

# Oil Production: A probabilistic model of the Hubbert curve.

Bertrand Michel\*

May 12, 2010

The specific distribution of hydrocarbon field sizes has a deep impact on the dynamics of the production inside a basin. This paper proposes a probabilistic model based on the field size distribution, taking into account the launching production process of the fields. This model can be seen as a “bottom-up” probabilistic method to simulate and predict oil production. The asymptotic results obtained allow us to propose relevant fitting to real production curves.

*Keywords* : Oil Production model, Oil reserves, Hubbert Curve, field size distributions.

## 1 Introduction

Estimating the amount of oil available in a basin is one of the most important debates in energy economics. Today, experts do not agree on the future of the world oil production. Even the date when the world peak oil will occur is the subject of much debate as was shown in Babusiaux [1]. At the end of the 1950s, the geologist K. Hubbert proposed to extrapolate the hydrocarbon production of the USA with a bell shape curve: the *Hubbert curve*. His prediction published in [2] turned out to be true: the maximum of the USA production was reached in 1970. His methodology however, did not convince all of the experts, especially in the economic community.

---

\*Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie, 175 rue du Chevaleret 75013 Paris, France. Tel : +(33) 1 44 27 85 68 Fax : +(33) 1 44 27 33 42, email : bertrand.michel@upmc.fr

Since Hubbert, the choice of an adequate family of curves to obtain good fittings on real production curves has been a recurrent question, see [3] for a recent work on this topic. The aim of our study is to specify the shape of aggregations of production profiles. More than giving accurate predictions on production, the main contribution of this paper is to give a mathematically justified method to fit specific curves on production profiles. Indeed this family of curves is obtained by probabilistic modelling that mimics the dynamics of the production inside a basin.

Since the Hubbert forecast, many econometric models have also been proposed to study the production of hydrocarbons in a basin. For instance, [4] and [5] propose production models based on times series. In [6] and [7], the production models combine the curve fitting approach of Hubbert with economics considerations too. The methods detailed in these papers attempt to highlight the relationships between production and explicative variables. These models, however, do not take into account the field size distribution in a basin. Now, oil is concentrated in conventional discrete fields, when mineral and unconventional oil and gas accumulations are sprayed on large areas. It therefore appears natural to us to build a model that is based on the following three elements:

- a description of reserves (field size distribution) ;
- the dynamics of the production ;
- and a model for field production.

The decision to launch a field depends heavily on its size. Thus, the distribution of oil reserves in a basin is also a key constraining factor in production. This new probabilistic model of the exploitation of a basin with  $n$  fields inside may be summarized as follows:

1. The field sizes  $(Y_1, \dots, Y_n)$  of the  $n$  fields have a Pareto distribution.
2. The date  $S$  at which each field starts to produce has a gamma distribution that depends on its field size  $Y$ .
3. Each field produces oil according to the following model:

$$\text{prod}(t, y) = \frac{y}{v(y)} K \left( \frac{t}{v(y)} \right)$$

where  $y$  is the field size and  $t$  is time. Function  $K$  is an *elementary profile* and  $v$  is a function which allows the profile shape to be deformed as a function of the field size.

4. The production of the basin is the sum of the individual productions.

The organization of the paper is as follows: in Section 2, the main criticisms of the Hubbert curve are given. This motivates the probabilistic model for basin production which is detailed in Section 3. To provide extrapolation of true productions, an asymptotic result of the model is given in Section 4. In Section 5 we study the production curve with simulations and Section 6 is devoted to an application on the North Sea.

## 2 Criticism of the Hubbert curve

### 2.1 The Hubbert curve

K. Hubbert published *Nuclear energy and the fossil fuels* in 1956. In this first work, he extrapolates the hydrocarbon production of the USA and of the world with a bell shape curve: the *Hubbert curve*. Actually, in [2], Hubbert only presents some graphical methods whereas he gives a more detailed methodology in [8]. The main idea of the Hubbert curve method is to fit a bell shaped curve on the production curve. Although Hubbert does not particularly defend the use of symmetrical curves, logistic or Gaussian curves have been currently used in practice to fit oil production (see [9] or [10] for instance). The area under the curve is exactly the amount of ultimate reserves. In a specific area, the ultimate reserves are the total oil that will be produced during the exploitation and, of course, this amount of oil can only be estimated. Thus, the Hubbert method is no more than the fitting of a bell shaped curve on the production curve, with an area above the curve imposed by the estimate of ultimate reserves.

The cumulative curves of discoveries and production usually have the same shape, with a constant delay  $\Delta$  between the first and the second. Since the discovery curve is better known than the production curve, the fitting is made on the first curve.

In [8], Hubbert proposes to model the cumulative discoveries of hydro-

carbons with a logistic curve

$$Q_D(x) = \frac{Q_\infty}{1 + a * e^{-bx}}, \quad (1)$$

Where  $Q_\infty$  and  $Q_D(t)$  respectively represent the ultimate reserves of the basin and the cumulated discoveries before  $t$ . As in (1),  $Q_D$  fulfills the logistic equation

$$\dot{Q}_D = r * Q_D \left(1 - \frac{Q_D}{Q_\infty}\right).$$

This is an elementary population model: the growth of the population is proportional to the product of the size of the population and the available resources. Taking the logarithm into (1) with oil production notations leads to

$$\ln \left( \frac{Q_\infty}{Q_D(t)} - 1 \right) = \ln(a) - bt. \quad (2)$$

Note that it is important to efficiently estimate  $Q_\infty$  to make a correct linear fitting on the time series  $\left(\ln\left(\frac{Q_\infty}{Q_D(t)} - 1\right)\right)_t$ . In his 1962 report, Hubbert took  $Q_\infty = 170\text{Gb}$  and proposed for the U.S. [see 8, p59] :

$$Q_D(t) = \frac{170 * 10^9}{1 + 46.8 * e^{-0.0687*(t-1900)}}.$$

Thus, the cumulated production is

$$Q_P(t) = \frac{170 * 10^9}{1 + 46.8 * e^{-0.0687*(t-1910.5)}}.$$

Finally, the peak is easy to evaluate since the production is the derivate of  $Q_P$ . As we saw, a correct estimate of the ultimate reserves is necessary. Hubbert was relying on proved reserves. Unfortunately, it is well known that summing the proved reserves of each field grossly underestimates the proved reserves of a country. Today, it is believed that 200Gb is a good estimate and with this estimate, the peak actually happens in 1970.

After Hubbert, other experts such as Deffeyes [10] have proposed applications and refinements of this method. Bartlett [9] proposes an analysis of U.S. and world oil production with Hubbert curves and more recently, Broto [11] has proposed a multi-logistic model in which the production is considered as a sum of independent logistic cycles.

## 2.2 Discussion about the Hubbert curve

There is a considerable amount of controversial literature (books, articles, websites...) on peak oil. Hubbert's method needs good estimates of the amount of oil trapped underground and this causes a confusion between the Hubbert's method and the debate on the remaining world reserves. These are two different problems and here our aim is to discuss Hubbert's method without taking part in the debate between the optimistic and pessimistic point of view. See [1, 12, 13, 14, 15, 16] for more information on peak oil and reserves debate.

Hubbert's method is easy to implement since it only requires a least square computation to estimate the parameters of a logistic or Gaussian curve. However, no element related to the dynamics of oil production justifies such a model. Hubbert recognizes himself that his methodology is empirical and as far as we know, no rigorous justifications for using this methodology have been proposed. One difficult issue in using the Hubbert curve is to choose a convenient curve for the fitting. For instance, are symmetrical curves realistic enough? In Section 4, we show rigorously that the shape of a basin production depends on how the fields are gradually launched. The way the fields are gradually developed in a specific area does not present any symmetry: the biggest fields are produced first. In fact, the oil companies try to delay the peak to mitigate the decline in an area's production, and this phenomenon has no "symmetry".

