Peak Oil: A Summary of Models and Predictions

Courtney Drew

September 20, 2017

Abstract

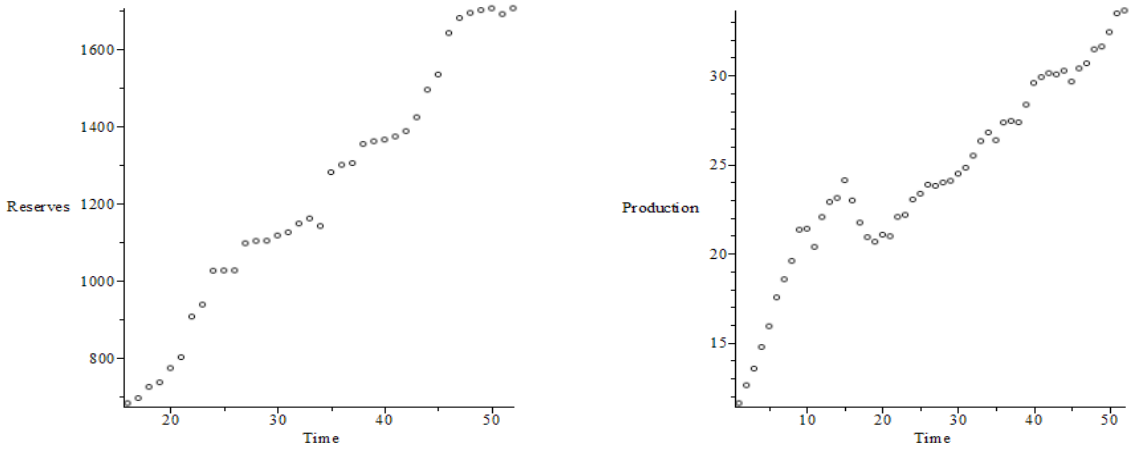
Ever since M. King Hubbert's accurate prediction in 1956 that US oil production would peak in 1970 [6], oil modeling has been given a lot of attention. This paper presents a summary of various models of oil production and reserves, as well as the methods used to find optimal parameter values when no exact solution could be found, as is the case with many inverse problems. Comparisons between models and analyses of the validity of the results are also discussed.

Keywords: Peak Oil, Parameter Estimation, Hubbert Model, Collage Coding

Contents

1 Introduction

In the United States, oil production began in Pennsylvania in 1859 [4, 6]. Although attention to alternative energy sources has been growing during the past decades, many countries still rely heavily on oil. If we were to run out of this resource tomorrow, a very harsh reality would set in, a reality similar to, but much worse than that which surrounded the 1973 oil crisis. But when will we run out of oil? In 1956, M. King Hubbert accurately predicted that a peak in oil production in the United States would occur in 1970 [6], and this prediction brought a lot of attention to this matter. The question "When will we run out of oil?" is one that many have asked themselves and which has led to various predictions of the timing of peak production and the ultimate depletion of oil reserves. This paper presents a summary of the various models that we developed as well as their respective predictions. Our first idea starts with a system of differential equations based on a logistic model and our second main idea is a modified Hubbert model with a varying carrying capacity. Modifications to these models as well as various methods used to solve them are also presented.



(a) Proved reserves data from 1980 to 2016

(b) Production data from 1965 to 2016

Figure 1: BP Proved Reserves and Production Data

Figure 1 presents the proved oil reserve data (in Gbbl) and production data (in Gbbl/year) from the BP Statistical Review of World Energy [13] that was released in June of 2017. In these graphs, 1965 corresponds to time $t = 1$. Production data begins in 1965 whereas the data for proved reserves only dates back to 1980. The term *proved reserves* refers to the "quantities that geological and engineering information indicates with reasonable certainty can be recovered in the future from known reservoirs under existing economic and operating conditions" (British Petroleum) [13]. From now on, the term *reserves* will be used interchangeably with *proved reserves*. Another expression that will often be used is *cumulative production*, or simply *cumulative*. This refers to the total amount of oil that has been pumped, up to a given year, since the beginning of production. Data for 1978 cumulative was taken from Gallagher's publication [4] to be 416.481 Gbbl and then BP's data was used to calculate the remaining cumulative data in the following way: $Q_i = Q_{i-1} + P_i$, where Q is cumulative and P is production.

Readers are encouraged to consult papers written by Hubbert [6], Maggio and Cacciola [11], Gallagher [4] and Brandt [3] for more information on previous models and predictions that were published.

2 Starting Model

Our first approach was to use a logistic-based model, modified in order to describe physical properties as well. It consists of a system of three nonlinear, first order differential equations relating production (P), reserves (R) and cumulative production (Q). Note that production is the rate of change of the cumulative.

$$\begin{aligned} P'(t) &= rP(R - bP) - \varepsilon Q \\ R'(t) &= -P + \alpha R(1 - \beta R) \\ Q'(t) &= P \end{aligned} \tag{1}$$

This model contains three variables (P, R and Q) dependent on time and five parameters (r, b, α, β, ε). Some of these parameters have a physical meaning. The first equation in this model has logistic-type growth that is decreased by an environmental concern proportional to cumulative, with proportionality constant ε. The constant b is an alarm rate, or the number of years left at current production before production declines (in the absence of any environmental concern). In the second differential equation, there is a $-P$ term since the reserves should decrease by the amount of oil pumped out. The second term in that equation represents the growth of the reserves due to new oil discoveries, thereby defining α as an effort parameter. With improving technologies, reserves have increased over time instead of decreasing by the amount pumped. In other words, the quantity of new oil found surpassed the amount produced. Another model will deal with this issue by modifying the reserve data.

The main difficulty with this model is that we were unable to solve it analytically; that is, no exact solution to the system of differential equations was found. We therefore proceeded with numerical solutions using Maple. With no known form of the solution, it is difficult to find parameters which optimally fit the known $P(t)$ and $R(t)$ data; however, the physical meanings of some of the parameters allowed us to impose restrictions on some of them. For instance, all of the parameters had to be positive. The alarm rate b has units of years, so a reasonable number of years had to be chosen. For example, an alarm rate of one year was not considered to be reasonable, but an alarm rate of at least 10 years was thought to be realistic.

We often considered the phase diagram, plotting reserves versus production from $t = 1$ to $t = n$, where n is the number of data points. We then chose parameters that appeared to fit the data well. However, we had no method to determine the best-fit parameters.

3 Environmental Concern Modification

In our starting model presented in section 2, the environmental concern term consisted of a proportionality constant applied to the cumulative production, but is that the most realistic model? Does the environmental concern really depend on all of the oil produced since the beginning of production? Possibly, but when considering this, we decided to change the environmental concern term to depend on production instead of cumulative. The rationale behind this is that if production increases, then the environmental concern will also increase accordingly, and hence changes in this term would depend more heavily on changes in production than on cumulative production. We can then reduce our model down to two equations.

$$\begin{aligned} P' &= rP(R - bP) - \varepsilon P = rP\left(R - bP - \frac{\varepsilon}{r}\right) \\ R' &= -P + \alpha R(1 - \beta R) \end{aligned} \tag{2}$$

3.1 Algebraic Simplifications

This modification simplified a little the system of differential equations, but still no analytic solution was found, forcing us to consider, once again, numerical solutions. The following manipulations were done in hopes of solving it exactly.

$$P' = rP\left(R - bP - \tfrac{\varepsilon}{r}\right) \text{ and } R' = -P + \alpha R(1 - \beta R)$$

$$\frac{dR}{dP} = \frac{-P + \alpha R(1 - \beta R)}{rP(R - bP - \frac{\varepsilon}{r})}$$

$$rP\left(R - bP - \frac{\varepsilon}{r}\right)\frac{dR}{dP} = -P + \alpha R(1 - \beta R) \tag{3}$$

$$\left(R - bP - \frac{\varepsilon}{r}\right)\frac{dR}{dP} = -\frac{1}{r} + \frac{\alpha R}{rP}(1 - \beta R)$$

Let $S(P) = R - bP - \frac{\varepsilon}{r}$
Then,

$$R = S(P) + bP + \frac{\varepsilon}{r}$$

$$\frac{dS}{dP} = \frac{dR}{dP} - b$$

$$\frac{dR}{dP} = \frac{dS}{dP} + b = S' + b$$

$$S(S' + b) = -\frac{1}{r} + \frac{\alpha}{rP}\left(S + bP + \frac{\varepsilon}{r}\right)\left(1 - \beta\left(S + bP + \frac{\varepsilon}{r}\right)\right)$$

$$SS' = -\frac{1}{r} - bS + \frac{\alpha}{rP}\left(S + bP + \frac{\varepsilon}{r}\right) - \frac{\alpha\beta}{rP}\left[S^2 + 2\left(bP + \frac{\varepsilon}{r}\right)S + \left(bP + \frac{\varepsilon}{r}\right)^2\right]$$

$$SS' = -\frac{\alpha\beta}{rP}S^2 - \left[\frac{2\alpha\beta}{rP}\left(bP + \frac{\varepsilon}{r}\right) + b - \frac{\alpha}{rP}\right]S - \left[\frac{\alpha\beta}{rP}\left(bP + \frac{\varepsilon}{r}\right)^2 + \frac{1}{r} - \frac{\alpha}{rP}\left(bP + \frac{\varepsilon}{r}\right)\right]$$

Let $\gamma = \frac{\alpha\beta}{r}$

$$SS' = -\frac{\gamma}{P}S^2 \underbrace{- \left[(2\gamma + 1)b + \frac{1}{rP}(2\gamma\varepsilon - \alpha)\right]}_{g(P)} S \underbrace{- \left[\frac{1}{P}\left(\gamma\left(bP + \frac{\varepsilon}{r}\right) - \frac{\alpha}{r}\right)\left(bP + \frac{\varepsilon}{r}\right) + \frac{1}{r}\right]}_{h(P)}$$

Let $S(P) = E(P)w(P)$ where $E(P) = e^{-\int \frac{\gamma}{P} dP} = e^{-\gamma ln(P)} = P^{-\gamma}$

Then, $ww'_P = F_1(P)w + F_0(P)$ where $F_1(P) = \frac{g(P)}{E(P)}$ and $F_0(P) = \frac{h(P)}{E(P)^2}$

$$F_1(P) = -P^{\gamma}\left[(2\gamma + 1)b + \frac{1}{rP}(2\gamma\varepsilon - \alpha)\right]$$

$$F_0(P) = -P^{2\gamma}\left[\frac{1}{P}\left(\gamma\left(bP + \frac{\varepsilon}{r}\right) - \frac{\alpha}{r}\right)\left(bP + \frac{\varepsilon}{r}\right) + \frac{1}{r}\right]$$

Let $x = P$

$$\begin{aligned} F_1(x) &= -x^{\gamma}\left[(2\gamma+1)b + \frac{1}{rx}(2\gamma\varepsilon - \alpha)\right] \\ &= -x^{\gamma-1}\left[(2\gamma+1)bx + \frac{(2\gamma\varepsilon-\alpha)}{r}\right] \end{aligned}$$

$$\begin{aligned} F_0(x) &= -x^{2\gamma}\left[\frac{1}{x}\left(\gamma\left(bx+\frac{\varepsilon}{r}\right)-\frac{\alpha}{r}\right)\left(bx+\frac{\varepsilon}{r}\right)+\frac{1}{r}\right] \\ &= -x^{2\gamma-1}\left[\left(\gamma\left(bx+\frac{\varepsilon}{r}\right)-\frac{\alpha}{r}\right)\left(bx+\frac{\varepsilon}{r}\right)+\frac{x}{r}\right] \\ &= -x^{2\gamma-1}\left[\gamma b^2x^2+\left(b\left(\frac{2\gamma\varepsilon-\alpha}{r}\right)+\frac{1}{r}\right)x+\left(\frac{\gamma\varepsilon-\alpha}{r}\right)\left(\frac{\varepsilon}{r}\right)\right] \\ &= -x^{2\gamma-1}\left[b^2\gamma x^2+\left(\frac{b\gamma\varepsilon}{r}+b\left(\frac{\gamma\varepsilon-\alpha}{r}\right)+\frac{1}{r}\right)x+\left(\frac{\gamma\varepsilon-\alpha}{r}\right)\left(\frac{\varepsilon}{r}\right)\right] \\ &= -x^{2\gamma-1}\left[b^2\gamma x^2+\frac{1}{r}(2b\gamma\varepsilon-b\alpha+1)x+\frac{1}{r^2}(\varepsilon^2\gamma-\alpha\varepsilon)\right] \end{aligned}$$

