Signal analysis and modelling of non-linear non-stationary phenomena

- from Human Voice to Financial Markets -

Dissertation

zur Erlangung des Doktorgrades

der Fakultät der Naturwissenschaften (Dr. rer. nat.)

der Bergischen Universität Gesamthochschule Wuppertal

vorgelegt von

Lorenzo Matassini

aus Montevarchi

Juli 2001

ii

To Maud, Leonardo and Marvin

Abstract

We address the problem of coping with non-stationarity in time series analysis. Very often the non-stationarity is quite weak and can be ignored for many purposes. But this is not the case here, since the systems generating the signals under analysis contain information on the non-stationarity themselves. In other words the variability of the dynamical regimes due to non-stationarity is the essential property and cannot be cut out.

The work consists essentially of three parts: *(i)* A theoretical approach to non-stationarity, showing that if one is interested in implicitly discovering the equations of motion of the system, then under quite general hypothesis the over-embedding allows one to solve this task, of fundamental relevance for prediction, noise reduction and data classification. *(ii)* Applications to human voice, where the non-stationarity is involved in the concatenation of consecutive phonemes, to be considered as signals with few degrees of freedom but non-constant parameters. In other words non-stationarity here means that the instantaneous dynamics can differ very significantly from phoneme to phoneme. Three problems of technological relevance are addressed, namely the noise reduction of human speech signals with a proper optimization scheme, a classification of vocal disorders and a software correction of voice patologies. *(iii)* Applications to financial markets, where the non-stationarity is related to the volatility of market prices, trends and seasonality. Primary tasks are the analysis of correlations in the absolute value of price differences, very useful in risk management, and the development of models where large numbers of units interact, giving rise to empirical observed phenomena like panic selling, herding behaviour, speculation.

Preface

One of the most common critique about research involves the gap between abstract and theoretical reasoning and the solution of practical problems. Very often there is almost no interaction between scientists and people working in the industry, the first being fashinated from the challenge of conceptual thinking, the second having mostly to do with strong deadlines and heavy economical constraints. Looking back at my education I can assert to belong to a strange class of researchers, since I graduated on electronic engineering and automatic control at the University of Technology of Florence to switch then to the study of non-linear time series analysis at the Max Planck Institut for the Physics of Complex Systems in Dresden, also called the Florence on the Elbe.

This work is the result of the evolution of my education and I think it will appear quite often going through the text that I am neither a pure physicist nor anymore a pure engineer. Working at the border of two such different worlds is very attractive and fruitful but simultaneously quite dangerous, since one has to find the right compromise and trying to be rigorous and practical. I have focused, in the framework of non-linear dynamics and non-stationarity, on the complementary problems of data analysis and system modelling. In order to maintain faith to the previously mentioned challenge, I have chosen to cope with human voices and financial markets, performing data classification, noise reduction, prediction, analysis and modelling always trying to show the similarities and the differences that interplay among these systems.

The goal is of course to obtain valuable results and performances that are comparable with the state of the art in the field, but also the idea of testing the applicability of concepts and tools derived from the theory of deterministic chaos has played a fundamental role. When approaching the noise reduction problem, for instance, it is

very satisfactory to get performance at least comparable, when not even better, to famous schemes where several people have been working for several years and with huge resources. But also the application of chaos theory outside the typical working hypothesis is alone a good reason to invest time in this direction. The discovery of an over-embedding that enables to perform an automatic separation of a sentence into the constituent phonemes is a result that goes beyond the applicability itself. The construction of a very realistic model for the stockmarket has not only the purpose of reproducing empirical features, but it also offers some indications about the collective behaviour of people competing for the same goal. This is again an example of working at the borders between two different disciplines, namely stochastic processes and psychology. Feature space analysis and software corrections of vocal pathologies provide direct applications in biomedicine and has a great technological impact.

In the following I am going to briefly focus on the perspectives and the limit of this work and of modern science in general, then I will discuss the problem of non-stationarity as an introduction for a better understanding of the two main parts of this contribution, namely the extraction of the essential features from the human voice and the analysis of up to which extent financial markets can be considered not to be random and therefore to posses deterministic structures. I am indebted to several people, but I would like to thank here Holger Kantz, Stefano Ruffo and Roberto Genesio, who are mainly responsible of the foundation and the maintenance of the link between Florence and Dresden, between physics and engineering.

<div align="right">Dresden, 2001</div>

Contents

List of Figures

Chapter 1

Dynamical systems and chaos

Prediction is difficult, especially of the future.

(Niels Bohr)

Nothing in nature is random, a thing appears random only through the incompleteness of our knowledge.

(Spinoza)

The scientist does not study nature because it is useful; he studies it because he delights in it, and he delights in it because it is beautiful. If nature were not beautiful, it would not be worth knowing, and if nature were not worth knowing, life would not be worth living.

(Henri Poincaré)

Nowadays' science believes that if an accurate mathematical description of a physical system can be found, then the potential for a deep understanding of the system's properties exists and therefore predictions of its evolution can be made. These assertions have been proved and used for a wide variety of phenomena, ranging from the motions of planetary bodies to the fundamental constituents of matter. However, it is quite easy to show that these notions are generally not true when dealing with non-linear phenomena and non-stationarity. This may sound as a surprise for many people, since it is common experience that although the fine details of the complete solutions of the full governing equations of any given system are not

known, approximate versions of the laws may be used to make robust predictions about its behaviour, predictions that are very often experimentally confirmed.

1.1 Predictability

Let us consider, as an illustration of this simplified scientific approach, the laws of planetary motion as formulated by Newton and Kepler. It is possible to predict accurately the orbits of the moon around the earth, for example, when the influence of other planets is ignored. Further, these predictions have been tested over centuries and are found to be robust. They are based on the sound principles which Newton established in his *Principia* over 300 years ago by considering the gravitational interaction of two planetary bodies. In this case there is an analytical solution to the mathematical formulation. Now, if a third smaller planet is introduced into Newton's mathematical description of the gravitational interaction of two massive bodies, this gives rise to an intractable three-body problem. Newton solved various restricted versions of the complete problem but he was unable to find a general solution to it. Two centuries later, Poincaré hinted that the motion of the third smaller planet orbiting in the gravitational field of two massive planets would, in general, be highly complicated. With the actual computational power we can obtain very precise numerical solutions of the three-body problem and it can be shown that the orbit of the third planet is indeed unpredictable in practice: Every small error in the initial condition settings would drastically reduce the time horizon of the prediction.

With such a quite simple and very well established example one is able to provide a situation that can give rise to both predictable behaviour and complicated unpredictable motions. It is of fundamental relevance to note that the appearance of the complications is not due to a breakdown in the validity of the equations, but rather it is just one of their properties. Indeed we can take the conceptually much simpler system of an excited pendulum and show a systematic progression from one type of behaviour to the other. Let us consider, for instance, a pendulum driven by a sinusoidal force and subject to friction, giving rise to the following model:

$$\ddot{x} = -\rho\dot{x} - \omega^2 \sin x + a \cos \Omega t. \tag{1.1}$$

The plane of Fig. 1.1 is a two-dimensional section through the extended phase space,
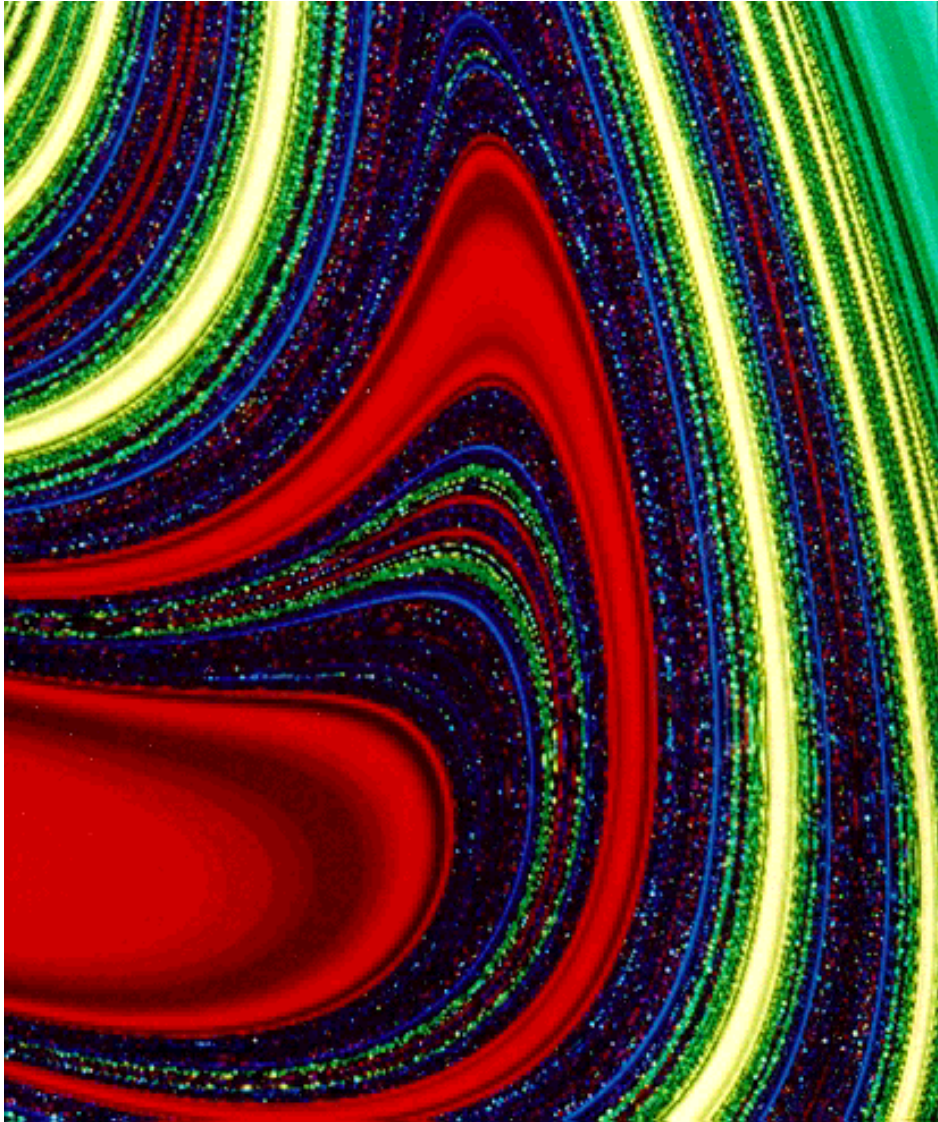
Figure 1.1: Even a simple pendulum can show a chaotic behaviour. Here the pendulum is driven by a sinusoidal force and it is subject to friction. The plane is a two-dimensional section of the three-dimensional extended phase space, whose components are position and velocity of the pendulum and the phase of the driving term. Taking the points as initial conditions, different colors code different asymptotic behaviours of the pendulum, showing a great variety of structures and fractal borders.

whose components are the position x of the mass, its velocity $\dot{x}$ and the phase Ω of the driving force. Different colors are associated with different asymptotic behaviours and very interestingly one can show that the borders between them present a fractal structure. The point we want to stress here is that in order to understand the appearance of the disordered behaviour we do not necessarily need to invent new physical laws. Rather we have to recognize that the presence of non-linear terms in the governing equations can give rise to qualitatively different types of behaviour, which are not present in the solutions of the linearized versions of the equations. Considering the linearized approximation can sometimes be of poor value, since it is like studying a different model that is not able to capture the essential features of the system.

