Fractals and Epidemic Process

Philippe Sabatier, Pierre-Michel Guigal*, Daniel M. Dubois**

* Unité BioInformatique, École Vétérinaire de Lyon,
1 Av. Bourgelat, 69280 MARCY L'ETOILE - FRANCE
E-mail: sabatier@clermont.inra.fr
** Institut de Mathématique, Université de Liège,
12, Grande Traverse, B-4000 LIEGE 1 - BELGIUM
E-mail: Daniel.Dubois@ulg.ac.be

Abstract

The spread of an epidemic can be studied on a discrete space into small cells arranged into a d_S-dimensional regular lattice [Durett & Levin, 1994]. Each sites are occupied by healthy individuals may be infected by neighbours, after which they recover completely, they recover and are subsequently immune, or they die. Such a model is a generalisation of the differential equation approach. It corresponds to a modification of the directed percolation problem, useful to describe a large number of disordered systems in physics and chemistry. A critical concentration separate a phase where the epidemic dies out after a finite number of time steps, from a phase where the epidemic can continue forever.

In the simplest models, we assume that the vicinity, in which the infection process takes place, is a small domain surrounding the healthy individual considered. This vicinity is made up of the first layers of $\mathcal{M} = 3^{d_S}-1$ cells surrounding the central cell considered (Moore neighbourhood). The purpose of this article is to generalise the dimension of the substrate by introducing a fractal distribution of the sites. For each distribution of infected individuals in this vicinity, there is a certain probability ξ of infection. Due to the self-similarity, the infection quantities are significantly modified on fractal substrate.

The fractal distribution of the sites can be related to the spatial distribution of the epidemic vector [Meltzer, 1991]. Vector distribution is a matter of suitable habitat, which is a sum of a wide range of environmental factors (humidity, soil moisture, ground temperature, parasitic-host population density, etc..). The distribution of the sites can be also related to the genetic distribution of the susceptibility of the host population. In a herd, the laws of inheritance form a discrete and recursive system which mixes and distributes the genes of susceptibility. We can propose an aggregation model of relatives around an individual, which is based on the direct inheritance.

Keywords

Fractal ; Zipf-Mandelbrot Law ; Coefficient of Kinship ; Genetic-dependence, Epidemic model.

1. Introduction

Various authors have already demonstrated that the potential spread of diseases, such as cattle diseases, can be anticipated by studying environmental and/or genetic factors which affect the diffusion of the vectors or the susceptibility of the host. The existing work focuses on identifying and mapping potentially suitable sites often using averages taken from time series (for example, climatic data). The issue of the fluctuations in the amount and distribution of suitable habitat sites, or susceptible host genotypes, has not be fully addressed. Such fluctuations could destabilise enzootically stable disease populations, or introduce the vectors and diseases into previously naive herds. In either case, the subsequent mortality rate among susceptible cattle could be significant. Even if there are no fluctuations in the suitable habitat sites, or susceptible host genotypes, knowledge of the area involved defines potential control sites. In order to study some epidemic processes, such as diffusion of scrapie which is a diseases with genetic susceptibility, we aim to analyse the organisation of a related population.

2. Distance of kinship between individuals

The studies of genetic relationships between relatives propose to connect distance and genetic resemblance. The principle is based on the notion of identity which forecasts, with reference to Mendel's laws of inheritance, the probability of gene similarity in two individuals. It is a complex function of the number of generations between them. This purpose is expressed in the methods of "*Analysis of genetic variability using genealogical data*" [Vu Tien Khang 1989].

The whole genealogical data which may be obtained for an individual (i.e. its ancestors, its pedigree) gives, retrospectively, the morphogenesis of specificity of this individual within the genetic variety of the related population. Any part of pedigree hold in common by two individuals involves that they are, more or less, similar (at least in probability). Anyway, the definition of kinship may be expressed by the following sentence : " Two individuals are related when they have part of their pedigree in common *(...) The fact that related individuals resemble one another was no doubt the starting point of all genetical thought. The most casual observation shows that children resemble their parents, or sometime are strikingly like a more distant relative, and that sibs resemble one another* " [Jacquard 1982]..

Theoretically, the number of ancestors of an individual increases as 2^n, where n is the number of generations. The measures of kinship are performed between individuals, taken two by two.

In a limited size population, it is obvious that two individuals must probably share several ancestors between them. The direct relatedness leads to illustrate this sharing by a graph in which any relationship is an oriented arc.