Let $y = rx$
Then,

$$x = \frac{y}{r}$$
$$y^2 = r^2x^2$$
$$x^2 = \frac{y^2}{r^2}$$

$\hat{F}_1(y) = F_1(\frac{y}{r}) = -\frac{y^{\gamma-1}}{r^{\gamma}}\left[(2\gamma+1)by + (2\gamma\varepsilon-\alpha)\right]$
$\hat{F}_0(y) = F_0(\frac{y}{r}) = -\frac{y^{2\gamma-1}}{r^{2\gamma+1}}\left[b^2\gamma y^2 + (2b\gamma\varepsilon - b\alpha + 1)y + (\varepsilon^2\gamma - \alpha\varepsilon)\right]$

$$ww'_x = ww'(y)r = \hat{F}_1(y)w + \hat{F}_0(y)$$
$$\frac{dw}{dx} = \frac{dw}{dy}\cdot\frac{dy}{dx} = \frac{dw}{dy}r$$

$$ww'_y = -\frac{y^{\gamma-1}}{r^{\gamma+1}}\underbrace{\left[(2\gamma+1)by+(2\gamma\varepsilon-\alpha)\right]}_{\tilde{F}_1(y)}w - \frac{y^{2\gamma-1}}{r^{2\gamma+2}}\underbrace{\left[b^2\gamma y^2+(2b\gamma\varepsilon-b\alpha+1)y+(\varepsilon^2\gamma-\varepsilon\alpha)\right]}_{\tilde{F}_0(y)}$$

$$ww'_y = -\frac{y^{\gamma-1}}{r^{\gamma+1}}\tilde{F}_1(y)w - \frac{y^{2\gamma-1}}{r^{2\gamma+2}}\tilde{F}_0(y)$$

Let $u = \frac{y^a w}{r^a}$
Then,

$$u' = \frac{1}{r^a}(ay^{a-1}w + y^aw') = \frac{au}{y} + \frac{y^aw'}{r^a}$$
$$\frac{y^aw'}{r^a} = u' - \frac{au}{y}$$

$$w' = \frac{r^a}{y^a}\left(u' - \frac{au}{y}\right)$$

$$ww' = \frac{r^{2a}}{y^{2a}}u\left(u' - \frac{au}{y}\right)$$

$$\frac{r^{2a}}{y^{2a}}\left(uu' - \frac{au^2}{y}\right) = -\frac{y^{\gamma-1}}{r^{\gamma+1}}\tilde{F}_1(y)\frac{r^a}{y^a}u - \frac{y^{2\gamma-1}}{r^{2\gamma+2}}\tilde{F}_0(y)$$

$$uu' - \frac{au^2}{y} = -\frac{y^{a+\gamma-1}}{r^{a+\gamma+1}}\tilde{F}_1(y)u - \frac{y^{2a+2\gamma-1}}{r^{2a+2\gamma+2}}\tilde{F}_0(y)$$

$$yuu' = au^2 - \frac{y^{a+\gamma}}{r^{a+\gamma+1}}\tilde{F}_1(y)u - \frac{y^{2(a+\gamma)}}{r^{2(a+\gamma+1)}}\tilde{F}_0(y)$$

Let $a = -\gamma$
Then,

$$yuu' = -\gamma u^2 - \frac{1}{r}\tilde{F}_1(y)u - \frac{1}{r^2}\tilde{F}_0(y) \tag{4}$$

$\tilde{F}_1(y) = b(2\gamma + 1)y + (2\gamma\varepsilon - \alpha)$
$\tilde{F}_0(y) = b^2\gamma y^2 + (2b\gamma\varepsilon - b\alpha + 1)y + (\varepsilon^2\gamma - \varepsilon\alpha)$

Eq. 4 is a simplified version of the Abel differential equation of the second kind that was started with (Eq. 3), but it was not found to be solvable.

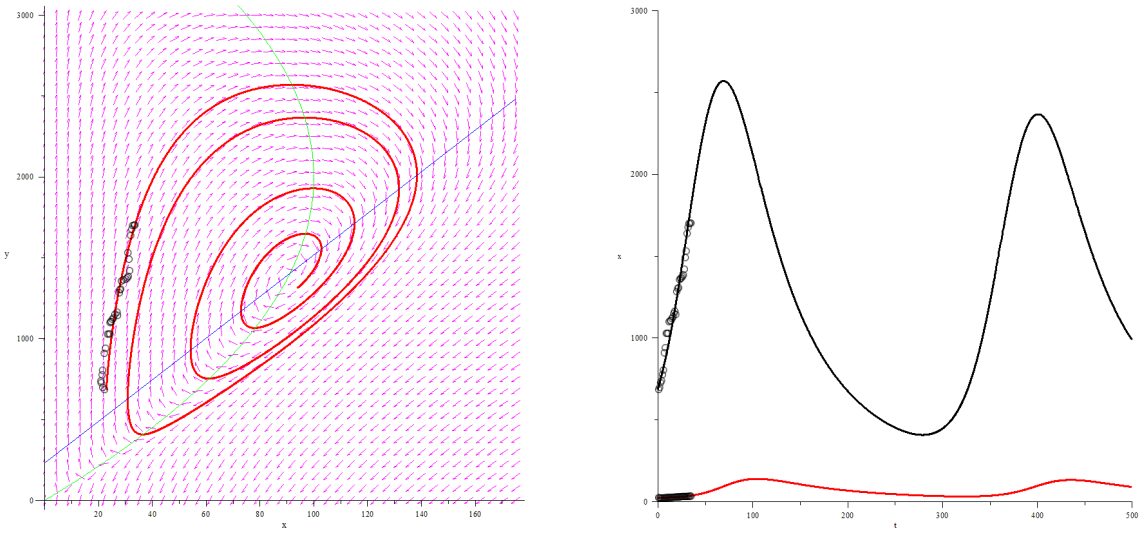
3.2 Results

By varying the parameter values, it is possible to obtain models that loop around a critical point of the system, also called a fixed point. Figure 2a presents one of these situations that models from time $t = 1$ (1980) to $t = 1000$. A scale of $\frac{3}{5}$ was applied to the model and parameters of $b = 12.86$, $e = 0.00805$, $\alpha = 0.1$, $\beta = 1/4000$ and $r = \frac{\left(-\alpha\beta\left(\frac{R_0^2}{P_0}\right)+\alpha\frac{R_0}{P_0}-1\right)}{\left(\frac{R[35]-R[1]}{P[35]-P[1]}(R_0-bP_0)\right)}$ were chosen. The data points are in black, the model is in red and the direction field is in magenta. The green and blue lines are the null lines of the steady state when $R' = 0$ and $P' = 0$. The fixed point is located at the intersection of the two null lines.

All of our results found so far using this model have been guesses. We have tried various parameter sets and varied each parameter until we obtained a model that appeared to fit the data well. In doing so, some models went to a fixed point right away, others looped around a fixed point and others yet crashed. In other words, reserves hit zero while we were still pumping out a lot of oil, which would make for a hard landing. However, we did not know that there wasn't another set of parameters, perhaps completely different, that would produce a better fit. Also, we didn't know which of these models was the most realistic. Hence, we proceeded with the collage method that can determine optimal parameters without a known form of the solution.

4 Collage Coding Method

This method is a fractal-based collage coding method that can determine optimal parameters for inverse problems like ours. Inverse problems often describe real-world situations for which data is commonly available. The problem with finding parameters for inverse problems is that most methods are based on iterative methods that are computationally heavy and that require a good



(a) 1980-2015 Reserves vs Production

(b) Reserves (black) and production (red) in time

Figure 2: Reserves and Production

starting point, which is not always known [5, p. 243]. A brief explanation of this method will be given, but readers are encouraged to consult [5] for a complete description. The general idea can be explained in three steps:

1. Find a polynomial basis representation for the data
2. Find a target to the polynomial basis
3. Find a fixed point approximation to the target

4.1 Mathematical Description

The basis of this method is to "express the problem in terms of a contractive map on a complete metric space" [8, p. 125] so that Banach's fixed point theorem can then be used. We will let $(\mathbb{X}, d)$ be a complete metric space and $Con(\mathbb{X})$ be a set of contraction maps on $\mathbb{X}$ [8]. Then, a contractive map is defined in the following way by Kunze, La Torre, Mendivil and Vrscay (2012) [5, p. 21]:

> A function $f : \mathbb{X} \to \mathbb{X}$ is a *contraction* if there is some $c \in [0, 1)$ with $d(f(x), f(y)) \leq c\, d(x, y)$ for all $x, y \in \mathbb{X}$. The smallest such constant c is the *contraction factor* for the contraction f.

This method then uses Banach's fixed point theorem which states that there is a unique fixed point $\overline{x} \in \mathbb{X}$ where $f(\overline{x}) = \overline{x}$ [5], [8]. Kunze and Heidler (2007) define the set of fixed points in the following way [8, p. 125]:

> Let $FP(\mathbb{X})$ denote the set of all fixed points of the contraction maps in $Con(\mathbb{X})$, i.e. $FP(\mathbb{X}) = \{\overline{x} \in \mathbb{X} \,|\, \overline{x} = f(\overline{x}) \text{ for some } f \in Con(\mathbb{X})\}$.

The goal of this method is then to find the best fixed point that approximates an element $x \in \mathbb{X}$, assuming one exists [8]. "Instead of searching for contraction maps whose fixed points lie close to y, [this method] search[es] for contraction maps that send y close to itself." [9] There are two types of errors in this method, a collage error, also called collage distance, $d(x, f(x))$, and an approximation error, $d(x, \overline{x})$. Collage coding aims to minimize the approximation error by minimizing the collage error [5, p. 17]. Finding the best fixed points is difficult, so this is where collage coding is used [8, p. 125]. Kunze and Heidler (2007) describe the collage theorem [8, p. 125]:

Let $x \in \mathbb{X}$ and $f \in Con(\mathbb{X})$ with fixed point $\overline{x}$ and contraction factor $c \in [0, 1)$. Then, $d(x, \overline{x}) \leq \frac{1}{1-c} d(x, f(x))$.

The proof of the collage theorem is fairly simple, relying on the triangle inequality as well as the definitions of a contraction and a fixed point.

$$\begin{aligned} d(x, \overline{x}) &\leq d(x, f(x)) + d(f(x), \overline{x}) \\ &= d(x, f(x)) + d(f(x), f(\overline{x})) \\ &\leq d(x, f(x)) + c\, d(x, \overline{x}) \end{aligned}$$

Thanks to Maple worksheets, specific to this model, that were provided to us, we were able to modify them to fit our model. We defined the system of differential equations to be the same as the modified environmental concern model presented in Eq. 2, and chose initial values that appeared to fit the data well. However, we were also able to leave the initial conditions as parameters, thus increasing the number of parameters to seven. We then assigned starting values to the parameters based on our previous models' results. We also chose the form of our differential equations; that is, which terms and cross terms we wanted. We were then able to find a polynomial basis, of desired order, that approximated either our actual data points or noisy data generated by the assigned parameters. A Riemann sum approximated the Euclidean, $\mathcal{L}^2$, distance and the best fit polynomial basis representation was determined by least-squares.