The idea that nothing is really linear but almost everything can be quite easily linearized has become very strong because of many great scientific breakthroughs achieved using approximate, linear methods for a wide range of problems. However, many natural processes across the whole spectrum of science are inherently strongly non-linear and simple adaptation of known methods may not be sufficient to resolve important issues, such as prediction of the weather or climate, for example. Therefore, we need to develop a new way of dealing with non-linear processes and this is the subject of much current research, including this small contribution. Many of these new research ideas are of course reinventions of older concepts but equally there are fresh outlooks on some classical problems and serious attempts to tackle difficult issues which have been brushed aside in the past.

The first practical demonstration of the phenomenon can be found in a work of Lorenz [61], where he carried out a numerical study of an extremely crude model of atmospheric convection and found that when the integrations of the equations were started with two slightly different initial conditions, very different outcomes were realized. This observation is an example of sensitivity to initial condition in that a change in the least significant digit of the starting conditions for the calculation will eventually lead to completely different outcomes despite the fact that the computations involve the representation of a deterministic law on a deterministic calculating machine. In other words, if two calculations were started with the same initial conditions to within the accuracy of the machine then there would be no divergence of the outputs, i.e. the unpredictability is not an effect of the accumulation of rounding errors, but rather something we cannot avoid.

Sensitivity to initial conditions is the first recorded indication for the appearance of a strange attractor in the solution space of the equations. One should anyway note that chaotic systems can have strange attractors, but they can also have trivial attractors, e.g. the tent map. In other words, strange is typically attributed to fractality, but there are also cases where the motion on the attractor is strange in view of the above introduced regular attractors. A simple physical analogy of an attractor is the following: Let us put a ball inside a circular tube and consider a two-dimensional cross-section of it. If we displace the ball to any point up the bottom half of the tube it will always roll back to a uniquely defined point at the bottom when it is released. If we now imagine that the motion of the ball is projected onto a sheet of paper below the tube then we can see that all trajectories of the ball will lead to a single point on the paper which we call the fixed point of the system. What is more, here it is called an attracting fixed point since any trajectory within the section of the tube will lead to this point. It is therefore the attractor for the system in this stationary state.

Now suppose we lay our tube flat on a turntable which is rotating at a constant speed. The ball will now run around the tube at some distance up its side which will depend on the selected speed of the turntable. The projection down onto the paper will now be a circle and we call this new attractor a limit cycle. If we again displace the ball from its attracting orbit then it will quickly return to it. In fact, if the ball is started at any point within the tube then it will eventually end up rotating on the uniquely defined attracting cycle for this system. We can then say that the basin of attraction for the cyclic attractor is the whole of the inner surface of the tube in this case. If we now add to the turntable a large up and down sinusoidal motion then the ball will orbit around the inside surface of the tube. If the frequency of the up and down motion is not commensurate with that of the rotation then the trajectory of the ball will eventually cover the whole of the inside of the tube. This new attractor is called a torus.

1.2   Strange attractor

We have now introduced all the necessary ingredients for the understanding of the
so-called strange attractor, where two competing effects are present: *(i)* Points
which are neighbourhood at some instant in time diverge exponentially fast due to
the sensitivity to initial conditions; *(ii)* Trajectories must remain in a finite region
of the solution space, otherwise the concept of an attractor has no meaning. An
immediate consequence of such a *stretching and folding* mechanism is that strange
attractor behaviour cannot exist in two dimensions or less for differential equations[1].
If we tried to construct a deterministic model whose solutions were confined to a
plane then trajectories could diverge from each other but they could never remain
in a finite region without intersect each other because deterministic rules do not
allow any intersection. Poincaré suggested to consider a geometrical structure of the
solutions of the equations of motion of dynamical systems, leading to the idea of
Poincaré map (Fig. 1.2), where the intersections between the trajectory of a system
and an hyperplane are considered. Analysing the phase space of a system of m
autonomous equations, we find that locally the direction tangential to the flow does
not carry much interesting information. The position of the phase space point along
this direction can be changed by reparametrising time. Therefore it is clear that
it has no relationship to the geometry of the attractor and does not provide any
further information about the dynamics. As an immediate consequence, one can
reduce the phase space dimensionality by 1 and turning the continuous time flow
into a discrete time map.

   In order to obtain a *Poincaré map*, the first step is to identify a suitable oriented
surface in phase space. Matematically it makes no difference which plane one uses,
provided it is not tangent to the trajectory. Then one can reconstruct an invertible
map on this plane simply by following the trajectory of the flow. The iterates of the
map are given by the points where the trajectory intersects the surface (only one
of the two possible intersections has to be considered; in Fig. 1.2 from above). The
relation between the discrete and the original time is not a proportionality, since
the time between two successive intersections depends on the actual path in the
reconstructed state space and on the chosen plane.

   In the context of time series analysis, it has become clear that traditional linear

---

[1]Chaotic behaviour can however be represented by lower-dimensional discrete maps.
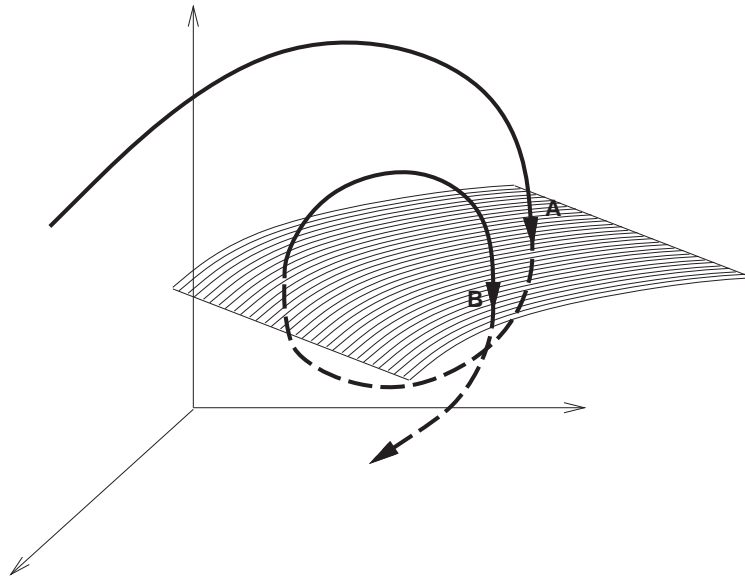
Figure 1.2: Construction of the Poincaré map, given by the intersections between the trajectory of the system and a given hyperplane (in this case just the points A and B). The orientation of the latter is qualitatively not essential, giving rise to different Poincaré maps having the same topological structure.

methods may give an unnecessarily complicated interpretation of the contents of a relatively simple signal, whereas the geometrical representation immediately gives a more informative description. This leads one to reconstruct a vector valued state space and the attractor therein from a scalar time series, called embedding. Any geometrical structure present in the reconstructed attractor from an apparently irregular time series implies that there is some low-dimensional behaviour present in the signal. This can be of enormous benefit when dealing with a signal from an unknown system whose governing equations are either speculative or are known to be high dimensional. In particular, if the reconstructed attractor has a geometry or topology which is also found in a more well-understood problem, then predictions about changes in the dynamics with variation in control parameters of the unknown system can be made with some degree of confidence.

Next, both qualitative and quantitative descriptions of the underlying solution structure which gave rise to the observed behaviour can be made. For example, unstable fixed points of a system can be identified and then quantitative estimates of dynamical information near these fixed points can be used to make predictions about,

as yet, unobserved behaviour. This feature is of immense value when investigating a system governed by partial differential equations and thus where the solution space is in principle infinite dimensional. The origins of observed highly complicated motions in one parameter range may be associated with the low-dimensional dynamical behaviour observed in another. Thus, identifying critical points about which theoretical analysis can be performed is often crucial in determining the important terms of the full equations which must be included in any model. In addition, quantitative estimates of the nature of the attractor can be made to establish the validity of finite-dimensional representation, the strength of the observed chaos and an upper bound on the number of modes which would be required in a model representation of the behaviour.

Last, but not least, it has been recognized in recent years that the limits of prediction can be increased in deterministic dynamical systems when they are in a chaotic regime. Such systems are of course predictable in principle since they are deterministic, but because of sensitivity to initial conditions practical prediction is delicate. However, we know that the reconstructed attractor has some geometrical structure which we can hope to exploit. Thus the trajectories in the reconstructed phase space must follow an approximately defined track and therefore there is a physical limit to the divergence of trajectories. The range of prediction can be extended beyond probabilistic correlations using the redundancy provided by the analysis of past values of the signal. It is worth remarking that chaos is just one feature of non-linear systems, albeit an important one. Another crucial property is the existence of multiple states with qualitatively different dynamics: At some prescribed value of the parameters of the system, there is the potential for several steady states, various types of periodic and non-periodic motion and chaotic states. The particular state which is realized in practice will depend on the history of its creation, e.g. the initial conditions and the speed of change of parameter to the preselected point or the variation of more than one parameter. Investigations of the interactions of various steady states give insight into the more complicated phenomena which are often found to be organized by the underlying steady solution structure. It should also be noted that the dynamical solutions can exhibit certain symmetry properties which are found in the steady regime such as the coexistence of pairs of states which break a geometrical symmetry. Finally, highly complicated motion can sometimes usefully be regarded as the manifestation of the presence of

multiple states where the system does not settle into any particular attractor but instead wanders through a range of them.