Due to the Central Limit Theorem (CLT), it is sometimes claimed that the production curve is approximatively a Gaussian curve. As explained in [17] this is not a convincing justification. The CLT says that under reasonable conditions, the distribution of a sum of random variables tends to a Gaussian distribution. In our context, an individual profile would play the part of the distribution and the basin profile would be the sum of the individual profiles. But the distribution of a sum of distributions is not the sum of the distributions; it is a convolution of them. Thus, the CLT can not be invoked to justify the choice of a Gaussian curve, even if Gaussian curves provide a good fit to the production profiles.

In the report [8] written in 1962 for the National Academy of Sciences Hubbert proposes justifications for the use of logistic curves claiming that the situation could be compared with demographic models. But this analogy with the theories of population growth is not credible. What could be the

interpretation of the logistic equation

$$\dot{Q}_D = r * Q_D(1 - Q_D/Q_\infty) ?$$

Finally, in [18] a quantity called *yield per effort* (YPE) is introduced to estimate the ultimate oil recovery in a studied area. It is defined as the amount of oil that is discovered per foot of well drilled. Fields that hold a large amount of oil are supposed to be found first. Thus the YPE should always be decreasing for all basins (assuming there is only one cycle). Of course, as shown in [6], in reality this is not exactly the case. Only probabilistic models can overcome this difficulty relaxing this assumption. For instance, in [19] a discovery process is proposed to model the sequence of the successive discoveries.

Even if Hubbert's fitting is able to propose extrapolations of oil productions, it is clear that a stronger model is necessary to explain the shape of the basin production profiles.

### 3 A probabilistic model for oil production

In this section we describe a new probabilistic model for oil production. It is based on the field size distribution, the dynamics of production and the shape of the individual profiles of the fields.

#### 3.1 Main Hypothesis

Oil production depends on many phenomena. See [17] for an exhaustive description of hydrocarbon exploration and production. Since all the factors influencing the production can not be taken into account, our model depends on the following hypotheses:

1. The most important factor for the production is the geological context, namely the field size distribution of the basin.
2. The area is a geological homogeneous area.
3. Production does not mix various hydrocarbons.

Several elements can be propounded to validate the first hypothesis. As we said before, oil concentration is not constant everywhere over the area.

Furthermore, due to the specific distribution of field sizes, the largest layers tend to be discovered and produced first. Thus, we believe that the geological realities constrain production so that the field size distribution must be privileged. However, economic parameters that influence the dynamics of production could be taken into account as well.

The second hypothesis is natural. First, a model based on field size distribution needs a geological homogeneity. Estimations procedure would fail to mix basins with dissimilar geological structure, and it would be hazardous to aggregate productions of areas exploited at different times, even in a same country. Note that petroleum databases are usually specific to the country, implying additional work to create basin databases. One convenient geological unit is the Petroleum System defined by the source-rock but this genetic classification is fairly new, while the old tectonic classification of geological basin is still widely used.

The last hypothesis is connected with [20] and [11], which propose to decompose the production as a succession of various cycles. For instance in the gulf of Mexico, the recent development of “deep oil” cannot be compared with the extraction of oil in the same place in the past since the two dynamics are different.

### **3.2 Field size distribution**

Field size distribution has been the subject of much study during the last decades. Allais [21] is the first to use Lognormal distributions for mineral resources and Kaufman [22] uses this distribution for a population of oil or gas fields in a petroleum basin. Later, in [19], a probabilistic model of hydrocarbon discoveries based on size-biased sampling on a sample of Lognormal or truncated Lognormal distributions is proposed. After Zipf [23], Pareto distributions have been commonly used too. For instance, in [24] and [25] field sizes are modeled with Truncated Shifted Pareto distributions. More recently, Lepez [26] also based his works on reserves estimation on the Pareto distribution. It is well known that beside an important layer, smaller “satellites” are usually discovered and the Pareto distribution is consistent with this kind of structure. More flexible power law distributions have also been proposed by Laherrère in [27, 28].

As it is explained in [19], it can be sometimes difficult to distinguish between Lognormal and Pareto distributions. The important implications on the reserve estimations of this choice has been underlined in [29]. Con-

sequently, which distribution should be chosen? An interesting compromise has been proposed by Lepez in [30]. The Lognormal distribution gives a good description of the discovered fields whereas the Pareto distribution models the field size distribution of the the whole reservoir. The exploration process first discovers the larger fields with a higher probability. Consequently, the small fields are under-represented in the discoveries and the empirical distribution of discovered fields under-estimate the proportion of small fields. We use a truncated Pareto distributions for the field size distribution. For the North Sea example we focus on in this paper, Lepez has shown that they efficiently fit discovered field sizes. Thus, a shift transformation is not necessary and fewer parameters have to be estimated. Note that our basin production model could be used with other field size distributions (lognormal or other power law distributions).

The Pareto distribution  $\mathcal{P}ar(\alpha, \underline{y})$  of parameter  $\alpha$  and threshold  $\underline{y}$  is the distribution with density:

$$h_{\alpha, \underline{y}}(y) = \frac{\alpha \underline{y}^\alpha}{y^{\alpha+1}} \mathbf{1}_{y > \underline{y}} .$$

The threshold  $\underline{y}$  can be interpreted as the ultimate level of profitability and visibility of the fields. Indeed, installations are very expensive, and so the field must contain enough oil to be profitable.

The Pareto distribution is also characterized by the fact that statistics of the order of a sample of independent random variables with this distribution have a linear trend in a loglog diagram with a slope of  $-\frac{1}{\alpha}$ . Figure 1 is the loglog-diagram of the North Sea Area. Observations of the largest fields respect this linearity, but the smaller fields do not. The under-representativeness of the small objects results in the bending of the curve on the right-hand side. Oil exploration can be seen as and the result of a sized-biased random sampling inside a population with a Pareto distribution. The largest fields are found first while only a short fraction of small ones have been discovered. One of the main difficulties in estimating the potential reserves of the basin is to correctly describe this bias.

If the basin is mature enough, the largest fields have all been discovered. This upper bound  $\bar{y}$  on the size of the fields means we can use *truncated Pareto* distributions of density

$$h_{\alpha, \underline{y}, \bar{y}}(y) = \frac{1}{\underline{y}^{-\alpha} - \bar{y}^{-\alpha}} \frac{\alpha}{y^{\alpha+1}} \mathbf{1}_{\underline{y} < y < \bar{y}} .$$

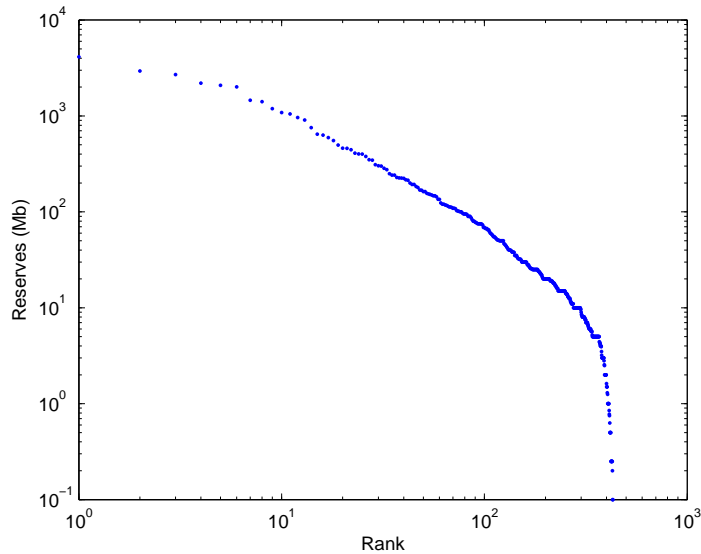


Figure 1: LogLog diagram for the sorted sizes of the North Sea fields: field size versus the size rank in decreasing order (sources: Wood Mackenzie 2003).

However, Houghton [24] also prefers to use truncated distributions to improve fittings. Taking into account the upper limit on the field size, we are now dealing with integrable distributions. This facilitates the study of the asymptotic behavior of the production in section 4. Let  $\mathcal{P}\text{ar}(\alpha, \underline{y}, \bar{y})$  be this truncated Pareto distribution.

