Floris Takens
Mathematisch Instituut, Universiteit Groningen
Postbus 800, 9700 AV Groningen
F. Takens@math.rug.nl

Afscheidsrede Floris Takens

From chaos to early warnings

Een van de grondleggers van het moderne vakgebied 'dynamische systemen', Floris Takens, werd onlangs zestig jaar. Tegelijkertijd ging hij met emeritaat. Ter ere van zijn bijdragen aan de internationale ontwikkeling van de dynamische systemen, werd in Groningen van 25–29 juni een workshop 'Global analysis of dynamical systems' georganiseerd.

Op 30 juni werd deze viering besloten met een publiek evenement in de Senaatszaal van de Universiteit Groningen. Jacob Palis, president van de International Mathematical Union, vriend en collega met wie Takens veel heeft samengewerkt, sprak in de eerste lezing zijn lof uit over het werk en de persoon van Floris Takens. De tweede lezing die dag was van Floris Takens zelf.

Het onderstaande artikel is een bewerkte versie van die rede. Takens bespreekt hoe de theorie van de niet-lineaire dynamische systemen wordt toegepast om bij een chemisch proces vroegtijdig voor klontering te waarschuwen, daar waar gewone lineaire tijdreeks-analyse faalt.

We first need to say a few words about chaos. I do not mean chaos here in the theological sense of 'the disorder of formless matter and infinite space which is supposed to have existed before the creation', but rather in the mathematical sense, which refers to a situation where the time evolution is strictly deterministic, but still, to some extent, unpredictable. I will illustrate this somewhat paradoxical state with an example: the game of billiards.

Billiards

This game is played on a table of approximately 250 cm by 125 cm, usually with 3 balls (although there are variants with 2 to 22 balls); when a ball hits a side of the table it is reflected back. The general idea is to hit one of the balls, giving it an initial velocity, such that it collides with other balls. The rules of the game give prescriptions about which balls should collide and in which order. With some experimentation one finds that it is not hard to hit a ball in such a way that it collides with a prescribed ball. What happens after this first collision, however, is harder to control. This is due to the fact that small differences in the direction of the initial velocity of the first ball lead to big differences after the first collision. This can be understood in the following way: if the distance between the first ball (which we hit directly) and the second ball, with which the first collision is to be made, is denoted by ℓ and if the diameter of the balls is denoted by d, then, in order to realize this first collision, we need to give the first ball an initial velocity whose direction is in a sector with opening angle approximately $2d/\ell$ (radians). This is illustrated in figure 1. So a deviation in the initial direction up to d/ℓ can cause the ball to go, after the first collision, anywhere; this means that ℓ/d is a measure of the amplification of small deviations in the velocity which occurs at the first collision. We shall call this amplification factor the *scattering factor*. Note that the value ℓ/d is only a rough estimate, since, among other things, we neglected the strength of the initial velocity and took only its direction into account.

The above argument can be repeated: after the second collision the effect of a deviation of the initial velocity is multiplied by the square of the scattering factor. Et cetera. This growth of the effect of a small deviation is the reason that skill is required to play this game, and that it could become a sport.

We can imagine a frictionless game of billiards where much more collisions are possible. After the n-th collision the effect of a deviation of the initial velocity, if sufficiently small, is multiplied by the scattering factor to the n-th power; this is called *exponential growth*. It turns out that exponential growth rapidly leads to unbelievable results.[2] For example, if one wants to hit a ball so that it collides with nine balls in a given order, then one has to take into account even the effect of the gravitational force between the player and the balls![3] In a situation like this, where the effect of small deviations in an initial state grow exponentially with time, one speaks of *sensitive dependence on initial states*. Systems with sensitive dependence on initial states are called *chaotic*.

The same arguments can be appied to the molucular motion in a gas. Even without quantum mechanics, this rapidly leads to complete unpredictability.

