An approach to growth curves analysis from a simplicial point of view

Eusebi Jarauta-Bragulat[1], Juan José Egozcue[2]

[1]Dept. de Matemàtica Aplicada III, E.T.S.E. Camins, Canals i Ports de Barcelona, Univ. Politècnica de Catalunya, Barcelona (Spain). E-mail: **eusebi.jarauta@upc.edu**
[2]Dept. de Matemàtica Aplicada III, E.T.S.E. Camins, Canals i Ports de Barcelona, Univ. Politècnica de Catalunya, Barcelona (Spain). E-mail: **juan.jose.egozcue@upc.edu**

Abstract

Many dynamical systems of interest in natural resources studies, geosciences, economics and many other scientific fields are studied using different models of the so called growth curves. The most applied models are negative exponential, logistic, Gompertz, Richard's, Weibull, Huber, etc. In general, growth (decay) can be thought of as evolution of a part of a system, growth curves describing the evolution of the absolute value of this part in time. Usually, standard viewpoint does not take into account other components of the complete system. This may cause difficulties when interpreting the fitted model and makes the selection of it somewhat arbitrary.

For instance, when exploiting an ore in a mining area, we can consider the total mass of ore in the deposit, decomposed into three parts, unknown mass, yet extracted mass, known but not extracted ore. The evolution on time of this system can be described as the evolution of total mass (constant in this case), and evolution of the proportions of the three parts. This kind of approach suggests modelling evolution of proportions from the compositional point of view, i.e. using dynamical systems modelled in the framework of Aitchison geometry and differential equations in the simplex.

The study starts with simple but classical examples with two components (exponential and logistic growth) which share a common linear model in the simplex for the parts, and a different behaviour of the total mass. Next step consists of studying simple evolution compositional models with different hypothesis on the total mass. The models obtained are compared with some traditional well-known growth curves.

Keywords: Simplicial dynamical systems, resources assessment, derivatives, compositional processes.

1. INTRODUCTION

A growth curve is an empirical model of the evolution of a quantity over time. Growth curves are widely used in natural resources studies, geosciences, economics or biology, and the studied quantities are diverse, e.g. population size, biomass, ore-mass, resources. Values of the measured property or variable are usually plotted on a graph as a function of time. When a mathematical function is fitted to the data it is called growth curve. There is an extensive literature on growth curves, frequently adapted to the specific field in which they are used. From a statistical point of view, the use of growth curves would require the following steps: (a) choice of the model adapted to the characteristics of the studied quantity and following the parsimony principle (simple models are preferred); (b) fitting the model by minimising-maximising some error criterion, leading to the estimation of parameters in the model; (c) validation of the fitted model and underlying hypotheses; (d) interpretation of the result taking into account statistical uncertainty in the estimation. These steps are seldom completely covered in practice and deeper sight is still needed. The present goal is to open up the range of usual models for growth curves (a). The main idea is to consider several parts of the system to be modelled in order to take into account their interactions. The system is viewed as a whole, here called total mass, decomposed into its parts or categories that can be described as proportions of the total mass. Both total mass and proportions may evolve in time. The model of total mass may be very simple (e.g. constant, exponential decay, etc.) and the evolution of proportions can be modelled using differential models in the simplex, where proportions take values. The kind of models to be used (total mass and proportions evolution) are also a hint for fitting the curve taking into account the characteristics and scale of proportions which are assumed compositional.

As a typical example, Figure 1-1(a) shows the growth data corresponding to USA oil yearly production from 1900 to 2008; Figure 1-1(b) shows the cumulative growth curve. Figure 2, shows the fitting of the yearly production in Figure 1-1(b) of a growth curve by P. Cathaimer (2008).

A motivating case is a classical example which has been controversial in the XIX century: the growth of a population in opinion of Thomas R. Malthus (1798), exponential growth, and Pierre F. Verhulst (1838), logistic growth, which represent the two most popular and ancient models. The equation of the exponential model (increasing if $k > 0$ and decreasing if $k < 0$) is $p(t) = p_0 e^{kt}$, where $p(t)$ is the population in time and p_0 is the population at t=0. The one-parameter logistic model is $p(t) = p^*(1+e^{-kt})^{-1}$. Figure 3 shows an example of these growth curves. They are examined under the compositional or simplicial approach proposed here.

