Biological modelling / Biomodélisation

By way of introduction: Modelling living systems, their diversity and their complexity: some methodological and theoretical problems

Alain Pavé

Centre national de la recherche scientifique, CNRS–Guyane, 16, av. André-Aron, 97300 Cayenne, France

Received 24 May 2005; accepted 9 September 2005

Available online 2 December 2005

Presented by Michel Thellier

Abstract

Some principles for a current methodology for biological systems’ modelling are presented. It seems possible to promote a model-centred approach of these complex systems. Among present questions, the role of mechanisms producing random or quasi-random issues is underlined, because they are implied in biological diversification and in resulting complexity of living systems. Now, biodiversity is one of our societies’ and scientific research’s main concerns. Basically, it can be interpreted as a manner, for Life, to resist environmental hazards. Thus, one may assume that biodiversity producing mechanisms could be selected during evolution to face to corresponding risks of disappearance: necessity of chance? Therefore, analysing and modelling these ‘biological and ecological roulettes’ would be important, and not only their outputs like nowadays by using the theory of probabilities. It is then suggested that chaotic behaviours generated by deterministic dynamical systems could mimic random processes, and that ‘biological and ecological roulettes’ would be represented by such models. Practical consequences can be envisaged in terms of biodiversity management, and more generally in terms of these ‘roulettes’ control to generate selected biological and ecological events’ distribution.

To cite this article: A. Pavé, C. R. Biologies 329 (2006).
© 2005 Académie des sciences. Published by Elsevier SAS. All rights reserved.

Résumé

En guise d’introduction : Modéliser les systèmes vivants, leur diversité et leur complexité : quelques problèmes méthodologiques et théoriques. Quelques principes d’une méthodologie actuelle de modélisation des systèmes biologiques sont présentés. Il semble maintenant possible de promouvoir une approche centrée modèle de ces systèmes complexes. Parmi les questions actuelles, le rôle des mécanismes produisant des événements aléatoires ou quasi-aléatoires est souligné, parce qu’ils sont impliqués dans la diversification des systèmes vivants et dans la complexité résultante de ces systèmes. Or, la biodiversité est une préoccupation majeure de nos sociétés et de la recherche scientifique. Fondamentalement, elle peut être interprétée comme une façon de résister aux aléas environnementaux, si bien que les mécanismes impliqués dans cette diversification auraient pu être sélectionnés au cours de l’évolution pour faire face aux risques de disparition : nécessité du hasard ? Il serait donc intéressant d’analyser et de modéliser ces « roulettes biologiques et écologiques » et pas seulement leurs « sorties », comme aujourd’hui, en utilisant la théorie des probabilités. Aussi est-il suggéré que les comportements chaotiques, engendrés par des systèmes dynamiques déterministes,
1. Introduction

Modelling in environmental sciences has been discussed during meetings organized by interdisciplinary research programs on environment and supported by the CNRS [1,2]. It has also been the case during the ‘Modelling and numerical simulation’ [3] program and workshop recently supported by the French Academy of Agriculture [4]. Actually, among the contributors, many were mainly interested by living systems’ modelling. In addition, at least half of a more recent paper has also been devoted to these topics’ discussion [5].

Living systems are complex, in the primary sense of the word: they are constituted of interrelated subentities at all levels of organization (structural complexity) and can exhibit amazing behaviours, such as chaotic or stochastic ones (behavioural or dynamical complexity) [6].1 They are therefore well adapted to study the relevance of these concepts, which have a wider field of interest.

All these initiatives proceed on an international effort illustrated by specialized journals’ success and by the position of international councils, such as the IUBS (e.g., [7]).

Anyway, living systems’ modelling, now well recognized, leads to multiplication and diversification of initiatives. It is not a surprise then to notice the important number of papers published in a lot of journals on these topics and by the diversity of modelling objectives. From a mathematical point of view, biology is a framework justifying mathematical developments; but from a biological one, the central question comes from biology, and the model is a tool that enables to answer this question. Furthermore, we know that modelling is a good way to analyse systems’ complexity, in particular living ones. In addition, biological and ecological engineering’s development, new agronomical techniques lead to an increasing demand in modelling approaches too.