General description

The fact that isolated systems, which are reasonably predictable over short periods, may be completely unpredictable over longer periods is known in other situations too. One may think of the weather. However, in the case of the weather we have a system,

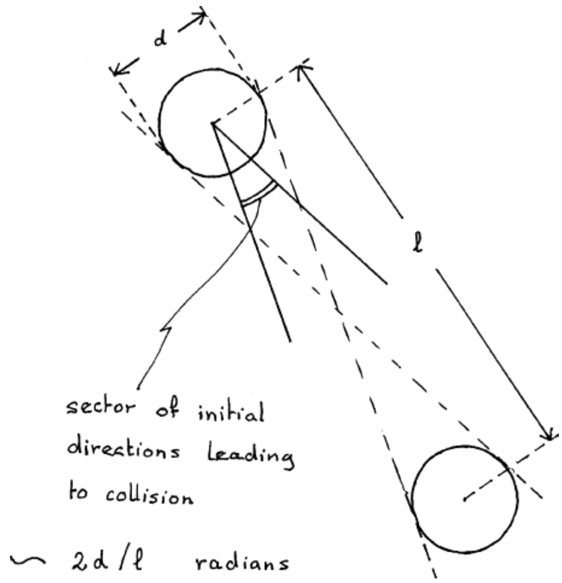

Figure 1 The game of billiards

the state of which has a much greater complexity (or dimension).

With this we mean the following. The state of a game of billiards is given by the positions of, say, three balls (that is, by six numbers) and their velocities (another, by six numbers), at least if we neglect the spinning of the balls; so the state of the game of billiards is 12 dimensional (given by 12 numbers). The state of the meteorological system, as it is represented in present day computer models, has a dimension of the order of 5,000,000. In principle, and this was the standard opinion up to a few decades ago, the unpredictability of the weather is related to the fact that its states have such a high complexity.

Around 1970, within the framework of the mathematical theory of dynamical systems, Stephen Smale and co-workers had developed a geometric theory of abstract dynamical systems with both low dimension states and with sensitive dependence on initial states.[4] On the other hand, already in 1963 the meteorologist Edward N. Lorenz had experimentally found sensitive dependence on initial states in an extremely simplified model of the meteorological system.[5] (In this model the states had only dimension three.) Years later it became known that these investations were dealing with the same phenomenon, only by different means. It has now become clear that the unpredictability of the weather, at least over longer intervals than a few days, is related to sensitive dependence on initial states. This effect is known as the *Butterfly effect* after the title of a paper by Lorenz: *'Can the flap of a butterfly's wing stir up a tornado in Texas?'*. Apart from this, it is questionable whether the meteorological system should be considered as strictly deterministic.

Deterministic processes with sensitive dependence on initial states were produced as laboratory experiments. For example, in nonlinear electronic devices, chemical reactions, and also in mechanical experiments involving resonances. Also, one is now convinced that this sensitive dependence plays a role in the dynamics of our solar system, but that is on a much longer time scale.[6] In general however, deterministic systems with low dimensional states are rare, especially outside the laboratory. Still one may expect that there are situations which can be realistically described in terms of a combination of sensitive dependence on initial states *and* some randomness.

Chaos, determinism, and predictability

The notions of determinism, chaos, and predictability can be illustrated in a schematic way by following the future development of an initial state or a set of initial states for a short time (say one unit of time) and also for a longer time interval, here denoted by $T \gg 1$. In figure 2 we compare the deterministic case with the stochastic case: that is the case where the state at time $t+1$ is not completely determined by the state at time t. The indeterminacy is supposed to be completely random. These diagrams suggest that, if we cannot distinguish between nearby states, the difference between the stochastic case and the chaotic case is only small, except that in the chaotic case the short term predictability is better. This short term predictability can still be somewhat extensive: for the meteorological system it is this better predictability on the short term that makes weather forecasts better than statistical predictions such as 'the weather of tomorrow will be like the weather of today with some correction in the direction of the climatological averages for the time of the year'.

In situations where it is interesting to know how predictable the future is, one may think for instance of the financial markets, one might want to know whether one deals with randomness or with sensitive dependence on initial states. As mentioned before, pure determinism is rare, so we expect no pure sensitive dependence on initial states. However, we should be able to detect the difference between the extreme cases: 'purely random'[7] and 'purely deterministic but with sensitive dependence on initial states'. This is a question in *time series analysis*.

A time series is a, usually long, sequence of measurements $\{x_1, x_2, \ldots\}$. In terms of the game of billiards without friction mentioned earlier one may think of the position of one of the balls measured at each of the successive collisions. The question is how to decide whether such a time series is generated by a deterministic process. (In the case of such a time series from the game of billiards it will be very hard to detect the deterministic principle.)

Consider the following two examples shown in figure 3. In this case one can find out, by an hoc method, that time series (a) is predictable: Plotting for each n the point with coordinates (x_n, x_{n+1}), one sees that x_{n+1}, for the first time series, is completely determined by x_n (within the accuracy of the figure) so that the system must be deterministic. One can even conclude that the scattering factor equals 5. Time series (b) in figure 3, on the other hand, does not reveal any dependence of x_{n+1} on x_n. In this way we detected the difference between the two time series very clearly. However, it is easy to produce examples where this ad hoc method does not work.