(a) (b)

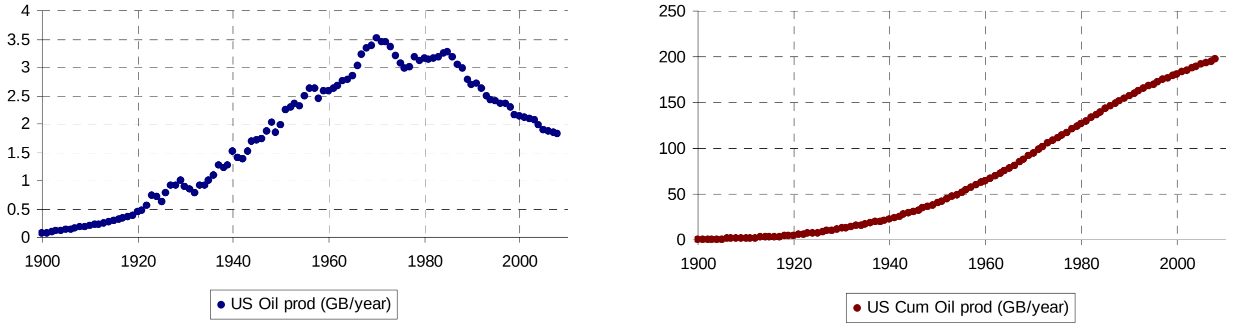

Fig.1. (a) USA oil production, 1900-2008. (b) Cumulative USA oil production, 1900-2008. (P. Cathaimer, 2008).

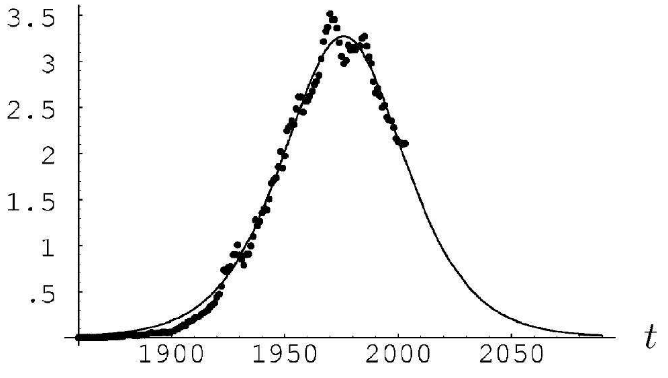

Fig.2. USA oil production 1900-2008, with a mathematical function fitting the data. (P. Cathaimer, 2008).

(a) (b)

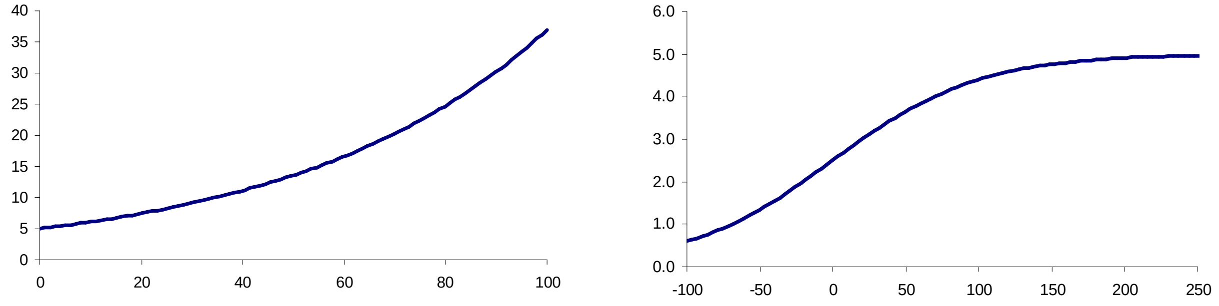

Fig.3. Exponential (a) and one-parameter logistic (b) mathematical models.

2. COMPOSITIONAL GROWTH MODELS

The classical approach of a growth curve typically does not consider all the components or parts of the problem. That is, evolution is studied as a single part of a system independently of the others parts. In some cases, this strategy may not make sense. In addition, the variables are studied "in mass" that is, in absolute units (number of individuals, number of oil barrels, number of trees in a forest). The alternative here presented is to consider the proportions between some observed or studied parts of the total mass changing in time. Additionally total mass may evolve in time as well.

We consider a system of *n* parts defined by a vector valued function

$$\vec{P}(t) = \left(P_1(t), P_2(t), ..., P_n(t)\right), \tag{2.1}$$

being the system total mass

$$M(t) = P_1(t) + P_2(t) + \cdots + P_n(t) = \sum_{j=1}^{n} P_j(t) \tag{2.2}$$

The vector of proportions can be defined as

$$\vec{p}(t) = \left(p_1(t), p_2(t), ..., p_n(t)\right) = \left(..., \frac{P_j(t)}{M(t)}, ...\right) \tag{2.3}$$

The sum of proportions is then

$$\sum_{j=1}^{n} p_j(t) = 1 \tag{2.4}$$

Exponential and logistic growth illustrates the compositional approach. Consider a population with exponential growth consuming $P_1(t)$ resource units. We can complete the system with a second component $P_2(t)$ representing presently available resource units which is assumed constant in this case. The equations of this two part system are (model 1)

$$\begin{aligned} &\vec{P}(t) = \left(P_1(t), P_2(t)\right) \\ &P_1(t) = P_1(0)\exp(kt);\ \ P_2(t) = C_2 \text{ (constant)} \\ &M(t) = P_1(t) + P_2(t) \end{aligned} \tag{2.5}$$