Now that we had a polynomial basis representation for our data, we created the Picard map and calculated the collage distance. Then, we used a least-squares approach to solve for the optimal parameters in our system of differential equations. The Picard operator T is defined as follows [8, p. 126]:

$$v(t) = (Tu)(t) = x_0 + \int_{t_0}^{t} f(t, y(t))dt \tag{5}$$

The collage distance is defined in the following way, assuming that $x(t)$ is the target function and $t \in [0, 1]$ [8, p. 126]:

$$\Delta = \left(\int_0^1 (x(t) - (Tx)(t))^2 dt \right)^{\frac{1}{2}} = \left(\int_0^1 \left[x(t) - x_0 - \int_0^1 f(x(s), s)ds \right]^2 dt \right)^{\frac{1}{2}} \tag{6}$$

4.2 Hubbert Model Ultimately Recoverable Reserves

We decided to test this method by finding the ultimately recoverable reserves for the Hubbert Model. The target was an 8th degree polynomial. Figure 3a presents the polynomial target in red, with coefficients determined from the minimal error, with the cumulative production data points in blue. The Hubbert Model is of the form $f(x) = c_1 x + c_2 x^2$. We minimized the collage distance to 13.96925195, and the values of c_1, c_2 and x_0 were found to be:

$$c_1 = 0.04693173036$$

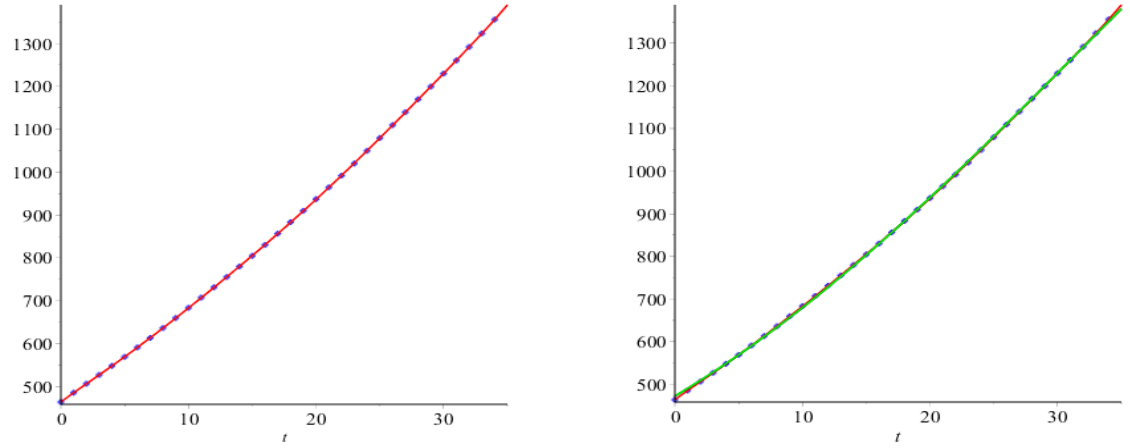
$$c_2 = -0.00001829722704$$

$$x_0 = 472.0746334$$

The fixed point approximation of the target was then found to be the following:

$$\frac{d}{dt} y(t) + 0.00001829722704 y(t)^2 - 0.04693173036 y(t)$$

Figure 3b presents the cumulative data in blue, the target in red and the fixed point approximation of the target in green.



(a) Cumulative Production data and 8th degree polynomial target

(b) Data, target and fixed point approximation for the Hubbert Model

Figure 3: Data, target and fixed point approximation for the Hubbert Model

Comparing the actual Hubbert Model, $Q' = rQ - \left(\frac{r}{Q_\infty}\right) Q^2$ to the fixed point approximation, $f(x) = c_1 x + c_2 x^2$, Q_∞ can be found to be similar to the expected value, thus validating our use of this method.

$$Q_\infty = -\frac{c_1}{c_2} = 2564.964093 \text{ Gbbl}$$

It is worth noting that the order of the target polynomial has a big effect on the results. For example, if the order is changed from 8th degree to 10th degree, Q_∞ changes to around 3852 Gbbl, but it is about 2558 Gbbl for a cubic target.

4.3 Results

Presented here are a few models that were obtained by varying the form of the differential equations as well as the basis order. The error associated with these models varies, and the signs of the parameters are not all correct. Note that the error given in the following sections refers to the collage distance and not the approximation error. Most notably, the environmental concern parameter, ε, is of the wrong sign, which does not, in our opinion, make sense. This parameter models efforts to reduce production for environmental reasons, so it would be inconsistent if it increased production by being of opposite sign. We will include our results from this method in this report for completeness, but we dismissed them due to their incompatability with our model.

4.3.1 First Model

Our first attempt consisted of a first order basis representation, without matching the initial conditions, and with P' having P, P^2 and PR terms, and R' having P, R and R^2 terms.

$$P' = -0.032148034168089058246P + 0.0043672166072849053384P^2 - 0.00005733931087698912980 3PR$$

$$P_0 = 19.974905832764888498$$

$$\text{Error} = 0.098571373241145769045$$

$$R' = 2.4365541015030441275P - 0.028618478363650151967R - 4.1087123300256113000 \cdot 10^{-20}R^2$$

$$R_0 = 668.23814470000002631$$

$$\text{Error} = 5.6910871869364278612 \cdot 10^{-14}$$

(a) Production basis and data

(b) Reserve basis and data

Figure 4: Production and reserve bases and data

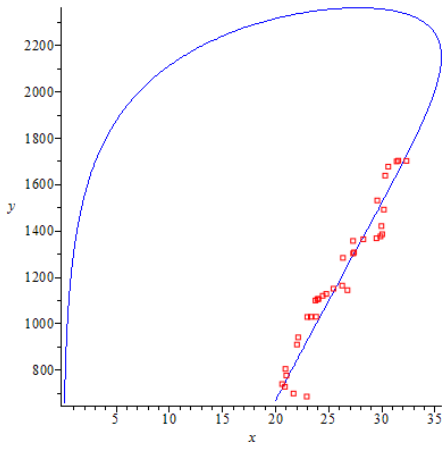


Figure 5: Reserve vs Production, 3.5·period

This model loops back with time, meaning that production peaks at around 35 Gbbl/year and then decreases without a hard landing.

4.3.2 Second Model

Our second attempt consisted of a 6th order basis representation, without matching the initial conditions, and with P' and R' both having P, R, P^2, PR, and R^2 terms.

$$P' = 0.62962430310702150210P - 0.012745570439809795537R - 0.062247491492436303380P^2$$
$$+0.0021366537600882580698PR - 0.000017553241686787580157R^2$$
$$P_0 = 25.504609052027645193$$
$$\text{Error} = 1.6106603462555350313$$

$$R' = 75.738886293239004981P - 1.6399274982927759915R - 4.7484205396204981535P^2$$
$$+0.14637618409692176092PR - 0.00092748551915847212905R^2$$
$$R_0 = 740.97444944179674915$$
$$\text{Error} = 67.230038152599616088$$

(a) Production basis and data

(b) Reserve basis and data

Figure 6: Production and reserve bases and data

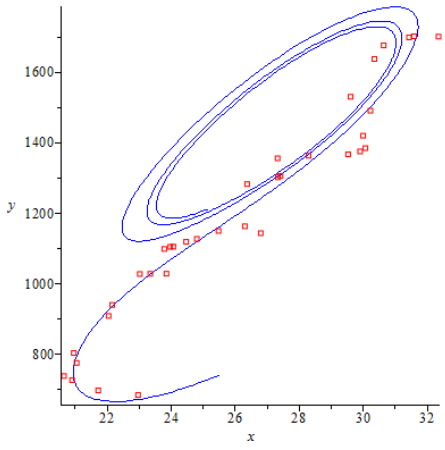


Figure 7: Reserve vs Production, 3.5·period

This model loops around a fixed point.

4.3.3 Third Model

Our third attempt consisted of a 12th order basis representation, without matching the initial conditions, and with P' and R' both having P, R, P^2, PR, and R^2 terms.

$$P' = 0.7384111896394200364 5P - 0.014730920028491451966R - 0.067951903846823814706P^2$$
$$+0.002250468139991858208 2PR - 0.000017655685271896626093R^2$$
$$P_0 = 25.044092671110087229$$
$$\text{Error} = 1.6517761046824717884$$

$$R' = 117.72362695675446678P - 2.5309202431232886777R - 7.7599907474222749563P^2$$
$$+0.239428339085145781 76PR - 0.001546751525226749601 0R^2$$
$$R_0 = 816.56075300456227973$$
$$\text{Error} = 118.07404935039705836$$

(a) Production basis and data

(b) Reserve basis and data

Figure 8: Production and reserve bases and data

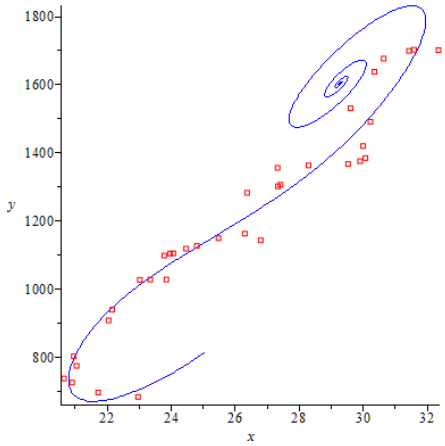


Figure 9: Reserve vs Production, 3.5·period

This model also loops around a fixed point, but the critical point is located at higher reserves and production than the previous model.

4.3.4 Fourth Model

Our fourth attempt consisted of a third order basis representation, without matching the initial conditions, and with P' having P, P^2 and PR terms, and R' having P, R and R^2 terms.

$$P' = 0.052192832694485568555P - 0.0039651889748620040889P^2 + 0.000053826553850260864593PR$$

$$P_0 = 21.156721261713759792$$

$$\text{Error} = 2.0697635239757511970$$

$$R' = 8.5858165431966003026P - 0.21395973006810814735R + 0.00004171659485654694193 9R^2$$

$$R_0 = 564.96577078317953614$$

$$\text{Error} = 98.917152505977687995$$

(a) Production basis and data

(b) Reserve basis and data

Figure 10: Production and reserve bases and data

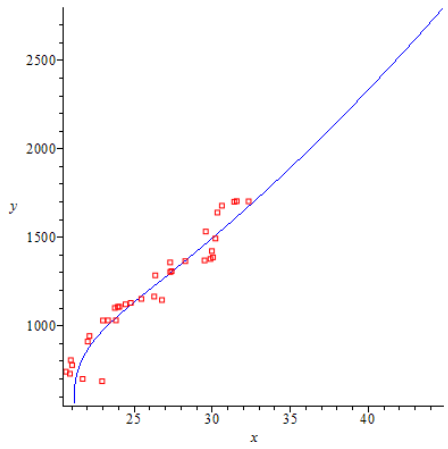
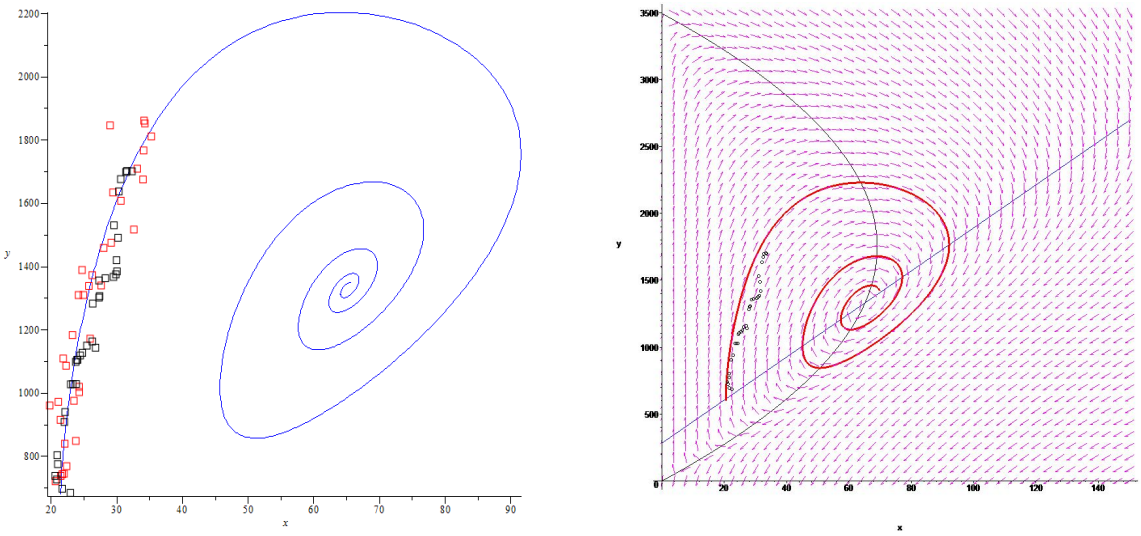


Figure 11: Reserve vs Production, 1.5·period

This model heads off to a fixed point right away without any looping behaviour.

