SPE 144790

Production Forecasting with Logistic Growth Models

A.J. Clark, L.W. Lake, T.W. Patzek, University of Texas at Austin

Abstract

With the commercial development of extremely low permeability oil and gas reservoirs, new challenges have arisen both from operational and reservoir standpoints. Reservoir models, which previously yielded reasonable results for reserves estimates and production forecasts, no longer do so. Various new models and techniques have been proposed to improve the accuracy and reliability of reserves estimates; however, none have gained widespread industry acceptance. This paper will propose a new empirical model for production forecasting in extremely low permeability oil and gas reservoirs based on logistic growth models. The new model incorporates known physical volumetric quantities of oil and gas into the forecast to constrain the reserve estimate to a reasonable quantity. The new model is easy to use, and it is very capable of trending existing production data and providing reasonable forecasts of future production. The logistic growth model does not extrapolate to non-physical values.

Introduction

One source of production to meet the demand for oil and gas has come from extremely low permeability oil and gas reservoirs, often referred to as unconventional resources. These unconventional resources typically exhibit permeabilities in the nanodarcy to microdarcy range. The production from these wells is characterized by very long and extended periods of transient flow before reaching the reservoir boundary, and entering into boundary-dominated flow. Thanks to drastic improvements in both horizontal drilling and hydraulic fracturing, these resources have been able to be successfully exploited in North America. Along with providing many new challenges in successfully drilling and completing unconventional wells, they have also provided new challenges in accurately forecasting the reserves.

The traditional empirical Arps' equation used to forecast reserves in conventional reservoirs, often gives over estimates of reserves when used in these extremely low permeability formations. Commonly the b values obtained are greater than 1, which results in an unrealistic production rate that never approaches 0. The purpose of this paper, however, is not to discuss the errors encountered when using the Arps' equation, as the problem has already been thoroughly discussed in the past. Lee (2010) and Ilk *et al.* (2008) have both provided good explanations of the problems encountered with the traditional models as have numerous other authors over the years. Various new models for use in production forecasting in tight reservoirs have been proposed including Maley (1985), Ilk *et al.* (2010), Kupchenko *et al.* (2008), and Valko (2009). The industry has been slow to adopt the new methods for forecasting reserves despite the apparent over estimation of reserves encountered when using the traditional decline curve models. This paper presents a new method for empirically forecasting production based on the logistic growth model.

Logistic Growth Models

Logistic growth curves are a family of mathematical models used to forecast growth in numerous applications. Originally developed by the Belgian mathematician Pierre Verhulst in the 1830s, logistic growth curves were used to model population growth. Verhulst based his ideas on the works of Malthus who believed that the population of a particular country or region would only be able to grow to a certain size before competition for resources would cause the growth to stabilize. Verhulst took this idea and by adding a multiplicative factor to the equation for exponential growth, created the logistic growth model.

The logistic growth model equation has a term referred to as the *carrying capacity*. This carrying capacity is the maximum size a population can grow to, at which point the size of the population will stabilize and the rate of growth will terminate.

The logistic growth models have come to be used in different fields for numerous things. Besides modeling population growth, they have modeled the growth of yeast, the regeneration of organs, and the penetration of new products into the market (Tsoularis and Wallace, 2001). They have even been used previously in the petroleum industry in the form of Hubbert's model. Hubbert's model (1956) was used to forecast production for entire fields or entire producing regions. The model being proposed in this paper differs from the Hubbert model in that it is used to forecast production for a single well.

The logistic growth models used across the various fields have taken on different forms; however, they have been combined by Tsoularis and Wallace into a generalized logistic growth model that takes on the form:

$$\frac{dN}{dt} = rN^{\alpha} \left[1 - \left(\frac{N}{K} \right)^{\beta} \right]^{\gamma}$$ (1)

Where

N	=	Population
r	=	Constant
α	=	Exponent
β	=	Exponent
γ	=	Exponent
K	=	Carrying Capacity

The model being proposed here is a special case of the generalized logistic growth model, which falls into the specific subcategory of hyperlogistics (Blumberg, 1968). It is commonly known that extremely low permeability oil and gas wells decline hyperbolically. The generalized logistic growth model is very flexible in nature and can accommodate numerous curve behaviors.