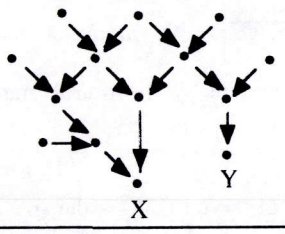

Fig. 1 : Graph of relatedness between *X* and *Y*

However complex such a graph may be, the genetic involvement of the connection between X and Y can be expressed as the probability that they harbour in common a gene which are copies of the same ancestor gene. The measures of kinship aim to quantify the probability of identity. The coefficient of kinship Φ_{XY}, which is one of these measures, is equal to the probability that, in any autosomal locus taken at random, one get the same gene in X and Y :

$$\Phi_{XY} = \sum_{A_i} \left(\frac{1}{2}\right)^{1+\lambda+\lambda'} \left[1 + F_{A_i}\right] \quad \text{(Eq. 1)}$$

where $\{A_i\}$ is a set of ancestors that two chains of descendants connect to X and Y, λ et λ' are the length of these chains.

This property, which defines ultrametrics, is characteristic of the hierarchical structures. In a hierarchical tree structure, the respective distances of leaves can be measure by the height of their first common root. If we consider a triplet of individuals, X, Y, Z, their respective distance of dissimilarity $d\Phi$, a kind of Hamming distance, form either equilateral triangles, or isocele triangles whose two equal sides are always longer than the third side. The distribution of the attractors, however, is not arbitrary.

3. Distribution of population attractors

We propose an aggregation model of relatives around an individual X. The process of iterative constitution of the X family group may be performed using a syntactic formalism [Dekking 1987] . We propose the following coding system in which R, B and F are individuals and k is a mean (or stochastic) number of descendants per individual :

- B is an individual which neighbouring is totally defined,
- R is an individual which roots remain to be defined,

- F is an individual which descendants remain to be defined.

At every iteration, the following substitution (a θ morphism) applies :

$\theta(R) = RRB(k-1)F$; $\theta(B) = B$: $\theta(F) = BkF$

At the first step, we start with an individual and its neighbouring :

$\left\{R, R, \underline{B}, \overbrace{F,...F}^{k\ times}\right\}$ = *{X father, X mother, X, 1° descendant,...,k° descendant}*. Then others aggregate by the transformation process. That gives a sequence of "words" {M(0),M(1), ...} (Fig. 2).

It may be observed that, as the size of the group increases, a property of autosimilarity seems to appear. Considering the asymmetrical tree, on which is based the family group constitution, a property of strict autosimilarity is not possible. Though, it can be defined autosimilarities by classes $F^{(1)},...F^{(n)}$, if a fractal $F^{(i)}$ can be a partition $\left\{F_j^{(i)}\right\}$ such as any part is similar to one of the $F^{(1)},...F^{(n)}$. The morphism which leads to the family group constitution, gives the properties of autosimilarity by classes. This syntactic formalism shows that a set which components organise themselves locally, may harbour global properties of autosimilarity. These properties, when they apply to an intensive value (such as an incidence rate), allows us to make a change of space for the approach of diffusion processes.

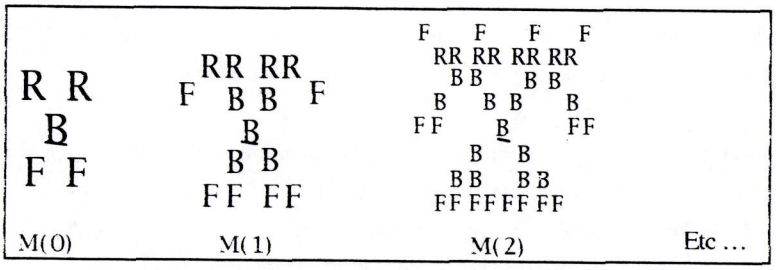

Fig. 2 : Graphic representation of the syntactic model application

In this model, a scale law associated to the autosimilar structure is revealed by a special behaviour of the kinship relation. If r is the number of relatives distant of n generations from X, and Φ their coefficient of kinship with it, the bi-logarithmic representation of r versus Φ shows a quasi-linearity. This linearity and the family group properties (revealed by the syntactic model) suggest that a fractal model of kinship may exist within a population.