1.3 Limits and perspectives

The popularity of chaos theory, judged by the number of scientific publications and by the attention of media, has a dangerous drawback in that it could be seen as the answer to everything, something that it is obviously not. An example of this oversimplification is to be found in the phenomenon of turbulence in fluid flows. Over the past decade the words chaos and turbulence have become almost synonymous in many popular accounts and yet the connection between the two is far from obvious even today. Common opinion is that the ideas of chaos have thus far added very little to the understanding of the phenomenon of turbulence. Also in the new field of econophysics it is not clear whether the non-linear science could contribute to discover new features or just introduce additional formalisms and complications [70]. However, chaos seems to have survived the fashionable phase and perhaps one reason is that the natural world is inherently non-linear. Therefore, one should expect to find chaos rather than order and perhaps we now have some tools for furthering our understanding of what was previously thought of as random noise. There are many deep mathematical ideas behind the dynamical systems approach to the study of non-linear phenomena which are aimed at describing and understanding the origins and structures of complicated behaviour.

Naturally, there has been a tendency to extend some of these ideas into fields where there is no rigorous justification for doing so. These are often resolute efforts to tackle very difficult problems with new scientific ideas. The very last that one can say is that non-linearity should play a key role in natural phenomena and therefore some of these modern concepts may well give a new insight into some unresolved problems. On the other hand it is also worth noting that an irregular time series formed from ice core samples or the monitoring of a bodily function for example need not necessarily be describable in terms of low-dimensional chaos. Therefore one must remain cautious about such studies for it is quite easy to misrepresent the above ideas by an imprecise application of techniques which have thus far only been successfully tested in well-controlled laboratory situations. However, if new insights into difficult areas are obtained using this approach, which amount to more than

putting common sense into fancy mathematical language, then a great deal has been achieved.

Chapter 2

Non-stationarity

The art of asking the right questions in mathematics is more important than the
art of solving them.
(George Cantor)

In order to study an unknown system, one needs to get some information about
it. A very useful way to perform this task consists on measuring some quantities
related to the system, taking into account that a scientific measurement of any kind
is useful as far as it is reproducible, at least in principle. One has to be sure that
the numbers obtained from the measurement device correspond to properties of the
system. The concept of reproducibility and therefore of meaningfullness is strictly
related to the *notion of stationarity.*

In field measurements, non-stationarity is ubiquitous, but even in laboratory
experiments there are always small fluctuations of system parameters. In the sta-
tistical framework, stationarity is defined as the time independence of all thinkable
joint probabilities. In the framework of time delay embedding, it means that the
measure in an arbitrarily high dimensional embedding space has to be time invariant.
For a dynamical system the trajectory has to lie on an invariant set. In practice,
when dealing with a finite amount of data, the time series length has to be also
sufficient to cover the whole invariant set in an ergodic way, so that the invariant
measure can be estimated from the data. Of course, all system parameters have to
be constant during the measurement time.

Stationarity means that all the parameters of the system remain constant during
the measurement but unfortunately in most cases one has no direct access to the

system and therefore it cannot be established with a good degree of confidence that its parameters are indeed constant. A good definition of stationarity has to be related to the available time series, which has to provide enough information in order to determine the quantity of interest without any doubt. A process is called stationary if all transition probabilities from one state of the system to another are independent of time, at least within the observation period. This is a stronger requirement than the constancy of parameters, including that phenomena belonging to the dynamics were contained in the time series a number of times large enough in order to infer properly all the transition probabilities. In the following sections we will distinguish between three sources of non-stationarity, with different signatures and requiring different tools:

- *Drift of parameters.* The control parameters of the dynamical system generating the time series are not constant. This situation, typical for the human voice, can be treated using the idea of reconstructed phase space, phase space average and over-embedding.

- *Diffusive properties.* The transition probabilities are constant, but the marginal probabilities spread out and therefore we get a lack of recurrence of the process. The analysis of the Hurst exponent is very helpful in this case.

- *Trends and seasonality.* These are typical features of financial time series and make the estimation of the Hurst exponent not reliable. If trends are additive, they can be easily removed in order to correctly characterize the (eventual) other type of non-stationarity. This idea is called detrended fluctuation analysis.

Detecting non-stationarity can be a very difficult task. One could even say that stationarity is a property which can never be positively established. The situation is comparable to the linearity and the difficulty to assert that a given device has a linear input-output behaviour. Due to non-stationarity all the tasks related to data analysis and system modelling become more difficult or even meaningless. On the other hand, in particular with the two specific case studies we are going to treat, non-stationarity can be a relevant, essential property of a system, process, signal: It is in fact responsible for the rich dynamics in financial markets and for several features contributing to differentiate between human voices; a completely

stationary voice would be empty of information and if no local dynamical difference between consecutive phonemes were present, it would sound quite artificial and maybe uncomfortable to hear.

As a first requirement in time series analysis, the data set should cover a stretch of time which is much longer than the longest characteristic time scale that is relevant for the evolution of the system. This quantity can be estimated as the inverse of the lowest frequency in the power spectrum containing significant power. One simple stationarity check consists in dividing the data set into several segments and computing, say, correlations and transition probabilities for all of them, with the requirement that they should not differ beyond their statistical fluctuations. Unfortunately it can happen that a parameter drift is not able to produce any visible drift in the measurements. In such cases one needs special non-linear dynamical relations and the quantity to be compared along the different sets of data can be, say, the prediction error with respect to a proper non-linear model.

We have said that in order to be considered stationary, the system producing the time series has to satisfy at least two conditions, namely the parameters should remain constant and the data set should be sufficiently sampled. The latter requirement may be tested observing the convergence of a given quantity when a larger and larger portion of the time series is used for its computation. A typical quantity used for this purpose is the correlation dimension, since it suffers considerably from non-stationarity. A drift of one parameter results in the increasing of the correlation dimension because it destroys all the fractal structures. On the other hand, almost all the other types of non-stationarity and incorrect sampling produce values of the correlation dimension smaller than the correct one.

2.1 Testing stationarity in time series

The power of non-linear signal processing has been already established during the last decade. Quantities like entropy, mutual information, fractal dimension and correlation integral provide very useful tools for modelling and predicting time series. The main drawback of all of them and of most other time series analysis concepts, however, is that they assume more or less implicitly the stationarity of the data set under observation. It is well known that detecting stationarity in a time series is not an obvious task. When dealing with natural systems the problem becomes even

more subtle, since they are marked by influences of several external processes, which might lead to non-stationarity and long-range correlations. Furthermore, in testing stationarity in observations of natural systems, only realizations of the system under study are known instead of the system itself, whereas stationarity is a property of the process.

A review of several tests is provided by [111], where the time series is considered as a set of random variables. Several attempts grasp the notion of stationarity from the viewpoint of dynamical systems, with the drawback that no really quantitative characterization of the data can be provided. Other tests have been developed in the framework of mathematical statistics, but they require very strong assumptions on the time series and therefore they are not perfectly suitable for natural systems where these conditions are difficult to check or even evidently violated (e.g. the condition of gaussianity of the random variables is not satisfied). Since the detection of stationarity requires an observational length which should be large in comparison to the typical time scales of the underlying process, the discussion about stationarity cannot be carried out without addressing the question of long-range correlation.

A typical difficulty is that usually only a single realization $\{x_t\}_{t=1}^n$ of the system is available. In order to get a set of data series for statistical tools, one can divide the original sequence into several parts and proceed by testing, as reported by [111], the independence of time of the one dimensional probability density and of the power spectral density. One has to pay attention to the following aspect: In order to check the stationarity of the data set, every subset has to be considered stationary. This condition poses an upper limit to the length of every subsegment but on the other hand they should be chosen to be long enough: Each of them should present all the essential properties of the main time series, with particular regard to the long-range correlations. The latter condition represents a lower limit to the size of subsets. Furthermore, the number of them should be not so small in order to let significant comparisons be possible. For more details about the problem of finding the best choice of the window lengths see [111].

We want now to briefly review the stationarity test using the marginal distributions of the time series segments. The comparison of the probability distribution of the window i and the window j is performed through a χ^2 test [45]. The elements of both windows are coarse grained with the same binning r, such that the kth bin of

the ith windows contains the elements $X_k^i = \{x_{\rho_1}, x_{\rho_2}, ..., x_{\rho_{R_k^i}}\}$. If the time series elements were uncorrelated, R_k^i (number of elements belonging to the bin k with respect to the window i) could be understood as a realization of a binomial random variable with variance σ^2 of the same order of magnitude as R_k^i. But in the general case of correlated time series, R_k^i's are not binomially distributed: This requires a direct estimation from the time series of this variance.