The form of the model proposed is adopted from an equation found in work by Spencer and Coulombe (1966), which was used to model the regrowth of livers. Livers naturally regenerate, so for their experiment rat livers were reduced to one third of their original size, and it was seen that they regenerated hyperbolically. They proposed a simple mathematical model to predict this growth, with the original size of the liver before reduction being used as the carrying capacity. For the purposes of forecasting production in oil and gas wells, the model has been altered after empirical analysis to the form:

$$Q(t) = \frac{Kt^{n}}{a + t^{n}}$$ (2)

Where

Q	=	Cumulative Production
K	=	Carrying Capacity
a	=	Constant
n	=	Hyperbolic Exponent
t	=	Time

The logistic growth model is as the name implies a growth equation. In this case the growth is cumulative oil or gas production. The derivative with respect to time can be taken to obtain the rate form.

$$q(t) = \frac{dQ}{dt} = \frac{Knbt^{n-1}}{\left(a + t^{n} \right)^{2}}$$ (3)

Where

q	=	Production Rate

While it doesn't appear to be so in this form, this is a specific case of the generalized logistic growth model (**eq. 1**) shown above where 'r' is equal to $n(K/a)^{1/n}$, 'α' is equal to $1-1/n$, 'β' is equal to 1 and 'γ' equals $1+1/n$. The logistic growth model matches production data for ultra-low permeability oil and gas wells very well (**Figure 1**).
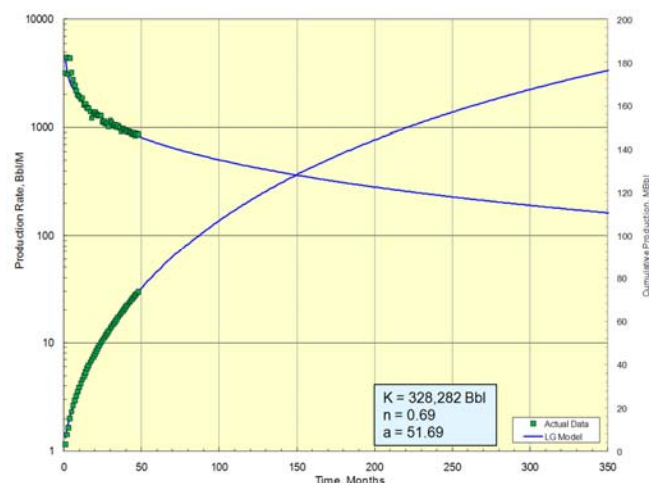
Figure 1 – Data from example Bakken Shale well fit with logistic growth model

Figure 1 shows an example of both rate and cumulative versus time production data for a Bakken Shale well being fit with the logistic growth model. The model is able to trend production data for this well and others extremely well.

Discussion of Parameters

There are either 2 or 3 unknown parameters in the logistic growth model that must be determined to obtain a fit to the production data. The three parameters are the carrying capacity K, the hyperbolic exponent n, and the constant a. The carrying capacity is the total amount of oil or gas recoverable from primary depletion in the well, independent of time or economic constraints. In other words, K is like the estimated ultimate recovery (EUR) for the well without economic constraints, and it is included in the model itself. The K acts as the term to constrain the cumulative production and to cause the rate to eventually go to 0. The cumulative production will approach the carrying capacity until it is eventually reached, at which point the rate will terminate. This parameter is also the one which determines whether there are 2 or 3 unknowns in the equation. It is possible to obtain the EUR from volumetric calculations prior to the deterministic forecast. If the EUR from volumetrics is not known prior to producing the well, it can also be used as a fitting parameter, and the optimal fit to the data will yield the K value. The logistic model is very flexible in nature, and also non-unique when the carrying capacity is not known ahead of time. It is possible to obtain multiple "good fits" to the data with various combinations of parameters. **Figure 2** is an example of a Barnett Shale gas well in which various carrying capacities were used, and all three instances yielded reasonable fits to the data.
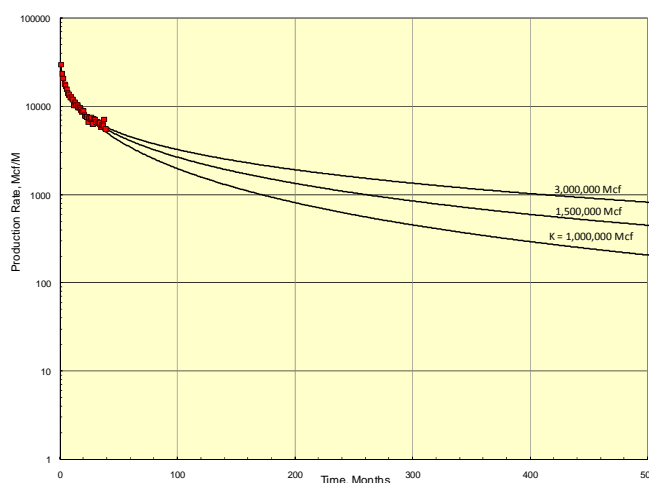

Figure 2 - Rate vs. time data fit with the logistic model using varying K values

Figure 2 uses the Barnett Shale as an example. The carrying capacity is set at 1, 1.5, and 3 Bcf for the three different forecasts. All 3 forecasts fit the production data very well.