The following model is considered. A basin is assumed to be composed of  $n$  fields of sizes  $(Y_1, Y_2, \dots, Y_n)$ . The random variables  $Y_i$  are independent with the same Pareto distribution  $\mathcal{P}\text{ar}(\alpha, \underline{y}, \bar{y})$ . At each moment  $t$  in the history of the basin development,  $\tilde{N} = \tilde{N}(t)$  fields have been discovered and only  $N = N(t) \leq \tilde{N}(t)$  of them have started to produce.

### 3.3 Dynamics of the production

In order to model the production of an oil basin, a good understanding of the mechanisms which explain the way fields are exploited is required. Oil basins are gradually exploited according to world demand and the ease of extracting oil in the considered area. In some places, oil traps were easy to locate, for instance in the United States and in the Middle-East where production began

a long time ago. On the other hand offshore production of the North Sea only began in the 70's. Thus, accessibility is an important factor of hydrocarbon production, but will be considered as constant everywhere in the basin.

Local taxation also influences oil exploration and production. Indeed, oil companies have to get the authorization to carry out exploration and production activities. Except in the United States, the basin is a state-owned property. States can also launch exploration campaigns supported by their national companies. To simplify matters, we will not integrate a taxation variable in this modeling work.

There is a fundamental and well-known principle in oil exploration: "*big stuff gets found first*", i.e. the largest layers are easy to find, are thus discovered first. Furthermore, geologists have to find the largest layers to declare the zone profitable to produce. In addition, the development of an oil producing area is a very ambitious financial undertaking, with only large companies able to assume the risk. Although the exploration stage is not the most expensive, it weighs heavily on the accounts of the companies; an exploration campaign is in fact a substantial investment. Consequently, to avoid a financial catastrophe, companies must initially develop the largest layers. For these geological and financial reasons, the production of a basin generally begins with the exploitation of the most significant fields. Then, the network of pipelines is developed from these first producing fields. In the North Sea, the Brent, Ninian, Forties and Oseberg layers were launched at the beginning of the production in the 70s. Generally, after a few years, the first fields reach their maximum, and soon begin to decline. Companies then begin to exploiting smaller layers which are located around the zones already in production.

It seems reasonable to base the model of the gradual launching of fields inside the basin on the dates of discovery. The left plot of Figure 2 shows the discovery process for the North Sea area. To define the production start-up date distribution, it is necessary to know what is the distribution of the lag time between discovery and the beginning of production. Distribution of such time lags have been studied by Lucki and Szkutnik [31] using time series. However, their results do not help in our context since our model describes the development of each field whereas these authors are interested in the total discoveries and productions in a specific area. The right plot of Figure 2 shows this delay as a function of the field size. The launch date of the production of a modest field can depend on the proximity of the pipeline network, or of an infrastructure making it possible to evacuate the production

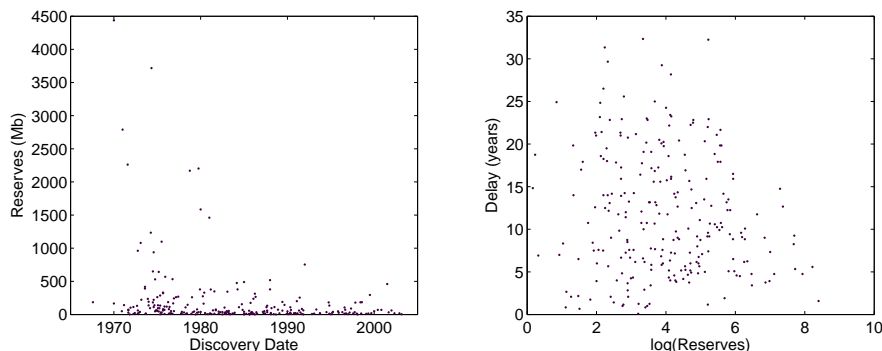


Figure 2: Discovery process (left) and delays before the production starts (right) in the North Sea fields (sources: Wood Mackenzie 2003).

at a reasonable cost. As the plot shows, it is difficult to correctly explain the time between discovery and the beginning of production by the size of the field alone. Since this is the only explanatory variable that we want to retain, we will directly study the dates on which the productions are launched.

At each time, the stock of available fields for production is composed of discovered fields which are not already producing. Since the largest fields are found first, they are in stock first and hence also produce first. As a consequence, the distribution of the launching production date  $S$  of a field depends of its size  $Y$ . The joint distribution of  $(S, Y)$  represents the dynamics of the production. Figure 4 presents the histogram of  $S$  by class of sizes for the North Sea. One can observe a common general shape for the density, with a displacement to the right of the maximum when the size of the fields decreases. Indeed, for each class of a given size corresponding to a period of production during which those fields will be developed in priority determined by size.

The previous description of the launching process can be compared with a queuing problem. Indeed, fields are gradually launched in the same way as people are served at a counter. Thus, the distribution of  $S$  knowing  $Y$  can be considered as a lifetime distribution. In reliability theory (see [32] for instance), Gamma distributions are commonly used to model this kind of phenomenon. Indeed, Gamma distributions appear naturally in homogeneous Poisson processes. The time of arrival  $S_n$  of the  $n$ th occurrence is the sum of  $n$  independent variables with the same exponential distribution of parameter  $b$ . Thus the distribution of  $S_n$  is the Gamma distribution with pa-

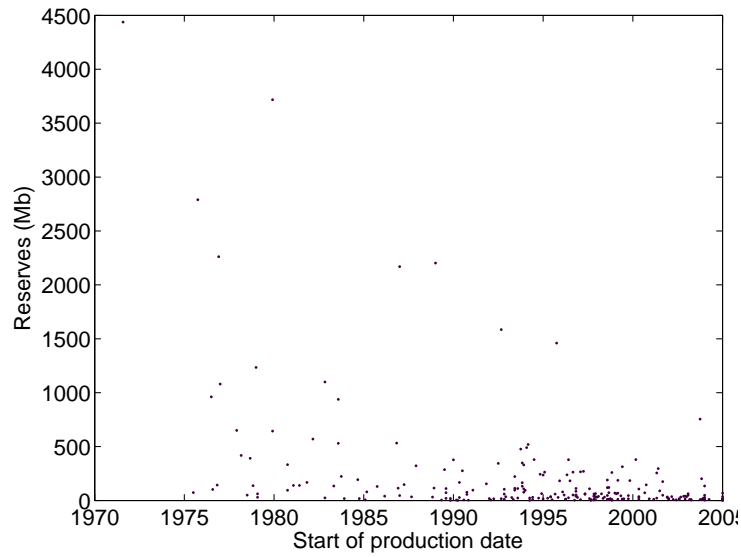


Figure 3: Launching production process of the North Sea fields (sources: Wood Mackenzie 2003).

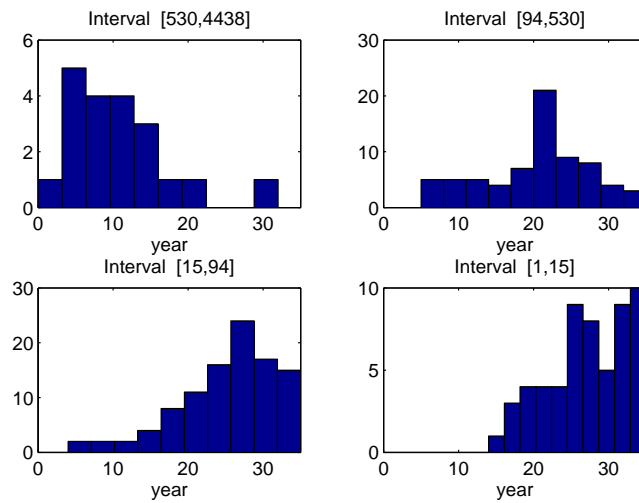


Figure 4: Histograms of launching production dates for different classes of field sizes in the North Sea (unit is Mb),  $S = 0$  corresponds to the beginning of the production in the Basin. In 2003, there was also 260 oil fields (namely fields with oil reserves including less than 1 Mb) discovered and undeveloped in this area.