4.3.5 Fifth Model

Although producing interesting models, our results obtained so far with collage coding method did not fit with our model, so we imposed restrictions on some of the parameters in hopes of getting more realistic parameters. We imposed -1 for the coefficient of the production term in the reserves differential equation, which can be explained physically by a decrease in reserves by the amount pumped out, and we were obliged to impose a value for ε, determined from our previous model. These restrictions led to the following results with a sixth order basis representation. The actual data is in black and the generated data is in red, presented in figure 12a. The parameter values are as follows: $\alpha = 0.079064049640094891359$, $\varepsilon = 0.00805$, $r = 0.000028390474956398582703$, $b = 16.077912738485876713$ and $\beta = 0.00028621790382326287875$. These parameters are fairly similar to our initial set of parameters in our first model. However, we imposed some restrictions based on that model as well, so it is not surprising that both sets of parameters are alike. Now that we have values for all of the parameters, we could then plot the direction field along with the model and the actual data, as shown in figure 12b.



(a) Reserve vs Production, 22.5·period

(b) Reserves vs Production with Direction Field

Figure 12: Fifth model reserves vs production

5 Modified Hubbert Model with Varying Carrying Capacity

5.1 Hubbert Model

Hubbert used a logistic function to model oil production (Eq. 7). However, the carrying capacity (Q_∞) that he used for world oil was approximately 1250 Gbbl [6, p. 17], and we now know that this greatly underestimated the total amount of oil that can be produced. Therefore, we decided to keep Hubbert's logistic model, but modify it so that the carrying capacity is no longer constant, but a logistic function as well.

$$P = Q' = rQ\left(1 - \frac{Q}{Q_\infty}\right) \tag{7}$$

5.2 Maximum Cumulative Data

Figure 1a presents the proved oil reserve data (in Gbbl) from the BP Statistical Review of World Energy 2017 [13]. The reserves increase almost every year during this time interval, which can be

explained not by an actual increase in the ultimate amount of oil in the ground since 1859, but by developing technologies that enable the discovery of new oil fields and more oil from existing ones. Therefore, our data for oil reserves does not reflect the actual amount of oil left in the ground that can be pumped out. With this in mind, we wished to modify the reserve data so that it would follow a decreasing path, based on an initial quantity that is not significantly increased, as was suggested by M. King Hubbert in 1956 [6, p. 4]. Defining a sequence of maximum cumulative production values (Q_{max}) to be the sum of the reserves at that time and cumulative production up to that year (Eq. 8), we can then fit these points to a logistic model. The carrying capacity, or limiting value, of this logistic equation, $Q_{max\infty}$, will hence be the total initial reserves at the beginning of time, according to this model.

$$Q_{max}(t) = R(t) + Q(t) \tag{8}$$

5.2.1 Maple's Nonlinear Fit for the Maximum Cumulative Data

The following model was obtained using Maple's Nonlinear Fit command (Eq. 9). Figure 13 presents the Q_{max} points (in Gbbl) as well as the logistic curve. Time $t = 1$ corresponds to 1965.

$$Q_{max} = \frac{5069.629978532037}{1 + e^{(1.928947798260190 - 0.046708278034852 37t)}} \tag{9}$$

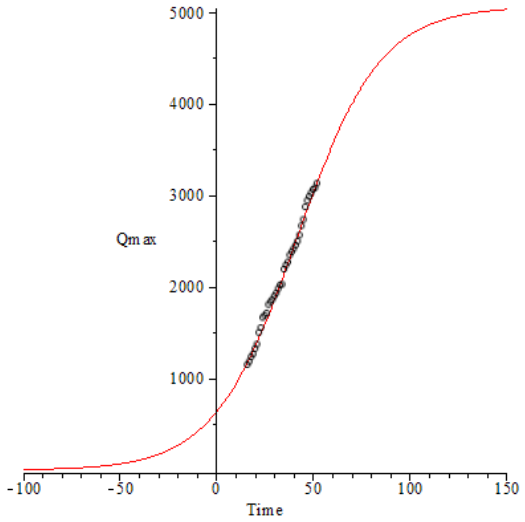


Figure 13: Q_{max} data from 1980 to 2016 and the logistic curve

Hubbert had predicted in 1956 that the ultimately recoverable reserves (URR) [11, p. 112] would be 1250 billion barrels [6, p. 17]. The URR, called here $Q_{max\infty}$, is now almost 5070 Gbbl, more than four times larger than Hubbert's prediction. In our opinion, this is a much more realistic value for the total reserves. The inflection point of this logistic function occurs at time $t = 41.298$, which corresponds to the year 2005. This implies that, according to this model, the growth rate of Q_{max} peaked about 12 years ago, when Q_{max} had a value of approximately 2534.815 Gbbl. This does not quite fit with the data that we have, but it is not an unrealistic value either in our opinion.

5.2.2 Root-Mean-Squared Error

In order to test the validity of Maple's Nonlinear Fit, we can determine $Q_{max\infty}$ that minimizes the root-mean-squared error. Results from the 1980 onwards data are shown here. This procedure was repeated with the 1965 onwards data, but due to the abnormal peak prior to 1980, the error was found to be much larger. That being said, the purpose of doing this was to validate Maple's results, and this goal was achieved. RMSE in Eq. 10 refers to the root-mean-squared error.

$$Q_{max} = \frac{Q_{max\infty}}{1 + e^{b-rt}}$$

$$b + rt = \ln\left(\frac{Q_{max\infty}}{Q_{max}} - 1\right)$$

$$b + rt := Y(t; Q_{max\infty})$$

$$r(Q_{max\infty}) = \frac{SS_{ty}}{SS_{tt}} = \frac{\sum t_i Y_i(Q_{max\infty}) - n\bar{t}\overline{Y}(Q_{max\infty})}{\sum t_i^2 - n\bar{t}^2}$$

$$b(Q_{max\infty}) = \overline{Y}(Q_{max\infty}) - r(Q_{max\infty})\bar{t}$$

$$RMSE = \sqrt{\frac{1}{n}\left(\sum_{i=1}^{n}\left(Q_{max}[i] - \frac{Q_{max\infty}}{1 + e^{b-rt}}\right)^2\right)} \tag{10}$$

Calculating these quantities using the maximum cumulative data found by using Eq. 8, the following graph of the root-mean-squared error as a function of $Q_{max\infty}$ was obtained. The total initial reserve value that minimizes the error is approximately 5133 Gbbl.

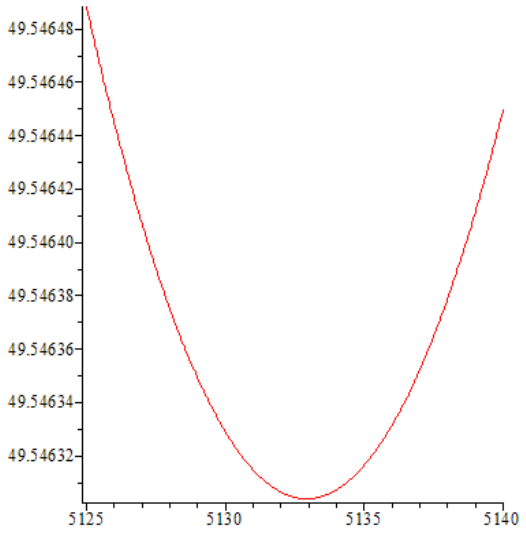


Figure 14: Root-mean-squared error as a function of $Q_{max\infty}$

The $Q_{max\infty}$ value obtained through Maple's Nonlinear Fit for 1980 onwards data ($t = 1$ being 1980) was about 5170 Gbbl compared to this value of 5133 Gbbl. The two values are not exactly the same, but they are close enough, in our opinion, to give validity to Maple's result.

5.3 Modified Reserve Data

It is now possible to define a modified reserve data sequence to be the cumulative production up to that time subtracted from the total initial reserves.

$$R_{mod}(t) = Q_{max\infty} - Q(t) \tag{11}$$

Figure 15 shows the modified reserve data with respect to time, 1965 being time $t = 1$. The reserves are now strictly decreasing from year to year by the amount produced. In our opinion, this data is more realistic since reserves should be decreasing with time, not increasing as the actual proved reserves data indicates.

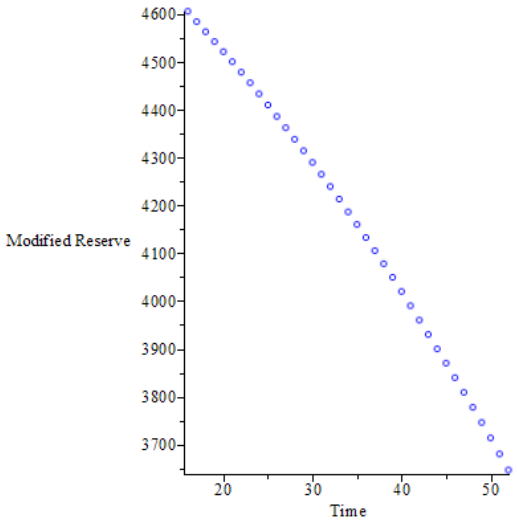


Figure 15: Modified reserve data

5.4 Model

Now that we have defined some new data sequences, we can create our new model based on Hubbert's logistic model (Eq. 7), but using a varying carrying capacity.

$$Q_{max}(t) = R(t) + Q(t)$$
$$R_{mod}(t) = Q_{max}(\infty) - Q(t)$$

$$Q'_{max}(t) = \hat{r}Q_{max}\left(1 - \frac{Q_{max}}{Q_{max\infty}}\right) \tag{12}$$

$$P = Q'(t) = rQ\left(1 - \frac{Q}{Q_{max}(t)}\right) \tag{13}$$

Eq. 12 is a logistic function with a known solution of the form $Q_{max} = \frac{Q_{max\infty}}{1+e^{b-\hat{r}t}}$. We can then substitute this solution into Eq. 13, the latter becoming a Bernoulli differential equation that can be solved.

$$P = Q' = rQ\left(1 - \frac{Q}{\frac{Q_{max\infty}}{1+e^{b-\hat{r}t}}}\right) = rQ\left(1 - \frac{Q(1+e^{b-\hat{r}t})}{Q_{max\infty}}\right) = rQ - \frac{rQ^2}{Q_{max\infty}}(1+e^{b-\hat{r}t})$$

$$Q' - rQ = -\frac{rQ^2}{Q_{max\infty}}(1+e^{b-\hat{r}t})$$
$$Q^{-2}Q' - rQ^{-1} = -\frac{r}{Q_{max\infty}}(1+e^{b-\hat{r}t})$$

We can now use the following substitution to solve the Bernoulli Equation:

$v = Q^{1-n}$ and $v' = (1-n)Q^{-n}Q'$

With $n = 2$, this becomes:

$v = Q^{-1}$ and $v' = -Q^{-2}Q'$

$$v' + rv = \frac{r}{Q_{max\infty}}(1 + e^{b-\hat{r}t})$$

This is now a first order, linear differential equation that can be solved using the integrating factor method.

$$\mu = e^{\int P(t)dt} = e^{\int r dt} = e^{rt}$$

$$\begin{aligned}
v(t) &= \frac{1}{\mu(t)} \int \mu(t) g(t) dt \\
&= \frac{1}{e^{rt}} \int e^{rt} \left(\frac{r}{Q_{max\infty}}(1 + e^{b-\hat{r}t}) \right) dt \\
&= e^{-rt} \int \frac{re^{rt}}{Q_{max\infty}} + \frac{re^{rt}}{Q_{max\infty}}(e^{b-\hat{r}t}) dt \\
&= e^{-rt} \int \frac{re^{rt}}{Q_{max\infty}} + \frac{re^{b+(r-\hat{r})t}}{Q_{max\infty}} dt \\
&= e^{-rt} \left(\frac{re^{rt}}{Q_{max\infty} r} + \frac{re^{b+(r-\hat{r})t}}{Q_{max\infty}(r-\hat{r})} + C \right) \\
&= \frac{1}{Q_{max\infty}} + \frac{re^{b-\hat{r}t}}{Q_{max\infty}(r-\hat{r})} + Ce^{-rt}
\end{aligned}$$

$$\begin{aligned}
\frac{1}{Q} &= \frac{1}{Q_{max\infty}} \left(1 + \frac{r}{r-\hat{r}} e^{b-\hat{r}t} \right) + Ce^{-rt} \\
Q &= \frac{1}{\frac{1}{Q_{max\infty}} \left(1 + \frac{r}{r-\hat{r}} e^{b-\hat{r}t} \right) + Ce^{-rt}}
\end{aligned}$$