Consider a second population consuming $P_1(t)$ resource units which evolves according to a one-parameter logistic model. Assuming that the total mass of resource units $M(t)$ is constant in time, the system is described (model 2)

$$\vec{P}(t) = \left(P_1(t), P_2(t)\right)$$
$$P_1(t) = P_1^*(1+\exp(-kt))^{-1};\;\; P_2(t) = M(t) - P_1(t) \qquad (2.6)$$
$$M(t) = M \;\; \text{(constant)}$$

Figure 4 shows the growth curves in mass of both systems: (a) exponential growth (model 1), (b) logistic growth (model 2). Clearly, the systems considered are different in mass (Fig. 4). When proportions of used and available resources are computed according Eq. (2.3), we get the compositional evolution of these models of growth. Figure 5 shows the compositional growth curves of both systems: (a) model 1, (b) model 2. As shown, the compositional model of growth curves for both systems are equal. This means that the Malthus-Verhulst controversy only relays on the total mass of resources and not on the mechanisms affecting the proportions of available-consumed resources.

To proceed with the development of compositional models of growth curves, we need some concepts that are summarized in the following section.

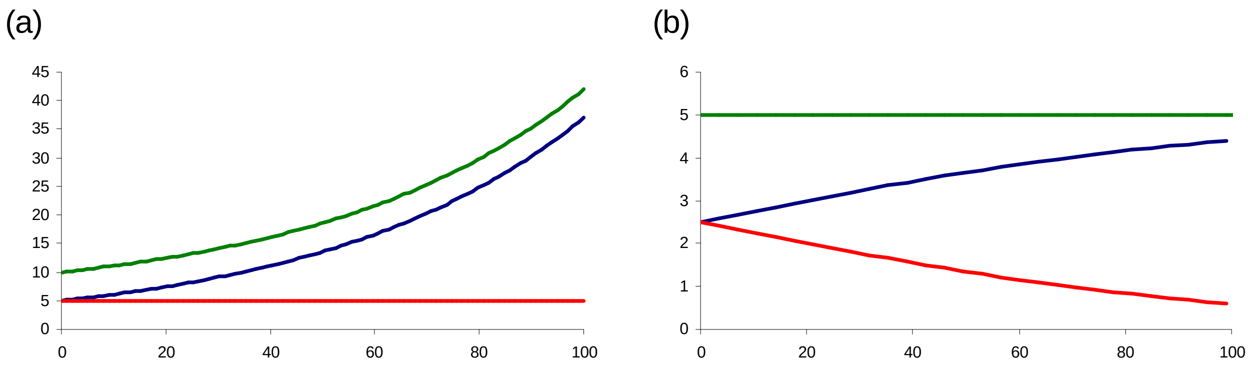

Fig.4. Growth curves (in mass) of systems in the example. Green: total resources; Blue: used resources; Red: available resources.

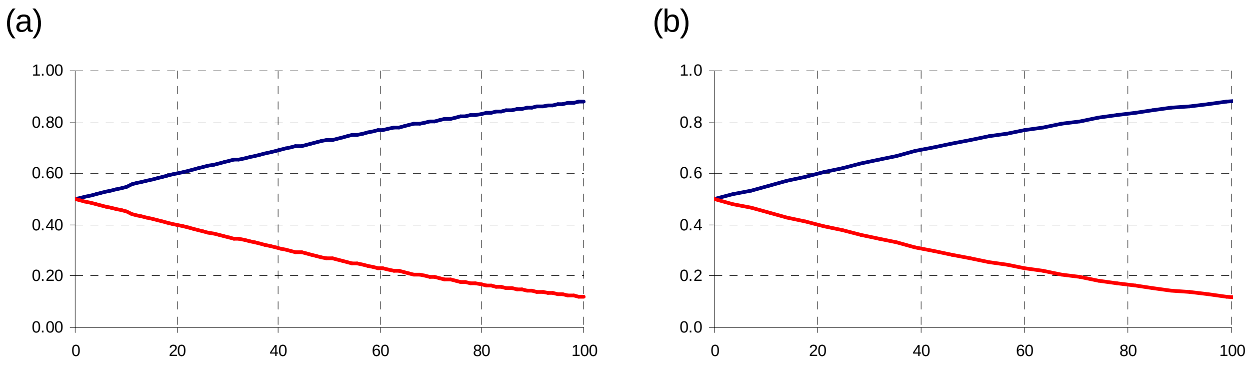

Fig.5. Growth curves (in proportions) of systems in the example. Blue: proportion of used resources; Red: proportion of available resources.

