



## The Derivation of "Logistic-shaped" Discovery

Posted by [Nate Hagens](#) on June 26, 2008 - 10:00am

Topic: [Geology/Exploration](#)

Tags: [gaussian](#), [logistic](#), [m. king hubbert](#), [original](#), [verhulst](#), [webhubbletelescope](#)  
[\[list all tags\]](#)

21  
diggs

digg it

*This is a guest post from [WebHubbleTelescope](#). The post addresses the origins and relevance (or lack thereof) of the logistic equation as it is commonly used in projecting/modeling oil production forecasts. As far as I can see, this is the first time anyone has succeeded in deriving the Logistic oil model from first principles. I will follow this with a post on the Maximum Power Principle next week, which in my opinion may shed light on the logistic curve from the perspective of oil 'demand' (as opposed to supply).*

Many people believe that the [Logistic equation](#) adequately models the Hubbert peak. This comes about for a few reasons:

1. We can (often/occasionally) get an adequate heuristic fit to the shape of the production data by matching it to a logistic sigmoid curve.
2. The logistic-growth formula  $dU/dt = U(U_0 - U)$  carries some sort of physical significance.
3. The logistic has hung around for a *long* time, in modern terms, therefore it must have some practical value.

I see nothing wrong with the first reason; scientists and analysts have used heuristic curves to fit to empirical data for years and a simple expression provides a convenient shorthand for describing the shape of a data set. In the case of the Hubbert peak, we get the familiar S-function for cumulative production, and a bell-shaped curve for yearly production -- both characteristics that describe the Hubbert peak quite nicely to first-order.

As for point #2, we usually see hand-wavy arguments that point to an exponential growth that causes the peak oil curve to rapidly increase and then levels off as a negative feedback term in the equation takes over. What I consider circular reasoning with respect to [Hubbert Linearization](#) supports the idea that a physical process must drive this effect -- perhaps something similar to the constrained growth arguments popularized by Verhulst:

[Verhulst showed in 1846 that forces which tend to prevent a population growth grow in proportion to the ratio of the excess population to the total population. The non-linear differential equation describing the growth of a biological population which he deduced](#)

[and studied is now named after him.](#)

Unfortunately, I have never seen a derivation of this idea to oil production, at least to my liking. Most proofs have simply asserted that the relationship fits our intuition and then the equation gets solved with the resulting sigmoid curve ([here](#) or [here](#)):

$$U(t) = 1 / (1/U_0 + 1/Ae^{Bt})$$

I have problems with these kinds of assertions for a number of reasons. First of all, the general form of the resulting expression above can result from all sorts of fundamental principles besides the non-linear differential equation that Verhulst first theorized. For one, [Fermi-Dirac statistics](#) show the exact same S-curve relation as described by the  $U(t)$  formula above, yet no respectable physicist would ever derive FD by using the  $dU/dt = U(U_0 - U)$  logistics-growth formula. Most physicists would simply look at the relationship and see a coincidental mathematical identity that doesn't help their understanding one iota.

Secondly, one can play the same kind of identity games with the Normal (gaussian) curve, which also gets used occasionally to describe the production peak. In the case of the gaussian, we can generate a similar non-linear differential equation  $dG/dt \sim -t \cdot G$  which also "describes" the curve. But this similarly says nothing about how the gaussian comes about (the central limit theorem and the law of large numbers), instead it only shows how a mathematical identity arises from its parameterized curvature. This becomes a tautology, driven more by circular reasoning than anything else.

The last point of the logistic having implicit practical value has the historical force of momentum. This may seem blasphemous, but just because Hubbert first used this formulation years ago, doesn't make it *de facto* correct. He may have used the formula because of its convenience and mathematical properties more than anything else. I have either tried to contradict the use of the Logistic or searched for a fundamental derivation for some time now, but since everyone has shown some degree of satisfaction with the logistic, I haven't had much success until now ...

The breakthrough I have come across uses the [Dispersive Discovery model](#) as motivation. This model doesn't predict production but I figure that since production arises from the original discovery profile according to the Shock Model, this should at least generate a first-principles understanding.