The present workshop’s aim is to keep track of these different scientific streams by proposing specific examples, from gene level to agro and technoeosystems2 ones. I am therefore grateful to this meeting’s participants and speakers, as well as to the authors of the articles of this thematic issue.

In this introduction, I would like to specifically evoke some problems. The main, theoretical, has however practical consequences on living systems’ management. It concerns the importance of inner random processes working into these systems and the need of modelling approaches taking these processes into account. Therefore, this introduction is not a presentation of author contributions, but inspired by these contributions; it is mainly an opening to future developments.

2. The increasing diversity of modelling tools

Today, if mathematical modelling still represents the main part of formal approaches, computer-based models, such as cellular automata or multi-agents systems, are more and more used. They generally enable a better representation of dynamical process generated structural complexity of living systems than mathematical ones but conversely do generate weaker theoretical results.

However, modelling success is strongly dependent on computer-tool developments. Clearly, simulation concerns computer-based models as well as mathematical ones (e.g., numerical solutions of ODE or PDE). High-Performance Computing (HPC) is not still crucial

---

1 One may note that living systems’ structures also are the consequence of processes governing their dynamics, particularly their developments for organisms or their spatial arrangement for ecosystems. Therefore, structural complexity is linked to dynamical complexity.

2 For that matter, one can underline the concept of agroecosystem comes from the introduction of ecological approaches in agronomy. We also propose to speak of technoeosystems for biotechnical processes, when ecological concepts and models are used in analysis, design and control of these processes. It is significant to note artificial ecosystems now concern ecology, devoted for a long time to natural spontaneous ones. It illustrates the efficiency of these concepts and models. That is one of the reasons why the idea of “ecological systems engineering” [8] was recently introduced.
in biological sciences, but recent works show it will be more and more used in a nearest future [9], in computational biology [10], for example.

The under-equipment in France, even in Europe, in comparison to the situation in the USA and Japan, becoming a problem [11], we must plan for equipment.

Another important point is to have easy-to-use software adapted to specific applications. If general ones such as Mathematica or Mappe are very powerful tools, they conversely often require effort and particular skills difficult to support or gain for a non-specialist.

We can be nostalgic of tools such as Dynamac, developed during the 1980s in the Edora Club framework supported by INRIA, but no further developed nor adapted.

3. Models in biological sciences versus models in physical sciences

In fact, living sciences’ models’ status and efficiency are quite different from those of models in physical sciences. The principal difference comes from theoretical bases, more substantial and formalized in Physics than in Biology. In the latter case, one speaks more of models’ class than of specific theory-derived models. There are exceptions though: for example, a great part of genetics and population dynamics is well formalized and the neutral theory of evolution’ success mainly comes from its mathematical developments [13].

Another important difference is related to the nature of objects and phenomena. Biological ones are more variable and difficult to measure than physical ones. It leads to methodological problems. In some cases, random processes appear to be inherent to biological systems. As mentioned above, biological systems also are complex, constituted of numerous interacting entities, hierarchically organized, changing over time. In addition living processes are generally nonlinear and do sometimes produce complex dynamics, chaotic or stochastic ones (op. cit. [6]). Simple models can however represent global dynamics, resulting from a large set of elementary units that may be individually complex. It is a kind of ‘emergence of global properties’. For example, the well-known logistic law gives good descriptions of population dynamics, where interaction between individuals may be as complex as they are in human population. The emergence concerns also new properties, unknown at lower organization levels. A way to analyse these emergences is to build detailed models describing these populations at the individual level and study the convergence conditions leading to global logistic behaviour. More generally, simultaneous manipulation of mathematical models and of computer-based ones may be a good solution to study emergence of global properties, but also, conversely, to test the reliability of a simulator. For example, in population dynamics, there are a lot of models where implicit hypothesis on relationships between individuals can be explicitly represented (e.g., Logistic’s, predator-prey’s, competition’s, etc. models). Then if the simulator works properly we must find a dynamic closed to that foreseen by the mathematical model (Fig. 1).

