so we don't know.

Before I launch into a stability analysis, I want to highlight one of the big caveats here in light of what we discovered above. The answer we got obviously depends on the fact that we fit to the bright green 1983-2004 region, which is the last of the pieces in the piecewise exponential model I showed above on the semilog plots. If that piece had had a different slope, we obviously would have got a different answer. Do we know if, in an alternative universe with the same geology but different economic history after 1860, that particular segment of the curve must have had that

The Oil Drum | Extrapolating World Production

slope? I'm not sure we do - our most compelling argument right now really comes down to "linearization has worked elsewhere" (which is true, but not 100% satisfying). And if we don't know that the slope must be what it is, then the extrapolation is more uncertain than the following stability analysis would suggest.

In particular, had we decided to linearize in 1983, we would have got the wrong answer based off that steep approximately linear region from 1973-1983 (the purple and orange regions in the graphs above). Just for kicks, let's combine those regions and do that bad linearization:



Hubbert linearization of global oil production (warning - this is a bad extrapolation for illustration purposes only. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

So we'd have thought that URR was 860gb, K was around 10.5%, so bad declines on the way, and that we were at 63% of the URR (what WesTexas calls  $Q_t$ ). How would we have known this was garbage? It doesn't look any worse than what we just did with Kuwait!

Well, if we'd looked at the whole curve, maybe we'd have said, "Oh, it looks like there's some risk we're fitting a noise feature, instead of something that matches the trend, we'd better wait and see." On the other hand, maybe we'd have said, "Gaussian's are not reliable before the peak anyway, let's just trust the linearization."



Semilog plot illustrating the situation in 1983 with the bad Hubbert linearization of global oil production (warning - this is for illustration purposes only). The bad linearization is based on the purple data. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

My sense is that the last two decades represent a much better basis to extrapolate than that decade, but still, bear in mind as we go into the stability analysis that it fundamentally assumes that extrapolating the 1983 onward linear region is a valid thing to do, and I think some sniff of question-mark should still attach to that until we have a stronger theoretical understanding than we do at present.

I suspect that we really need evidence of peak from outside the linearization itself to tell us whether it is likely to be valid or not. For example, if almost all known discoveries are in production, oil is over \$60, and the price is still going up, that might be suggestive:-)

There's another big caveat coming, but let's do the stability analysis. First is the value of K (which is the intercept of the y-axis in the linearization). K controls the width of the peak in the logistic curve, and also the speed of growth in the past and declines in the future (which asymptotically approach K in both cases, though with opposite sign).

If we start varying the start and end date of our linearization and see how K changes, we get the following surface:



Start year

Stability surface for K in the Hubbert linearization of global oil production as a function of the start and end year of fitting. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

So the stability surface is that small flat area in the front center of the picture, with start dates between 1983 and about 1990, and end dates from 1994 on. Not exactly the broad plains of the <u>US stability analysis</u>, but perhaps we can set up a small farm there on the side of the mountain. Obviously, the linear region only begins in 1983, so if we start before that our K estimate climbs rapidly to the "bad" linearization with K=10.5%. On the other hand, if we work with less than a decade, we get into major problems of fitting the noise instead of the trend.

So this is a density plot just of the actual portion I used for the stability estimate:



Density plot of stability surface for K in the Hubbert linearization of global oil production. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Scale is 0% (blue) to 10% (red), with 5% being green. Contours are 0.1% apart. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

The blue triangle at the bottom is excluded because it corresponds to places where the linear fit is less than a decade in length. If we look at the sample deviation in the K estimate over this region, we find it that it is 0.16 percentage points, which I doubled to give the 0.32% error bar in the outset.

Turning now to the estimate of ultimate recovery, here's the surface: There's more fluctuation in the URR (because that long extrapolation forward can change it's intercept with only modest changes in slope).



Stability surface for URR = ultimately recovered reserve in the Hubbert linearization of global oil production as a function of the start and end year of fitting. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

Using the same stability region as we did for K, we get the density plot:



Density plot of stability surface for URR in the Hubbert linearization of global oil production. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Scale is 0 (blue) to 3000gb (red), with 1500gb being green. Contours are 30gb apart. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

The deviation over this region is 130gb, which I double to give a two sigma error bar of 260gb.

To translate these stability estimates into more convenient form, I did the following. Firstly I assumed that the highest URR estimate would correspond to the lowest K estimate (this is pretty close to true, as the main uncertainy in the linearization corresponds to the slope of the line, not it's vertical position). Next, I fixed the date of peak in the production versus time graph by insisting that all models match the correct value for the end of 2004 cumulative production of 1059gb. Ie my constraint insists that the area under the model before 2004 must be the same as the area under the real production versus time graph before 2004.