Linear time series

There is a systematic statistical theory dealing with time series.[8] It describes such series in terms of autocovariances (and average values). This is what one usually calls linear time series analysis. These autocovariances contain information about the average dependence between successive elements of the time series, but these autocovariances are absolutely unable to give information about the question whether the time series is generated by a deterministic process or not. In fact, whatever the autocovariances are, one can make a deterministic process generating time series with the same autocovariances. When the original time series was generated by a stochastic process, the corresponding deterministic process has sensitive dependence on initial states. For example, the two time series in figure 3 have equal autocovariances, the first being generated by a deterministic process with sensitive dependence on initial states, while the second is purely random.

This means that additional methods for the analysis of time series are needed if one wants to get an indication whether a given time series is generated by a deterministic process or not; such methods can be expected to also give information about the short term predictability of the time series.

Nonlinear time series and predictability

The main systematic way to analyse a time series from the deterministic point of view makes use of the so called correlation integrals $C^k(\varepsilon)$, with k a positive integer and ε a positive real number. The number $C^k(\varepsilon)$ is the probability that two randomly chosen segment of length k differ by at most ε. Alhough this definition was motivated by the problem of analysing deterministic systems by the time series which they produce, one can see directly that these quantities should give information in general about the predictability of a time series in the following way:

Suppose we recorded a quantity during a (long) period so as to obtain a time series $\{x_1, \ldots, x_N\}$. The question is 'how to predict the next value?' and 'how reliable is our prediction?'. We try to predict by analogy: we search for a segment in the past which equals, up to a small error, the final segment $x_{N-k+1}, \ldots, x_N$.

We have to fix here the length k of the segment and the error ε which we are willing to allow; these quatities are to be chosen by a trial and error method. If we have found such a segment $x_m, \ldots, x_{m+k-1}$ in the past, we base our prediction on the assumption that: 'what happened (approximately) in the past will happen (approximately) again'. This is why we predict that x_{N+1} will be close to x_{m+k}. Now the question is: how reliable is such a prediction? An indication can be obtained from the correlation integrals $C^k(\varepsilon)$. Namely, in making the prediction we assume that two segments of length k, which were close, say within distance ε, will stay close if they are both prolongated to length $k+1$. The probability that this will indeed happen is equal to $C^{k+1}(\varepsilon)/C^k(\varepsilon)$. So if this quantity is close to 1 then the prediction has a high probability to be correct within an error of ε, but if this quantity is close to zero, the prediction has only a very small probability of being correct within an error of ε.

The above method of prediction is rather primitive and can be improved. Still the quantities $C^{k+1}(\varepsilon)/C^k(\varepsilon)$ give a good indication of the predictability. In fact it turns out that very often $C^{k+1}(\varepsilon)/C^k(\varepsilon)$ becomes practically constant for increasing values of k (and ε fixed). So one defines the *ε-entropy* $H(\varepsilon)$ as

$$H(\varepsilon) = -\log(C^{k+1}(\varepsilon)/C^k(\varepsilon)).$$

The effect of '$-\log$' is that $H(\varepsilon) \geq 0$ and that the smaller $H(\varepsilon)$ is, the more reliable the predictions are. So $H(\varepsilon)$ is a measure for the *unpredictability*. For a time series generated by a deterministic system $H(\varepsilon)$ stays finite for decreasing ε, whereas for a time series generated by a random system, $H(\varepsilon)$ tends to infinity as ε goes to zero.

So the unpredictability, or entropy, can be estimated in terms of the way $C^k(\varepsilon)$ decreases as k increases.

Similarly, the *dimension* $D(k)$ of the states can be estimated in terms of the way $C^k(\varepsilon)$ decreases as ε tends to zero: if

$$C^k(\varepsilon) \sim \varepsilon^{D(k)}$$

for fixed k, then $D(k)$ is an indication for the dimension of the states of the system. For a time series generated by a deterministic system $D(k)$ stays bounded for increasing k and approaches the 'real dimension'; for a time series generated by a random system, the values of $D(k)$ increase in an unbounded way. The values of $D(k)$ are also related to unpredictability: if $D(k)$ is big then $C^k(\varepsilon)$ decreases very fast with decreasing ε, which means that one needs an extremely long time series in order to find segments in the past which are sufficiently close to the 'final segment'.