In its general form, keeping search growth constant, the dispersive part of the discovery model produces a cumulative function that looks like this:

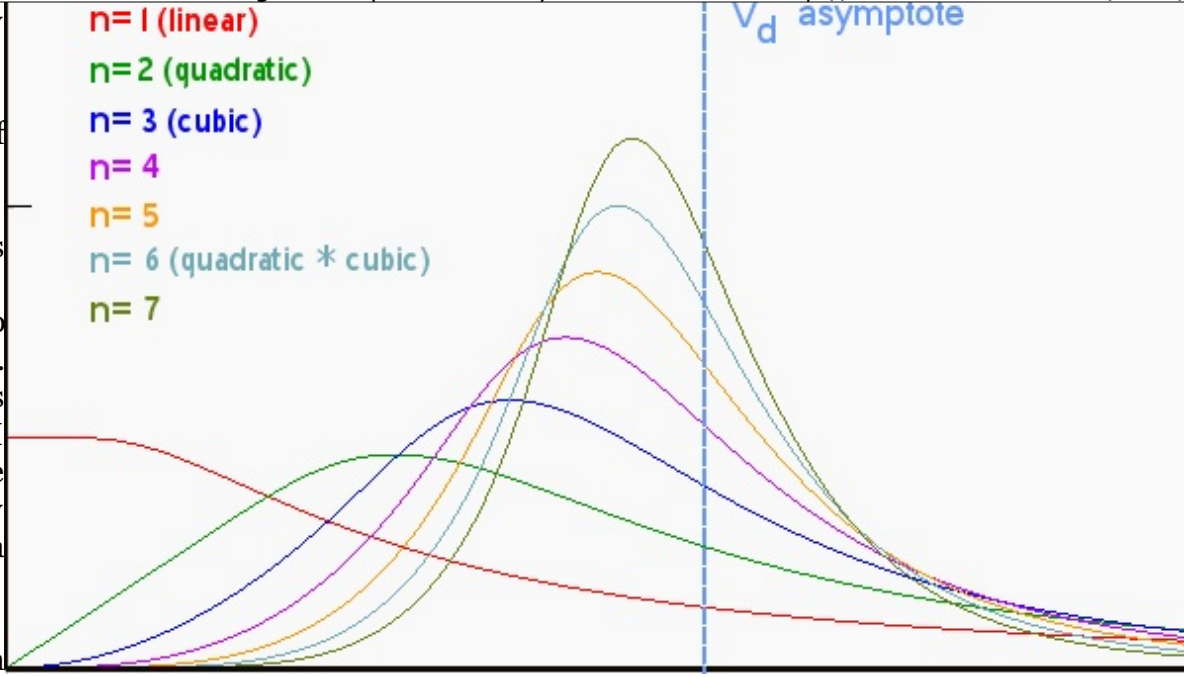
$$D(x) = x * (1 - \exp(-k/x))$$

The instantaneous curve generated by the derivative looks like

$$dD(x)/dx = c * (1 - \exp(-k/x)) * (1 + k/x)$$

Adding a growth term for  $x$  and we can get a family of curves for the derivative: I generated this set of curves

simply by applying growth terms of various powers, such as quadratic, cubic, etc, to replace  $x$ . No bones about it, I could have just as easily applied a positive exponential growth term



here, and the characteristic peaked curve would result, with the strength of the peak directly related to the acceleration of the exponential growth. I noted that in an [earlier post](#):

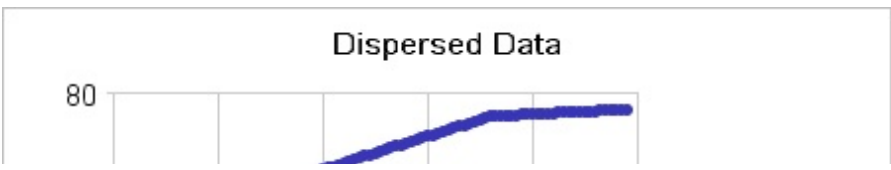
As for as other criticisms, I suppose one could question the actual relevance of a power-law growth as a driving function. In fact the formulation described here supports other growth laws, including monotonically increasing exponential growth.