3. SOME CONCEPTS ON CALCULUS IN THE SIMPLEX

The simplex of *n* parts or *n*-part simplex, designated S^n, is the set of positive-component vectors of constant sum and it is defined as

$$S^n = \left\{ \vec{x} = \left(x_1, x_2, \ldots, x_n\right) \in \mathbb{R}_+^n, \forall j = 1, \ldots, n \ x_j > 0, \sum_{j=1}^{n} x_j = 1 \right\} \tag{3.1}$$

The elements of the *n*-part simplex are called compositions. The sum of the components can be another constant different from one; for example is 10^6 if parts per million are considered. However, the sum can be reduced to one dividing each component by this constant. The operation of dividing by the sum of the components is called closure and it is denoted by $\mathcal{C}$.

Obviously, the operations of ordinary Euclidean space are not feasible in the simplex, so it must be furnished of a vector space structure by other operations (Aitchison, 1986; Pawlowsky-Glahn and Egozcue, 2001). These are the perturbation ($\oplus$) and powering ($\odot$), defined as

$$\begin{aligned} \vec{x} \oplus \vec{y} &= \left(\ldots, \frac{x_i y_i}{\sum x_j y_j}, \ldots. \right) = \mathcal{C} \exp\left(\log(\vec{x}) + \log(\vec{y})\right) \\ \lambda \odot \vec{x} &= \left(\ldots, \frac{x_i^{\lambda}}{\sum x_j^{\lambda}}, \ldots. \right) = \mathcal{C} \exp\left(\lambda \log(\vec{x})\right) \end{aligned} \tag{3.2}$$

In that expressions, exponential (exp) and natural logarithm (log) functions are applied componentwise. The neutral element of perturbation operation is the composition

$$\vec{1}_n = \mathcal{C}(1,1,\ldots,1) = \left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n} \right) \tag{3.3}$$

The opposite element of a composition is calculated by

$$\ominus \vec{x} = \left(\ldots, \frac{1/x_i}{\sum (1/x_j)}, \ldots. \right) = \mathcal{C} \exp\left((-1)\log(\vec{x})\right) \tag{3.4}$$

The difference, i.e. the sum (perturbation) with the opposite element, is calculated

$$\vec{x} \ominus \vec{y} = \vec{x} \oplus (\ominus \vec{y}) = \left(\ldots, \frac{x_i / y_i}{\sum (x_j / y_j)}, \ldots. \right) = \mathcal{C} \exp\left(\log(\vec{x}) - \log(\vec{y})\right) \tag{3.5}$$

With perturbation and power transformation operations, the n-part simplex has the structure of real vector space of dimension $n-1$. A structure of Euclidean space has been also studied (Pawlowsky-Glahn and Egozcue (2001)) thus providing a distance (Aitchison distance and the corresponding metrics). There are several isomorphisms between the simplex S^n and the ordinary Euclidean space $\mathbb{R}^{n-1}$. The application of such an isomorphism allows us to work in the ordinary Euclidean space with the transformed values called coordinates. Then, well-known methods and procedures can freely applied to coordinates. Finally, we can retrieve values in the simplex using the inverse transformation. Here the so-called isometric logratio transformation (ilr), defined by Egozcue et al. (2003) is used to get coordinates of vectors of proportions. In the case with n=3, ilr coordinates, also known as balances, are

$$ilr\left(\vec{x}\right) = ilr\left(x_1, x_2, x_3\right) = \left(\frac{1}{\sqrt{2}}\log\frac{x_2}{x_1}, \sqrt{\frac{2}{3}}\log\frac{x_3}{\sqrt{x_1 x_2}}\right) = \left(u_1, u_2\right) \tag{3.6}$$

The inverse ilr transformation is

$$ilr^{-1}(u_1, u_2) = \mathcal{C}\exp\left(-\frac{1}{\sqrt{2}}u_1 - \frac{1}{\sqrt{6}}u_2, \frac{1}{\sqrt{2}}u_1 - \frac{1}{\sqrt{6}}u_2, \sqrt{\frac{2}{3}}u_2\right) = \left(x_1, x_2, x_3\right) \tag{3.7}$$

Here we deal with positive-component vector-valued functions (mass of parts) of real variable (time), denoted f(t); and simplex-valued functions (proportions) of real variable (time) denoted F(t). The relationship between these two kinds of functions is the closure, this is

$$\vec{f}(t) = \mathcal{C}\vec{F}(t) \tag{3.8}$$

The simplicial derivative of a simplex-valued function (Aitchison and Egozcue, 2005; Jarauta-Bragulat and Egozcue, 2008) is defined as

$$D^{\oplus}\vec{f}(t) = \lim_{h\to 0}\left(\frac{1}{h}\odot\left(\vec{f}(t+h)\ominus\vec{f}(t)\right)\right) \tag{3.9}$$