parameter  $a = n$  and  $b$ , where  $b$  is the constant intensity of the process. Indeed, Gamma distributions are related with homogeneous Poisson processes. The time of arrival  $S_n$  of the  $n$ th occurrence of a homogeneous Poisson process is the sum of  $n$  independent variables with the same exponential distribution of parameter  $b$ . Thus the distribution of  $S_n$  is the Gamma distribution with parameter  $a = n$  and  $b$ , where  $b$  is the constant intensity of the process. This model is adapted to our case in the following way. Let  $a$  and  $b$  be two functions of the field sizes, we set:

$$\forall y \in [\underline{y}, \bar{y}], (S | Y = y) \sim \Gamma(a(y), b(y)) . \quad (3)$$

The density of the distribution  $\Gamma(a(y), b(y))$  is denoted

$$g_{a(y), b(y)}(s) := \frac{s^{a(y)-1}}{b(y)^{a(y)} \Gamma(a)} \exp\{-s/b(y)\} \quad (4)$$

When  $b$  is a constant, since the biggest fields are produced sooner, in this case  $a = a(y)$  is necessarily a decreasing function. This allows Gamma distributions to be used to take into account the priority of the largest fields in launching of the production process. Concerning the family of smallest fields, the distribution of  $S$  is more difficult to describe because it is difficult to evaluate how they will be produced in the future. Figure 4 shows that the North Sea area has just begun to produce its small fields. This problem will be discussed in section 6.

### 3.4 Individual field production model

Oil of an individual field is not produced at constant rate. In theory, companies adapt the number of fields in activity to optimize the profitability. In reality, production starts to decline quite soon after the maximum is reached. As shown in [33], the reserves variable can be used to predict profiles. Profiles get thinner when field size decreases, but they all present the same kind of shape, but with different time scales and amplitude. The following model approximately predicts the production of an individual field containing an amount  $y$  of oil in its reservoir:

$$\text{prod}(t, y) = \frac{y}{v(y)} K\left(\frac{t}{v(y)}\right) . \quad (5)$$

Function  $K$  is an *elementary profile* and function  $v$  controls the distortion due to the reserve variable. We assume that  $K$  is a positive *kernel*:  $K(u) du =$

1. In our application,  $v$  is a power function. We describe the estimation procedure of  $K$  and  $v$  in Appendix A.

Note that this elementary field production model is deterministic whereas the launching production process proposed is probabilistic. Indeed, we believe that randomness in basin production is essentially contained in the way the exploration happens.

### 3.5 Basin production

The joint distribution of  $(Y, S)$  together with the individual field production model allows to define a complete basin production model. Suppose that  $\mathcal{B}_n$  is a basin which contains  $n$  fields. The exploitation of the basin is assumed to start at  $t = 0$ . As before,  $(S_1, \dots, S_n)$  are the i.i.d random variables corresponding to the dates the productions start, and  $(Y_1, \dots, Y_n)$  are the i.i.d random variables of the amount of oil in each of the fields. We assume that

$$\forall i, Y_i \sim \mathcal{P}\text{ar}(\alpha, \underline{y}, \bar{y}) \quad \text{and} \quad (S_i | Y_i = y) \sim \Gamma(a(y), b(y)) . \quad (\mathbf{H0})$$

Let  $\text{prod}_i$  be the production of field  $i$ , then the basin production is defined by

$$\begin{aligned} \text{Prod}_n(t) &:= \sum_{i=1}^n \text{prod}_i(t) \\ &= \sum_{i=1}^n \frac{Y_i}{v(Y_i)} \text{K} \left( \frac{t - S_i}{v(Y_i)} \right) , \end{aligned}$$

where  $\text{K}$  and  $v$  are defined as in (5). This model is a combination of three mathematical tools: the field-size distribution (distribution of  $Y$ ), the dynamics of the production (distribution of  $S$ ) and the field production model (5). Estimation procedures for each of these elements are detailed in Section 6 and Appendix A.

## 4 Limit of the production

A basin is composed of several thousands of fields, hundreds of which are producing fields. In the North Sea example for instance, in 2009, about one

thousand fields have been discovered, and about 600 have been exploited (many of which still currently produce). Thus some law of large number arguments can be used to fit a *limit curve* on the known part of the production. This is the main advantage of the new model presented in this paper.

We make two additional assumptions:

$$\sup_{t \in [0,1]} K(t) := \kappa < \infty , \quad (\mathbf{H1})$$

$$w(y) := \frac{y}{v(y)} \text{ is non decreasing on } [\underline{y}, \bar{y}] , \quad (\mathbf{H2})$$

where  $K$  and  $v$  are defined for the individual field production model (5). Function  $K$  satisfies a Lipschitz condition: there exists a constant  $k \geq 0$  such that for all  $x_1$  and  $x_2$  in  $\mathbb{R}^+$ ,

$$|K(x_1) - K(x_2)| \leq k|x_1 - x_2|. \quad (\mathbf{H3})$$

These hypotheses allow us to easily obtain convergence of the basin production. Furthermore, they are relevant in the context of oil production. Indeed, an idealized production profile should be regular thus **(H1)** and **(H2)** are obviously realized in our context. Hypothesis **(H2)** states that the maximum of profiles is not a decreasing function of  $y$ , which is also natural considering real production curves.

The following result shows that the production curve converges almost surely to a limit curve as  $n$  tends to infinity.

**Proposition 1.** *Under **(H0)** – **(H3)**,*

$$\lim_{n \rightarrow \infty} \sup_{t > 0} \left| \frac{1}{n} \text{Prod}_n(t) - \mathbb{E}[\text{prod}(t - S, Y)] \right| = 0 \quad a.s.$$

The proof of the proposition is given in Appendix B.

According to the last proposition, it is relevant to fit the mean curve (also called *limit curve*)

$$t \rightarrow \mathbb{E}[\text{prod}(t - S, Y)]$$

on the real curve. An illustration of this fitting is given in Section 6.

Most of the basins of the world have been explored and data are available. Thus, estimating the parameter  $\alpha$  for the distribution of the reserves is now possible. Even if it could be improved, the production profile model

(5) provides a good approximation of real profiles and is easily computed. In any case, it does not have a large effect on the basin curve since it affects production only locally. This is not the case for the distribution of  $S$ , which is difficult to estimate. In fact, we change the problem of directly extrapolating the basin curve into the problem of extrapolating the distribution of  $S$ . Indeed, in a mature basin, the distribution of  $S$  is well-known for classes of large and medium fields. A major difficulty is to correctly evaluate the sizes of the fields that will be produced in the future. It must be said that at the end of the production basin, our model is quite optimistic since it tends to assume that many little fields are in production. This weakness in the model can be moderated because the contribution of small fields is not considerable. This is discussed for the North Sea area application in Section 6.

## 5 Simulation studies

In this section, we present simulations of the basin production to evaluate the impact of the distributions of  $Y$  and  $S$  on the production curve. According to the estimation method detailed in Appendix A, we can use the following version of the individual production field model (5) :

$$\text{prod}(t, y) = y^{1-\beta} \text{K} \left( \frac{t}{y^\beta} \right) \quad (6)$$

where  $\beta = 0.27$ . The elementary production profile  $\text{K}$  is estimated using spline functions (see Appendix A). A given distribution of  $(S, Y)$  together with (6) allows us to simulate the basin production using

$$\text{Prod}_n(t) = \sum_{i=1}^n Y_i^{1-\beta} \text{K} \left( \frac{t - S_i}{Y_i^\beta} \right). \quad (7)$$

### 5.1 Impact of the field size distribution

To compare the shapes of production curves with different distributions for  $Y$ , the same functions  $a$  and  $b$  defining the distribution of  $S$  are used for all the simulations. The distribution of  $(S|Y)$  is defined by the two functions  $a$  and  $b$  with  $b = 10$  constant and  $a$  a piecewise constant function. Namely  $a$