After it has been determined whether or not the carrying capacity is known ahead of time, we must determine the hyperbolic decline exponent n. The n serves to control the decline behavior of the model, and it allows it to be more flexible to better fit the production data. To exhibit the effect of the n parameter on forecasts, dimensionless rate and dimensionless cumulative

terms were created. The dimensionless rate, q_D, is the production rate over the peak production rate, and the dimensionless cumulative, Q_D, is the cumulative production normalized by the carrying capacity. The dimensionless cumulative in this form is the fraction of total oil or gas recovered. As the well approaches the EUR, the Q_D will approach 1. **Figure 3** shows the dimensionless rate versus dimensionless cumulative for various values of the decline exponent.
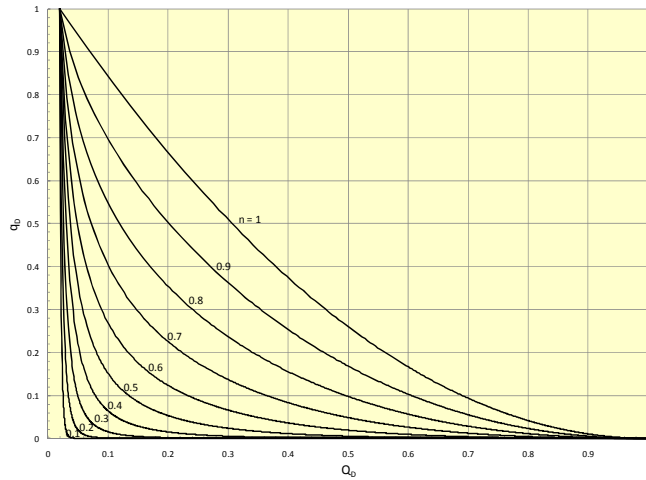

Figure 3 - Dimensionless type curve for varying n values.

Figure 3 shows the behavior of the model for n between 0 and 1. The K and a values used in this example are arbitrary. The 'n' value controls the steepness of the decline. At smaller n the well will decline at a high rate for a short period of time before stabilizing at a low rate and declining more slowly, while at higher values the well exhibits a more gradual decline through its life. When the n exceeds 1, the model will have an inflection point, where rate increases for a short period of time before decreasing. This does not cause the forecast to be in error, and can in fact be used to match data for wells whose initial rate is not their peak rate.

The third parameter, a, is the time to the power n at which half of the carrying capacity has been produced. This should not be confused with half the time it will take for the well to reach the carrying capacity. **Equation 4** shows that as time to the n approaches a, the logistic model approaches half the carrying capacity.

$$\lim_{t^n \to a} \left[\frac{Kt^n}{a+t^n} \right] = \frac{K}{2} \tag{4}$$

This causes the a to behave much like the initial decline parameter, D_i, for the Arps' equation. The lower the a value, the more quickly the rate will decline before stabilizing. Conversely, the higher the a value, the more stable the production will be for the life of the well. In other words if the a value is very low, the well will produce at a high rate, and quickly recover half of the oil or gas, then steeply decline and slowly produce the rest of the oil or gas at a low rate for a long period of time. **Figures 4** and **5** show the dimensionless rate versus time and dimensionless cumulative versus time respectively for varying values of a.
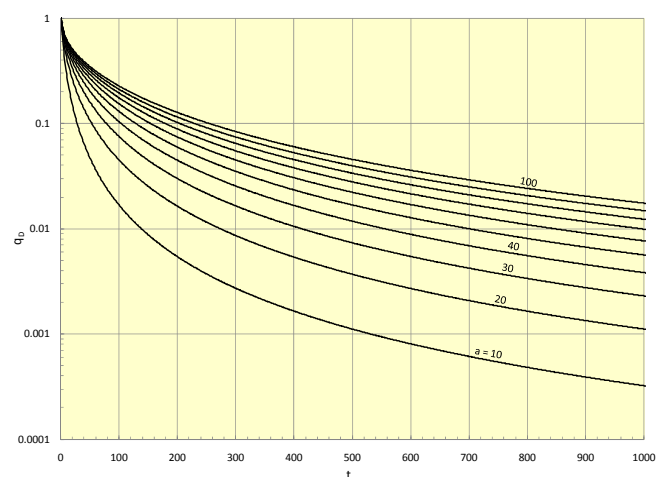
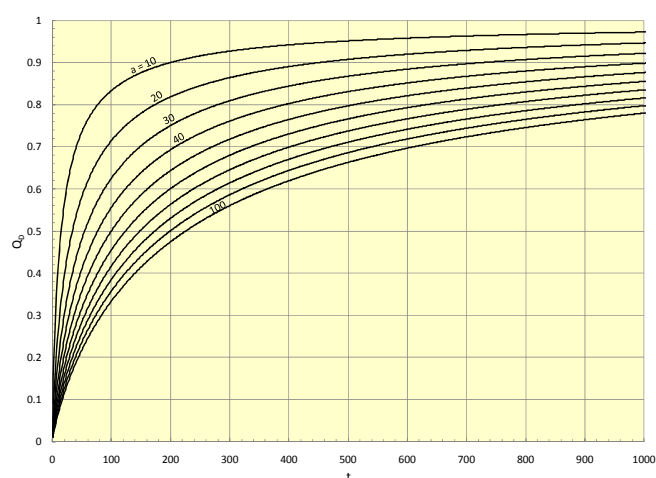**Figure 4 - Dimensionless rate vs. time for varying 'a' values**