The Zipf law, used in the complex systems analysis, connects an intensive value Φ (frequency) and a value of counting r (rank) :

$\Phi . r^{\gamma} = K$ (Eq. 2)

An extended power law, called this equation Zipf-Mandelbrot law, generalise the rank-frequency relation [Le Méhauté, 1974] :

$\Phi.(r + B_0)^{\gamma} = K$ (Eq. 3)

where B_0 gives the degree of organisation of the system. When $r_{max} \gg B_0$ (r_{max} = maximum rank in the set) there is a high dependency to the model. On the opposite the system is totally free [Le Méhauté, 1977] .

This transformation produces a family of curves (Fig. 4). For $B_0 = 0$, the Zipf-Mandelbrot law returns to the pure Zipf law. For the last ranks the Zipf-Mandelbrot distribution tends to the Zipf law values.

The equation (Eq. 3) may be replaced with a logarithmic one :

$-Ln(\Phi) = \gamma Ln(r+B_0) - Ln(K)$ (Eq. 4)

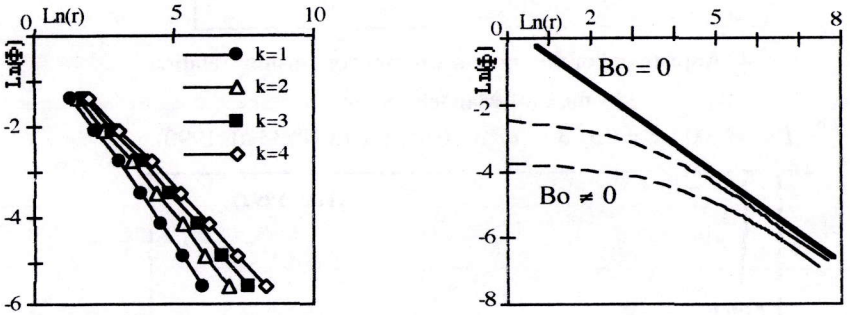

Fig. 3 : Bi-logarithmic representation of the relation $\Phi : r$.
(a) Zipf law. (b) Zipf-Mandelbrot law.

This conjecture have been tested with real genealogical data, obtained from 51 breeds of sheep, which pedigree were well known [Guigal, 1995] . The coefficient of kinship, which may be viewed as an intensive property, is connected to the rank of any relationship in the ordered set of the whole kinship identified. The range of size of these sets was 300 to 16000. The figure 4 shows an example of adjustment of the kinship relations between the young's of the year and all adults within a breed. In a large majority of the cases (39/51), the coefficient of correlation between the data and the Zipf-Mandelbrot law was over 0.95. For the remaining cases, we must put forward a new hypothesis which allows us to use another rank-size law (called the "Broken-Stick rank-size law").

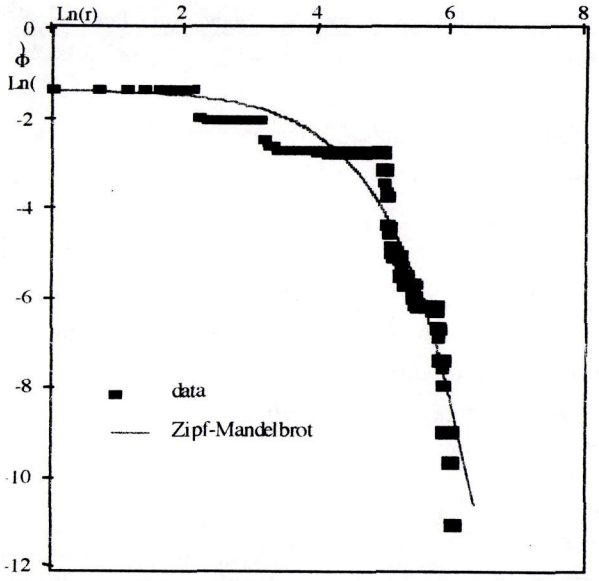

Fig. 4 : Adjustment of the "rank-coefficient of kinship" relation by the Zipf-Mandelbrot law :
B_0 =2100 ; γ = 26,73 ; *Corr.* = 0,91. (Herd N° 3370. 1990).