The computation of simplicial derivatives of a function is done using

$$D^{\oplus}\vec{f}(t) = \mathcal{C}\exp\left(\frac{d}{dt}\log\vec{f}(t)\right) = \mathcal{C}\exp\left(\ldots, \frac{\frac{d}{dt}f_j(t)}{f_j(t)}, \ldots\right) \tag{3.10}$$

Similarly, higher order derivatives are computed as

$$D^{\oplus k}\vec{f}(t) = \mathcal{C}\exp\left(\frac{d^k}{dt^k}\log\vec{f}(t)\right), \; k = 2,3,\ldots \tag{3.11}$$

An important property for the computation of simplicial derivatives is that simplicial derivative commute with the closure operation

$$D^{\oplus}\vec{f}(t) = D^{\oplus}\mathcal{C}\vec{F}(t) = \mathcal{C}D^{\oplus}\vec{F}(t) = D^{\oplus}\vec{F}(t) \tag{3.12}$$

At this point, we can express the property of exponential and logistic growth (models 1, 2, equations (2.5, 2.6)). Simplicial derivative is expressed in both cases as

$$D^{\oplus}\vec{P}(t) = \mathcal{C}\exp\left(\frac{k}{2}, -\frac{k}{2}\right) \tag{3.13}$$

showing that both systems are compositionally identical.

4. COMPOSITIONAL GROWTH MODEL FOR WORLD OIL

As an illustration, we develop a compositional growth curve model to study the evolution of oil in the world. The data set on which we base our analysis consists of a series of world oil production data and world oil proven reserves; the series is 1980-2009 (source: British Petroleum, BP). Figure 6 shows this data set, and oil production is computed and represented cumulative.

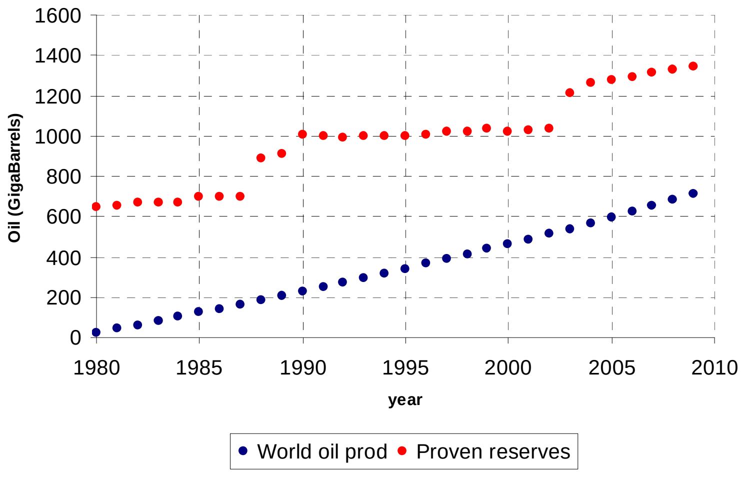

Fig.6. World oil production and proven reserves, 1980-2009 (British Petroleum).

To develop our model, we consider the vector-valued function $\vec{P}(t) = (P_1(t), P_2(t), P_3(t))$ of real variable *t*, being

t, time (years) from t = 0 (1980) to t = 30 (2009);

$P_1(t)$, cumulative world oil production (GBl/year), GBl = gigabarrels = 10^9 barrels oil;
$P_2(t)$, world oil proven reserves (GBl/year);
$P_3(t)$, unknown reserves (GBl/year) plus world oil consumed or stored at t = 0.

For each year *t*, the sum $M(t) = P_1(t) + P_2(t) + P_3(t)$ is the total mass of oil in the world, and it is assumed constant. The third component $P_3(t)$ is computed under different hypothesis on M(t), because it is unknown. Observed evolution of P1(t) and P2(t), and P3(t) under three hypotheses different hypothesis: (1) M = 4000 GBl, (2) M = 8000 GBl, (3) M = 12000 GBl are shown.

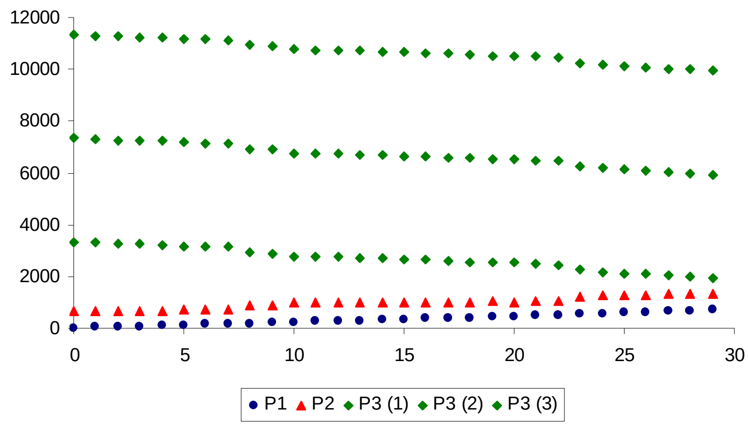

Fig.7. World oil evolution. Blue: consumed oil; Red: proven reserves; Green: unknown reserves under different hypothesis.