Figure 5 - Dimensionless cumulative production vs. time for varying 'a' values

The a values in both **figure 4** and **figure 5** vary between 10 and 100, while the K and n values used are arbitrary. It can be observed that the low a values decline very steeply initially, before stabilizing at a much lower decline rate, while the higher a values have a smoother and more gradual decline.

Method

There are 2 preferred methods for fitting the logistic growth model to the data. The first method involves optimizing the parameters with a numerical scheme, and the second method involves linearizing the equation to plot the data and obtain the parameters. Either method will yield the parameters for the logistic growth model to obtain a production forecast. The logistic growth model is a cumulative production equation as opposed to a rate. Our experience is that matching to the cumulative data as opposed to the rate data provides better fits to the data. Operating conditions and various other occurrences can cause the rate data to have extreme outliers. The integral of the rate, the cumulative production, provides a smoother trend and serves to reduce the effect of these outliers when matching the data. However, the model parameters can be obtained from matching either rate or cumulative production.

The first method involves using a non-linear regression to obtain the parameters. If the carrying capacity is known ahead of time this can be used to obtain the 2 remaining parameters, and if it is not known, then all 3 may be obtained with this method. This can be done several ways, the easiest of which involve using least squares regression in Excel coupled with the Solver add-in. The user must provide initial estimates for the parameters, and then take the difference between the actual production data and the model production squared. The Solver add-in can then be used to minimize the sum of these differences and the parameters can be obtained. **Figure 6** shows an example of production data for a Barnett Shale gas well being fit with this method.
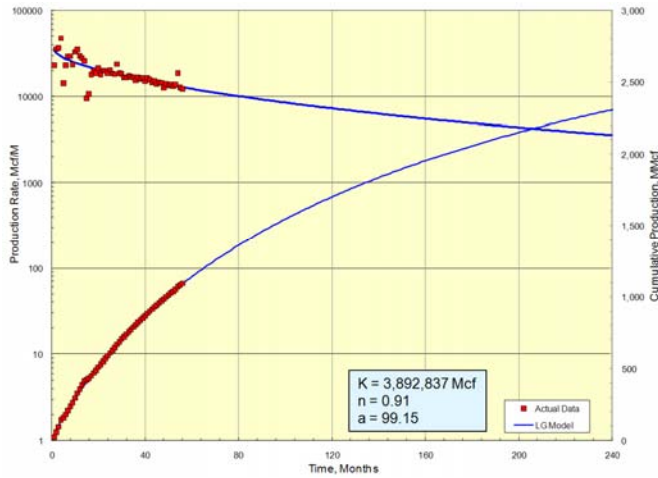
Figure 6 - Example of logistic model fit with least squares regression

If the forecaster has higher level software, built in functions such as MATLAB's 'nlinfit' can be used to obtain the parameters. In this method, initial guesses for the parameters values must also be set. An example of the cumulative production versus time data for a Barnett Shale well being fit with this method is in **figure 7**. This method can be more effective for fitting very large sets of data because of automation capabilities.
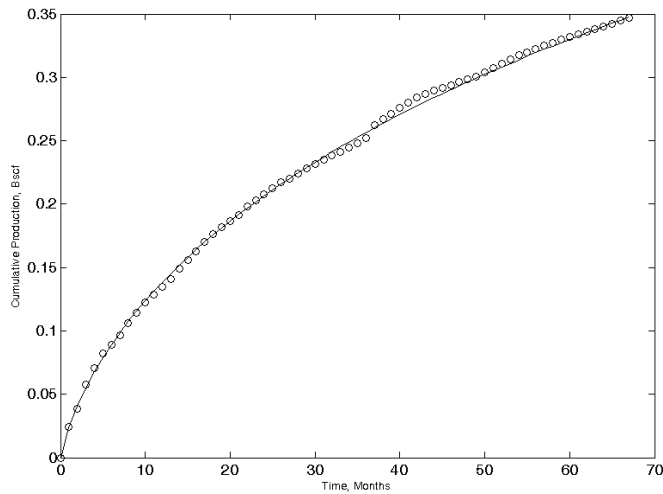