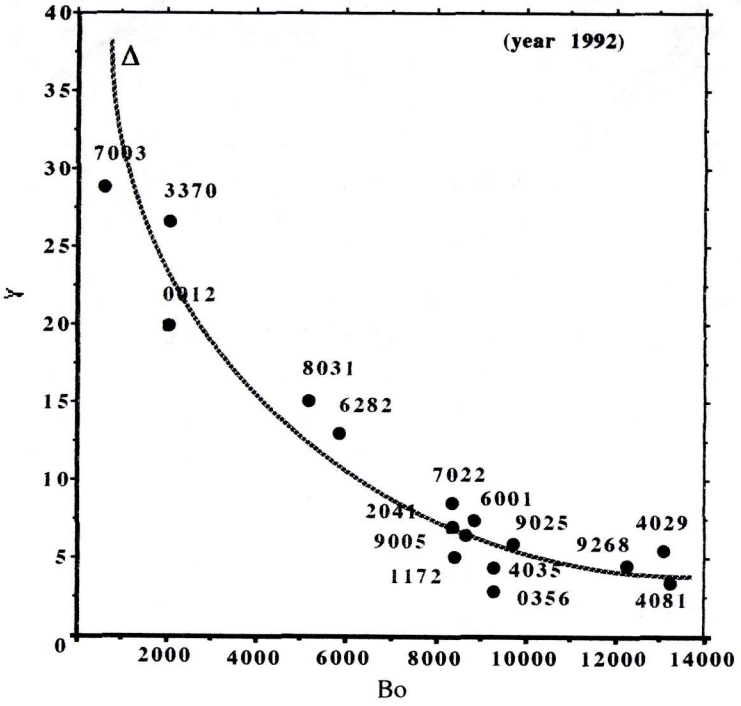

Fig. 5 : Localisation of the herds in a control plane (B_0, γ).
(B_0 : [600, 14000], γ : [3, 43], *Corr.* : [0,70, 0,93])

In a control plane (B_0, γ), the experimental data, reveal (Fig. 5) :

- an adjustment of the herds by an exponential Δ ;
- the discrimination of the herds according to the values of the parameters B0, γ :
  * B_0 increase with the size of the herds, and with the replacement rate ;
  * γ increase with the dispersal of the sires (males and females lines).

4. Potential use of fractals in epidemiology

The majority of epidemic models are formulated in terms of either differential equations or stochastic processes (Bailey, 1975). In an SIR model, based on disease status, the individuals are divided into three disjoint groups :

(S) the susceptible group, i.e. those individuals who are not infected but who are capable of contracting the disease and become infective;

(I) the infective group, i.e. those individuals who are capable of transmitting the disease to susceptibles; and

(R) the removed group, i.e. those individuals who have had the disease and are dead, or isolated, or have recovered and are permanently immune.

The possible evolution of an individual may, therefore, be represented by the following transfer diagram : $S \xrightarrow{p_i} I \xrightarrow{p_r} R$ where p_i and p_r denotes, respectively, the probability of being infected and the probability to be removed. The spread of the disease is governed by the first following rule (i) : susceptibles become infective by contact, i.e. suceptibles may become infective if and only if, it is in a neighbourhood of an infective. This hypothesis neglect latent periods, i.e. an infected susceptible become directly infective. But there is another rule (ii) : an individual selected at random may move to a vertex also chosen at random. If the chosen vertex is empty, the individual will move, otherwise the individual will not move. The set in which the vertex is randomly chosen depends on the range of the move.

More precisely, during one time step, the probability of a susceptible having z infected neighbours become infected is $(1-(1-p_i.I/N)^z)$. $x=p_i.I/N$ is the probability that at a time t a susceptible is infected by an infective located at a specific neighbouring site. Then, if z is the number of neighbouring vertices of a given vertex, $(1-p_i.x)^z$ is the probability that such an events does not occur, and finally $f(x)=(1-(1-p_i.x)^z)$ is the probability that such an event occurs at any neighbouring site. Note that, within the framework of this approximation, the interaction terms are not bilinear as in most models (Bailey, 1975).

Non-bilinear interactions have recently been shown to exhibit very different dynamic behaviour (Busemberg and Van den Driessche, 1991).

The shaky basis of the SIR model is found on these two rules by reference with the « chemical law of mass action ». All individuals are assumed to « move » randomly and to « contact » other individuals of various types in proportion to their density; upon contact the infective agent is transmitted with a certain probability, i.e. given a « collision » the « reaction » takes place with a certain probability. The classical SIR model of an epidemic given by the following equations (Kermack and McKendrick, 1927) :

$$\begin{cases} dI/dt = \beta(I).I.S - \gamma.I = E(I) - S(I) \\ dS/dt = -\beta(I).I.S = -S(S) = -E(I) \end{cases} \quad \text{(Eq. 5)}$$

$$I(t) + S(t) + R(t) = N$$

where $E(I)$, $S(I)$, et $S(S)$ denote the input and the output of infected I, and susceptible S individuals. And R denote the number of infected individuals who have been removed from a community of total size N. In brief, β is the infection rate, IS is the number of possible contact-pairs between susceptibles and infectives, and γ is the death or removal) rate of infectives. And, for convenience, we define $\rho = \gamma/\beta$ as the relative removal rate.