We now have explicit equations for cumulative production, production and reserves that depend on time. Our model is therefore the following:

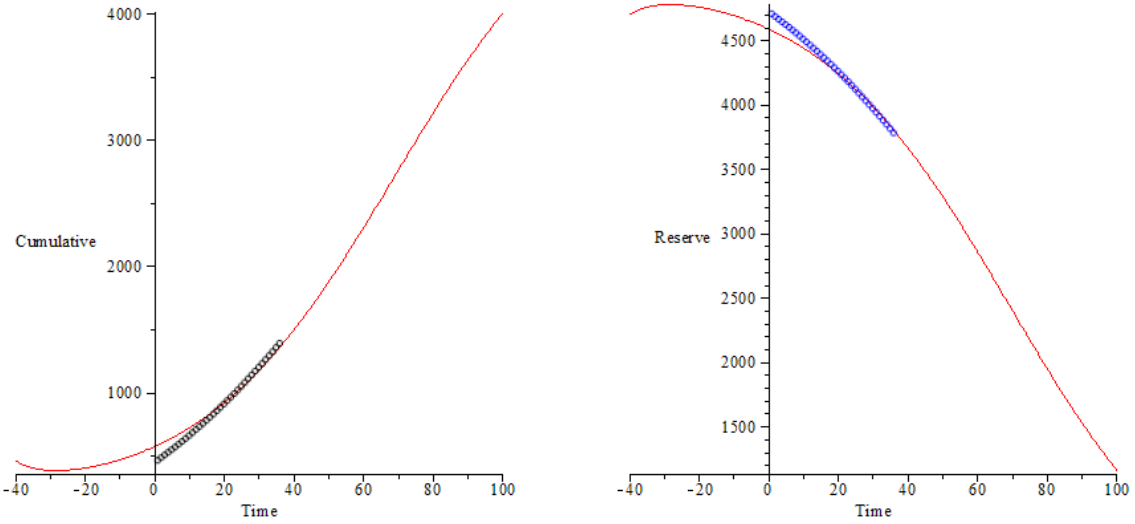
$$Q(t) = \frac{1}{\frac{1}{Q_{max\infty}} \left(1 + \frac{r}{r-\hat{r}} e^{b-\hat{r}t} \right) + Ce^{-rt}} \tag{14}$$

$$Rmod(t) = Q_{max\infty} - Q(t) = Q_{max\infty} - \frac{1}{\frac{1}{Q_{max\infty}} \left(1 + \frac{r}{r-\hat{r}} e^{b-\hat{r}t} \right) + Ce^{-rt}} \tag{15}$$

$$P = rQ \left(1 - \frac{Q(1 + e^{b-\hat{r}t})}{Q_{max\infty}} \right) \tag{16}$$

5.5 Maple's Nonlinear Fit for the Cumulative Data

Using the values of $Q_{max\infty} = 5169.805767754234$, $b = 1.204234580691476$ and $\hat{r} = 0.04600654508814379$, and knowing the form of the cumulative function, we can use Maple's Nonlinear Fit to determine the values of the parameters r and C. Imposing parameter ranges ($r = [0, 0.5]$ and $C = [-1, 1]$), values of $r = 0.04214247362704284$ and $C = 0.008570049495435710$ are found to be the optimal parameters. It is then possible to graph cumulative, reserve and production data as well as their time-dependent functions defined in Eqs. 14, 15 and 16. Figures 16 and 17 present these results, where time $t = 1$ corresponds to 1980.



(a) Cumulative data (black) and function with respect to time

(b) Modified Reserve data (blue) and function with respect to time

Figure 16: Cumulative and Modified Reserves

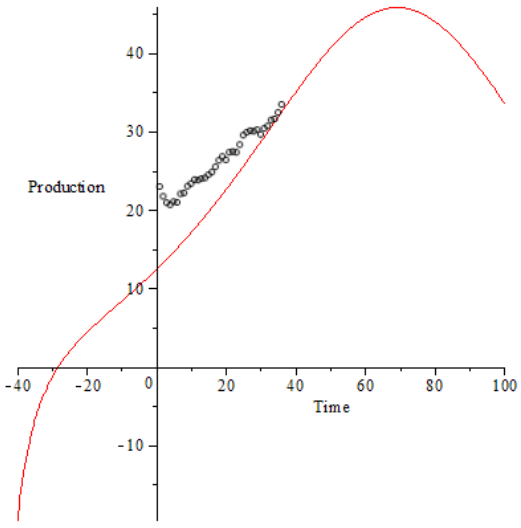
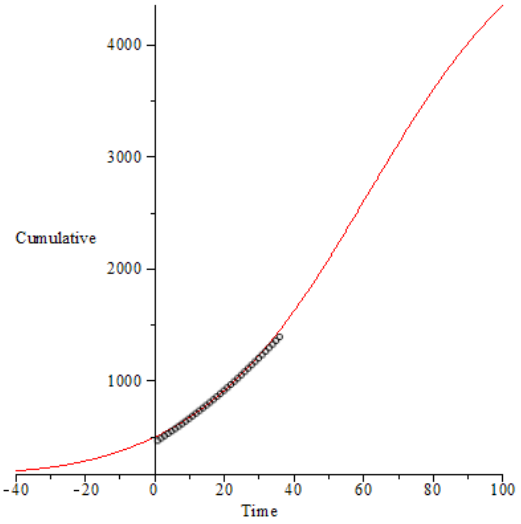


Figure 17: Production data (black) and function with respect to time

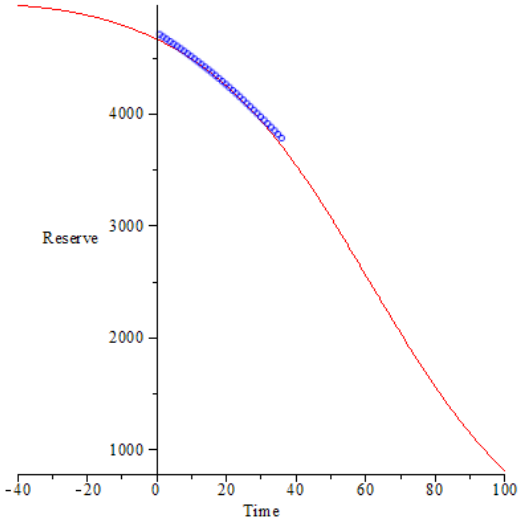
This model fits the second half of the cumulative and modified reserve data acceptably, but it does not fit the first half of this data or the production well at all. Note that in the way that we defined the modified reserve data and the cumulative data, these are directly linked. Hence, if the

model fits one of these data sequences, it will automatically fit the other.

It is important to note that imposing different parameter ranges will produce different optimal parameters. For example, another set of parameters that was obtained is $r = 0.05141658892710616$ and $C = -0.004326009345853979$. Figures 18 and 19 present these results. These parameters appear to fit the data a bit better. However, in both cases, the production model does not appear to fit the data very well. With the first set of parameters, production peaks at around time $t = 70$, so 2049, and in the second case, peak production (of 52.21783410854544 Gbbl) occurs at time $t = 61.391259942$, which corresponds to around 2040. In the second model, reserves would be down to about 33 Gbbl (roughly the same amount as produced in 2015) at time $t = 178$, so in 2157. That being said, according to this model, production decreases after 2040, so the yearly production in 2157 would only be about 1.44 Gbbl.



(a) Cumulative data (black) and function with respect to time



(b) Modified Reserve data (blue) and function with respect to time

Figure 18: Cumulative and Modified Reserves

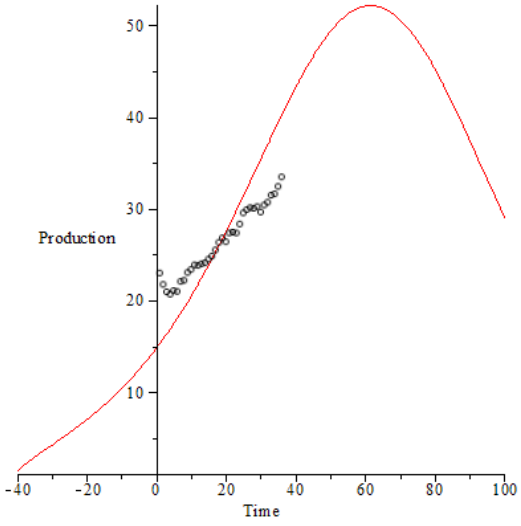
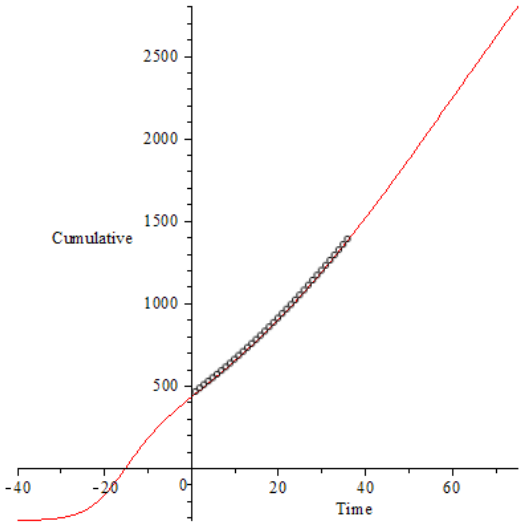


Figure 19: Production data (black) and function with respect to time

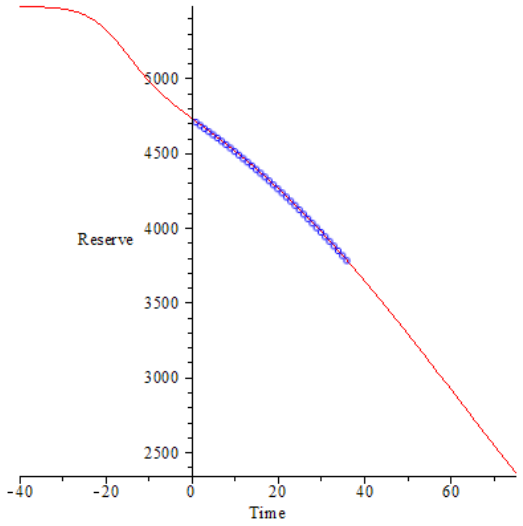
5.6 Maple's Nonlinear Fit for the Production Data

Since these choices of parameters do not fit the production data very well, we decided to fit the parameters to the production data instead of the cumulative data, and then compare these results to those previously found. Using the model defined by Eqs. 14, 15 and 16, and using Maple's Nonlinear Fit command, the following parameters were determined to be the best fit: $Q_{max\infty} = 5169.805767754234$, $\hat{r} = 0.02895916368539971$, $b = 1.635853154191023$, $r = 0.2597530951305678$ and $C = 0.00002360557342466189$.

These parameters fit the production data very well, but they do not fit the cumulative or reserve data. However, a simple shift downwards of 315 Gbbl for the cumulative function, and a shift of 315 Gbbl upwards for the reserve data produces a very good fit. The arbitrary constant can be explained by a constant of integration. Figures 20 and 21 show the adjusted cumulative and reserve graphs as well as the production graph, obtained by adding a constant of -315 Gbbl in the cumulative function.



(a) Cumulative data (black) and function with respect to time



(b) Modified Reserve data (blue) and function with respect to time

Figure 20: Cumulative and Modified Reserves

So far, data from 1980 onwards has been used. There is a decrease in production in the early 1980s which causes the model to peak previously to 1980, as can be seen in figure 21. However, this is one of the problems with this model. It fits the data very well, but the large peak before 1980 does not fit the data. When considering the 1965 onwards data as in figure 1b, there is a peak in oil production in 1979 when production reached approximately 24 Gbbl/year, but this is not nearly as large as the peak in our model of about 37 Gbbl/year. We therefore considered 1965 onwards production data to see if this would produce better results.