Figure 7 - Example of logistic model fit with MATLAB 'nlinfit' function

The other method for obtaining the parameters involves linearizing the equation and plotting it on a Cartesian grid to obtain the desired values. This method does not involve any numerical optimization; however, it does require an a priori knowledge of the carrying capacity. A straight line only allows for 2 degrees of freedom, and thus only 2 parameters can be determined through this method. By taking the reciprocal of **eqn. 2** and performing some algebraic manipulation, the equation can be rewritten in this form.

$$\log\left(\frac{K}{Q}-1\right) = \log a - n \log t \tag{5}$$

In this form, the logarithm of carrying capacity over cumulative production minus one plotted versus the logarithm of time on a Cartesian grid is a straight line. The slope and intercept of the resulting line can then be used to determine the n and a parameters respectively. The negative of the slope will be n, while 10 raised to the y-intercept will be a. **Figure 8** shows an example of the linearized logistic growth model.
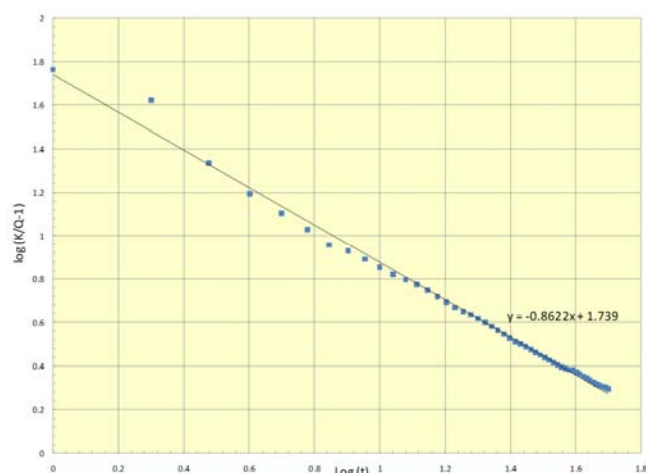
Figure 8 - Linearized logistic growth function

In this example, the carrying capacity was assumed to be 1.5 Bcf. Figure **8** shows that the cumulative production data follows a good linear trend. The resulting value of n obtained is 0.86. The resulting production forecast from the parameters obtained through this method is in **figure 9**.
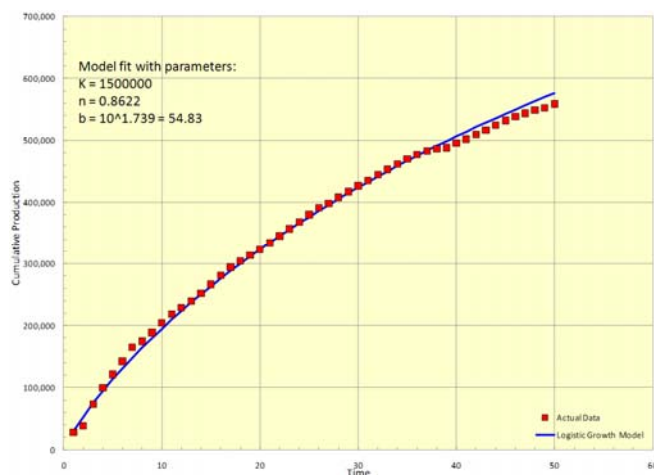

Figure 9 - Resulting fit of model with parameters obtained through linearization

Figure 9 shows that the model once again trends the actual data very well, and that a forecast of future production can be extrapolated forward in time with this model.

Example Bakken Shale Well

The use of the carrying capacity to constrain estimates to known physical volumes is one of the most powerful aspects of the logistic growth model. In his 1944 paper, "Analysis of Decline Curves" Arps' refers to the difference between deterministic methods and volumetric methods for estimating reserves when he says, "This assumption puts the extrapolation method on a strictly empirical basis and it must be realized that this may make the results sometimes inferior to the more exact volumetric methods." (Arps, 1944). This is an interesting insight from Arps' considering that the logging and coring technology which determines the volumetric estimates are still imperfect today, but still far better than they were 70 years ago when his model was introduced. Regardless, the logistic growth model has provided a way to combine both the volumetric and extrapolation methods to obtain what could be a more reliable forecast.

An example of this can be seen with a Bakken Shale well. To obtain the volumetric estimate of recoverable oil, six things must be known. These are porosity, thickness, oil saturation, formation volume factor, drainage area and recovery factor. The porosity, thickness and oil saturation can all be estimated from log data. The drainage area is a more uncertain factor in this equation, but it can be estimated. Finally, the recovery factor can be determined from the material balance equation.