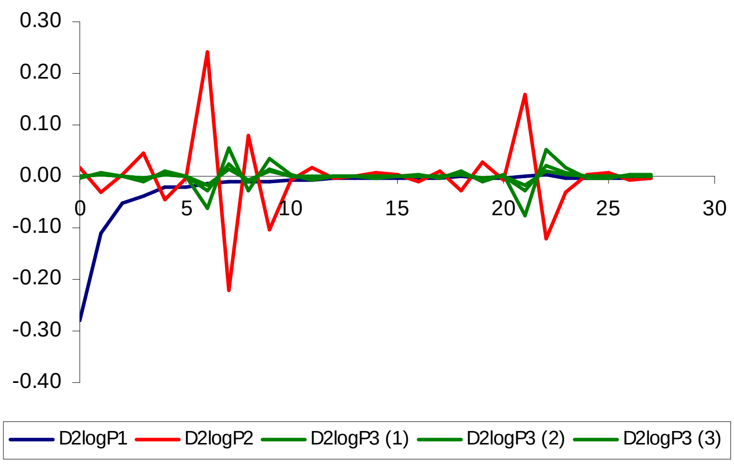

Fig.8. The second logarithmic derivatives of components under different hypothesis.

The first and the second simplicial derivatives of data have been numerically computed (Fig. 8). We conclude that the second derivative is approximately the neutral element of the simplex, that is $\mathcal{C}(1,1,1)$. Therefore, the first derivative is approximately $D^{\oplus}\vec{P}(t) = \vec{a}$ (constant).

At this point, we remark that the first simplicial derivative should be interpreted as the multiplicative rate of increasing-decreasing of each part. The solution of this simple differential equation is

$$\mathcal{C}\vec{P}(t) = \vec{K}_0 \oplus (t \odot \vec{K}_1); \;\; \vec{K}_0, \vec{K}_1 \in S^3 \tag{4.1}$$

which is readily identified with a linear function in the simplex being $\vec{K}_0$ a reference point and $\vec{K}_1$ the direction of the line. In order to estimate coefficients $\vec{K}_0, \vec{K}_1$ we fit a linear model to ilr-coordinates of data values in the 2-dimensional Euclidean space. The ilr-coordinates are

$$ilr\left(\vec{P}(t)\right) = ilr\left(P_1(t), P_2(t), P_3(t)\right) = \left(\frac{1}{\sqrt{2}}\log\frac{P_2(t)}{P_1(t)}, \sqrt{\frac{2}{3}}\log\frac{P_3(t)}{\sqrt{P_1(t)P_2(t)}}\right) = \left(U_1, U_2\right) \tag{4.2}$$

Note that the first ilr-coordinate U_1 can be computed using actually observed data P_1(t) and P_2(t) while U_2 contains P_3(t) which is hypothetical. Figure 9 shows ilr-coordinates of data and the linear fitted model. Using standard least squares, the fitted linear regression models for ilr-coordinates are

$$\begin{cases} U_1^* = -0.0301t + 1.2655 \\ U_2^* = -0.0503t + 2.0127 \end{cases} \tag{4.3}$$

where (U_1^*, U_2^*) are the ilr-coordinates of the predictor. The R-squared values are, respectively, $R^2(U_1) = 0.912$, $R^2(U_2) = 0.975$. Note that standard statistical methods of regression can be used here to check the fitted model although they are not discussed here. Compositional constants $\vec{K}_0, \vec{K}_1$ in Eq(4.1) can be computed as

$$\begin{cases} \vec{K}_0 = ilr^{-1}\left(1.2655, 2.0127\right) = \left(0.025, 0.151, 0.824\right) \\ \vec{K}_1 = ilr^{-1}\left(-0.0301, -0.0503\right) = \left(0.347, 0.333, 0.320\right) \end{cases} \tag{4.4}$$

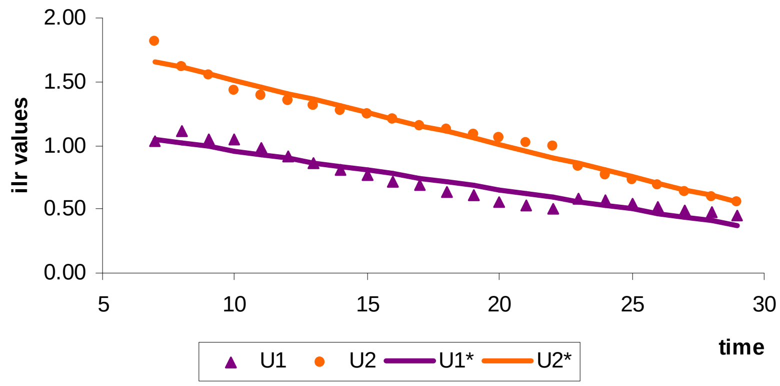

Fig.9. Linear model fitted to ilr-coordinates of data.