Although the question of predictability was central in the above discussion, and also played an important role in the development of the theory of nonlinear time series analysis, it now seems that the more important applications go in a somewhat different direction. There were no cases outside the laboratory where unexpected predictability was found. There have been some claims that the financial markets allowed better short term predictions with these 'nonlinear' methods, but these claims were not really substantiated. And that was to be expected one way or the other: if the method would have worked, it would have become known, many people would have used it, and that would have changed the dynamics of the markets. But in turn this would have ruined the method since it is based on the assumption of stationarity (what happened in the past will happen again).

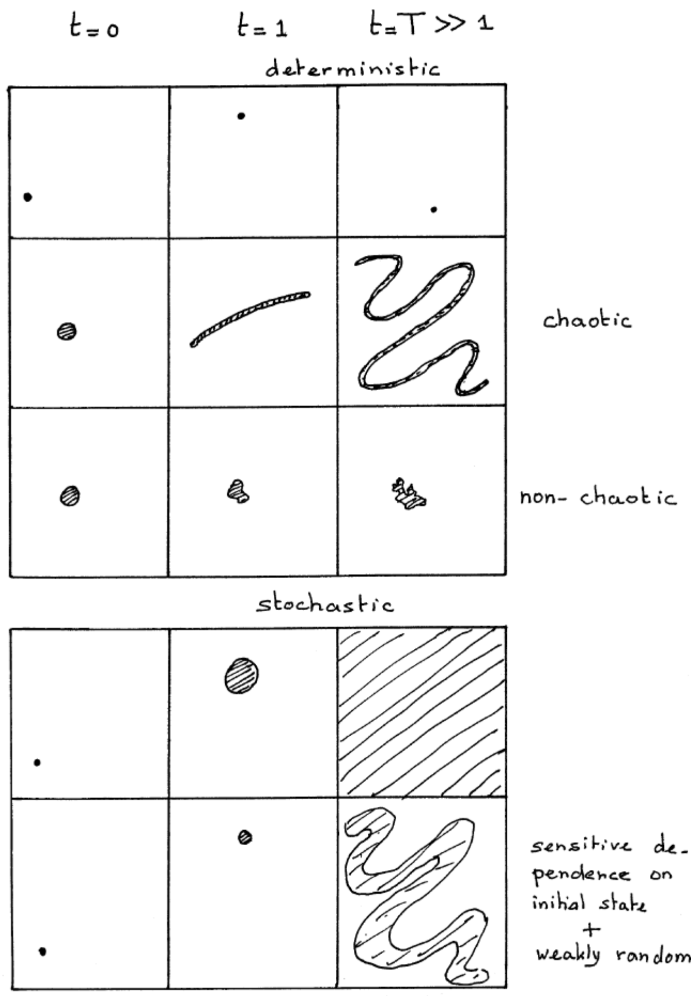

Figure 2 The behaviour of deterministic and stochastic systems

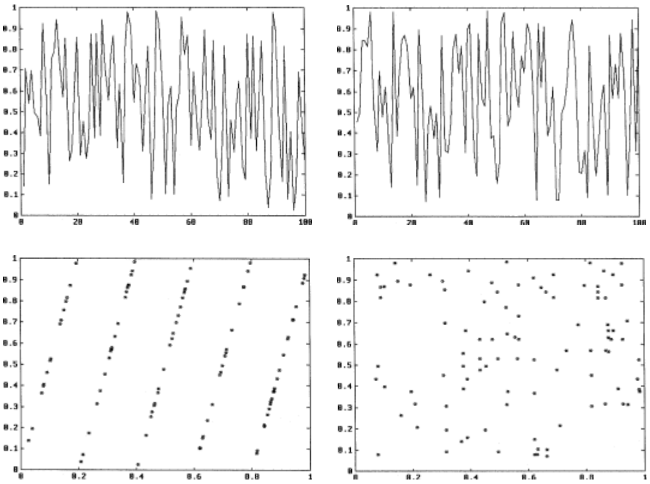

Figure 3 On top, on the left a deterministic time series, and on the right a chaotic one. Down, x_{n+1} plotted against x_n, for both cases.