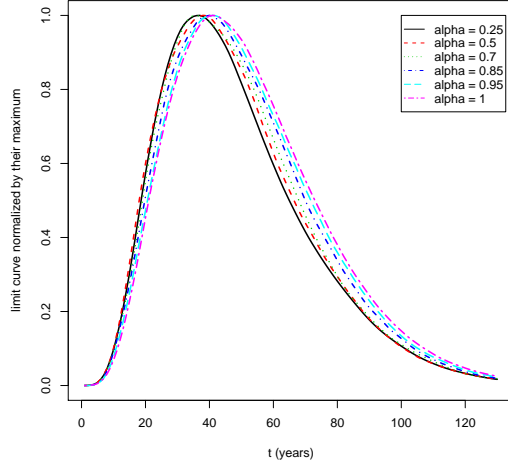


Figure 5: Limit curves normalized by their maximum. Each curve corresponds to a particular choice for the parameter  $\alpha$  of the Lévy-Pareto distribution modeling the field sizes.

is constant on the intervals of the partition  $m$ :

$$\begin{cases} m : & [1; 20; 100; 400; 4500] \\ a = & [6; 4; 3.5; 3] \end{cases} .$$

The distribution of the field size variable  $Y$  is  $\mathcal{P}ar(\alpha, 1, 5000)$ . The choice of parameters corresponds to a realistic situation. For different values of  $\alpha$  in the distribution of  $Y$ , the limit curve  $t \rightarrow \mathbb{E}[\text{prod}(t - S, Y)]$  is estimated by a Monte-Carlo procedure. In order to compare the curve shapes, each limit curve is normalized by its maximum. The normalized limit curve obtained this way are plotted on Figure 5. We note that the shape of the normalized limit curve does not depend much on  $\alpha$ . This remark is important regarding the production curve fitting procedure proposed in Section 6.2.

## 5.2 Impact of the distribution of $S$

For this second simulation study, we take  $Y \sim \mathcal{P}ar(\alpha, 1, 5000)$  for all the simulations. Consider the following functions  $a_0$  and  $b_0$  with  $b_0 = 8$  a constant

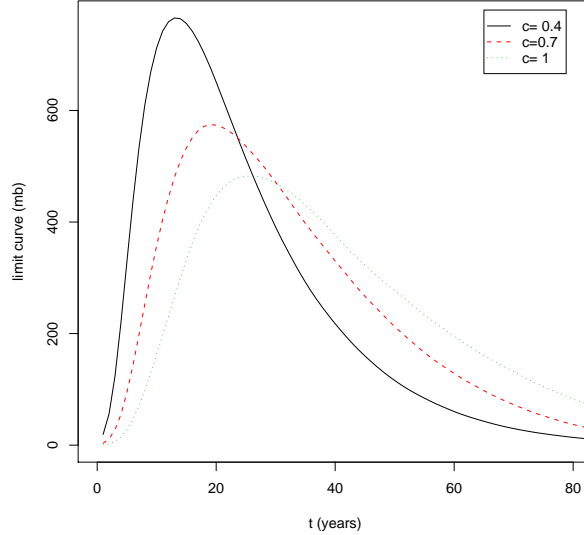


Figure 6: Limit curves corresponding to  $(S|Y) \sim \Gamma(ca_0(y), b_0)$  with  $c = 0.4$ ,  $c = 0.7$ , or  $c = 1$ .

function and  $a_0$  a piecewise constant function on the partition  $m$ :

$$\begin{cases} m : & [1; 20; 70; 400; 4500] \\ a_0 = & [8; 4; 3.5; 2] \end{cases} .$$

Figure 6 shows the limit curve obtained for different distributions for  $S$ :  $(S|Y) \sim \Gamma(ca_0(y), b_0)$ , with  $c = 0.4$ ,  $c = 0.7$ , or  $c = 1$ . This illustrates the importance of the distribution of  $S$  on the limit curve. Thus, for applications of this model, it is important to properly estimate this conditional distribution.

## 6 Application to the North Sea area Basin

This section is devoted to the application of our model to the North Sea area. It contains several hundreds of fields for which the data available is highly reliable.

## 6.1 Estimation of the distribution of $(Y, S)$

Wiorkowski [34] has pointed out the instability of past estimation procedures of remaining volumes of oil resources. However, for the oil production curve fitting procedure detailed in this section, we only need an estimation of the parameter  $\alpha$  of the Lévy-Pareto law. As was already stated, the statistics of the order of a sample of independent random variables with the Lévy-Pareto distribution has a linear trend in a loglog diagram with a slope  $-\frac{1}{\alpha}$  (see Figure 1). Thus  $\alpha$  can be easily estimated using a linear fitting on the subpopulation of the largest fields for which the linear behavior can be observed (see [25] for instance). The procedure gives  $\hat{\alpha} = 0.75$ . For the North Sea area, the lower threshold  $\underline{y} = 1\text{Mb}$  is realistic. Furthermore, the giant fields of the North Sea area have already been found, thus  $\bar{y} = 4437\text{ Mb}$ , which corresponds to Statfjord field. We use these estimates as the field-size distribution.

The conditional distribution of  $S$  is

$$\forall y > \epsilon, (S|Y = y) \stackrel{\mathcal{L}}{=} \Gamma(a(y), b(y)).$$

Let  $t_0$  be the present time. The history of the exploitation is only known on the period  $[0, t_0]$ , thus we only observe the censored variables  $(S^*, Y^*)$ . Denote  $f^*$  the density of  $(S^*, Y^*)$ . Then,

$$\begin{aligned} f^*(s, y) &= \frac{f(s, y)}{\mathbb{P}(S < t_0)} \\ &= h_{\alpha, \underline{y}, \bar{y}}(y) g_{a(y), b(y)}(s) \left\{ \int_{\underline{y}}^{\bar{y}} h_{\alpha, \underline{y}, \bar{y}}(y) \int_0^{t_0} g_{a(y), b(y)}(s) ds dy \right\}^{-1} \mathbf{1}_{0 < s < t_0} \\ &= h_{\alpha, \underline{y}, \bar{y}}(y) g_{a(y), b(y)}(s) \left\{ \int_{\underline{y}}^{\bar{y}} h_{\alpha, \underline{y}, \bar{y}}(y) G_{a(y), b(y)}(t_0) dy \right\}^{-1} \mathbf{1}_{0 < s < t_0} \end{aligned}$$

where  $G_{a(y), b(y)}$  is the distribution function of  $(S|Y = y)$ . We use a likelihood maximization method to estimate the parameters. The log likelihood of the model is

$$\log L(a, b) = \sum_{i=1}^n \log h_{\alpha, \underline{y}, \bar{y}}(Y_i) + \sum_{i=1}^n \log g_{a(Y_i), b(Y_i)}(S_i) - n \log \int_{\underline{y}}^{\bar{y}} G_{a(y), b(y)}(t_0) dy. \quad (8)$$

It is not easy to choose the functions  $a$  and  $b$ , because there is no obvious way to plot the shape of these functions. A possible solution is to propose

a partition  $m = \{I_1, \dots, I_k\}$  on the field sizes and to suppose as in previous section that  $a$  and  $b$  are piecewise constant on the partition  $m$ . A model selection procedure could be used to chose an optimal partition  $m$  for a particular statistical criterion. In this paper, for a first presentation of our basin production model, we prefer to fix a partition without making any partition selection :  $m = [1; 20; 100; 400; 4438]$ . The North Sea area sample is composed of 250 fields for which production has started. The minimization of  $\log L$  leads to the following estimates :

$$\begin{cases} m : & [1; 20; 100; 400; 4438] \\ \hat{a} = & [5.94; 3.26; 3.52; 2.93] \\ \hat{b} = & [10.2; 30.7; 9.09; 4.55] \end{cases}$$

The distribution of  $(S|Y = y)$  for small fields is difficult to estimate because only a small fraction of these ones have started to produce. However, an accurate estimation of  $(S|Y = y)$  for small values of  $y$  is less important then for large values of  $y$ , since the proportion of reserves contained in little field is minor. For instance, if  $Y \sim \mathcal{Par}(0.75, 1, 5000)$ , then around 70% of the basin reserves are contained in the fields greater than  $100Mb$ .