Let us call the variance of the number of occurrences R_k^i of the time series in the kth bin with respect to the ith window $\sigma^2(R_k^i)$; this quantity can be estimated from the variance of the index number distances with respect to the elements of the kth bin inside the ith window in the following way:

$$\sigma^2(R_k^i) = c\sigma^2[D_k^i(m)_{m=1}^{R_k^i-1}], \tag{2.1}$$

where $\sigma^2[D_k^i(m)_{m=1}^{R_k^i-1}]$ is the variance of the index number distances $D_k^i(j) = \rho_{j+1} - \rho_j$. The autocorrelation function $\rho(t,s) = \langle(X_t - \mu)(X_s - \mu)\rangle/\sigma^2$ (with $\langle X_t \rangle = \mu$, $\langle(X_t - \mu)(X_t - \mu)\rangle = \sigma^2$ and $\{X_t\}$ with $t \in N$ a stochastic process) has to depend only on the relative time delay $\tau = t - s$, i.e. $\rho(t,s) = \rho(t-s) = \rho(\tau)$, since we have assumed the stationarity within a time series segment. The variable c depends on the window length n_w and on the number of elements in the bin R_k^i. In the case of an uncorrelated time series the value of c can be evaluated analytically:

$$c = \frac{\sigma^2(X_R)}{\sigma^2(X_D)} = \frac{n_w p(1-p)}{(1-p)/p^2} = n_w p^3, \tag{2.2}$$

where X_R are binomially distributed random numbers characterized by the parameters (n_w, p). The variance $\sigma^2(X_R)$ of this random variable reads $\sigma^2(X_R) = n_w p(1-p)$. The random number X_D of the distances of trials of X_R falling in the kth bin is geometrically distributed with the same parameters (n_w, p) and therefore $\sigma^2(X_D) = (1-p)/p^2$. p can be estimated from a realization by $\hat{p} = R_k^i/n_w$. The interesting aspect of this treatment is that also c can be estimated from a realization thanks to the following relation:

$$\hat{c} = (R_k^i)^3/n_w{}^2. \tag{2.3}$$

It has been empirically checked in [45, 111] that Eq.(2.3) holds for several types of correlated time series. The χ^2 test statistic is then:

$$t_{A,l} = ln_w^2 \sum_{k=1}^r \sum_{i=1}^{n_w} \frac{(R_k^i - R_k/li)^2}{(R_k+1)^3\sigma^2[D_k(m)_{m=1}^{R_k^i-1}]}, \tag{2.4}$$

where R_k denotes the number of elements in the kth bin and $\sigma^2(D_k)$ is the variance of their index number distances. The quantity $t_{A,l}$ is χ^2 distributed with $r(n_w - 1)$ degrees of freedom. By comparing the probability densities of the windows theoretically one tests the time independence of all the central moments $M_k^{t_1 t_2 \dots t_k} = \langle (X_{t+t_1} - \mu)(X_{t+t_2} - \mu) \dots (X_{t+t_k} - \mu) \rangle$. Of course practical problems like finite-size effects and coarse graining may represent a severe limit for the correct testing of high order moments.

Having tested the independence of time of the one dimensional probability density, the second test involves the independence of time of the power spectra density. We transform now the data into samples of the spectral distribution densities, i.e. , with respect to each window we get a set of data $f_{i\,i=1}^{k}$ that are identically and independently distributed with the spectral distribution. The autocorrelation functions and the corresponding Fourier transformation are estimated ($k = 0, \dots, n_\rho$) through:

$$\rho^j(k) = \frac{\sum_{\tau=1}^{n_w - r}(x_\tau^j - \mu^j)(x_{\tau+k}^j - \mu^j)}{(n_w - k)\sigma^j} \tag{2.5}$$

$$P^j(f) = \sum_{r=-n_\rho}^{n_\rho} \rho^j(r)cos(2\pi f r). \tag{2.6}$$

Similarly to the previous case, a χ^2 test statistic is applied introducing the following quantity:

$$t_{B,n_w} = l \sum_{i=1}^{l} \sum_{k=1}^{r} \frac{(R_k^i - R_k/l)^2}{R_k}, \tag{2.7}$$

which is χ^2 distributed with $(r-1)(l-1)$ degrees of freedom if $R_k = \sum_i R_k^i$. If the hypothesis of time independence is rejected then the sequence can be considered non-stationary. One can further compare mutually the samples $f_{i\,i=1,\dots,n_\rho}^j$ with respect to different windows. In this way it is possible to detect whether the structure of the data series is generally inhomogeneous as in transient states or whether there are only some parts (windows) with a special structure, like a burst.

It is also possible to consider a mixed statistic, given by the simple relation:

$$t_l = t_{A,l} + t_{B,l}, \tag{2.8}$$

where the addition is due to the introduction of χ^2 statistics. t_l is again χ^2 distributed and the number of degrees of freedom is simply the sum of the degrees of freedom of $t_{A,l}$ and of $t_{B,l}$.

The power of this statistical test, developed using the notion of stationarity used in both matematical statistics and the theory of dynamical systems, is that it can be applied to a time series consisting of a few thousand elements. The analysis of several examples performed in [45] demonstrates clearly the necessity of both tests for testing stationarity. Of course the method cannot be used in a black-box fashion, since particularly the window length must be in accordance with the correlation length.

2.2 Reconstructed phase space

The study of dynamical systems comprises the introduction of a phase space, where the time evolution of the system can be defined. Since the determinism implies that once fixed the present state, the states at all future times can be known, specifying a point in the phase space is equivalent to specifying the state of the system. The concept of state of a system is powerful even for non-deterministic systems, like stochastic Markov processes for which the transition rules are given in the form of a set of transition probabilities and the future state is randomly selected according to these probabilities.

In time series analysis, most often only scalar measurements are performed, i.e. not the full phase space vectors $\mathbf{x}$ are measured, but an observable $s = g(\mathbf{x})$, which is a (non-linear) projection of the phase space to the reals. Embedding techniques are employed in order to reconstruct vector valued time series from scalar time series. This sequence of vectors can be interpreted as a sample of a trajectory of a dynamical system in a reconstructed phase space, which is related to the unknown space of the underlying dynamical system by some smooth coordinate transform, if the measurement function was smooth. Hence, its invariants such as attractor dimension, Lyapunov exponents and entropies are the same. A scalar time series $\{s_n = g(\mathbf{x}_n)\}$ is a sequence of measurements equidistant in time. The constancy of the time intervals in between successive measurements is the precondition of Takens theorem and its generalizations [79, 97, 106]: The Takens time delay embedding method allows us to reconstruct a sequence of vectors, from the sequence of scalar observables, which are topologically equivalent to the unobserved state vectors.

The theorem can be formulated in the following way. Given is a dynamical system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ in a phase space $\Gamma \subset \mathbf{R}^d$, a measurement function $h : \mathbf{R}^d \to \mathbf{R}$, and a

sampling interval Δt. Let the trajectory $\mathbf{x}(t)$ be confined to an $\mathbf{f}$-invariant set $A \subset \Gamma$, with the box-counting dimension D_f. Denote the scalar measurements obtained through the sampling by $s_n := h(\mathbf{x}(t = n\Delta t))$. Consider the delay embedding space spanned by delay vectors:

$$\mathbf{s}_n = (s_n, s_{n-\tau}, s_{n-2\tau}, ..., s_{n-(m-1)\tau}). \qquad (2.9)$$

If:

$$m > 2D_f, \qquad (2.10)$$

then there exists a unique smooth map from A into the delay embedding space, which is invertible and has non-zero derivative on the image of A in $\mathbf{R}^m$. A is then said to be immersed in $\mathbf{R}^m$. This holds for generic h, generic $\mathbf{f}$, almost all Δt, and every $\tau \in \mathbf{N}$.

Takens' original version of the theorem did not require the trajectory to live in an invariant set, but supported only the embedding of the full d-dimensional phase space in an $\mathbf{R}^m$ with $m \geq (2d + 1)$. Sauer and collaborators [97] could not only reduce the embedding dimension to $m > 2D_f$ but showed also the general validity of the theorem in the sense of prevalence concerning the measurement function h.

Hence, the m-dimensional delay embedding space is equivalent to the original unobserved phase space of the dynamical system, since in particular the dynamics of s is deterministic. The trajectory at successive time steps can be written as a sequence of non-linear transformations of the state vector $\mathbf{x}(t)$ at an arbitrary time $t = n\Delta t : \mathbf{x}(t = (n + k)\Delta t) = \mathbf{F}^k(\mathbf{x}(n\Delta t))$. The m successive measurements $s_{n+1}, ..., s_{n+m}$ define m different equations for the d unknowns in $\mathbf{x}(n\Delta t)$, which one can formally solve, if $m > 2d$. The delay embedding theorem does not require knowledge of neither the measurement function h nor the integrated dynamics $\mathbf{F}$, but the constancy of the sampling interval guarantees that the set of equations for every n of the time series has exactly the same structure, and a one-to-one relationship between $\{\mathbf{x}_n\}$ and the series of embedding vectors exists.

The requirement of $m > 2D_f$ is related to geometry and guarantees that self intersections of the reconstructed sets due to non-linearities are non-generic. When dealing with time series, usually one does not know the underlying system and therefore the dimension of the set in phase space D_f and thus the correct embedding dimension m are unknown. They have to be determined self-consistently from the
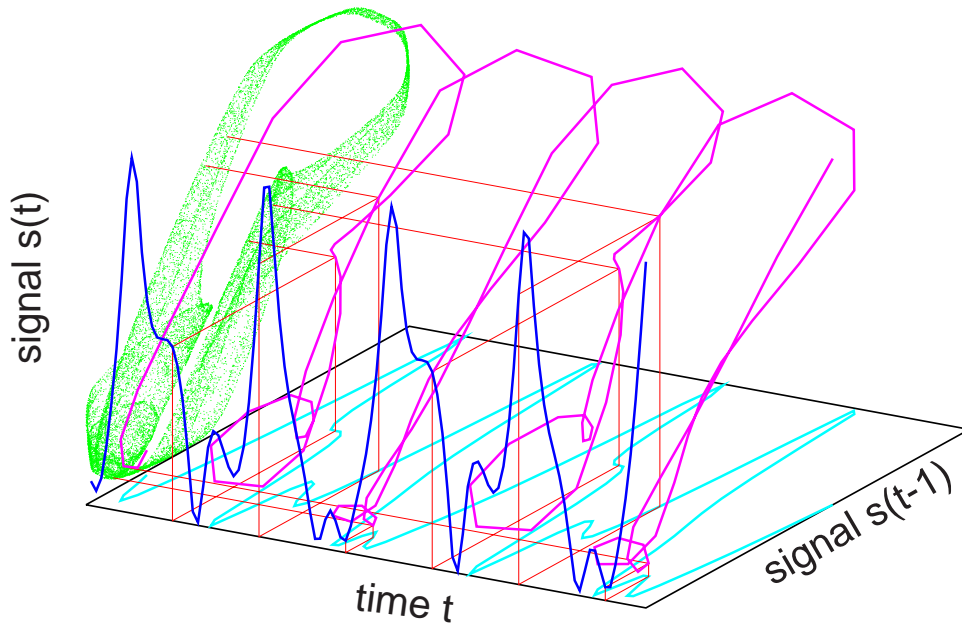
Figure 2.1: Sketch of the time delay embedding method for data from a physical laboratory experiment, courtesy of Ralf-Peter Kapsch. The time series s_n is plotted in the vertical and the horizontal plane, with a relative time lag of unity. The sequence of 2-dimensional vectors (s_n, s_{n+1}) is shown as a trajectory in the 3-d space. The projection of these (and many more data from the same data set) onto the "$x - z$"-plane reveals non-trivial structure, showing that the data are subject to complicated but evident dynamical constraints.

data. The time lag τ does not play a role in the mathematical theorem, which relies on data with mathematical (infinite) precision, but its proper choice is essential for noisy data with strong correlations.