Early warnings

The general goal of a system for early warnings is to detect from the recordings of a dynamical process whether some undesirable change is to be expected. We first consider the case of a chemical reactor of which a characteristic variable, like the pressure, is measured. This refers especially to so called fluid bed reactors as studied for example by the group of Van den Bleek at the Technical University of Delft. Such chemical reactors can sometimes get stuck, due to agglomeration of the reacting particles. If such an event is not avoided, then, especially in the case of big industrial reactors, it may be a big job to clean the reactor and the finantial consequences can be considerable. So one is interested in a system that produces a warning, as early as possible, whenever the process shows a tendency of going wrong. False alarms should be reduced to a minimum at the same time. The obvious strategy is to analyse whether the general characteristics of the measured signal change. It turned out that the analysis in terms of autocovariances, and related quantities, do not lead to reliable results. Also these pressure signals are not predictable over time intervals which are sufficiently long for the predictions to be used directly for a warning system. What, however, turns out to be a much better characteristic of the signal is the predictability itself: a change in the entropy is a good indication that the size of the particles changes (due to agglomeration), at least for reactors of laboratory size which were used for the investigations so far.

Another situation where these correlation integral methods are used is that of the analysis of the electro encephalogram (EEG) recordings of epileptic patients. One of the objectives there is to be able to give an early warning prior to a seizure, i.e. an epileptic attack. In the last decade groups in both Germany and France have made considerable progress in predicting these seizures by applying methods based on correlation integrals.[9] In this case, however, it is not the entropy but the dimension which serves as the most important indicator.

In both these cases, and possibly in other such applications, these correlation integral methods are applied to time series from systems which can certainly not be considered as both deterministic and having low dimensional states. This means that the mathematical theory of correlation integrals for time series from deterministic but chaotic systems is not valid. Still, it is not unreasonable to expect that these quantities, like entropy and dimension, are meaningful beyond the domain of validity of the mathematical theory; of course, this needs experimental validation in each case. The reason why in one case (chemical reactions) one has to use the entropy and in another (epilepsy) the dimension is essentially unknown but should have its origin in the dynamical structure of these processes. This means that the construction of these early warning systems is, to a large extent, a matter of experimentation. The main mathematical contribution is the construction of quantities which are relevant in special and simple, but relevant, situations and, hence, are candidates for useful indicators in the context of these more complicated dynamical processes.

Epilogue

After this exposé related to my own mathematical investigations in the last decades, I want to say a few words about some problems which seem to threaten the mathematical profession. I am thinking of the dramatic decrease of the number of students over the last fifteen years. This is a development which is not only restricted to mathematics, neither is it restricted to the Netherlands. There are some general reasons for this decline which cannot be influenced by the mathematical (or general scientific) community, such as:

- research in mathematics, and other exact sciences, involves a lot of painstaking work, or, as we call this in Dutch, monkish work; this does not seem to fit well with the spirit of the age.
- since the end of the cold war, the urge of being ahead in high technology and basic research does not exist any more — design has become more important.

Recently, there have been many initiatives to promote the interest in mathematics of future students, which led to fruitful cooperation between universities and the institutions for secondary education. Alhough these initiatives are very worthwile, and should certainly be continued, the effect does not seem to be sufficiently strong. In my opinion there are reasons to look also at other aspects of this problem. Without giving concrete solutions I want to point to two events in the recent years which, I think, one should keep in mind when thinking about the image of mathematics.

Secondary education

The first event I am thinking of is the discussion about the reform of the program of secondary schools in the Dutch parlement in 1997. In the law implementing this reform so-called profiles were introduced defining the different types of final examinations giving entrance to the university. These profiles were, and are, in order of increasing emphasis on mathematics:

1. society and culture,
2. society and economics,
3. nature and health, and
4. nature and technology.

In this discussion one of the members of parlement expressed the concern that the first profile would be too easy and, hence, would attract many students looking for an easy way out. In the answer from the goverment a suggestion was proposed: an amount of mathematics could be added to this first profile and, in order not to expand the number of mathematics teachers needed, this amount would be subtracted from the fourth profile. This led to a discussion in the newspapers. The representatives of professions, for which this first profile was intended, protested on the basis that mathematics was not needed for making this profile more serious and that mathematics could be easily replaced by more relevant subjects, such as Latin. The representatives from mathematics and the exact sciences protested because they did not want to see mathematics diminished. So the whole idea disappeared.

Still the story shows something. Even the responsible politicians seem to have the idea that one of the main purposes of mathematics is to make a curriculum heavier. Given this fact it is no surprise to see that people are often proud *not* to know any mathematics: they were smart enough not to become a victim of such artificial obstacles.