Finally, a methodological scheme associating these approaches can be drawn, both being obviously related to reality.

Models of environmental systems associate sub-models of biological and physicochemical components of these systems. Furthermore, human actions have sometimes to be included. Even when restricted to biophysical models, it appears, for example, in aquatic systems, that the efficiency and reliability of biological components’ sub-models are weaker than models of physicochemical ones [14,15].

While parameters of physical models are often measurable, in most cases, parameters of biological ones are not and have to be estimated from experimental data.

---

3 B. Rousseau developed this software during his thesis [12]. It was devoted to an easy study of ODE (Ordinary Differential Equations) and recurrent equations. The graphic interface was particularly sophisticated and well adapted to user with basic knowledge about these kinds of mathematical objects.
Although, optimization procedures exist, parameters’ values are often empirically determined. As for validation, it is rarely a central preoccupation of many authors.

The problems of coupling different models also deal with scales’ and concerned system’s precise definition’s compatibility [16]. For instance, limits of a system designed from physicochemical components may be different from bio-ecological ones (a watershed is a ‘natural’ domain for hydrological problem, but a forest, as ecological system, may belong to more than one watershed).

4. Model-centred approach

Progressively in life sciences, from an additional and ornamental object, a way to conclude a scientific paper (it was frequently the case before the 1970s) or an alibi for mathematical development, the model is becoming an unavoidable tool and modelling an essential methodology entering in a scientific approach, strongly connected to experiments and observations. However, in order to be efficient, it has to follow some rules, as described in Fig. 2. This is what we might call a model-centred approach. It proceeds from classical hypothetico-deductive method, where the model becomes the central tool. It can be elaborated very early from initial knowledge, data and hypothesis, and may change over time to fit well with additional information. This method is efficient both in problem-oriented or object-oriented approaches.

However modelling is obviously not the unique approach, at least because it is not always possible to draw up an efficient model, or may not be necessary. Fig. 2 gives a general (and ideal) diagram of this approach. If it is not specific to life sciences; however these sciences have significantly contributed to elaborate it.
5. The processes’ generating quasi-random events in living systems fundamental role

In recent paper and book [17,18], the role of chance in living entities, from gene to ecosystem, has been analyzed: we proposed inner processes generating spontaneously appeared random events have been selected during evolution. The interpretation is simple:

- a high level of biodiversity is the way to ensure survival of biological systems confronted to environmental hazards. In the diversity of organisms, there are chances to have ones well-adapted to a particular environment and to changes of this environment, even drastic ones. Life’s history shows such catastrophic situations and biodiversity explosions following the disasters [19,20];
- necessity of chance? These systems’ diversification, as well as biodiversity’s maintenance are mainly the consequence of random processes produced by inner mechanisms, that can be called ‘biological roulettes’, and work at all organization levels, from gene to ecosystem. Examples are given below;
- like Monte Carlo methods and genetic algorithms developed by computer scientists to solve complex problems, in particular optimization ones, we can assume that stochastic processes have appeared spontaneously and have been selected over evolution to ‘insure’ ‘life sustainability’ within an uncertain and unpredictable environment. Moreover, such mechanisms enable to produce organisms well adapted to a particular environment. Producing a large diversity of organisms, potentially adapted to a large set of environmental conditions is an analogue to find an optimum for a complex and multidimensional response function.