There are several areas of uncertainty in the volumetric calculation. The Bakken Shale consists of 3 distinct facies. There are the upper and lower Bakken shales, and sandwiched between them is the dolomitic middle Bakken which is the target of the

completion. It is commonly believed that neither the upper nor lower Bakken contributes to the production. If this is assumed true, then a reasonable value for the thickness of the middle member can be determined. The drainage area is also difficult to determine. If the concept of stimulated reservoir volume (Mayerhofer *et al.*, 2008) is believed to be true, then the known wellbore length, and an assumed fracture half length can be used to determine drainage area.

For a particular Bakken shale well, from log data it is determined that the porosity for the well is 7%. The thickness of the middle Bakken member is found to be 34 ft, and the oil saturation is determined at 62%. The lateral length is 10,000 ft, and a relatively generous fracture half length of 1500 ft can be assumed. Based on prior analysis by Clark (2009) recovery factor in the Bakken Shale was shown to be between 5 – 7%. For this example, an even more generous recovery factor of 10% is assumed. Finally, the formation volume factor can be found to be 1.4 res. bbl/STB. Based on these parameters, the volumetrically available oil for this well to produce is roughly 275,000 bbls. Using this as the carrying capacity, an estimate can be obtained and compared with one obtained using the Arps' equation. **Figure 10** shows the results of this analysis.
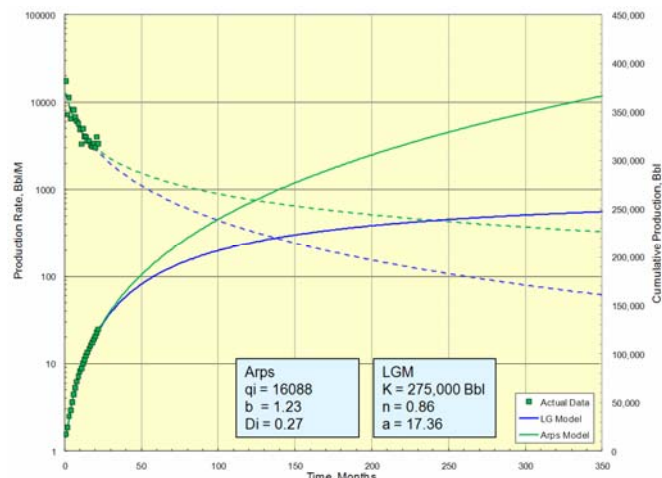

Figure 10 - Comparison of Arps' and logistic models. The Arps model extrapolates to infinite cumulative production for b>1.

It can be seen in the figure that both the Arps' model and the logistic model yield reasonable fits to the production data. The logistic model estimate is constrained using the volumetric estimate of 275,000 bbls, while the Arps' equation fits with a b value of 1.23 and is extrapolated into time to obtain the EUR estimate. In this case, after 30 years of production the Arps' model shows the well producing roughly 360,000 bbls of oil, while at that time the logistic model shows the well having only produced 250,000 bbls. In this particular example, despite volumetric analysis showing only 275,000 bbls to be available, the Arps' model predicts more than 40% greater recovery from the well.

Statistical Analysis of Parameters

To better understand what ranges of values can be expected from the logistic model in shale reservoirs, a MATLAB script was written to automate the model fitting process, and obtain a large sample of parameters. In this case, the logistic model was fit to roughly 600 Barnett Shale wells, and the resulting distributions of the parameters were analyzed. The initial sample of data contained production data for 1,000 horizontal wells completed between January 2004 and December 2006. The nonlinear regression function was used with MATLAB, and after obtaining the best fit for the data, all wells which contained unreasonable values for any of the parameters were discarded. This served to remove wells which did not exhibit reasonable decline trends for the automated forecasting.

After fitting the logistic model to the 600 wells the difference between the actual cumulative production data and the model data were reviewed. The results have been tabulated and can be seen in **Table 1**.

Table 1 – Results of MATLAB fitting of logistic model to Barnett Shale data

	Actual	LGM
Total production	555,494,337	531,423,708
Absolute average cum difference		58,665
Absolute average % error		6.36%
Average EUR 30 year estimate		1,438,363
Total EUR 30 year estimate		834,250,517

All values are in Mcf, except for the percent error. The total production for the approximately 600 wells estimated by the logistic model was very close to that of the actual data, but it did predict slightly less. The absolute average of the difference between the model and the actual data is approximately 59,000 Mcf. To calculate this value, the difference between the actual data and the model data was determined. When using the model, some of the estimated values were above the actual data while others were below it. To prevent these positive and negative values from cancelling each other out and masking potential errors, the absolute value of the difference was taken, and then the mean was calculated. The absolute percent error difference between the model and actual data was a very low 6.4%, showing that the model was able to very accurately trend the data. After extrapolating the model out 30 years, it was determined that the average Barnett Shale well, at least from this sample set, had an average EUR of 1.4 Bcf.