## 6.2 Fittings on the production curve

Using the previous estimations of the joint distribution  $(Y, S)$ , and the estimations related to model (6), it is now possible to plot the production limit curve  $\mathbb{E}[\text{prod}(t - S, Y)]$  with these parameters. According to [34], it is difficult to obtain accurate estimations of the number of fields  $n$  in the basin. Thus, we prefer to directly fit a rescaling of the limit curve (7) to the real curve. Three kinds of fittings are proposed in Figure 7: a fitting based on the least squares criterion, on the maximum of the true production and a fitting based on the current production.

The difference around 1990 between fittings and the true curve is mainly due to the Piper disaster in July 1988 which caused a decrease in the production for several years. Such phenomena cannot be taken into account by the model and except for this accident, the true production and the fit have had similar behaviors. The fittings gives possible extrapolations of the true production. Note that the production decreases quite quickly after the peak, despite of the fact that these scenarios are quite optimistic since they mobilize many fields. Indeed, the few giants are more productive then the aggregation

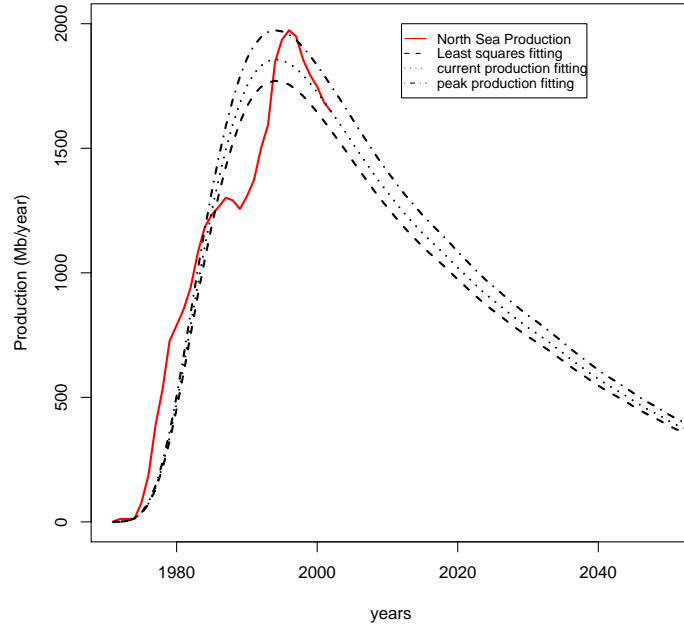


Figure 7: North Sea oil production and three fittings on the limit curve based on the least squares criterion, on the maximum of the true production and on the current production.

of all the many small ones. Considering the fact that the field distribution is the Pareto distribution, this observation is not a surprise. Even at the end of the production, the remaining part of the production of the medium and large fields has a major influence on the behavior of the global production. Figure 8 illustrates this phenomenon plotting the part of the production limit curve which comes from the fields greater than 20 or 50Mb.

## 7 Conclusion

Since Hubbert, the choice of adequate families to obtain good fittings on real production curves has been a recurrent question. The methodology developed in this paper is a way of getting around this problem by proposing a new probabilistic model which provides the asymptotic behavior of the

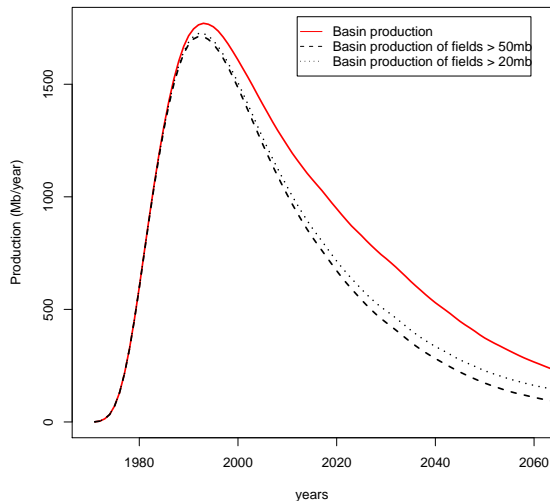


Figure 8: Production limit curve (rescaled as in Figure 7) and parts of this production which comes from the fields greater than 20Mb or 50Mb. The basin production is led by the productions of medium and large fields.

production curve. Thus, there is a mathematical justification to use our specific limit curve rather than other curves to fit to real production curves. We use Gamma distributions to model the start of production dates. This modelling and the estimation procedure proposed in Section 6.1 can probably be improved.

The basin curves obtained with simulations in Section 5 look more like Gamma curves than Gaussian or logistic curves. These simulations show that the distribution of the start of production dates have a deep impact on the basin production curve whereas the parameter  $\alpha$  of the Lévy-Pareto distribution has little effect on the shape of the basin production curve (except on the amplitude of course). One difficult question is to infer how little fields will be produced in the future. Small fields are numerous but they contain only a small part of the basin reserves. The exploitation of many little fields can only mitigate the decrease of the basin production after the peak.

This model is based on the assumption that the field size distribution is the major factor in production of a exploited area. Of course, the author does not claim that this is the only important factor. We mentioned that the many

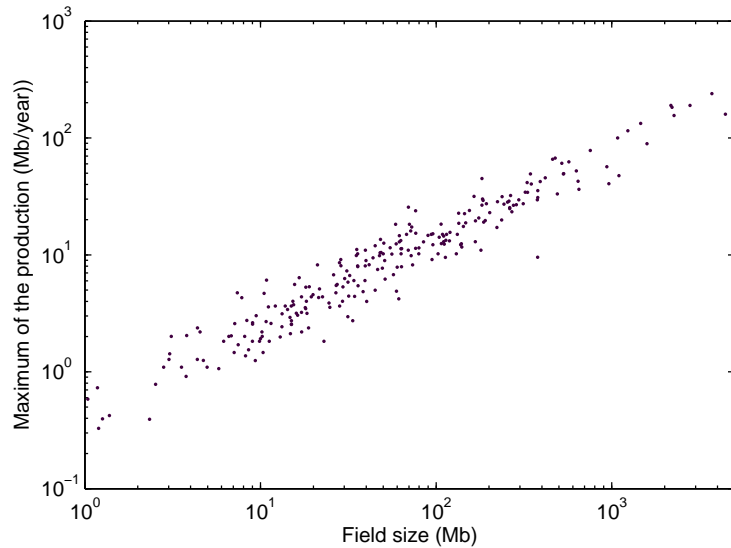


Figure 9: Plot in log scales of maximums of production versus field sizes for a sample of 250 fields in the North Sea (sources: Wood Mackenzie 2003).

econometric models of oil production have also been proposed. These works take into account interaction between the economic factors and production, but ignore the impact of the specific field-size distribution on production. It would be an interesting challenge to propose a model grouping together the “aggregated production” model of this paper with the “time series” point of view of these econometric works.

## Appendix A Individual production model estimation

Figure 9 is the plot in log scales of maximums of production versus field sizes for a sample of 250 fields in the North Sea. A linear regression can easily be fitted to these points, which means that that the production speed is a power of the size fields. This suggests the use of a power function for the distortion function  $v$  in (5).

$$\mathbb{E}(\text{prod}(t, y)) = y^{1-\beta} K \left( \frac{t}{y^\beta} \right) \quad (9)$$

where  $\beta$  is a constant,  $\text{prod}$  is now assumed to be a random variable and  $y$  is known and deterministic. The final estimation of  $K$  shown in Figure 10 validates the model. Considering a model with multiplicative noise in (9) gives better results than considering a model with additive noise, thus we assume in the following that

$$\frac{1}{x} \text{prod}(y, t) = \frac{1}{x^\beta} K \left( \frac{t}{y^\beta} \right) \tilde{\epsilon}_{x,t} \quad (10)$$

where  $\epsilon_{y,t} := \log(\tilde{\epsilon}_{y,t})$  is a random centered gaussian variable of variance  $\sigma^2$ .