Fig. 2.1 is a sketch of the embedding method for data from a physical laboratory experiment (a non-linear electric resonance circuit) which represent an attractor of dimension $D_f = 2.1$. Evidently, for shorter time lags τ, the resulting structure in embedding space would be more and more squeezed towards the bisectrix. Measurement noise will then extinguish all structure perpendicular to the bisectrix. This structure can be revealed even in the presence of noise when the extension in this direction is enlarged by a larger time lag. Too large a time lag will combine almost uncorrelated measurements such that visible (and algorithmically useful) structure is much more complicated and on much smaller scales, which again are easily hid-

den by noise. Unfortunately, there is no theoretical argument of how to determine the optimal time lag, and only guidelines exist (see [1, 46, 52]). For a sinusoidal oscillation, the optimal time lag for a two-dimensional embedding is one quarter of the period, but already for higher embedding dimensions even this is unclear, such that τ has to be optimized empirically for every single application.

For non-autonomous systems, the embedding theorem is not valid, except for the case that the fluctuating driving terms are known [103]. If, however, the system is periodically driven, the periodic driving can be interpreted as the limit cycle solution of a deterministic system itself. Hence, the non-autonomous system can be modelled autonomously by two sub-systems with a unidirectional coupling between them. For instance, a sinusoidal driving force, $p = a\cos\omega t$, can be created by $(\dot{y}_1, \dot{y}_2) = (y_2, -\omega^2 y_1)$. The reconstruction of these two additional variables from the time series requires at most 5 additional dimensions of the delay embedding space, but since they are confined to the one-dimensional limit cycle solution, 2 additional dimensions are generally sufficient. For fixed driving force p, the system creates an attractor with dimension D_p which might be smaller than D, the dimensionality of the phase space. When p oscillates periodically, in the worst case one has to reconstruct $D+1$ variables by the time delay embedding, which means $m > 2D+2$. In fact, for experimental data from a periodically driven electric resonance circuit, both autonomous 4-dimensional maps and 2-dimensional periodically driven maps were successfully constructed from the data [35].

2.3   Over-embedding

When the driving term varies in a non-deterministic way, there is no way to rewrite the equations of motion in an autonomous form. However, in many applications the knowledge of the instantaneous equations of motion, i.e. the equations with the actual parameter settings, is sufficient. In such a situation, under the assumption that parameters vary on much longer time scales than those which rule the instantaneous dynamics, and have only rare additional sudden changes, over-embedding solves the problem. As an illustration, let us consider the Lozy map:

$$x_{n+1} = 1 - a|x_n| + bx_{n-1} \ . \tag{2.11}$$

If a time series of x_n itself is recorded, a two-dimensional embedding allows to uniquely reconstruct the equation of motion, Eq.(2.11), from the data. If one parameter, say a, changes slowly in time, a two-dimensional embedding will be insufficient, since among the neighbours of a delay vector, $\mathbf{s}_n = (x_n, x_{n-1})$, there are in general neighbours with different settings of a contained in the time series (at least in regions of the $\mathbf{R}^2$ where the attractors for different a overlap), and uniqueness of the image is lost. However, the knowledge of the triple (x_n, x_{n-1}, x_{n-2}) allows in principle to first solve for a_{n-1} by:

$$a_{n-1} = \frac{x_n - 1 - bx_{n-2}}{|x_{n-1}|} \ , \tag{2.12}$$

and, under the assumption that $a_n \approx a_{n-1}$, to predict x_{n+1} as:

$$x_{n+1} = 1 - \frac{x_n - 1 - bx_{n-2}}{|x_{n-1}|} x_n - bx_{n-1} \ , \tag{2.13}$$

such that close neighbours of (x_n, x_{n-1}, x_{n-2}) can be used for the (local linear) reconstruction of this equation from data. Eq.(2.12) proves that similar delay vectors with $m=3$ are necessarily related to similar values of a. One can thus use the three-dimensional embedding for a mere selection of neighbours and then use these in an only 2-dimensional embedding, if desirable. Eq.(2.13), however, is a universal model of the dynamics of Eq.(2.11), valid for all quadruples generated with fixed a, irrespective of the actual value of a.

The above reasoning can be easily generalized for more than one fluctuating parameter. The accuracy to which we want an instantaneous deterministic equation of motion to hold delimits the tolerable speed of parameter variations. When we reconstruct equations of motion from observed data, a residual error remains even for stationary data, so that it appears to be sufficient for all practical purposes that the parameter variation per time step is below the percent level. When a sudden change of parameters occurs, there is, for a few time steps, no instantaneous deterministic rule in the time delay embedding space, and in particular we will generally find no neighbours for a delay vector covering the transition between two such different episodes. However, before and after this parameter jump, the above reasoning again applies, and we will show in the application to voice data (Sec. 3.3) that in fact sudden parameter changes do not destroy the usefulness of this concept.

We can generalize Eq.(2.10) using the following proposition: If a D-dimensional deterministic dynamical system depends on P parameters with slow time depen-

dence, then delay vectors of sufficient embedding dimension are approximately confined to a $(D_f + P)$-dimensional manifold, which can be reconstructed (see [37]) in a space of dimension:

$$m > 2(D_f + P). \tag{2.14}$$

Approximately means that the confinement is violated on length scales of the order of the standard deviation of the data times the average parameter change per time step. It can be shown that the way in which parameters vary is irrelevant for the embedding property: It holds irrespectively of whether the change is periodic, with a drift or stochastic. The reason lies in the slowness of their change: When the time lag of the delay embedding is adjusted according to the time scales of the instantaneous dynamics, it is much too small for a reconstruction of the dynamics of the varying parameters. Their attractor is not at all unfolded and they contribute in high precision only with a P-dimensional subspace.

2.4   Recurrences

When thinking of geometry in phase space, non-stationarity introduces a tendency according to which if points are close in space they are also close in time (in other words, points far in time are typically not neighbours if a high-dimensional embedding space is considered). A step into the direction of not just detecting but also characterizing the kind of non-stationarity was successfully done by recurrence analysis. In the spirit of determinism and phase space reconstruction, in [73] the idea of time dependent driving forces and their recurrence is strongly used: Time series are cut into segments and in the plane of segment indices a point is drawn if the distance between time series segments is smaller than some threshold. These *meta-recurrence plots* allow one to identify time series segments with similar or identical evolution equations and parameters. In [14] it has been shown that by help of recurrence analysis, one can identify the time dependence of driving forces under certain conditions.

   After having pointed out the power of recurrence analysis, let us introduce the basic graphical tool that evaluates temporal and phase space distance of states, namely the classical recurrence plot (RP), firstly introduced in [18]. In the time-
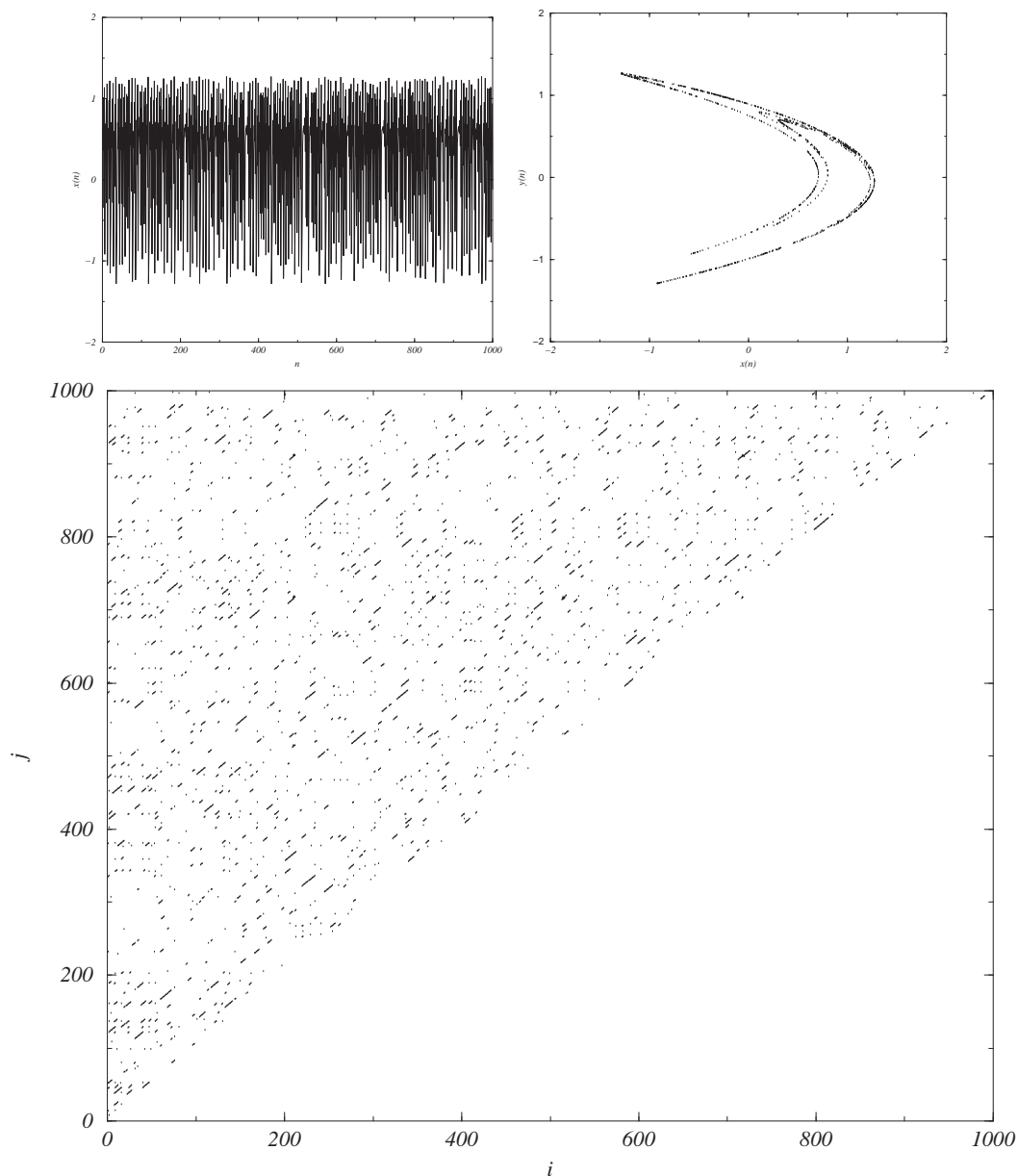
Figure 2.2: Upper left panel: The Hénon time series x_n, 1000 samples. Upper right panel: The two-dimensional embedding of the Hénon system. Lower panel: Section of a recurrence plot for the Hénon map: In the plane of indices i, j a dot is printed, whenever the two-dimensional delay vectors fulfill $|\mathbf{s}_i - \mathbf{s}_j| < \epsilon$. The presence of short but quantitative significant lines is a clear signature of determinism.