This is a positive point of view about random events, while chance is generally considered in a negative way, as an unavoidable constraint disturbing elsewhere deterministic processes or experimental measures. For example, the well-known formula $y_i = f(x_i) + e_i$ links the effective measure of a state variable $y_i$ with the expected value $f(x_i)$ calculated for a given value $x_i$ of an independent variable $x$, where $e_i$ is an error assumed to be a particular value of a random variable with a null average and a variance $\sigma^2$. Function $f$ is analytic: it is the formal deterministic model of the relationship between $x$ and $y$. It can be called classical modelling approach, where chance is viewed as disturbing term resulting from uncontrolled factors.

In fact, it is possible to identify ‘biological roulettes’ or at least their results, from genome to ecosystem level. For instance, at the genome level, if spontaneous punctual mutations may be attributed to external events (e.g., radiations, chemical mutagens), conversely genome structure’s modifications by insertion, deletion or transfer of DNA pieces are consequences of inner biochemical mechanisms. At the organism level, one can retain the continuous synthesis of possible antibody proteins from a limited number of genes, but with particular transcription rules from DNA to mRNA, that look like a combinatory process. Infectious agent’s structure being a priori unpredictable, the solution is to produce a great diversity of such proteins synthesized by some kind of random process. At the organism level once more, for sexed organisms, there are a lot of gene transfers between chromosomes, such as crossing-over, during the meiotic phase producing gametes. These make the genetic structure of produced gametes different from that of the ‘original’ organism.

At the population level then, partner’s choice for sexual reproduction is casual, at least for one part. It is still another diversity source for the offspring. Finally, at ecosystem level, natural systems are generally diversified and organisms quasi randomly distributed in the ecosystem’s geographic space (e.g., trees in intertropical forests).

If the negative point of view about chance was to be true, it would be difficult to explain that, on the one hand, evolution has produced many precise and sophisticated systems and processes, but has conserved needless even worst ones on the other one. If chance is to be useful, even necessary, to produce diversity that ensures living systems sustainability as a whole, it is reasonable, as announced above, to assume that mechanisms producing random processes have been selected during evolution.

6. Modelling ‘biological roulettes’, from chaos to chance

Anyway, we do observe random events inner processes produced into living systems. We therefore have a good and efficient theory for randomly distributed events: the theory of probabilities. However, we have few theories and models regarding mechanisms and processes producing these events from mechanical laws, like the mechanical roulettes of casino, and not for biochemical, biological and ecological processes producing stochastic-like events. Now algorithms producing pseudo-random processes are not explainable in terms of biological mechanisms. We also know that chaotic
At first blush, the temporal series at the left of schemes (1) and (2) are similar. But in a convenient representation: \( x_{n+1} = f(x_n) \), they can be distinguished. (1) corresponds to the discrete time logistic model \[22\]. It can exhibit a chaotic behaviour. It is the case here. In the plane \((x_n, x_{n+1})\) values are distributed along a parabola. (2) corresponds to a stochastic uniform process on \([0, 1]\) simulated by a pseudo-random generator. Values are uniformly distributed in the plane \((x_n, x_{n+1})\). The graph (a) is generated by the formula \(z_n = x_n - y_n\), where \(x_{n+1} = rx_n(1 - x_n)\), \(y_{n+1} = ry_n(1 - y_n)\), and \(r = 3.98\), with slightly different initial conditions: \(x_0 = 0.5000\) et \(y_0 = 0.5001\). The graph (b) corresponds to \(ζ_n = ξ_n - ψ_n\), where \(ξ_n\) and \(ψ_n\) are values of stochastic variables uniformly distributed on \([0, 1]\). At the beginning, the chaotic system is foreseeable (if the value of \(x\) is known, \(y\) is also known with a good precision). But the sensitivity to initial condition leads to a divergence between the series, and \(z\) also becomes chaotic. Conversely, there is no interval where stochastic series are closed (b).

Dynamics can somewhere mimic stochastic processes and the frontiers between these two classes of processes can be very blurry and porous. It is however reasonable to assume, at least for one part, such processes result from chaotic dynamics \[21\].