The « incidence » referring to the number of new cases per unit of time per unit of area (when the spatial domain is two-dimensional) is :

$$E(I) = \sum_r P_r(1-(1-p_i\frac{I}{N})^z)$$

where r is the number of infected neighbour, and x is the number of neighbouring vertices of a given vertex. For z=2, when p_i is small and when we are interested at a time scale at which θ, the number of low-tides per unit of time, is very large, we take for the incidence (which is a rate) the expression

$$E(I) = \beta.\frac{1}{N}.I.S \quad where \quad \beta = 2p_i\theta$$

For the epidemiology of the scrapie, the following expression of the « incidence » have been proposed :

$$E(I) = S.z\sum_r P_r p_i \frac{I}{N} = S.z.P_r\sum_r(\alpha\Phi_r + \beta)$$

We can propose an agregation model of relatives around an individual, which is based on the direct inheritance. The study of the system has been performed with simulations. The rate of prevalence for initial conditions is 6%, wich correspond to introduce six sheeps in the herd. The rate of prevalence obtained to the endemic equilibrium is accordance with the epidemiological data of the publications between 0% and 12%. When this equilibrium

happen, all the individuals seem to be germ carrier, but only someone become contagious. This equilibrium depend on the kinship. The closing of the reproductive line control the emergence of the scrapie. The simulation SIM5, show a prevalence rate of 0% for an closed situation, and 12% for an open line (SIM1). If we consider a ten year simulation, we cf. that the results evolve from an open situation, to a closed situation. We can see that the initial infection disappear imediatly, and next come back next to an endemic equilibrium. a closed situation to (Fig. 6)

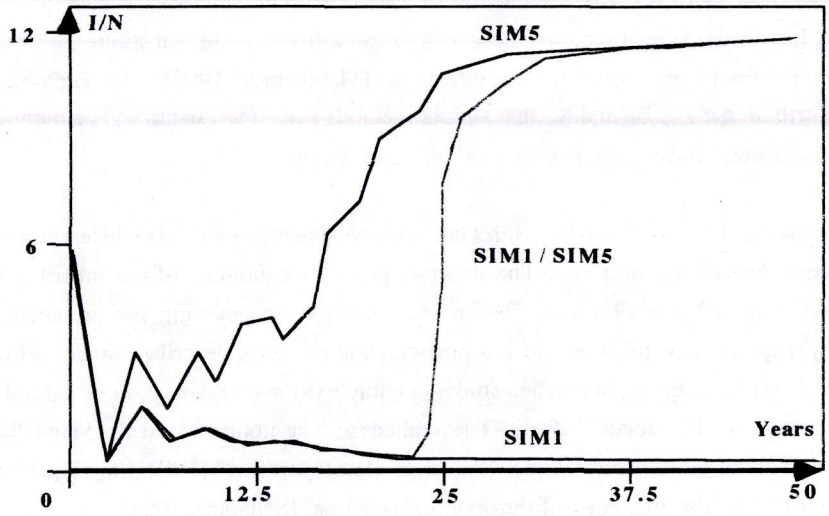

Fig. 6. Cinetics of the prevalance of the scrapie disease in two simulated herds (SIM1 & SIM5).
(α=2.85, β=0.05, μ=0.25, Z=30) (Guigal, 1995)

The equations of Kermack and McKendrick do generate useful qualitative predictions about possibles modes of behaviour. At the start of an epidemic, let $I(0) = I_0$; $S(0) = S_0$; $R(0) = R_0$, and we can see from (Eq. 5) that at time $t = 0$: $dI/dt = \beta I_0(S_0 - \rho)$. So an epidemic can only build up (i.e. $dI/dt > 0$) if $S_0 > \rho$. Thus $S_0 = \rho$ defines a deterministic threshold density of susceptibles below which an epidemic can not develop, since infectives are removed at faster rate than new infectives can be product. But these results are on the influence of the degree of mixing of the individuals from their diffusive motion. The spatial correlation created by the application of the first rule (i) can be partially undestroyed according to the degree of mixing of the population from the application of the first rule (ii).