5.7 1970s Energy Crisis and 1980s Oil Glut

When analyzing the production data, an unusual pattern is observed in the 1970s and 1980s. Total production increases rapidly until 1979, after which it decreases until 1983, is almost stagnant during the next two years, then appears to start a progressive increase that we classify as normal. When inquiring into this matter, we believe that it can be explained by the 1970s energy crisis and the 1980s oil glut that drastically affected oil prices and production in the world. Figure 22 shows the 1965-2016 production data by region. From this graph, we concluded that the abnormal peak in production in the late 1970s was greatly due to changes in the Middle East's production since

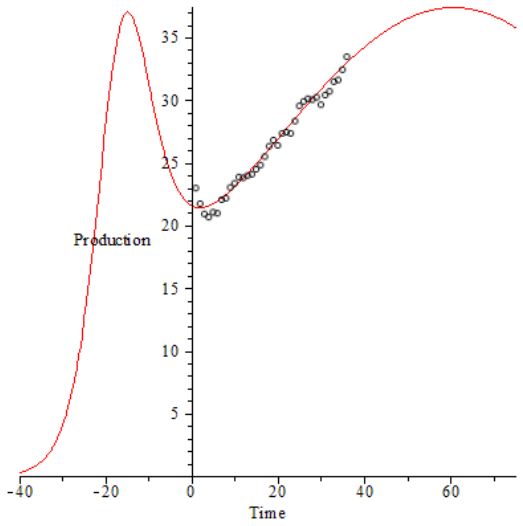


Figure 21: Production data (black) and function with respect to time

the other regions do not appear to follow as abnormal paths as the Middle East.

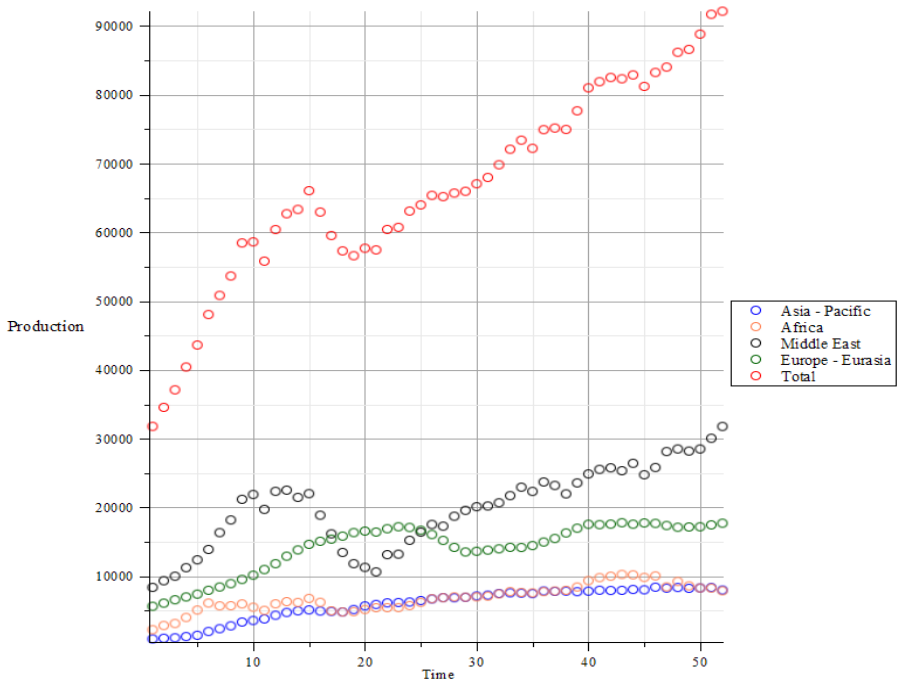


Figure 22: 1965-2016 Production Data by Region

The Organization of the Petroleum Exporting Countries (OPEC) was founded in September 1960 by Iran, Iraq, Kuwait, Saudi Arabia and Venezuela [12]. At the time, the "Seven Sisters" were largely controlling the global oil market [12]. In 1968, the "Declaratory Statement of Petroleum Policy in Member Countries" came into effect, which "emphasised the inalienable right of all countries to exercise permanent sovereignty over their natural resources in the interest of their national development" [12]. During the next decade, OPEC became more powerful and influential on international oil markets, due in part to their rapid increase in production. On October 6th, 1973,

Egypt launched its attacks on Israel and the Yom Kippur War began, the "fourth of the Arab-Israeli wars", that would continue for the next 20 days [10, p. 223]. After the United States and other countries showed support to Israel, the Arab countries of the Organization of the Petroleum Exporting Countries (OPEC) imposed an embargo that started with a five percent cut of oil exports and a doubling of the price, but later developed into a complete embargo. "Clearly, oil was being used as a weapon of political pressure, and very successfully."[2, p. 42] "U.S. imports of oil dropped from 1.2 million barrels to 19,000 barrels per day." [2, p. 18] The United States attempted to keep up with demand through its own production of oil, but was unable to do so. Many countries imposed energy-saving restrictions such as lower speed limits and even "car-free Sundays" in Europe [2, p. 42], and alternative energy sources were starting to be considered with more seriousness. In the early 1970s, oil prices had started to rise, and the price (in US dollars of the day per barrel) jumped from $3.29 in 1973 to $11.58 in 1974 [13]. The embargo ended in March of 1974, but prices were still high [10, p. 228]. However, the times of trouble of the oil industry were not over. In 1979, another oil crisis developed with the Iranian Revolution due to disturbances in Iran's oil production [10, p. 236-237]. Oil prices (in US dollars of the day per barrel) went from $14.02 in 1978 to $31.61 in 1979 [13]. Consumption decreased with the increase in price which then led to a surplus of oil supplies. By 1986, the cost of a barrel of oil in US dollars of the day was back down to $14.43. In the 1980s, OPEC members finally agreed to follow quotas and limit their production, hence helping to stabilize oil production.

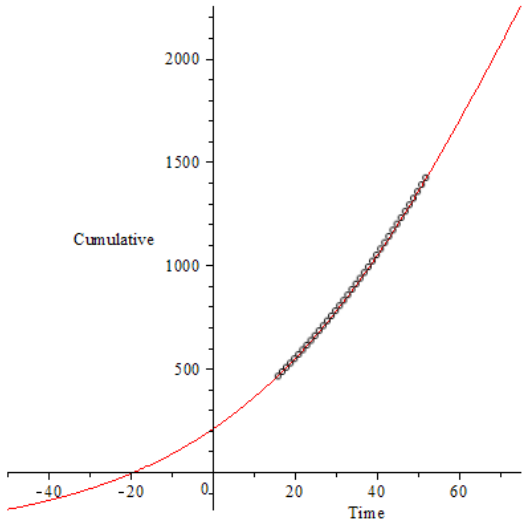
This short summary of the historical events that greatly impacted the production of oil in the 1970s and 1980s leads us to consider these years as having abnormal production data, and for us to construct a model that will not necessarily agree with the data during this time period.

5.8 Production Data Modifications

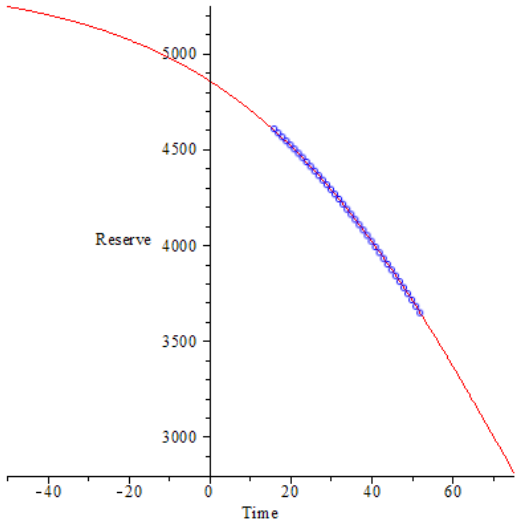
Instead of considering our model to be wrong, perhaps it is the data that is actually misleading since major historical events are not included in our model. Our proved reserves data only begins in 1980, but the BP Statistical Review of World Energy has production data from 1965 onwards. We had previously only been using 1980 onwards data, but we thought that the previous 15 years might help Maple to get a better nonlinear fit for the production. However, doing so did not really change the values of the parameters obtained. Then, for reasons explained in the historical section, we decided to analyze our model and nonlinear fit without putting a lot of importance on the fit in the 1970s and 1980s, and putting more importance into a good fit for the second half of the data. With this in mind, and using the same modified Hubbert model with varying carrying capacity, the following parameters were obtained to be maximal. Note that the parameters that were previously found were used as starting points and the modified reserve data points were generated using the appropriate ultimate reserves. Figures 23 and 24 show the data and the nonlinear model. The arbitrary constant added to the cumulative function was -310 Gbbl this time and the best parameters were found to be: $Q_{max\infty} = 5169.805767754234$, $\hat{r} = 0.02895916368539971$, $b = 1.635853154191023$, $r = 0.2597530951305678$ and $C = 0$.

The cumulative and reserve data fit extremely well, and the production data fits fairly well for the second half of the data, which is, in our opinion, the most valid data. Since the integration constant C is zero, the curve does not peak before 1980 as was the case previously. This makes sense to us since we expect that production would have increased gradually from the time that oil started to be pumped. According to this model, peak production occurs at time $t = 75.57$, with time $t = 1$ being 1965. Therefore, peak production is predicted to occur just after half way through 2039, and production is predicted to be approximately 37.42831 Gbbl at that time. Note that production in 2016 was 33.63475 Gbbl, so according to this model, maximum production will only be about 3.79 Gbbl/year more than current production. Since our models are logistic curves, production and reserve will never reach their limit value of zero. For a reference point though, production will be down to 3 Gbbl/year by 2174.

Something else that we considered was to approximate the data from 1969 to 1982 instead



(a) Cumulative data (black) and function with respect to time



(b) Modified Reserve data (blue) and function with respect to time

Figure 23: Cumulative and Modified Reserves

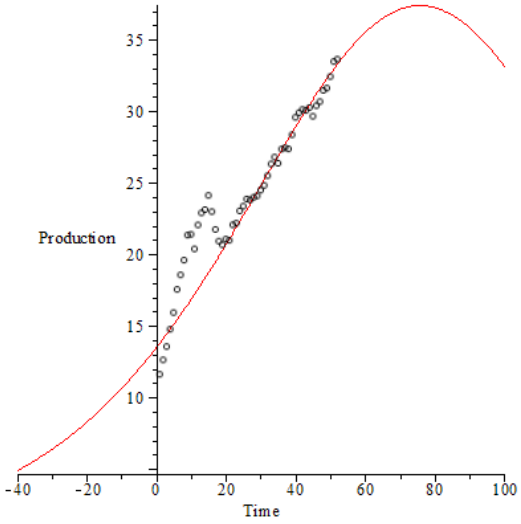
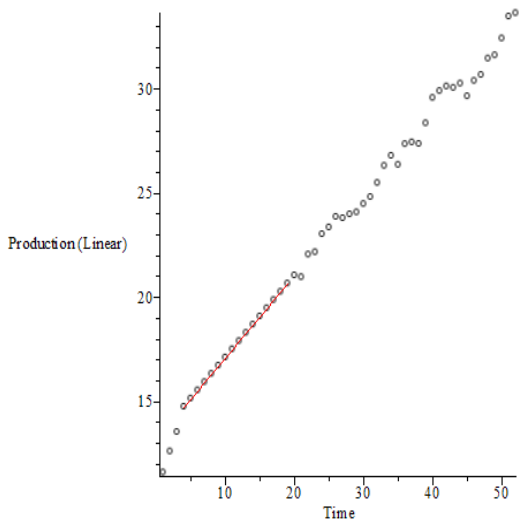
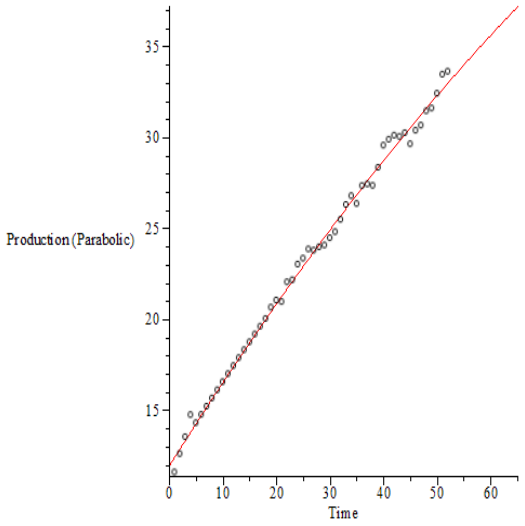


Figure 24: Production data (black) and function with respect to time

of removing it for the nonlinear fit. We started with a linear approximation (as shown in figure 25a, and then we used a quadratic approximation, presented in figure 25b. In other words, we fit a quadratic function to the production data without the 1969-1982 data points, and then replaced this data by points on the parabola. In doing so, we were then able to use Maple's nonlinear fit to find optimal values for all of the parameters. This set of parameters fits the modified production data fairly well and it fits the reserve and cumulative data very well, even though the nonlinear fit was to the production data. An arbitrary constant of -335 was added to the cumulative function to produce the graphs shown in figures 26 and 27, with the following parameters: $Q_{max\infty} = 5121.774404844959$, $\hat{r} = 0.02852232391417219$, $b = 1.997507796623311$, $r = 0.2246625827308103$ and $C = -0.00004824093352972945$. Note that the modified reserve data points were created using this $Q_{max\infty}$ and not the initial value of about 5069 Gbbl.