We now present the main principles of the estimation of  $K$  and  $\beta$ . Let  $y_i$  be the size of field  $i$  and  $t_{ij}$ ,  $1 \leq j \leq n_i$  the years of observed productions of field  $i$ . Taking the logarithm of (10) gives

$$Z_u = -\beta \ln x_u + H(\tau_u^{(\beta)}) + \epsilon_u, \quad (11)$$

with  $H := \ln K$ ,  $Z_u := \ln \frac{\text{prod}(y_i, t_{ij})}{x_i}$ ,  $y_u := x_i$  and  $\tau_u^{(\beta)} := \frac{t_{ij}}{y_i^\beta}$ . Note that the few null productions observed have been removed in order to enable this logarithmic transformation. Natural cubic splines are used to estimate the function  $H$ . Natural cubic splines are piecewise polynomial functions which are used in interpolation problems, see for instance [35]. For a fixed family of knots  $(\xi_r)_{r=1, \dots, l}$ , let  $m = (m_r)_{r=1, \dots, l}$  be a base of the linear space  $\mathcal{S}_{cn}(\xi)$  composed of the natural cubic splines defined on the knots  $\xi$ . Thus, a function  $H \in \mathcal{S}_{cn}(\xi)$  can be expressed as  $H = \sum_{r=1}^l \theta_r m_r$ , with  $\theta \in \mathbb{R}^l$ . In order to control the roughness of the estimated natural cubic spline function, we consider the following penalized least squares criterion [see 35, p.127] for  $H \in \mathcal{S}_{cn}(\xi)$  and  $\beta > 0$ :

$$\text{crit}(H, \beta, \lambda) = \sum_{u=1}^N \{Z_u + \beta \ln x_u - H(\tau_u^{(\beta)})\}^2 + \lambda \int_0^{\tau_{\max}} \{H(\tau)\}^2 d\tau. \quad (12)$$

We use the procedure `smooth.spline` in the R software to estimate  $\hat{H}^{(\beta, \lambda)}$  for fixed  $\beta$  and  $\lambda$  and next the optimal coefficient  $\beta$  is estimated numerically. The smoothing parameter  $\lambda$  can be chosen by a cross-validation procedure. Here the sample is large enough to split the data into a first block  $B_1$  for the estimations and a second block  $B_2$  to minimize the prediction risk

$$\hat{R}(\lambda) = \sum_{u \in B_2} \left[ Z_u + \hat{\beta} x_u - \hat{H}^{(\hat{\beta}, \lambda)}(\tau_u^{(\hat{\beta})}) \right]^2.$$

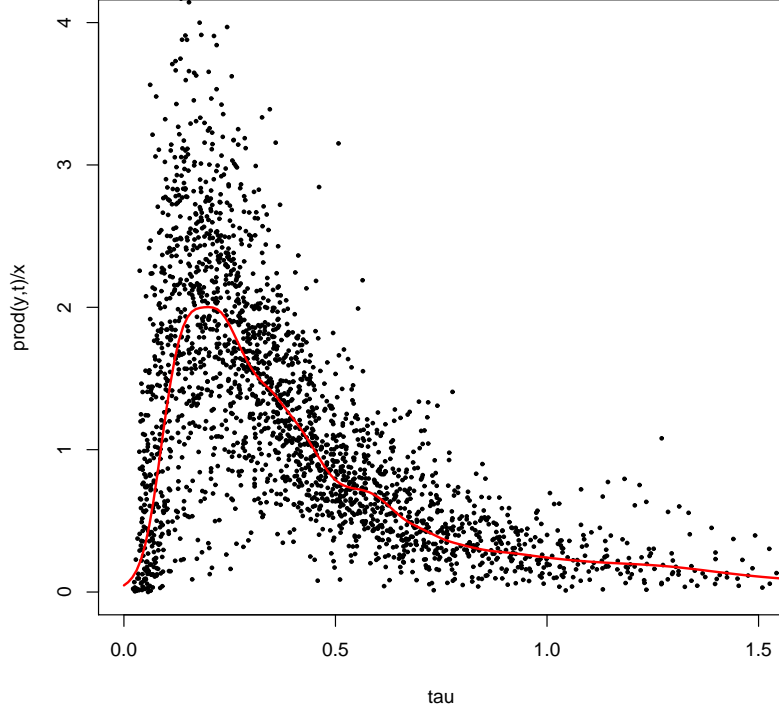


Figure 10: Plot of the normalized productions  $\frac{\text{prod}(y_i, t_{ij})}{x_i}$  versus the rescaled times  $\tau_u^{(\hat{\beta})} = \frac{t_{ij}}{y_i^{\hat{\beta}}}$ . The Elementary profile is estimated on these points by a smoothing spline method.

The complete estimation procedure is detailed in [36, chapter 7]. Applied to the oil productions of fields in the North Sea (Wood Mackenzie 2003 database) with natural cubic splines defined on 50 regularly spaced knots, it leads to  $\hat{\beta} = 0.27$ . The elementary profile is estimated by  $\hat{K} = \exp(\hat{H})$  and  $\hat{\sigma} = 0.53$ . Figure 10 is the plot of the normalized productions  $\frac{\text{prod}(y_i, t_{ij})}{x_i}$  versus the rescaled times  $\tau_u^{(\hat{\beta})} = \frac{t_{ij}}{y_i^{\hat{\beta}}}$ . The estimated elementary profile  $\hat{K}$  is also plotted on Figure 10. Note that at the end of the profile the decline is mitigated. This phenomenon is due to recovery methods applied to part of the fields in the sample.

## Appendix B Proof of Proposition 1

Let  $\mathcal{G}$  be the following class of functions

$$\mathcal{G} = \left\{ g_t(s, y) = w(y) K\left(\frac{t-s}{v(y)}\right), t \in \mathbb{R}^+ \right\}$$

where  $w(y) = \frac{y}{v(y)}$  is defined as in **(H2)**. Classical results about empirical processes allow us to prove the proposition. Indeed, proposition 1 says that  $\mathcal{G}$  is a Glivenko-Cantelli class for the distribution  $P$  of  $(S, Y)$ . We recall some definitions related to this topic. Let  $P$  be a probability distribution on a measurable space  $(\mathcal{X}, \mathcal{A})$ . Let  $\mathcal{H}$  be a class of measurable functions  $h : \mathcal{X} \rightarrow \mathbb{R}$ . Given two functions  $l$  and  $u$ , the bracket  $[l, u]$  is the set of all the functions  $h$  with  $l \leq h \leq u$ . An  $\varepsilon$ -bracket in  $L^1(\mathbb{R})$  is a bracket  $[l, u]$  with  $\int (l - u) dP \leq \varepsilon$ . The bracketing entropy allows us to measure the size of a class  $\mathcal{H}$ .

*Proof.* According to theorem 19.4 in [37], it is sufficient to prove that

$$\forall \varepsilon > 0, \quad \mathcal{N}_{[\cdot]}(\varepsilon, \mathcal{G}, L^1(P)) < \infty. \quad (13)$$

Let  $\varepsilon > 0$  and  $\delta = \varepsilon / \kappa w(\bar{y})$  where  $\kappa$  is the supremum of the function  $K$ . There exists  $t^* > 0$  with  $\mathbb{P}(S > t^* - v(\bar{y})) < \delta$ .

1. We check that  $[0, \kappa w(\bar{y}) \mathbf{1}_{[t^*-v(\bar{y}), \infty)}(s)]$  is an  $\varepsilon$ -bracket for  $\{g_t, t > t^*\}$ . First, for  $t > t^*$ ,

$$\begin{aligned} 0 \leq g_t(s, y) &\leq w(\bar{y}) \kappa \mathbf{1}_{0 < t-s < v(\bar{y})} \\ &\leq w(\bar{y}) \kappa \mathbf{1}_{s > t^* - v(\bar{y})}. \end{aligned}$$

Secondly,

$$\begin{aligned} \mathbb{E}(\kappa w(\bar{y}) \mathbf{1}_{[t^*-v(\bar{y}), \infty)}(S) - 0) &= \kappa w(\bar{y}) \mathbb{P}(S > t^* - v(\bar{y})) \\ &\leq \kappa w(\bar{y}) \delta = \varepsilon. \end{aligned}$$

2. Following [37](19.6 p.271) we show that  $\mathcal{N}_{[\cdot]}(\varepsilon, \{g_t \in \mathcal{G}, t \leq t^*\}, L^1(P)) < \infty$ .