time plane of a RP, one represents the couple of time indices (i, j) by a dot, if:

$$|\mathbf{s}_i - \mathbf{s}_j| < \epsilon. \tag{2.15}$$

A dot with coordinates (i, j) implies that the distance between the delay vector centered in i and the delay vector centered in j is smaller than a given threshold ϵ. A possible different way to define a RP is setting the number k of neighbours instead of the threshold ϵ. According to this alternative definition, a pair (i, j) is called a recurrence if $\mathbf{s}_i$ is one of the k-th nearest neighbours of $\mathbf{s}_j$, for some predefined value of k. RPs obtained following a different definition are not identical, since not all points have k neighbours contained in an ϵ-neighbourhood. On average, the relation between ϵ and k is the following: For extremely small values of the threshold one does not get any neighbour, while the number of the latters saturates if the value of ϵ is comparable to the size of the reconstructed phase space. The role played by k and by ϵ is qualitatively the same on average. Since $\mathbf{s}_i$ are delay embedding vectors, the resulting RP depends strongly on the embedding parameters. A RP is then generated by computing all recurrences at a given neighbourhood order k or resolution ϵ. In the remaining we will always consider the definition through the threshold ϵ.

In [109] different parameters for the statistical quantification of RPs have been proposed, suggesting that they can be a useful starting point for the analysis of non-stationary sequences if the relevant information is extracted in a suitable way. The most detailed account of these techniques has been provided in [14, 18], where it is shown by a scaling argument that for a dynamical system with time varying parameters, the RP in the limit of small ϵ, large N and sufficient m approaches the RP of the fluctuating parameter. However the task of extracting the time variation of a parameter from its RP is in general very difficult.

As an example, let us consider the map given by Hénon in 1976:

$$\begin{cases} x_{n+1} & = & a - x_n{}^2 + b y_n \\ y_{n+1} & = & x_n \end{cases} \tag{2.16}$$

It yields irregular solutions for many choices of the two parameters a and b. For $|b| \leq 1$ there exist initial conditions for which trajectories stay in a bounded region but, for example, when $a = 1.4$ and $b = 0.3$, a typical sequence of x_n will not be periodic but *chaotic*. In Fig. 2.2 we can see a recurrence plot for the Hénon
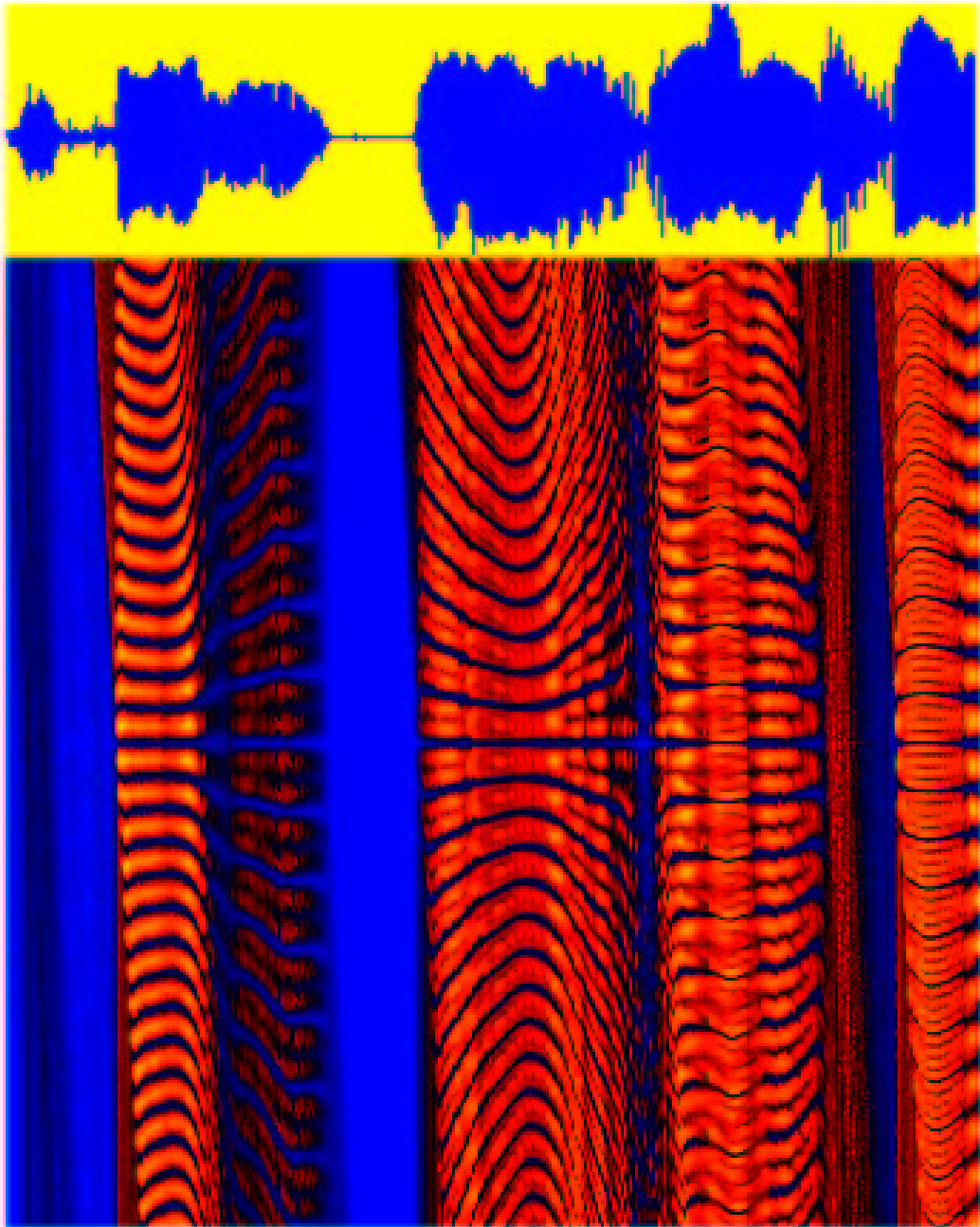
Figure 2.3: Section of a recurrence plot: In the plane of indices i, $(j - i)$ a dot is printed, whenever the delay vectors fulfill $|\mathbf{s}_i - \mathbf{s}_j| < \epsilon$. Different colors code different distances. Upper panel: The speech signal underlying the RP.

map. Due to the chaoticity of the system, trajectories originated by similar initial conditions tend to diverge but since the phase space is limited and the rules are fully deterministic, if two trajectories are similar (i.e. vector points are neighbours) they tend to remain similar, at least for some points. Therefore the presence of small lines in recurrence plots is a clear signature of determinism. One can further investigate the reason for the given extension of the lines, having in mind the two extreme cases: A periodic signal produces a RP where the lines are the longest possible, uncorrelated noise has RPs with no line at all. We have already seen the exponentially fast divergence of trajectories typical of chas. The properly averaged exponent of this increase is characteristic for the system underlying the data and quantifies the strength of chaos. It is called the *Lyapunov exponent*. Let β_{n_1} and β_{n_2} be two points in state space with distance δ_0. Denote by $\delta_{\Delta n}$ the distance at time Δn between the two trajectories emerging from these points. Then the maximal Lyapunov exponent λ is determined by:

$$\delta_{\Delta n} \approx \delta_0 e^{\lambda \Delta n}. \tag{2.17}$$

If λ is positive, this means an exponential divergence of nearby trajectories, i.e. chaos. It can be shown that the bigger λ is, the shorter are the lines in RPs [67, 113].

We want now to introduce a non-stationarity in the Hénon map and to show its effect in a RP. To this purpose, we build a time series composed by 3 segments, each one characterized by a different set of parameters a and b. The RP of such a time series is reported in Fig. 2.4. For every single segment it is qualitatively the same as in Fig. 2.2, in particular with respect to the line structures. If considered as a whole, anyway, a new effect appears, namely the clusterization of the RP. There is almost no point involving the first and the third segment and just a few cross-recurrences between the first and the second. The second and the third segment present a little bit more recurrences, but there is no line structure and therefore they can be considered almost random (they do not reflect any determinism).