Overall, the curves have some similarity to the Logistic sigmoid curve and its derivative, traditionally used to model the Hubbert peak. Yet it doesn't match the sigmoid because the equations obviously don't match -- not surprising since my model differs in its details from the Logistic heuristics. However, *and it starts to get really interesting now*, I can add another level of dispersion to my model and see what happens to the result.

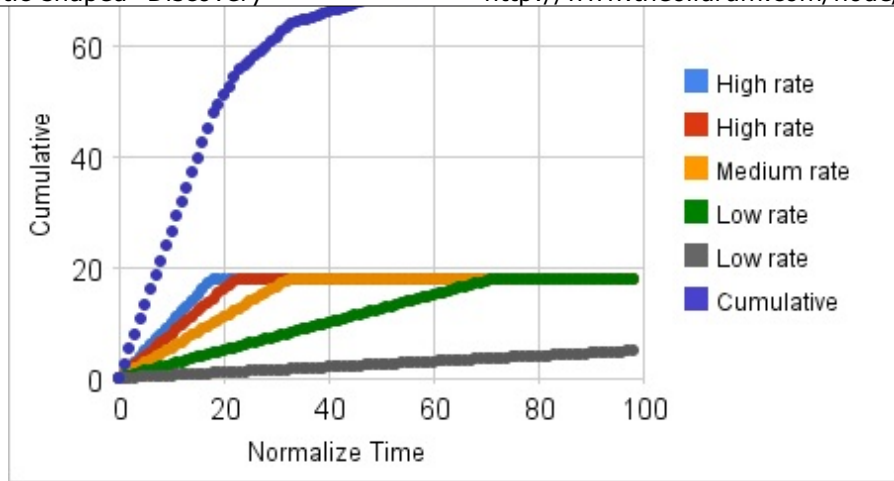
I originally intended for the dispersion to only apply to the variable search rates occurring over different geographic areas of the world. But I hinted that we could extend it to other stochastic variables:

We have much greater uncertainties in the stochastic variables in the oil discovery problem, ranging from the uncertainty in the spread of search volumes to the spread in the amount of people/corporations involved in the search itself.

So I originally started with a spread in search rates given as an uncertainty in the searched volume swept, and locked down the total volume as the constant  $k=L_0$ . Look at the following graph, which show several parts of the integration, and you can see that the uncertainties only reflect in the growth rates and not in the sub-volumes, which shows up as a clamped-asymptote below the cumulative asymptote: I figured that adding uncertainty to this term would make the result more messy than I would like to see at this expository level. But in retrospect, I should have taken the extra step as it does give a



very surprising result. That extra step involves a simple integration of the constant  $k=L_0$  term as a stochastic variable over a damped exponential probability density function (PDF) given by  $p(L)=\exp(-L/L_0)/L_0$ . This adds stochastic uncertainty to the total volume searched, or more precisely, uncertainty to the fixed sub-volumes searched, that when aggregated provide the total volume.



The following math derivation I extended from the original dispersive discovery equation explained in my TOD post ["Finding Needles in a Haystack"](#) (read this post if you need motivation for the general derivation). The first set of equations derives the original dispersive discovery which includes uncertainty in the search depth, while the second set of equations adds dispersion in the volume while building from the previous derivation.

Dispersive Discovery - fixed container L, dispersed depth  $\lambda$

$$P(x) = \frac{1}{\lambda} \cdot e^{-x/\lambda}$$

$$\bar{D} = \int_0^{\infty} x \cdot P(x) dx = \int_0^{\infty} x \cdot \frac{1}{\lambda} \cdot e^{-x/\lambda} dx$$

$$D(\lambda, L) = \int_0^L x \cdot \frac{1}{\lambda} \cdot e^{-x/\lambda} dx + \int_L^{\infty} L \cdot \frac{1}{\lambda} \cdot e^{-x/\lambda} dx$$

$$D(\lambda, L) = \lambda \cdot (1 - e^{-L/\lambda})$$

Double Dispersive Discovery - dispersed container size  $L_0$

$$p(L) = \frac{1}{L_0} \cdot e^{-L/L_0}$$

$$DD(\lambda) = \int_0^{\infty} D(\lambda, L) \cdot p(L) dL$$

$$DD(\lambda) = \int_0^{\infty} \lambda \cdot (1 - e^{-L/\lambda}) \cdot \frac{1}{L_0} \cdot e^{-L/L_0} dL$$

$$DD(\lambda) = \frac{1}{\frac{1}{L_0} + \frac{1}{\lambda}}$$

In the next to last relation, the addition of the second dispersion term turns into a trivial analytical integration from  $L=0$  to  $L=\infty$ . The result becomes the simple relation in the last line. Depending on the type of search growth, we come up with various kinds of cumulative discovery curves.