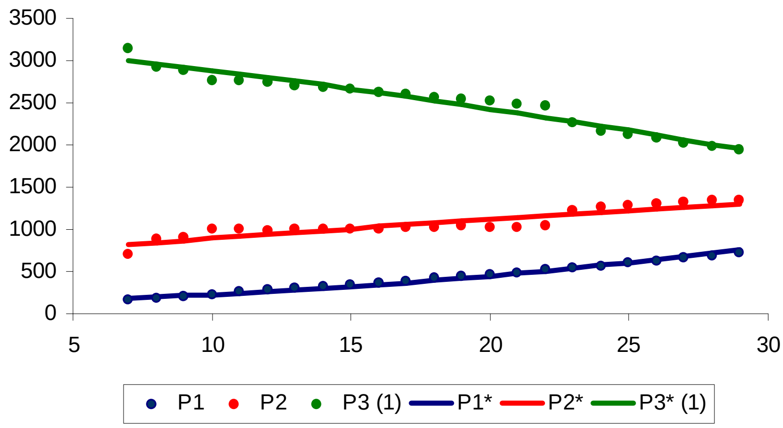

Fig.10. Oil-data and recovered model applying ilr back-transformation.

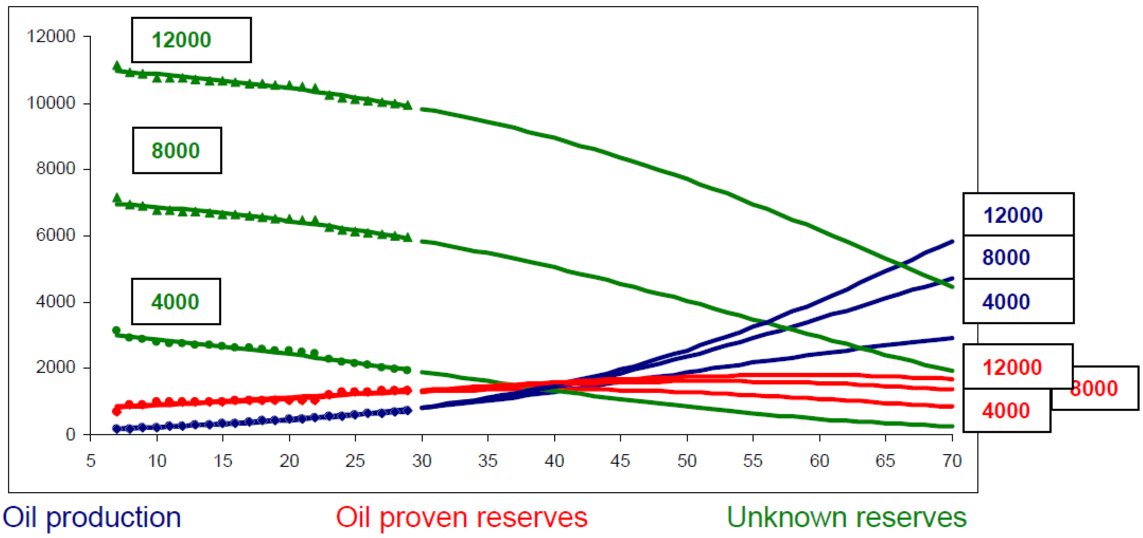

Fig.11. World oil components evolution prediction applying the fitted model.

Also the predictor values U_1^*, U_2^* can be back-transformed into the simplex using Eq(3.7). Oil estimated values for each component are recovered as follows:

$$\vec{P}^*(t) = \left(P_1^*(t), P_2^*(t), P_3^*(t)\right) = \left(M\,p_1^*(t), M\,p_2^*(t), M\,p_3^*(t)\right) \tag{4.5}$$

These recovered values are represented together with data values in Figure 10, corresponding to the first hypothesis M = 4000 GBl. Using the fitted model, predictions can be done (statistical uncertainty has been ignored for simplicity). These predictions are represented in Figure 11, under the three hypotheses for total oil (Fig. 7). By definition, curves in Figure 10 are cumulative curves; yearly values are obtained computing numerically the derivative (ordinary sense) of cumulative mass curves. Yearly curves are shown in Figure 12. Yearly curves permit a prediction about peak oil, i.e. the time when the maximum yearly rate of global petroleum extraction is reached and after which the rate of production enters terminal decline (Hubbert, 1956). Estimated peak oil ranges from year t = 48 (2028) to t = 70 (2050) under the three hypothesis about total oil. These estimations ignore statistical uncertainty and should be considered rough estimations.