On the one hand, we have drawn the well-known graphs of Fig. 3 to illustrate the difference between chaotic and random behaviours, or more precisely the results of algorithms simulating these behaviours. These graphs show two different structures of successive values: a simple one, a parabola, for chaotic simulation, and a largely distributed one (uniform distribution), for the random simulation. We also observe a divergence between two chaotic dynamics obtained from the same discrete time logistic model, from two different initial values. This is sensitivity to initial conditions: at the beginning dynamics are closed, but divergence rapidly appears, while difference immediately appears and always has the same amplitude for simultaneous stochastic dynamics. This property of sensitivity to initial conditions can be used to produce quasi-independent series with the same value of model parameters (Fig. 4).

On the other hand, we have compared stochastic and chaotic variables’ behaviours to illustrate chaotic dynamics’ quasi-random behaviours. Therefore, the ‘central-limit theorem’ is a keystone theorem of the theory of probabilities and statistics. Very schematically, it says that a linear combination of stochastic variables generally tends toward a Gaussian distribution, independently of each variable’s specific distribution. Can this behaviour also be observed for chaotic variables’ linear combination? We have chosen to study on one side the sum of chaotic variables, each one generated by the well-known discrete time logistic model. On the other side, we will study the sum of stochastic variables, each one uniformly distributed on \([0, 1]\). Fig. 5 shows results obtained by simulations. Convergence towards a Gaussian distribution, overall for sums of chaotic vari-
Fig. 4. Comparison between two simultaneous chaotic (at the left) and stochastic (at the right) processes. As previously, the chaotic one is generated by the discrete time logistic model ($r = 3.98$, $x_0 = 0.2$ and $y_0 = 0.1$). The stochastic one is generated by the ALEA() function from Excel. Hundred values have been computed for $x$ and $y$; in both cases, $x$ and $y$ are not correlated. The chaotic process exhibits a U distribution, with an accumulation on the edges, while the random one shows a quasi-uniform distribution, as expected. A simple chaotic process where two variables are generated simultaneously, but with very different initial conditions, begins to mimic a random behaviour.

ables, as well as its speed was a surprise. Then chaotic variables can exhibit properties that simulate those of stochastic ones, even for a strong formal result such as the central-limit theorem.

We have shown the likeness between simple chaotic processes and stochastic ones, and also between their statistical properties. It is not a formal demonstration, but it suggests that deterministic and mechanistic models and then biological and ecological processes can be represented by such models, can be precisely ‘biological or ecological roulettes’. The challenge is now to analyse these processes and build their models (Fig. 6).

It is quite different from classical approaches. In fact, these approaches analyse the output of these ‘biological roulettes’, considered as ‘black-boxes’. Then, the theory of probabilities gives efficient tools for such analyses and also to build models of the output processes. However, if we open the black-boxes, it does not solve the problem, no more than for a mechanical roulette. Practically, for games of chance, the output can be adjusted empirically, because experience shows that a well-equilibrated roulette, or die, lead to a quasi-uniform distribution. In biological or ecological roulettes’ cases, it can be of some interest to analyse, not only their outputs, but also how mechanisms’ modifications can affect results’ distribution. Model should then be efficient to simulate such modifications and their consequences.

Finally, biological and ecological roulettes generate complexity. Usually, logical complexity and stochastic complexity are opposed [23]. However, algorithms generate pseudo-random sequences of numbers, and, at least sometimes, chaotic dynamics look like stochastic processes and exhibit statistical properties similar to stochastic ones. Is this opposition necessary or useful? Would it be a progress to analyse their likeness, even their relationships? From our practical point of view, it would be interesting to show some convergences because it would be an encouragement to progress in the application of Fig. 6.