This event shows in my opinion that many hours of mathematics in the secondary school curriculum is not necessarily a good thing, neither for the students nor for the mathematical profession.

Quiz master problem

Another event is the public discussion in newspapers, not only in the Netherlands, but also in the United States and Germany, about the so called *Quiz Master Problem*. I am grateful to Niels Kalma for providing copies of a substantial part of the newspaper discussion. The problem can be stated as follows:

In a quiz a candidate is confronted with three identically looking doors; behind one of the doors there is a prize. The candidate is invited to select one of the doors, which however is not yet opened. Then the quiz master opens one of the two doors which were not selected by the candidate, showing that behind that door there is no prize. Now the canditate is allowed to change the door of his preference. Then that door is opened, and if the prize is behind that door, the candidate will win it. The question is: what is the better stategy, to change or not to change ones door of preference after the quiz master has opened one door behind which there was no prize.

A correct solution of the problem as given in the German weekly *Die Zeit* is the following.

There are three possibilities with equal probablity:

- *the prize is behind the selected door;*
- *the prize is behind non-selected door number one. In this case the quiz master has to open the non-selected door number two;*
- *the prize is behind non-selected door number two. In this case the quiz master has to open the non-selected door number one.*

It follows that in two of these three cases changing door leads to success; not changing door is only successful in one case. Hence changing door gives a greater probability to win.

Main sources of confusion in the discussion were arguments such as:

'Due to the quiz master opening one door, the probability of the prize being behind that door drops from $1/3$ *to* 0*; this probability moves to the door that is not opened and not selected.'*

This is an unclear argument, which happens to give the right answer. It was countered by an equally unclear argument, which happens to give a wrong answer:

'Why doesn't the probability move to the selected door? By symmetry one would expect that half of this probability goes to each of the closed doors, so that for both the new probability is $1/2$*.'*

What is missing in such arguments is a clear investigation of the notion of 'probabilities moving due to extra information' and the rules according to which these probabilities move. Such a discussion can be given, but it needs great care, since the notion of probability itself is aready not free of problems. In the solution as given in *Die Zeit* such unclear notions are avoided.

Both in the discussion in Germany and in the Netherlands the general confusion was comparable. A difference was that the German *Die Zeit* resolved the discussion with a good explanation, not only of the right solution, but also explaining the reason of the confusion. In the Dutch daily newspaper *NRC Handelsblad*, which is considered to be a 'quality paper', this was beyond the available intellectual level and there the discussion was closed with the observation:

'The misunderstanding between common sense and mathematicians is clearly unbridgeable.'

This remark makes it clear that the journalist in question seems to have no idea that the essence of a mathematical argument is not that it deviates from common sense, but that it is absolutely conclusive, and hence does not make use of notions like 'probability changing due to extra information' without such notions being fully understood. Such conclusive arguments, sometimes called proofs, play a role in mathematics similar to experiments in the natural sciences. Though it seems much more solid to build on experiments than on arguments, it is a fact that the arguments used by Euclid, some 2500 years ago, are still essentially considered valid.

It shows that even those who are not proud of knowing no mathematics may not really care about the arguments they use and, hence, convey a disdain for mathematical reasoning.

References

1. Henk Broer, Bernd Krauskopf and Gert Vegter (Eds.), *Global Analysis of Dynamical Systems: Festschrift dedicated to Floris Takens for his 60th birthday*, Institute of Physics Publishing, 2001.
2. D.L. Meadows, *The limits of growth — A report of the Club of Rome on the predicament of mankind*, Universe Books, New York, 1972.
3. I. Ekeland, *Das Vorhersehbare und das Unvorhersehbare*, Harnak Verlag, 1985, p. 87.
4. S. Smale, 'Diffenrential dynamical systems', *Bull. AMS* **73** (1967), 747–817.
5. E.N. Lorenz,'Deterministic nonperiodic flow', *J. Atmos. Sci.* **20** (1963) 130–141.
6. See the *Bernoulli Lecture* by J. Laskar, Groningen, 2001
7. We did not define here this notion. It should be interpreted as generated by a stationary Gaussian process, or by an i.i.d. process.
8. M.B. Priestley, *Spectral analysis and time series*, Academic Press, 1981.
9. See C.J. Stam, *Klinische neurofysiologie, pleidooi voor een dynamische toekomst*, inaugural lecture at Vrije Universiteit Amsterdam, 2001.