The individual parameters of the logistic model were analyzed to determine what kind of values would be expected when using the logistic model for production forecasting. The statistical analysis of the results can be seen in **Table 2**.

Table 2 – Statistical analysis of results of fitting the LGM to Barnett Shale wells

	mean	st. dev	min	median	max
K	1,782,382	1,309,691	139,360	1,386,216	9,102,227
n	0.90	0.11	0.65	0.90	1.32
a	33.07	19.37	7.66	27.63	152.98

The first parameter looked at was the carrying capacity, K. The resulting distribution of the K parameters can be seen in **figure 11**. The values appear to be log normally distributed. This is to be expected from the wells in any reservoir. The majority of the values obtained are close to the mean value, while the high productivity wells are much less likely to occur.
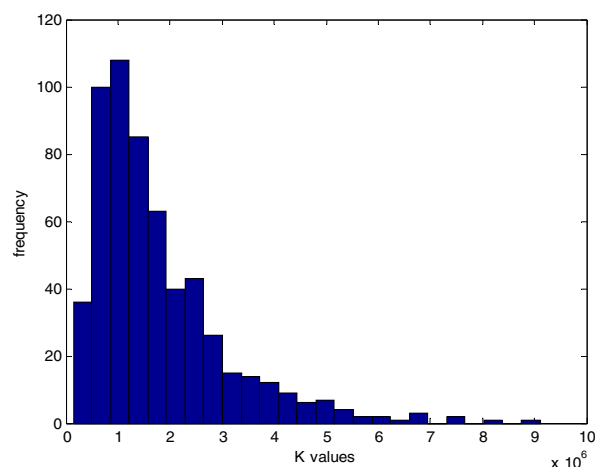

Figure 11 – Histogram of 'K' values for logistic growth model

Table 2 shows that the average carrying capacity is roughly 1.8 Bcf with a standard deviation of 1.3 Bcf. This is a very large standard deviation, caused by the log normal nature of the distribution. The minimum K value obtained was 0.14 Bcf while the maximum value was over 9 Bcf. This shows that there is a large range of possible recoveries that can be obtained in the Barnett Shale. It should be noted that the average carrying capacity is a bit higher than that of the average EUR calculated above. The carrying capacity represents the amount of oil or gas that the well will be able to recover from primary depletion if allowed to produce until exhaustion. The EUR calculations cut off the production at an arbitrary 30 years of production.

The next value looked at was the a value. This value as stated before behaves very similarly to the D_i value of the Arps' equation. It can be seen in **figure 12** that the resulting distribution is also log normal, much like the carrying capacity. The same distribution can be expected from the initial decline value of the Arps' equation. The mean value for the a was found to be roughly 33 months with a standard deviation of 19 months. This implies that after 33 months, the average Barnett Shale well is expected to have recovered half of its recoverable gas. In other words, in a well with a 30 year expected life, half of all the gas should be recovered in the first 3 years, while the second half of the gas will be recovered over the next 27 years. This is typical of all unconventional oil and gas wells. They tend to initially produce at very high rates, but then decline very quickly until they stabilize at a much lower rate where the decline becomes more gradual.
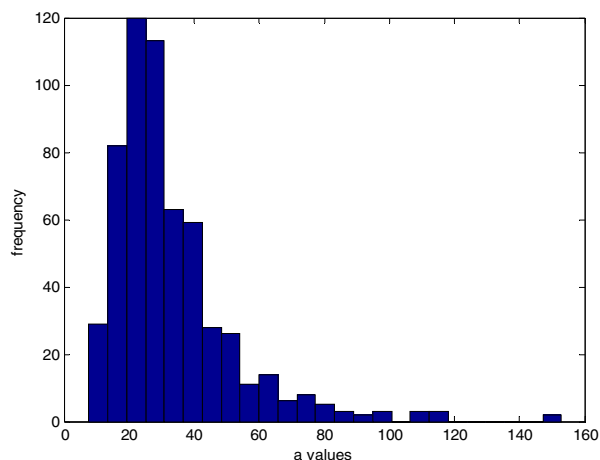
Figure 12 – Distribution of 'a' values for logistic growth model

The minimum a value obtained was approximately 8 months and the maximum was about 152 months. This is a very large range of values, and is caused by the decline behavior of the particular wells. The well with the very low a value would have been a well with a high initial rate, followed by an extremely sharp decline with a low stabilized rate. The high a value would be for a well with a more exponential decline, implying that the well would recover half of its gas approximately halfway through its producing life.