5. Discussion

Vector distribution can be a matter of suitable habitat, which is a sum of a wide range of environmental factors (humidity, soil moisture, ground temperature, parasitic-host population density, etc..) [Meltzer, 1991]. The distribution of the sites can be also related to the genetic distribution of the susceptibility of the host population. In a herd, the laws of inheritance form a discrete and recursive system which mixes and distributes the genes of susceptibility [Guigal, 1995].

What we have performed is a reduction of the complexity. The Zipf-Mandelbrot law, which has been used to adjust observed data, allows us to substitute a Pareto's distribution for a complex set of binary relations [Mandelbrot, 1990]. The parameters of this distribution are γ, B_0 and K, that simple methods permit to quantify. The number of these parameters may reduced because of their correlation.

This adjustment by the Zipf-Mandelbrot law and the syntactic model are different aspects of a same class of organisation . The structure given by continuity (direct inheritance) is analogous to an hyperbolic tree. The entropy principle (concerning the probability of identity) superpose on this tree. Such a phenomenon has been described, under different formulations, by some authors when studying complexity and organisation of natural and artificial systems [Schroeder, 1991]. They called it : "the notion of cost" (Mandelbrot), "the principle of least effort" (Zipf), "the principle of partition" (Hill), "the equilibrium law" (Orlov) or "the invariance of the lexicographic tree" [Schapiro, 1994]

The autosimilar organisation constrains hardly the material of the system and gives it a hierarchical structure. This minimises the general entropy of the system. Its characterisation allows, using a change in space of parameters, to reduce complexity from a high number of discrete components to a few global values. In the study of infectious diffusion process within a limited size population and concerning diseases with genetic susceptibility, It might be very useful to perform this change of space. In heterogeneous systems it is often important to substitute for averages (like the mean rate of consanguinity). The notions of organisation give a new approach where global values may be a way of simplification of the diffusion processes [Sugihara & May, 1990].

References

Busemberg, S., van den Driessche, P., (1991), Nonexistance of periodic solutions for a class of epidemiological models. In Busemberg, S., Martelli, M. (eds) Biology, Epidemiology, and Ecology. Lect. Notes Biomath. Springer : Berlin. vol. 92, pp. 70-79

Dekking, F.M. (1987), Constructions de fractals et problèmes de dimension. *In Fractals : dimensions non entières et applications*, Paris : Masson, p. 132-149.

Durrett, R. (1995), Spatial Epidemic Models. *In Epidemic models.Their Structure and relation to Data.* Mollison, D. (Ed.). Cambridge : Cambridge Univ. Press. Vol. 417 p..

Guigal, P.M. (1995), Modélisation de la propagation infectieuse dans un réseau organisé d'individus : apport de la prétopologie et de la géométrie fractale. *Univ: Institut National des Sciences Appliquées de Lyon* , Thèse d'université, 270 p.

Jacquard, A. (1982), Genetic relationships between relatives. *Biomathematics* , Vol-5. p. 103-140.

Le Méhauté, A. (1974), Application à la hiérarchie départementale française d'un modèle d'adaptation des individus à une structure de désutilité en arbre. *Ann. Cent. Rech. Urb. Paris* , Vol-1, p.141-186.

Le Méhauté, A., Appleby, A.J. (1977), A parametric analysis of the structure of an international energy consumption. *Energy* , Vol-2, p. 105-114.

Mandelbrot, B.B. (1990), *The fractal geometry of nature*. New-York : Freeman. 468 p.

Meltzer, M.I. (1991), The potential use of fractals in epidemiology. *Prev. Vet. Med.* Vol. 11 p. 255-260.

Schapiro, B. (1994) An approach to the physics of complexity. *Chaos, Solitons and fractals* , Vol-4, p. 115-123.

Schroeder, M.R. (1991), *Fractals, chaos, power laws*. New-York : Freeman. 335 p.

Sugihara, G., May, R.M. (1990), Applications of fractals in ecology. *Trends. Ecol. Evol.* Vol. 5 p. 79-86.

Vu Tien Khang, J. (1989) Analyse de la variabilité génétique à partir des données généalogiques. In *La gestion des ressources génétiques des espéces animales domestiques.* . Paris, 1989, p. 52-60 .