Note that the exponential term from the original dispersive discovery function

disappears. This occurs because of dimensional analysis: the dispersed rate stochastic variable in the denominator has an exponential PDF and the dispersed volume in the numerator has an exponential PDF; these essentially cancel each other after each gets integrated over the stochastic range. In any case, the simple relationship that this gives, *when inserted with an exponential growth term such as  $A \cdot e^{Bt}$* , results in what looks exactly like the [logistic sigmoid function](#):

$$\text{Power Law growth} \quad \lambda = k \times t^N$$

$$\text{Exponential growth} \quad \lambda = A \times e^{Bt}$$

Power Law discovery	$DD(t) = \frac{1}{\frac{1}{L_0} + \frac{1}{k \times t^N}}$
Exponential discovery	$DD(t) = \frac{1}{\frac{1}{L_0} + \frac{1}{A \times e^{Bt}}}$

That essentially describes the complete derivation of a discovery logistic curve in terms of exponential growth and dispersed parameters. By adding an additional stochastic element to the Dispersive Discovery model, the logistic has now transformed from a cheap heuristic into a model result. The fact that it builds on the first-principles of the Dispersive Discovery model gives us a deeper understanding of its origins. So whenever we see the logistic sigmoid used in a fit of the Hubbert curve we know that several preconditions must exist:

1. It models a discovery profile.
2. The search rates are dispersed via an exponential PDF
3. The searched volume is dispersed via an exponential PDF
4. The growth rate follows a positive exponential.

This finding now precludes other meaningless explanations for the Logistic curve's origin, including birth-death models, predator-prey models, and other ad-hoc carrying capacity derivations that other fields of scientific study have traditionally incorporated into their temporal dynamics theory. None of that matters, as the Logistic -- in terms of oil discovery -- simply models the stochastic effects of randomly searching an uncertain volume given an exponentially increasing average search rate. As an aside, you have to remember that Verhulst did not have the benefit of modern probability theory and the use of stochastic processes in the early 1800's, and came up with a very deterministic view of his subject matter. As a matter of fact, the theory and application of stochastic processes only became popularized to Western audiences in the mid 20<sup>th</sup> century (with classical English books on the subject by [Feller and Doob](#) appearing in the 1950's) and for someone like Hubbert to make the connection would in retrospect have seemed very prescient on his part.

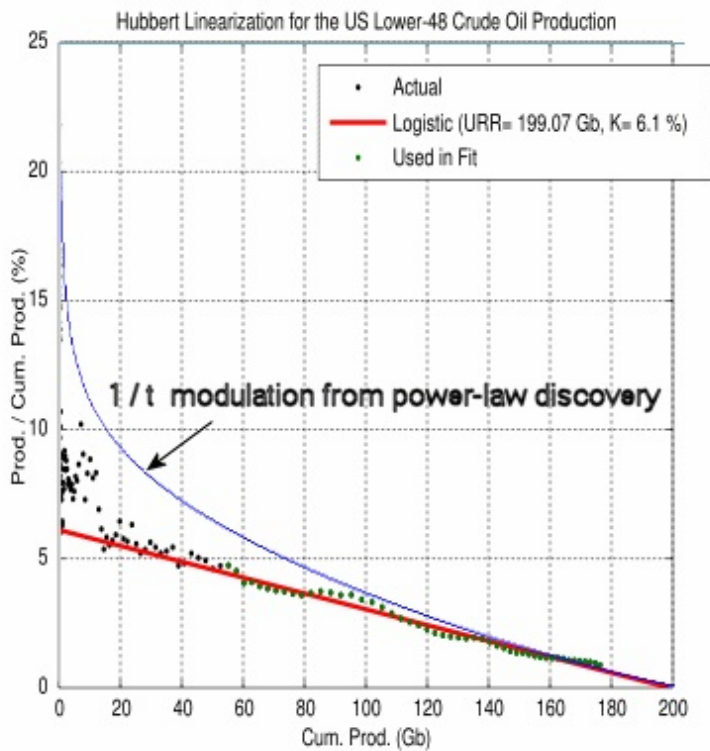
In the end, intuitive understanding plays an important role in setting up the initial premise, and the math has served as a formal verification of my understanding. You have to shoot holes in the probability theory to counter the argument, which any good debunking needs to do. As a very intriguing corollary to this finding, the fact that we can use a Logistic to model discovery means that we **cannot** use only a Logistic to model production. I have no qualms with this turn of events as production comes about as a result of applying the [Oil Shock model](#) to discoveries, and this essentially shifts the discovery curve to the right in the timeline while maintaining most of its basic shape. In spite of such a surprising model reduction to the sigmoid, we can continue to use the Dispersive Discovery in its more general form to understand a variety of parametric growth models, which means that we should remember that the Logistic manifests itself from a specific instantiation of dispersive discovery. But this specific derivation might just close the book on why the Logistic works at all. It also supports the unification between the Shock Model and the Logistic Model that [Khebab has investigated last year](#).