The final parameter looked at is the hyperbolic exponent n. The n value determines the behavior of the decline. In this case, as can be seen in **figure 13**, the distribution was normal, unlike that of the other parameters. The average n value was 0.9 with a standard distribution of 0.1. The minimum value was 0.65 while the maximum value was 1.35. This is a much smaller range of values than those obtained from the other 2 parameters. This implies that the n value will be far more likely to be very close to the mean value of 0.9 in the Barnett shale. It should also be noted that despite the inflection point which develops when the n value exceeds 1, it was still determined that the model was able to fit the data well.
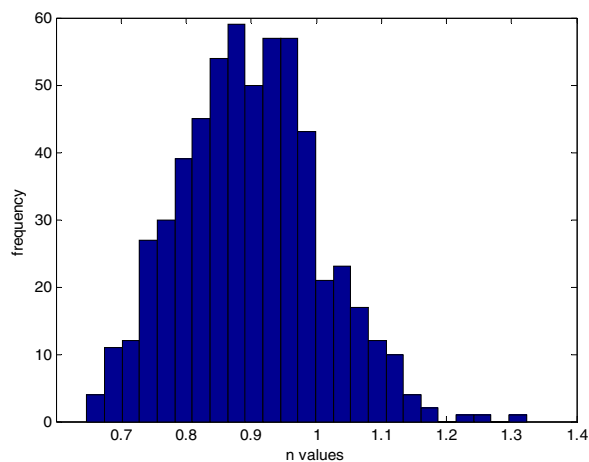

Figure 12 – Distribution of 'a' values for logistic growth model

Conclusion
This work develops a new model for use in forecasting production in extremely low permeability oil and gas reservoirs. The common problems encountered with the conventionally used Arps' model when wells exhibit extended transient flow behavior does not occur when using the logistic model. The logistic model can very accurately trend the decline behavior of these wells. The expected ultimate recovery values predicted with this model are also found to be slightly more modest than those obtained with the Arps' model. The logistic model uses the concept of the carrying capacity to constrain the total predicted cumulative production value.   The extrapolated production rate will always eventually reach 0. The carrying capacity can be used to constrain the forecast to known volumetrically available quantities of oil and gas to provide a realistic forecast based on physical characteristics of the well. The logistic growth model provides a new alternative method for empirically forecasting production in unconventional reservoirs.

References

Arps, J. J. (1945). "Analysis of Decline Curves."

Blumberg, A. A. (1968). "Logistic growth rate functions." Journal of Theoretical Biology **21**(1): 42-44.

Clark, A. J. (2009). Determination of Recovery Factor in the Bakken Formation, Mountrail County, ND. SPE International Student Paper Contest. New Orleans, Louisiana, USA.

Clark, A.J. (2011). "Decline Curve Analysis in Unconventional Resource Plays Using Logistic Growth Models." M.S. Thesis, The University of Texas at Austin, August 2011.

Hubbert, M. K. (1956). Nuclear Energy and the Fossil Fuel. Drilling and Production Practice, American Petroleum Institute.

Ilk, D., J. A. Rushing, et al. (2008). Exponential vs. Hyperbolic Decline in Tight Gas Sands: Understanding the Origin and Implications for Reserve Estimates Using Arps' Decline Curves. SPE Annual Technical Conference and Exhibition. Denver, Colorado, USA.

Kupchenko, C. L., B. W. Gault, et al. (2008). Tight Gas Production Performance Using Decline Curves. CIPC/SPE Gas Technology Symposium 2008 Joint Conference. Calgary, Alberta, Canada.

Lee, W. J. and R. E. Sidle (2010). Gas Reserves Estimation in Resource Plays. SPE Unconventional Gas Conference. Pittsburgh, Pennsylvania, USA.

Maley, S. (1985). The Use of Conventional Decline Curve Analysis in Tight Gas Well Applications. SPE/DOE Low Permeability Gas Reservoirs Symposium. Denver, Colorado.

Malthus, T. R. (1872). An essay on the principle of population: or, A view of its past and present effects on human happiness; with an inquiry into our prospects respecting the future removal or mitigation of the evils which it occasions, Reeves and Turner.

Mayerhofer, M. J., E. Lolon, et al. (2008). What is Stimulated Rock Volume? SPE Shale Gas Production Conference. Fort Worth, Texas, USA.

Spencer, R. P. and M. J. Coulombe (1966). "Quantitation of hepatic growth and regeneration." Growth, Development and Aging **30**(3): 277-284.

Tsoularis, A. and J. Wallace "Analysis of logistic growth models." Mathematical Biosciences **179**(1): 21-55.

Valko, P. P. (2009). Assigning value to stimulation in the Barnett Shale: a simultaneous analysis of 7000 plus production histories and well completion records. SPE Hydraulic Fracturing Technology Conference. The Woodlands, Texas.

Verhulst, Pierre-François (1838). "Notice sur la loi que la population poursuit dans son accroissement". *Correspondance mathématique et physique* **10**: 113–121. http://books.google.com/?id=8GsEAAAAYAAJ&q=.