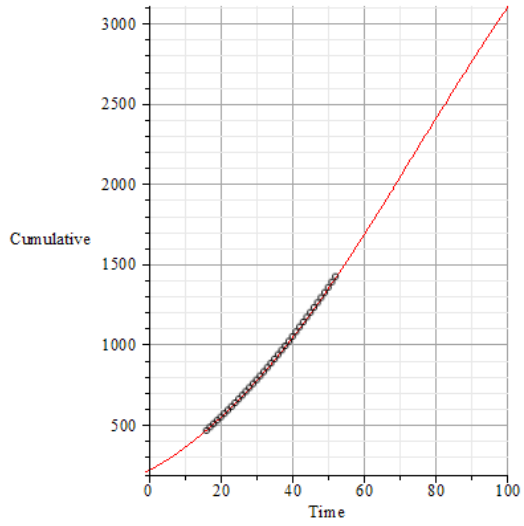


(a) Modified production data using a linear approximation for the 1969-1982 data

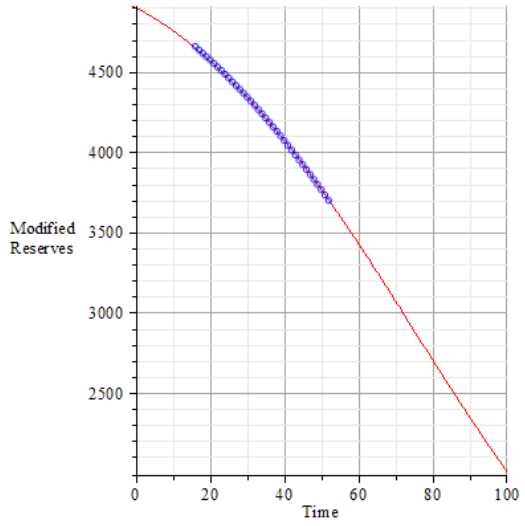


(b) Modified production data using a quadratic approximation for the 1969-1982 data

Figure 25: 1969-1982 Production Data Approximations



(a) Cumulative data (black) and function with respect to time



(b) Modified Reserve data (blue) and function with respect to time

Figure 26: Cumulative and Modified Reserves

This model predicts peak production to occur at approximately $t = 74.8$ which corresponds to around the end of 2038. At that time, production is expected to be about 36.52 Gbbl/year, which is about 2.89 Gbbl/year more than what was produced in 2016.

Since Maple's Nonlinear Fit was limiting the number of major iterations, we manually imposed the maximum number of iterations to be 1319, above which Maple would not produce results since the object was too large. Taking the results that are produced from the largest number of iterations to be the most accurate results, here are the optimal parameters as well as their graphs in figures 28 and 29. $Q_{max\infty} = 5255.928050886900$, $\hat{r} = 0.02805196169698855$, $b = 1.991200028978748$,

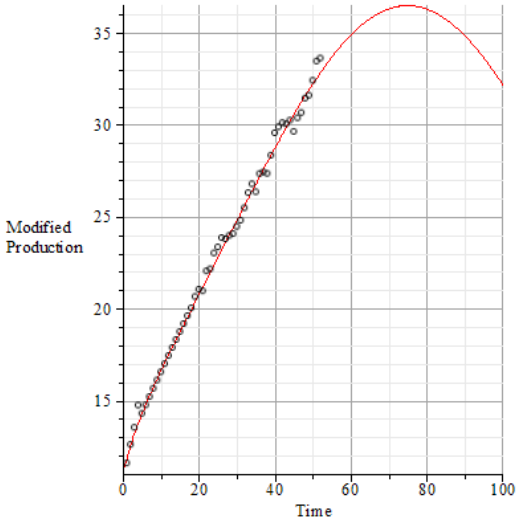
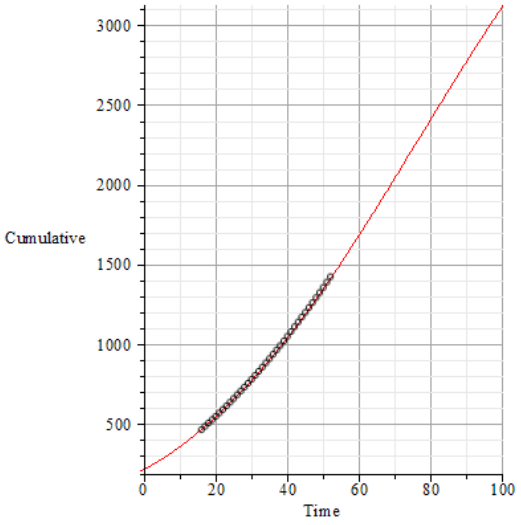
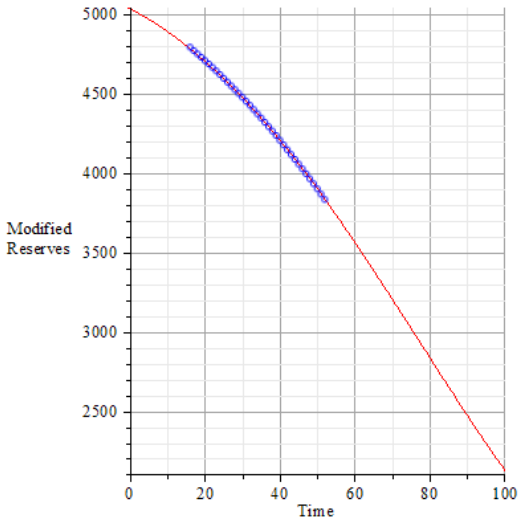


Figure 27: Modified production data (black) and function with respect to time

$r = 0.2093228582783033$ and $C = -0.00005110988626128131$.



(a) Cumulative data (black) and function with respect to time



(b) Modified Reserve data (blue) and function with respect to time

Figure 28: Cumulative and Modified Reserves with 1319 iterations

These parameters produce very similar results to those presented in figures 26 and 27. Peak production occurs at around $t = 76.11$, or at the beginning of the year 2040, and production at that time is expected to be approximately 36.86 Gbbl/year. That is about 3.23 Gbbl more than current production. $Q_{max\infty}$ is about 5256 Gbbl, which is slightly larger than what was found in section 5.2.1 with the nonlinear fit to the Q_{max} points, but only by less than 200 Gbbl.

We consider these last results to be the most accurate model presented in this paper. We will therefore present an analysis of the error associated with this model. We wish to quantify the estimation error between the actual data points (or modified data points in some cases) and the model's predicted values. We calculated the sum of squared error (SSE), the mean squared error

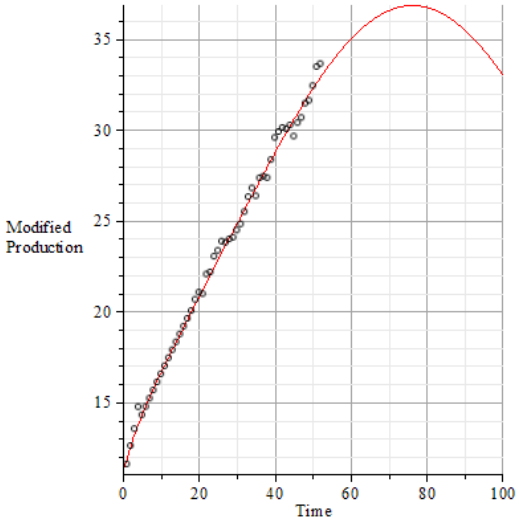


Figure 29: Modified production data (black) and function with respect to time, with 1319 iterations

	SSE	MSE	RMSE
Cumulative	327.00155	8.83788	2.97286
Modified Reserves	327.00155	8.83788	2.97286
Production	8.66461	0.16663	0.40820

Figure 30: SSE, MSE and RMSE Error Estimates

(MSE) and the root-mean-squared error (RMSE) for all three quantities, cumulative, modified reserves and production, and they are presented in figure 30. It makes sense that the error for the cumulative and the modified reserves is the same since the modified reserves data is simply a mirror image of the cumulative data. These quantities were calculated as shown in Eqs. 17, 18 and 19, where Y_i is the data point and $y(i)$ is the function evaluated at time $t = i$. There are 37 data points for cumulative and modified reserves whereas $n = 52$ for production.

$$\text{SSE} = \sum_{i=1}^{n} (Y_i - y(i))^2 \tag{17}$$

$$\text{MSE} = \frac{\text{SSE}}{n} = \frac{\sum_{i=1}^{n} (Y_i - y(i))^2}{n} \tag{18}$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - y(i))^2}{n}} \tag{19}$$

We also considered the mean absolute error (MAE). As its name indicates, the MAE calculates the average distance from the data point to the predicted value of the function. It was calculated as shown in Eq. 20 and figure 31 presents the error estimates for the three functions. This error estimate as well as the MSE were calculated in order to give an idea of the magnitude of the error, but since they are both scale-dependent, they cannot be compared with the error of other quantities [7].

$$\text{MAE} = \frac{\sum_{i=1}^{n} |Y_i - y(i)|}{n} \tag{20}$$

	MAE
Cumulative	2.19416
Modified Reserves	2.19416
Production	0.31724

Figure 31: MAE Error Estimates

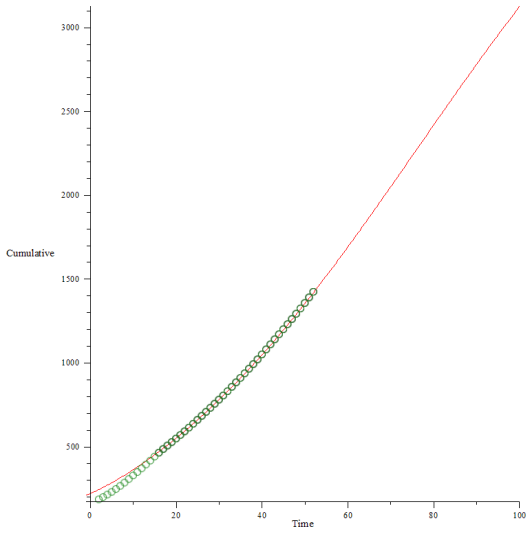
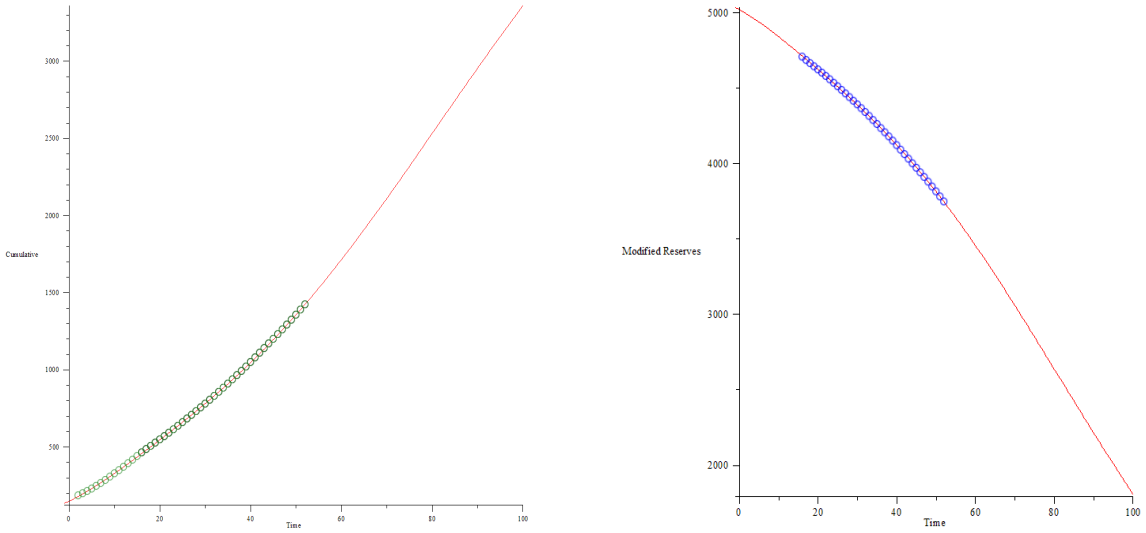


Figure 32: 1966-2016 Cumulative Production

Since Gallagher [4] also provides cumulative production data from 1966 to be 185.226 Gbbl, we decided to see if our model fits the previous cumulative data. When doing so, we found that our model predicts higher cumulative from 1966 to 1979 than the actual data points indicate. Figure 32 presents these results with the data points and the model. Time $t = 1$ corresponds to 1965. However, there is no drastic deviation from the model; it gradually leaves the curve. Also, we consider the more recent data to be the most accurate, so we will not consider this deviation from the model to be a major drawback, especially seeing as the model was not created with this earlier data.