This last scheme is a proposal for a modelling approach of living systems that combines deterministic and stochastic aspects. Stochasticity is assumed to come from inner mechanisms generating chaotic results, which look like random variables. Then the classical modelling approach that uses stochastic process models is viewed as a phenomenological one that represents the output, but does not give a mechanistic repre-
Fig. 5. Comparison between distributions of linear combinations of chaotic variables and of stochastic variables. Like above, chaotic dynamics are generated by the formula: $x_{n+1} = rx_n (1 - x_n)$, with $r = 3.98$ and different initial conditions that ensure to obtain quasi-independent series of values although successive ones in a series are self-correlated. The chosen value for $r$ corresponds to a chaotic domain where $x$ covers almost the interval $[0, 1]$. Stochastic dynamics are obtained with the generator ALEA, an Excel® function, based on an arithmetic congruence method. The first line shows histograms of the distributions obtained for a single variable (it is asymmetric and has a 'U' form for the chaotic variable and quasi-uniform for the stochastic one). The other histograms have been obtained by weighted summation (to ensure values distributed between 0 and 1) of 2 ($S_2$), 4 ($S_4$) and 8 ($S_8$) independent chaotic variables (-Chaos, left column) and stochastic ones (-ALEA, right column). For each case, the Gaussian curve fitted to data is drawn and the basic parameter of the distribution is given ($m$, the average, and $s$, the standard deviation). The convergence towards a Gaussian distribution was expected in the stochastic case, but not in the chaotic one.

<table>
<thead>
<tr>
<th>Chaotic variables</th>
<th>Stochastic variables</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Histogram" /></td>
<td><img src="image2" alt="Histogram" /></td>
</tr>
<tr>
<td>$m = 0.541 ; s = 0.335$</td>
<td>$m = 0.497 ; s = 0.292$</td>
</tr>
<tr>
<td><img src="image3" alt="Histogram" /></td>
<td><img src="image4" alt="Histogram" /></td>
</tr>
<tr>
<td>$m = 0.538 ; s = 0.236$</td>
<td>$m = 0.494 ; s = 0.202$</td>
</tr>
<tr>
<td><img src="image5" alt="Histogram" /></td>
<td><img src="image6" alt="Histogram" /></td>
</tr>
<tr>
<td>$m = 0.547 ; s = 0.165$</td>
<td>$m = 0.506 ; s = 0.145$</td>
</tr>
<tr>
<td><img src="image7" alt="Histogram" /></td>
<td><img src="image8" alt="Histogram" /></td>
</tr>
<tr>
<td>$m = 0.551 ; s = 0.116$</td>
<td>$m = 0.502 ; s = 0.103$</td>
</tr>
</tbody>
</table>

The presentation of how these results are obtained. The choice of the modelling way depends on objectives. It is not always necessary to detail and to model the inner mechanisms, as in games, where the importance is to have uniform and uncorrelated issues and not to have a precise roulette model.
Fig. 6. Modelling of biological systems by taking into account deterministic and stochastic components. For the latter, it is proposed to complete the classical probabilistic approach by a mechanistic one representing inner processes generating stochastic-like events. These models are generally written in terms of Ordinary Differential Equations (ODE), Partial Differential Equations (PDE) or Recurrent Equations (RE). For instance, the discrete time logistic model used above is an example of a recurrent equation.

7. Conclusion

Modelling approaches are more and more used in life and environmental sciences. It is therefore important to agree on a common methodology. The debate being open for at least 30 years; we have proposed here an ideal scheme of a ‘model-centred’ approach, which is a kind of ‘good-practice’ synthesis. It has also been proposed to examine more acutely the problem of mechanisms and models that can exhibit random outputs. They indeed have probably been selected during evolution to produce a large diversity of living systems that can work and live through a large variety of environmental conditions. The fact underlining the importance of chance in generating biodiversity is obviously not new. But to assume a selection of mechanisms producing chance is not standard. The question of biodiversity being important today, it is suggested that particular efforts should be made in this way. Finally, it would be ideal to converge towards a genuine theory of biodiversity, in the same idea defended by Hubbell [24], but by adopting a wider approach from ‘gene to ecosystem’ and by integrating all diversification and maintenance aspects of the resulting biodiversity mechanisms [18].

References