According to **(H3)**, there exists a positive constant  $k$  such that

$$\forall x > 0, \forall h, \quad |K(x) - K(x+h)| \leq k|h|.$$

Let  $\varepsilon > 0$  and  $\rho = \varepsilon \frac{v(\bar{y})}{k w(\bar{y})}$ . We set  $r = \lceil t^*/\rho \rceil$  and for  $i = 0 \dots r$ ,  $t_i = i\rho$ . First, we show that the family of brackets

$$\left\{ \left[ g_{t_i} - k \frac{w(\bar{y})}{v(\bar{y})} \rho, g_{t_i} + k \frac{w(\bar{y})}{v(\bar{y})} \rho \right], i = 0 \dots r \right\}$$

covers  $\{g_t \in \mathcal{G}, t \leq t^*\}$ . Let  $t \in [0; t^*]$  et  $t_i$  such that  $|t - t_i| < \rho$ . Then,

$$\begin{aligned} g_t(s, y) - g_{t_i}(s, y) + k \frac{w(\bar{y})}{v(\bar{y})} \rho &= w(y) \left( K \left( \frac{t-s}{v(y)} \right) - K \left( \frac{t_i-s}{v(y)} \right) \right) + k \frac{w(\bar{y})}{v(\bar{y})} \rho \\ &\geq k \frac{w(\bar{y})}{v(\bar{y})} \rho - w(y) k \frac{|t-t_i|}{v(y)} \\ &\geq 0. \end{aligned}$$

The proof is the same with the upper bracket. Finally it is easy to check that the size of the bracket is exactly  $k \frac{w(\bar{y})}{v(\bar{y})} \rho = \varepsilon$ .

Points 1 and 2 together allow us to cover  $\mathcal{G}$  with a finite family of  $\varepsilon$ -brackets. Thus, (13) is true and the proof is completed.  $\square$

*Acknowledgements:* The author is grateful to Primoz Skraba (INRIA Saclay), G. Celeux (Paris XI University), D. Babusiaux and Y. Mathieu of the Economics Studies Division and Geology Division of the IFP.

## References

- [1] D. Babusiaux. Quelles productions et quels prix à l'avenir ? Petroles & Techniques, (456), 2005.
- [2] M.K. Hubbert. Nuclear energy and fossil fuels. Am. Petrol. Inst. Drilling and Production Practice, pages 7–25, 1956.
- [3] A.R. Brandt. Testing hubbert. Energy Policy, 35:3074–3088, 2007.
- [4] J.R. Moroney and M.D. Berg. An integred model of oil production. The Energy Journal, 20(1):105–124, 1999.
- [5] A. Kemp and S. Kasim. An econometric model of oil and gas exploration development and production in the uk contiental shelf: A systems approach. The Energy Journal, 24(2):113–136, 2003.

- [6] C. Cleveland and R.K. Kaufmann. Forecasting ultimate oil recovery and its rate of production : Incorporating economic forces into the model of m. king hubbert. The Energy Journal, 12(2):12–46, 1991.
- [7] R.K. Kaufmann. Oil production in the lower 48 states: Reconciling curve fitting and econometric models. Resources and Energy, 13:111–127, 1991.
- [8] M.K. Hubbert. Energy resources. National Research Sciences, 1962.
- [9] A.A. Bartlett. An analysis of u.s. and world oil production patterns using hubbert-style curves. Mathematical Geology, vol32(1):1–17, 2000.
- [10] K.S. Deffeyes. Hubbert’s peak: the impending world oil shortage. Princeton University Press, 2001.
- [11] E Broto. Infinite limit of r/p ratio and rate of depletion. available at [www.reservoir.org/doc/poster1.pdf](http://www.reservoir.org/doc/poster1.pdf), 2006.
- [12] R.W. Bentley. Global oil and gas depletion - a letter to the energy modelling community. IAEE Newsletter, pages 6–14, Second Quarter 2006.
- [13] R.W. Bentley. Global oil and gas depletion: an overview. Energy Policy, 30:189–205, 2002.
- [14] M.C. Lynch. The new pessimism about petroleum resources: Debunking the hubbert model (and hubbert modelers). Mineral and Energy, 18(1), 2003.
- [15] M.C. Ryan. Hubbert’s peak: deja vu all over again. IAEE Newsletter, (26):9–12, Second Quarter 2003.
- [16] J. Laherrère. Oil and natural gas resource assessment: Production growth cycle models. Encyclopedia of Energy, 2003.
- [17] D. Babusiaux and et. al. Oil and gas exploration and production, reserves, costs, contracts. Technip, Paris, 2004.
- [18] M.K. Hubbert. Degree of advancement of petroleum exploration in the united states. (Bulletin 51):2207–2227, 1967.

- [19] G.M. Kaufman, Y. Balcer, and D. Kruyt. A probabilistic model of oil and gas discovery, 1975.
- [20] J. Laherrère. Multi-hubbert modeling. [www.hubbertpeak.com/laherrere](http://www.hubbertpeak.com/laherrere), 1997.
- [21] M. Allais. Methods of appraising economic prospects of mining exploration over large territories. Management Science, 3:284–357, 1957.
- [22] G.M. Kaufman. Statistical decision and related techniques in oil and gas exploration. N.J.:Prentice Hall, Englewood Cliffs, 1963.
- [23] G.K. Zipf. Human behavior and the principle of least-effort. Addison-Wesley, Cambridge, 1949.
- [24] J.C. Houghton. Use of the truncated shifted pareto distribution in assessing size distribution of oil and gas fields. Mathematical Geology, vol20(8):907–937, 1988.
- [25] A. Crovelli and C. Barton. Fractals and the pareto distribution applied to petroleum accumulation-size distributions. Technical Report 91-18, USGS, 1993.
- [26] V. Lepez. Potentiel de réserves d’un bassin pétrolier : modélisation et estimation. PhD thesis, Université Paris Sud, 2002.
- [27] J. Laherrère. Distributions de type “fractal parabolique” dans la nature. C. R. Acad. Sci. Paris Sér. IIa, 322(7):535–541, 1996.
- [28] J. Laherrère and Sornette D. Stretched exponential distributions in nature and economy “fat tails” with characteristic scales. European Physical Journal, (B-2):525–539, 1998.
- [29] E.D. Attanasi and R.R Charpentier. Comparison of two probability distributions used to model sizes of undiscovered oil and gas accumulations: does the tail wag the assessment? Mathematical Geology, vol34(6):767–777, 2002.
- [30] V Lepez. Modélisation de la distribution de la taille des champs d’un système pétrolier, LogNormale ou Fractale ? Une approche unificatrice. Les cahiers de l’économie, (44), 2001.

- [31] Z. Lucki and Z. Szkutnik. Distribution of time lags between successive petroleum exploration activities. Mathematical Geology, vol22(5):523–556, 1990.
- [32] D.L. Lawless. Statistical models and method for lifetime data. John Wiley and Sons, New York, 1982.
- [33] C. Maugis and B. Michel. Data-driven penalty calibration: a case study for Gaussian mixture model selection. ESAIM: P & S., 2010. To appear.
- [34] J.J. Wiorkowski. Estimating volumes of remaining fossil fuel resources: a critical review. Journal of the American Statistical Association, 76:534–559, 1981.
- [35] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The elements of statistical learning. Springer Series in Statistics. Springer-Verlag, New York, 2001. Data mining, inference, and prediction.
- [36] B. Michel. Modélisation de la production d’hydrocarbures dans un bassin pétrolier. PhD thesis, University Paris XI, 2008. Available at <http://tel.archives-ouvertes.fr/tel-00345753/en/>.
- [37] A.W. Van der Vaart. Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge Univ. Press., Cambridge, 1998.