(One drawback of this is that since the logistic tails are systematically too high, because the tails are really Gaussian, it then has to make up the difference in the rest of the fit).

That gave me two models, which I call Logistic-High (with URR= 2510gb, K=4.6%, and a peak in late 2011), and another Logistic-Low (with URR=1990gb, K=5.2%, and the peak of the smooth curve in the middle of 2002). Those bound the two-sigma region for the center of the model evolution, given the linearization uncertainties.

So if Professor Deffeyes is right about November 2005, it's because he got lucky! The error bars really are quite significant.

The next graph shows everything in a semilog plot to 2020. The middle red line is the best fit

logistic, and the upper and lower ones are the logistic-high and logistic-low ones. Note that these are **not** error bars on annual production. They are error bars on where the center of the model would go, if it was truly logistic, and if extrapolation of this last region of linearity in the linearization is valid. Annual production can have significant noisy excursions above and below whatever the true line turns out to be.



Average annual oil production on a semilog plot with quadratic (ie Gaussian) fit, central, high, and low logistics, and piecewise exponentials. 1860-2020. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

Moving back into regular old production numbers, rather than the semilog plot, here's the various models extrapolated to 2040:



Average annual oil production with quadratic (ie Gaussian) fit, central, high, and low logistics, and piecewise exponentials. 1860-2040. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

Again, it's very important to remember that the high and low models are error bars on where the center of the model would go, if it was truly logistic, and if extrapolation of this last region of linearity in the linearization is valid. Annual production can have significant noisy excursions above and below whatever the true line turns out to be.

The Gaussian peak is in 2024 at 95mbpd, but I don't trust that extrapolation. The linearizationbased logistics are a lot more likely to be correct in my opinion. Notice that the last couple of year's production are a big spike above the logistic models. But that's ok, because the spike seems to be ending in the <u>recent production plateau</u>.

Let's bring it back to growth rates. Here's a year-on-year growth rate graph again, with all models included. The black line is the same Gaussian model from previous graphs (achieved by fitting a quadratic to the logarithm of production). The **grey** line is a direct linear fit to all the growth rates. That is a Gaussian curve too, but you can see it's much more pessimistic if you do it this way. This is further evidence of the instability of Gaussian prediction at this pre-peak stage, in my opinion. However, that most pessimistic model still takes till 2038 to hit 5% decline rates on a sustained basis.



Hubbert linearization of global oil production. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Color coding of data ranges matches the previous plots. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

Note that the world economy achieved 4-5% annual reductions in oil usage between 1979-1983 without collapsing. So I continue to believe that all this modeling suggests the future decline rates are within the adaptive capacity of the economy -- it's a <u>slow squeeze</u>, as I put it last month. I'm not saying that there won't be major economic hard times, but it does appear to me that peak oil is something that society can handle for quite some time to come, unless these models are just worthless.

A last caveat. One of the major reasons for a linearization extrapolation to go wrong is that there's a big chunk of discovery that isn't even seriously started production yet. I do not think there are any such discoveries in the conventional oil world (the Caspian is quite small on the world scale, and I think deepwater is well under way). However, there are trillions of possible barrels of LQHCs (low quality hydrocarbons), such as tar sands, extra-heavy oils, coal-to-liquids, and then biofuels. That stuff can't be ramped in a hurry, but it will probably get ramped up eventually (depending on the climate wild card). The linearization is not taking account of those things. So if I had to guess, my scientific, wild-ass guess about what will happen is something like the yellow curve in the following:



Average annual oil production on a semilog plot with quadratic (ie Gaussian) fit, central, high, and low logistics, and piecewise exponentials, together with scientific wild-assed guess (SWAG) as to the extrapolation of the piecewise exponential. 1860-2040. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

In production space, that looks like:



Average annual oil production with quadratic (ie Gaussian) fit, central, high, and low logistics, piecewise exponentials and SWAG extrapolation in yellow. 1860-2040. Click to enlarge. Believed to be all liquids, but excluding refinery gains. Data sources: API, <u>ASPO</u>, and <u>BP</u>.

However, please don't take the specific numbers on that yellow curve too seriously - it's just intended to illustrate a general qualitative idea of what might happen.

© SUMERIGENER RESERVED This work is licensed under a <u>Creative Commons Attribution-Share Alike</u> 3.0 United States License.