In Fig. 2.3 the RP of a human voice signal is plotted. In this case we have used a slightly different version of RP: Along the x-axis we still have one time index i, but along the y-axis we report the difference between the two indexes $(j - i)$. Geometrically this is equivalent to a 45 degree rotation, so that now the expected lines should be horizontal. Furthermore, in Fig. 2.3 the ϵ is used as color coded parameter: Instead of printing a dot whenever the delay vectors fulfill $|\mathbf{s}_i - \mathbf{s}_j| < \epsilon$,

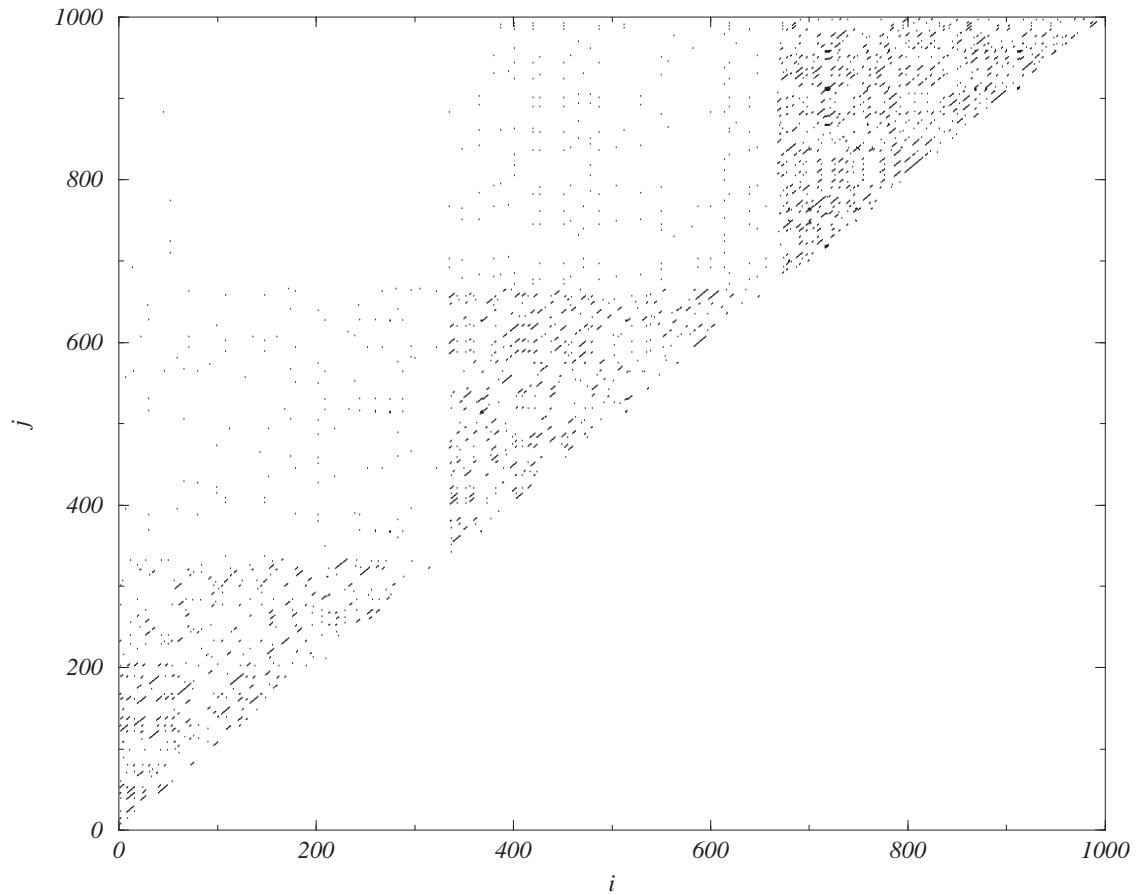Figure 2.4: Section of a recurrence plot for the non-stationary Hénon map: A dot in position (i, j) means that $|\mathbf{s}_i - \mathbf{s}_j| < \epsilon$. Along the 1000 points of the time series, 3 different sets of parameters are used, giving rise to a non-stationary Hénon map. The effect is a lack of cross-recurrences between two different segments, detectable as a clusterization of the recurrence plot.
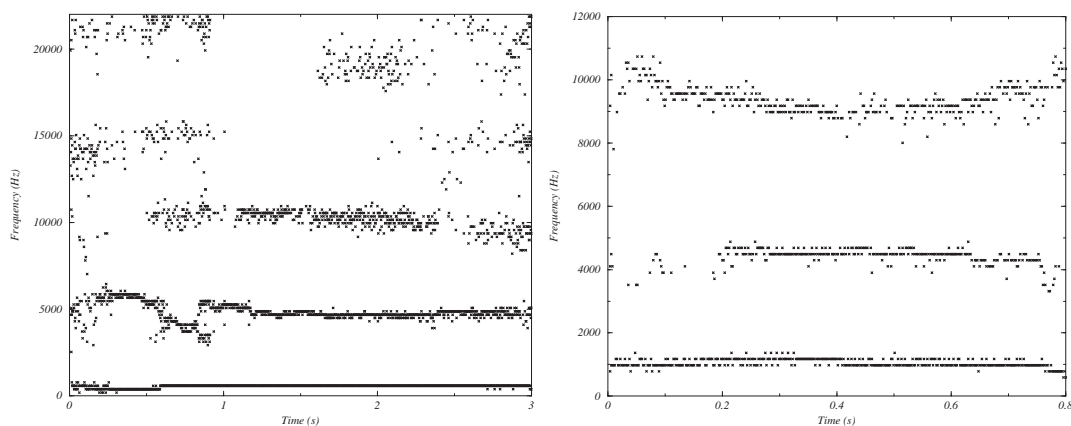
Figure 2.5: Left panel: Spectrogram of a long word. Along the 3 seconds phonation, only the fundamental frequency, related to the basic mechanism of speech, is always present. This is not the case for higher frequencies, showing that even within a single word different types of dynamics are present or, alternatively, the parameters of the dynamics are non-stationary. Right panel: Spectrogram of an elongated phoneme. The signal is almost stationary, apart the effects involving the higher frequency (*U shape*), due to the concatenation with the previous and the following phonemes.

we have computed the distance $|\mathbf{s}_i - \mathbf{s}_j|$ and plotted a dot with a color that depends on it.

The non-stationarity in a time series can also be revealed with the help of *spectrograms*, graphs where the time series index is reported along the x-axis and the power spectrum along the y-axis. A time series is divided into overlapping segments (typical values are windows of 2048 points with an overlap of 1024 points) and for each segment a symbol is plotted at the frequencies where the power spectrum assumes a relative maximum. In Fig. 2.5 we can see the spectrogram of a word and of an elongated phoneme. With the help of this tool one can identify the time scales that are relevant in terms of the non-stationarity: A word, concatenation of phonemes, is non-stationary, as the different frequency spectrum indicates; on the other hand a single phoneme can be considered stationary, at least in its central part, where the effects of the concatenation are not so evident.

2.5 ARCH and GARCH models

In the last section we have discussed the analysis of deterministic systems with slowly time dependent parameters. Here we introduce the opposite setting, namely stochastic processes where the non-stationarity comes from a second time scale. Over the last twenty years ARCH (Autoregressive Conditional Heteroskedasticity) models have been widely applied in the financial literature and specifically in the option pricing literature. They have been first proposed by Engle [20]; some years later Bollerslev [8] has introduced a generalized version called GARCH (Generalized ARCH). These models assume conditionally normally distributed price returns (see Sec. 4.2) with a time-varying conditional variance, and appear to reconcile the typical properties of financial time series better than more classic time series models, such as ARMA models.

We suppose that our dependent variable y_t is generated by:

$$y_t = x_t^{'}\gamma + \epsilon_t \qquad\qquad t = 1, 2..., T, \qquad\qquad (2.18)$$

where x_t is a k-vector of lagged endogenous variables (state variables) and exogenous variables (external driving inputs), and γ is a k-vector of parameters. The ARCH model characterizes the distribution of the stochastic error term ϵ_t conditionally on a set of lagged variables $\Psi_{t-1} = \{y_{t-1}, x_{t-1}, y_{t-2}, x_{t-2}, ...\}$. In his original model, Engle [20] assumes a conditional normal distribution of the error term,

$$\epsilon_t | \Psi_{t-1} \approx \mathbf{N}(0, h_t)^1, \qquad\qquad (2.19)$$

where

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + ... + \alpha_q \epsilon_{t-q}^2, \qquad\qquad (2.20)$$

with $\alpha_0 > 0$ and $\alpha_j \geq 0$ for $j = 1, ..., q$, in order to ensure a positive conditional variance. Here, q is the order of the ARCH process.

The appeal of this model is that the conditional variance h_t depends on the past Ψ_{t-1} and is a positive function of the size of past errors in absolute terms. Thus a large positive or negative error tends to be followed by a large (in absolute terms) error, and similarly a small error tends to be followed by a small error. The order q determines the length of the period during which a disturbance persists in

---

[1]$\mathbf{N}(0, h_t)$ is the Gaussian distribution with zero mean and variance h_t^2.

conditioning the variance of the following disturbances. The larger is q, the longer
the periods of volatility clustering (see Sec. 4.4). Another important property of
ARCH processes is that the mixing produced by the changing conditional variance
induces additional kurtosis in the unconditional distribution. In fact, the parame-
terization does not impose a priori the existence of unconditional moments, which
allows the model to be consistent with Mandelbrot [64], who found evidence that
the distribution of financial asset returns may well have infinite variance.

It rapidly became apparent in applied work that the specification of the con-
ditional variance as an ARCH(q) called for a large number of lags and therefore
the estimation of numerous parameters subject to inequality constraints. In conse-
quence, Bollerslev [8] proposed a generalization of the ARCH model, called GARCH,
that allows for a parsimonious representation of a high-order ARCH model. The
conditional variance function of a GARCH(p,q) model has the following form:

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + ... + \alpha_q \epsilon_{t-q}^2 + \beta_1 h_{t-1} + ... + \beta_p h_{t-p}, \qquad (2.21)$$

with $\alpha_0 > 0$, $\alpha_j \geq 0$ for $j = 1, ..., q$ and $\beta_k \geq 0$ for $k = 1, ..., p$. The last constraints
about β_k ensure a positive conditional variance. Engle and Bollerslev [9] introduced
the IGARCH process (Integrated GARCH), a GARCH process in which

$$\sum_{j=1}^{q} \alpha_j + \sum_{k=1}^{p} \beta_k = 1. \qquad (2.22)$$

In this case, a contemporaneous shock persists indefinitely in future conditional
variances. For agents in the options market, the degree of persistence of the shocks
on the variance is an essential element. In effect, they will be prepared to pay a
higher price for long-lived options if they perceive that the shocks are large and
sufficiently permanent relative to the life of the options.

The finding of a very high degree of persistence for financial data is not univer-
sal. Although Engle and Mustafa [22] obtained a very high degree of persistence for
several individual stocks of large firms on the US stock market and for the *S&P500*
index, the results of Engle and Gonzalez-Riviera [21] suggest that the degree of per-
sistence depends on the size of the firm, smaller firms exhibiting a lower degree of
persistence than larger ones. Furthermore, according to Lamoureux and Lastrapes
[55], the high degree of persistence that is observed could be due to a misspecification
of the conditional variance. They suggest that structural changes in the uncondi-
tional variance of the process produce volatility clusters that result in a high degree

of persistence. Allowing the constant in the conditional variance function to vary on different sub-periods of their sample, they obtain a lower degree of persistence as compared to that of a model without structural changes.