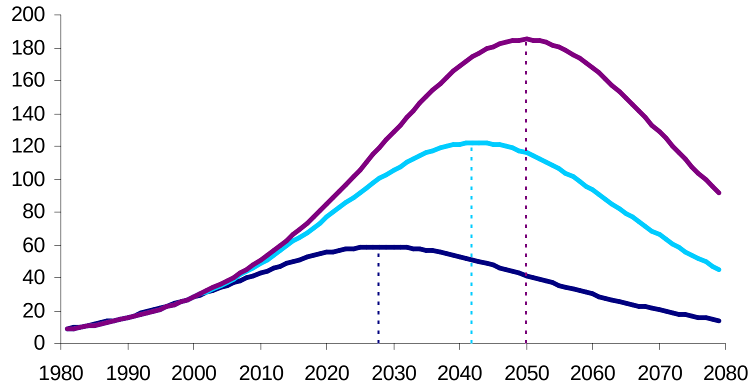

Fig.12. Peak oil prediction applying the fitted model.

5. CONCLUSIONS

1. Growth curves models can be studied from compositional point of view separately of the total mass evolution. In the studied cases, growth compositional models are simple and interpretable. Simplicial derivatives provides a tool to check the validity of proposed models.

2. Different growth curves models in mass may correspond to the same compositional model thus hiding simple common features of mass growth curves.

3. World oil data can be studied under compositional point of view. The simplest compositional model, the linear one, seems to explain correctly the world oil production-consumption.

4. In a preliminary estimation the Hubbert peak oil may range from 2028 to 2050.

REFERENCES

1. Aitchison, J. (1986). *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability*. Chapman & Hall Ltd., London (UK). (Reprinted in 2003 with additional material by The Blackburn Press). 416 p.

2. Aitchison, J., C. Barceló-Vidal, J. J. Egozcue, and V. Pawlowsky-Glahn (2002). *A concise guide for the algebraic-geometric structure of the simplex, the sample space for compositional data analysis*. In U. Bayer, H. Burger, and W. Skala (Eds.), Proceedings of IAMG'02, the eight annual conference of the International Association for Mathematical Geology, Volume I and II, pp. 387{392. Selbstverlag der Alfred-Wegener-Stiftung, Berlin, 1106 p.

3. Aitchison, J. and J. J. Egozcue (2005). *Compositional data analysis: where are we and where should we be heading?* Mathematical Geology 37 (7), 829-850.

4. Caithamer, P. (2008). *Regression and time series analysis of the world oil peak of production: another look*. Math Geosciences 40: 653 – 670.

5. Egozcue, J. J., V. Pawlowsky-Glahn, G. Mateu-Figueras, and C. Barceló-Vidal (2003). *Isometric logratio transformations for compositional data analysis*. Mathematical Geology 35 (3), 279 – 300.

6. Egozcue, J. J., E. Jarauta-Bragulat, J.L. Díaz-Barrero (2011). *Calculus of simplex-valued functions*. Submitted. In press.

7. Fekedulegn, D., M.P. Mac Siurtain, J.J. Colbert (1999). *Parameter estimation of nonlinear growth models in forestry*. Silva Fennica 33(4): 327 – 336.

8. Jarauta-Bragulat, E. and J.J. Egozcue (2008). *Compositional evolution with mass transfer in closed systems*. In Daunis-i-Estadella, J. and Martín-Fernández, J.A. (Eds.), Proceedings of CODAWORK'08, The 3rd Compositional Data Analysis Workshop, May 27-30, University of Girona, Girona (Spain), CD-ROM.

9. Pawlowsky-Glhan, V. and J.J. Egozcue (2001). *Geometric approach to statistical analysis on the simplex*. Stochastic Environmental Research and Risk Assessment (SERRA) 15(5), 384 – 398.

10. Verhulst, P. F. (1838). *Notice sur la loi que la population poursuit dans son accroissement*. Correspondance mathématique et physique 10, 113-121.

11. British Petroleum, http://www.bp.com/. Statistical Review of World Energy 2010.

ACKNOWLEDGEMENTS

Departament de Matemàtica Aplicada III and E.T.S. Eng. de Camins, Canals i Ports de Barcelona (Universitat Politècnica de Catalunya), for its economic support for attending the IAMG'2010 conference.

This research has been supported by the Spanish Ministry of Science and Innovation under the projects ENE2007-67033-C03-01 and CODA-RSS, MTM2009-13272. Also supported by the Agència de Gestió d'Ajuts Universitaris i de Recerca of the Generalitat de Catalunya, under the project Ref: 2009SGR424.