That being said, we considered fitting the model to this extended set of cumulative data, and we used Maple's Nonlinear Fit to determine the best parameters for the cumulative data. We were obliged to impose one of the parameters and only leave four free, so we chose to set the ultimate reserves to be about 5170 Gbbl. We considered this parameter to be the one which was the best known based on our previous models. Doing so produced very good fits for the cumulative and modified reserves data, but a less good fit for production, as shown in figures 33 and 34. These parameters fit the second half of the production data fairly well though, and it vaguely approximates the production data during the unusual peak around 1979. However, it consistently overestimates the data from 1965-1972. According to this model, peak production will occur around $t = 81.153075$, during the year 2045. This is five years later than our previous prediction, but it is still around the same time frame. However, peak production is expected to be approximately 42.58 Gbbl/year, which is more than 5 Gbbl/year larger than our previous model. This model therefore predicts that production will increase by nearly 9 Gbbl from now until peak production in 29 years, which is a lot more than the 3.23 Gbbl increase in the previous model. From 1988 to 2016 (also a 29-year time period), production increased by 10.6 Gbbl. One might expect a similar increase during the next 29 years as well, or production increases might diminish while approaching peak oil and the 3.23 Gbbl increase might be plausible.

We then calculated the error associated with this model in order to compare it to the previous



(a) Extended cumulative data (green) and function with respect to time

(b) Modified Reserve data (blue) and function with respect to time

Figure 33: Extended Cumulative and Modified Reserves

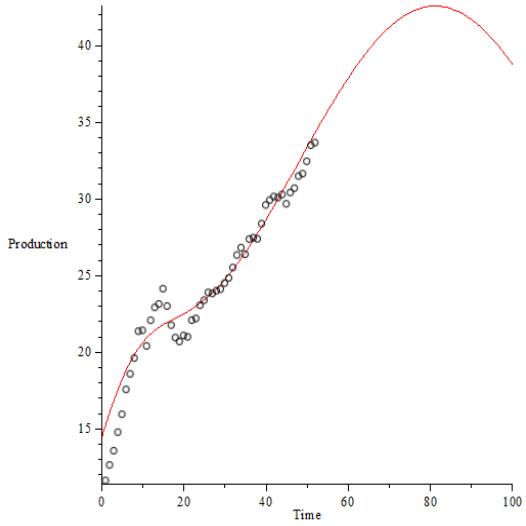


Figure 34: Production data (black) and function with respect to time

model's error. Since this model was obtained from a fit to the cumulative data instead of the production data, we expect the cumulative and modified reserves to have a smaller error whereas production would have a larger error than the previous model. Also, we now have more cumulative data points than modified reserves, so the error estimates of these quantitiese will no longer be the same. Figure 35 presents the SSE, MSE and RMSE error estimates.

We also calculated the MAE for the cumulative, modified reserves and production models. Figure 36 presents these error estimates. As was the case with the SSE, MSE and RMSE error estimates, the MAE for the cumulative and the modified reserves decreased whereas it more than tripled for production. In general, with the model for the extended cumulative data, the error increased proportionally more for production than it decreased for the other quantities, so we will consider our other model with 1319 iterations to be the best fit.

	SSE	MSE	RMSE
Cumulative	316.87893	6.21331	2.49265
Modified Reserves	204.20466	5.51904	2.34926
Production	90.56073	1.74155	1.31968

Figure 35: SSE, MSE and RMSE Error Estimates

	MAE
Cumulative	2.08577
Modified Reserves	1.92445
Production	0.97541

Figure 36: MAE Error Estimates

One might think that the increase in the production error is only due to the earlier data. Although it is true that the model deviates the most from the data in the earlier half of the data, the latter part of the data also has an increased error. Taking into account only the data from 1983 to 2016 and ignoring the earlier data, the RMSE goes from 0.466 for the production fit with 1319 iterations to 0.799 for the extended cumulative fit; hence, there is still an increased error for the later data as well. Also, we used the modified production data for the production fit model, and we replaced some of the data by points with no noise, hence reducing the error quite a bit. In fact, for the data from 1969 to 1982 that we replaced with a quadratic approximation, the MAE is 0.11 whereas it is 1.14 for the same time period in the extended cumulative model with the original production data. Also, the extended cumulative fit with the quadratic approximation data has a MAE of 3.56 from 1969 to 1982 and a MAE of 1.63 from 1965 to 2016, which is why we chose to use the original production data in that model.

Since production data and reserves data have a large range of values (different orders of magnitude) we also considered the mean absolute percentage deviation MAPD, or mean absolute percentage error MAPE which is a scale-independent error estimate (see Eq. 21). The production fit with 1319 iterations has a MAPD of 1.35 whereas the extended cumulative fit with the original production data has a MAPD of 5.07, more than three times larger than the production fit error. This also contributes to justifying our choice of the production fit model. However, we must keep in mind that we are using modified reserves and production data instead of our original data. Moreover, the MAPD has a few drawbacks. The first is that it becomes undefined if there is a zero data point. That being said, no such data points are in our data set, so this should not be an issue. Hyndman adds that another problem with the MAPD is that it "puts a heavier penalty on positive errors than on negative errors" [7, p. 45].

$$\text{MAPD} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{Y_i - y(i)}{Y_i}\right| \tag{21}$$

We also considered the mean absolute scaled error (MASE), a more recent error estimate by Hyndman and Koehler (2006) [7] calculated as shown in Eq. 22. This error estimate has the advantage that it avoids some of the issues of other error estimates. It is scale-independent and it is not affected by zero values. In fact, the only situation where the MASE would encounter a problem is if all data points were the same value [7, p. 46]. The 1319 iteration production fit has a MASE of 0.63197 whereas the extended cumulative fit has a MASE of 1.47870, more than double the production fit error. Hyndman claims that MASE values are usually smaller than one [7, p. 46]. This again supports our choice of the 1319 iteration production fit as being the better model.

$$\text{MASE} = \frac{\sum_{i=1}^{n} |e_i|}{\frac{n}{n-1}\sum_{i=2}^{n} |Y_i - Y_{i-1}|} \tag{22}$$

5.9 Comparisons with Other Models

Now that we have our results, we can compare them to other predictions presented in the litterature. Gallagher presents the year 2014 as an average of 12 previously published peak oil predictions and 2011 as the median year [4, p. 790]. This author also presents his results of peak production occuring in 2009 with production being 30.16 Gbbl in that year and the URR being 2240 Gbbl [4, p. 800]. That being said, this paper was published six years ago in 2011, and production data since 2009 does not indicate that peak oil has occurred yet. Another topic that Gallagher discusses is an asymmetrical model that would have a sharper decline than incline. This is also suggested by Berg and Korte [1, p. 228] and it could be due to developing technologies that greatly increased production before peak oil that would then lead to a lack of resources after peak production [4, p.700-791]. This is something that we could further consider. Our models have been constructed based on the data leading up to peak production and have looked only a little at what path production would follow after peak oil. Gallagher's results differ from ours quite significantly. The URR is less than half of our value for the reserves, and our peak oil prediction is more than 30 years later. Furthermore, Maggio and Cacciola [11] also presented different results than what we obtained. According to their predictions, peak oil would occur somewhere between 2009 and 2021 with production peaking between 29.2 and 31.6 Gbbl/year. We have already surpassed these production levels with production reaching 33.6 Gbbl in 2016. Their estimate for the URR is between 2250 and 3000 Gbbl. These results were published in 2012 and are very similar to Gallagher's results, but they have a larger range for the occurence of peak oil. Thus, peak oil could still possibly occur in this time period if production were to peak within the next five years. Finally, Berg and Korte [1] present peak oil predictions to be from 2005 to the 2030s, which is also sooner than our predictions.

6 Conclusion

Our models all predict different occurences of peak oil production, but many suggest that it will occur in the not-too-distant future. The path that the phase plane of oil reserves vs production will follow is hard to predict though. Some models predict a hard landing, others show us heading off to a fixed point and many predict a looping behaviour. The many unknowns, political, economical and environmental just to name a few, make oil production a difficult topic to model. Our models have included several parameters in hopes of encompassing various social and environmental influences, but they are in no way comprehensive. That being said, based on our models, we predict peak oil to occur somewhere around 2040, and for production to be approximately 36.86 Gbbl/year at that time. Based on this model, production and reserves will gradually decrease without ever running out of oil, thus creating a happy ending to the age of oil.

References

[1] P. Berg and S. Korte. Higher-order Hubbert Models for World Oil Production. *Petroleum Science and Technology*, 26:217–230, 2008.

[2] Giovanna Borasi and Mirko Zardini, editors. *Sorry, out of gas : architecture's response to the 1973 oil crisis*. Canadian Centre for Architecture, Montreal, 2007.

[3] Adam R. Brandt. Review of mathematical models of future oil supply: Historical overview and synthesizing critique. *Energy*, 35:3958–3974, 2010.

[4] Brian Gallagher. Peak oil analyzed with a logistic function and idealized Hubbert curve. *Energy Policy*, 39:790–802, 2011.

[5] Franklin Mendivil, Herb Kunze, Davide LaTorre and Edward R. Vrscay. *Fractal-Based Methods in Analysis*. Springer, New York, 2012.

[6] M. King Hubbert. Nuclear Energy and the Fossil Fuels. Technical Report publication no. 95, Shell Development Company, Exploration and Production Research Division, Houston, Texas. June 1956.

[7] Rob J. Hyndman. Another Look at Forecast-Accuracy Metrics for Intermittent Demand. *Foresight*, 4:43–46, June 2006.

[8] H. Kunze and K. Heidler. The collage coding method and its application to an inverse problem for the Lorenz system. *Applied Mathematics and Computation*, 186:124–29, 2007.

[9] H. E. Kunze and E. R. Vrscay. Solving inverse problems for ordinary differential equations using the picard contraction mapping. *Inverse Problems*, 15:745–770, 1999.

[10] Kwasi Kwarteng. *War and gold : a 500-year history of empires, adventures, and debt*. PublicAffairs, New York, 2014.

[11] C. Maggio and G. Cacciola. When will oil, natural gas, and coal peak? *Fuel*, 98:111–124, 2012.

[12] Organization of the Petroleum Exporting Countries. Brief History. `http://www.opec.org/opec_web/en/about_us/24.htm`. Accessed: 2017-06-26.

[13] British Petroleum. BP statistical review of world energy - underpinning data, 1965-2016, June 2017.