A different question to ask: Does the exponential-growth double dispersive discovery curve (the "logistic") work better than the power-law variation? Interesting that the power law discovery

curve does not linearize in the manner of Hubbert Linearization. Instead it generates the following quasi-linearization, where  $n$  is the power in the power-law curve:

$$dU/dt / U = n/t * (1 - U/URR)$$

Note that the hyperbolic factor (leading  $1/t$  term) creates a spike near the  $U=0$  origin, quite in keeping with many of the empirical HL observations of oil production. I don't think anyone has effectively explained the hyperbolic divergence typically observed. Although not intended as a fit to the data, the following figure shows how power discovery modulates the linear curve to potentially provide a more realistic fit to the data. It also reinforces my conjecture that these mathematical identities add very little intuitive value to the derivation of the models -- they simply represent tautological equivalences to the fundamental equations.



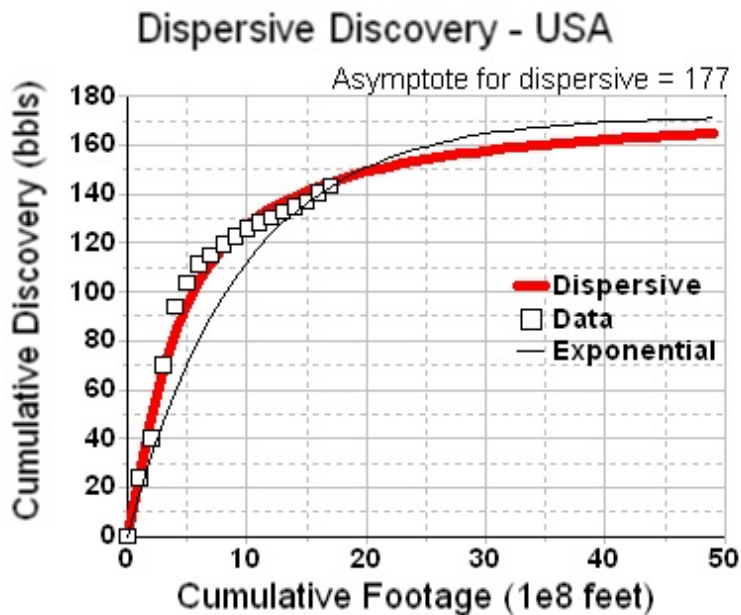
As another corollary, given the result:

$$D(x) = 1/(1/L_0 + 1/x)$$

we can verify another type of Hubbert Linearization. Consider that the parameter  $x$  describes a constant growth situation. If we can plot cumulative discovered volume ( $D$ ) against cumulative discoveries or depth ( $x$ ), we should confirm the creaming curve heuristic. In other words, the factor  $L$  should remain invariant allowing us to linear regress a good estimate of ultimate volume :

$$L_0 = 1/(1/D - 1/x)$$

It looks like this might arguably fit some curves better than [previously shown](#).



## 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](#)



This work is licensed under a [Creative Commons Attribution-Share Alike 3.0 United States License](#).