2.6 Non-recurrent processes

The archetypical feature in the theory of stochastic processes is Brownian motion. Already in 1827 the botanist R. Brown had reported the observation of a very irregular motion displayed by a pollen particle immersed in a fluid. It was the *kinetic theory of gases*, dating back to David Bernoulli (1700-1792) which would provide the basis for Einstein's (1879-1955) and Smoluchowski's (1872-1917) successfull treatment of the Brownian motion problem in 1905. Through the work of Maxwell (1831-1879) and Boltzmann (1844-1906), *statistical mechanics*, as it grew out of the kinetic theory of gases, was the main area of application of probabilistic concepts in theoretical physics in the 19th century.

In the Brownian motion problem and all its variants, one deals with a phenomenon that is the outcome of many unpredictable and sometimes unobservable events, which individually contribute a negligible amount to the observed phenomenon, but *collectively* lead to an observable effect. The individual effects cannot sensibly be treated in details, but their statistical properties may be known, and they determine the observed macroscopic behaviour. A problem closely related to the Brownina motion is that of random walker, introduced into science by Karl Pearson in a letter to Nature in 1905:

> *A man starts from a point 0 and walks l yards in a straight line: He then turns through any angle whatever and walks another l yards in a straight line. He repeats this process n times. I require the probability that after these n stretches he is at a distance between r and $(r + \delta r)$ from his starting point 0.*

The solution to this problem was provided in the same volume of Nature by Lord Rayleigh (1842-1919), who told him that he had solved the problem 25 years earlier when studying the superposition of sound waves of equal frequency and amplitude but with random phases. Let us assume that a walker can sit at regularly spaced positions along a line that are a distance Δx apart. In this case one can label the

positions by the set of integer numbers. Furthermore, we require the walker to be at position 0 at time 0. After fixed time intervals Δt the walker either jumps to the right with probability p or to the left with probability $(1-p)$. What is the probability to find the walker at position n after N steps, with the obvious constraint $n < N$? The computation is quite easy, since the probabilities are independent. One has just to evaluate the probability to jump $(N+n)/2$ time to the right and $(N-n)/2$ times to the left and to count how many such paths there are. One ends up with the result that the probability of being at position n after N jumps is given by the following binomial distribution:

$$p(n, N) = \frac{N!}{(\frac{N+n}{2})!(\frac{N-n}{2})!} p^{\frac{1}{2}(N+n)}(1 - p)^{\frac{1}{2}(N-n)}. \qquad (2.23)$$

Knowning this probability distribution, one can evaluate all the moments of n at fixed time N. Let us denote the number of jumps to the right as $r = (N + n)/2$. Simple computation lead to:

$$E[r] = Np, \qquad Var[r] = Npq, \qquad (2.24)$$

with $q = (1 - p)$. Translating the results for the number of steps to the right r into the position of the random walker, one gets:

$$E[n] = 2N \left(p - \frac{1}{2}\right), \qquad Var[n] = 4Npq. \qquad (2.25)$$

If the probability of jumping to the right equals the probability of jumping to the left, then Eq.(2.25) reduces to *free diffusion*:

$$E[n] = 0, \qquad Var[n] = N. \qquad (2.26)$$

This is again a kind of non-stationarity, although quite different from the two previously presented situations. In this case, in general, there is no variation in the control parameters of the system and no time dependence in the transition probabilities. But the fact that the more data we get, the larger $Var[n]$ is (Eq.(2.26)), has as consequence the lack of recurrences. As already said, many important phenomena can be descrived as the compound effect of many small influences. The observable quantities are most often the sum of a very large number of random events, like in the case of the pressure exerted by a gas on a piston, which is a thermodynamical (read macroscopic) variable made up from all the collisions of the gas particles (microscopic scale) with the piston.

Mandelbrot and Van Ness [65] have introduced *fractional* Brownian motion as a generalization to processes which grow at different rates t^H, where $0 \leq H \leq 1$ is called Hurst exponent (see Sect. 2.7):

$$Var[n] = N^{2H}. \tag{2.27}$$

Successive increments of a fractional Brownian motion are called fractional Gaussian noise and an important feature of them is that correlations decay slower than in the standard case, so that the resulting fractional Brownian motion exhibits long memory effects. Correlations are positive for $H > 0.5$ (*persistence*) and negative for $H < 0.5$ (*antipersistence*) as shown in Fig. 2.6.

As for standard Brownian motion, all fractional Brownian motion series are self-affine, meaning that the series appears statistically identical under rescaling the time axis by some factor a and the displacement by a^H. Hence, a fractional Brownian motion lacks any characteristic time scale and when generating or sampling a series, an arbitrary step length of one unit may be used without loss of generality [112]. Self-affine signals can be described by a *fractal* dimension D which is related to the Hurst exponent through:

$$D = 2 - H \tag{2.28}$$

for fractional Brownian motion [94]. The fractal dimension D can be loosely interpreted as the *number of dimensions* the signal fills up. For example, notice that in Fig. 2.6 the $H = 0.1$ signal *fills in* significantly more space than $H = 0.9$ and, consequently, has a higher fractal dimension. For a detailed treatment of the interconnections between fractal dimension, box-counting dimension and correlation dimension see [46].

The power spectrum (defined as the amplitude-squared contributions from the frequencies $\pm f$, $S(f) \equiv |F_H(f)|^2 + |F_H(-f)|^2$ where F_H is the Fourier transform of the displacement [93]) of fractional Brownian motion also demonstrates scaling behaviour. The exact spectrum is difficult to compute but for low frequencies it can be approximated by a power law $S(f) \sim 1/f^{2H+1}$ (see Fig. 2.7) which corresponds to long-term spatial correlations[2]. Fractional Brownian motion has been criticized because it lacks a physical interpretation and because the process has an unrealistic

---

[2] *Flicker* or $1/f^\alpha$ noise with $\alpha \approx 1$ is ubiquitous in nature and some of it may be attributable to long-memory fractional Brownian motion processes.
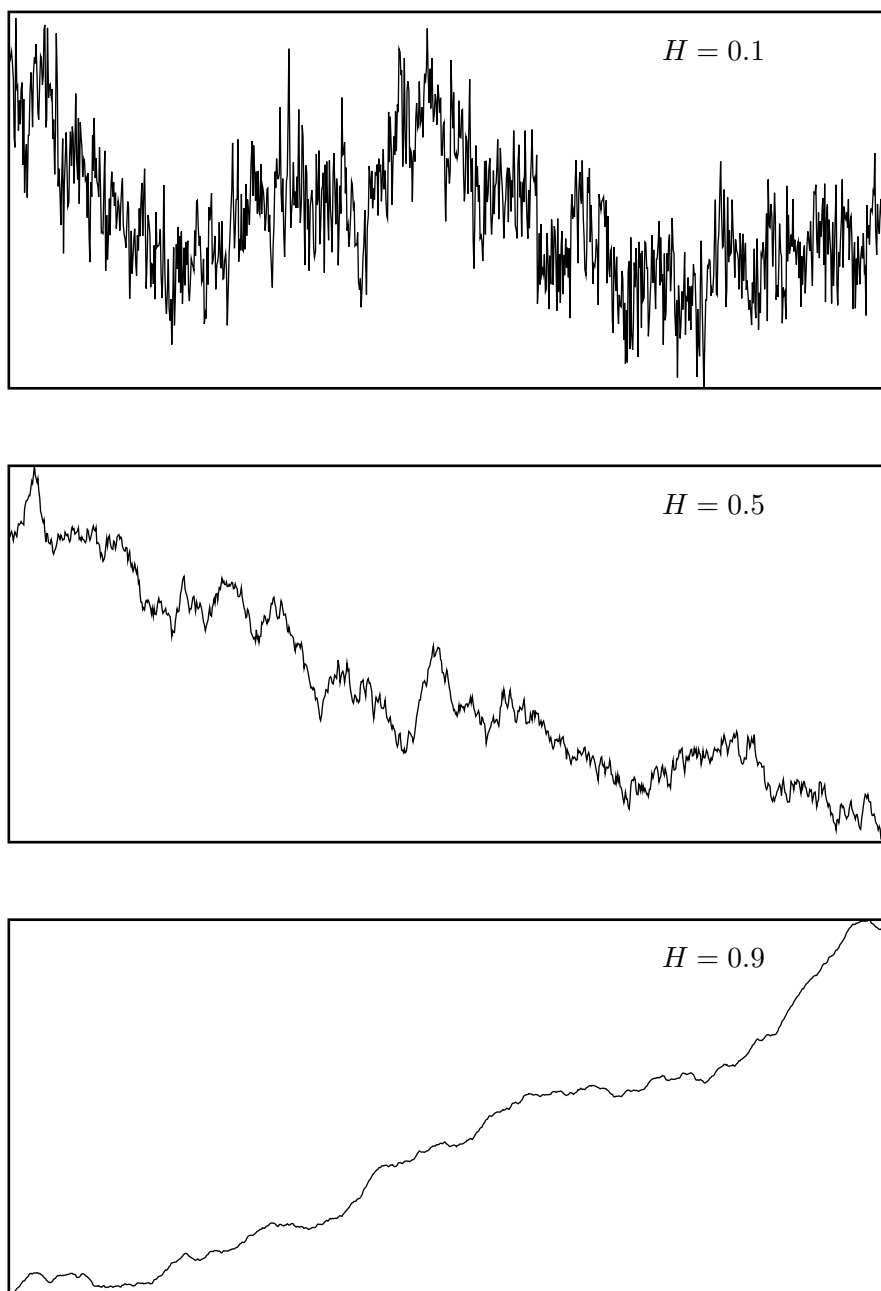
Figure 2.6: Sample fractional Brownian motion time series with different Hurst exponents: Antipersistent $H = 0.1$ (top) has negative long-range correlations, uncorrelated $H = 0.5$ (center) is standard Brownian motion, and persistent $H = 0.9$ (bottom) has positive long-range correlations.
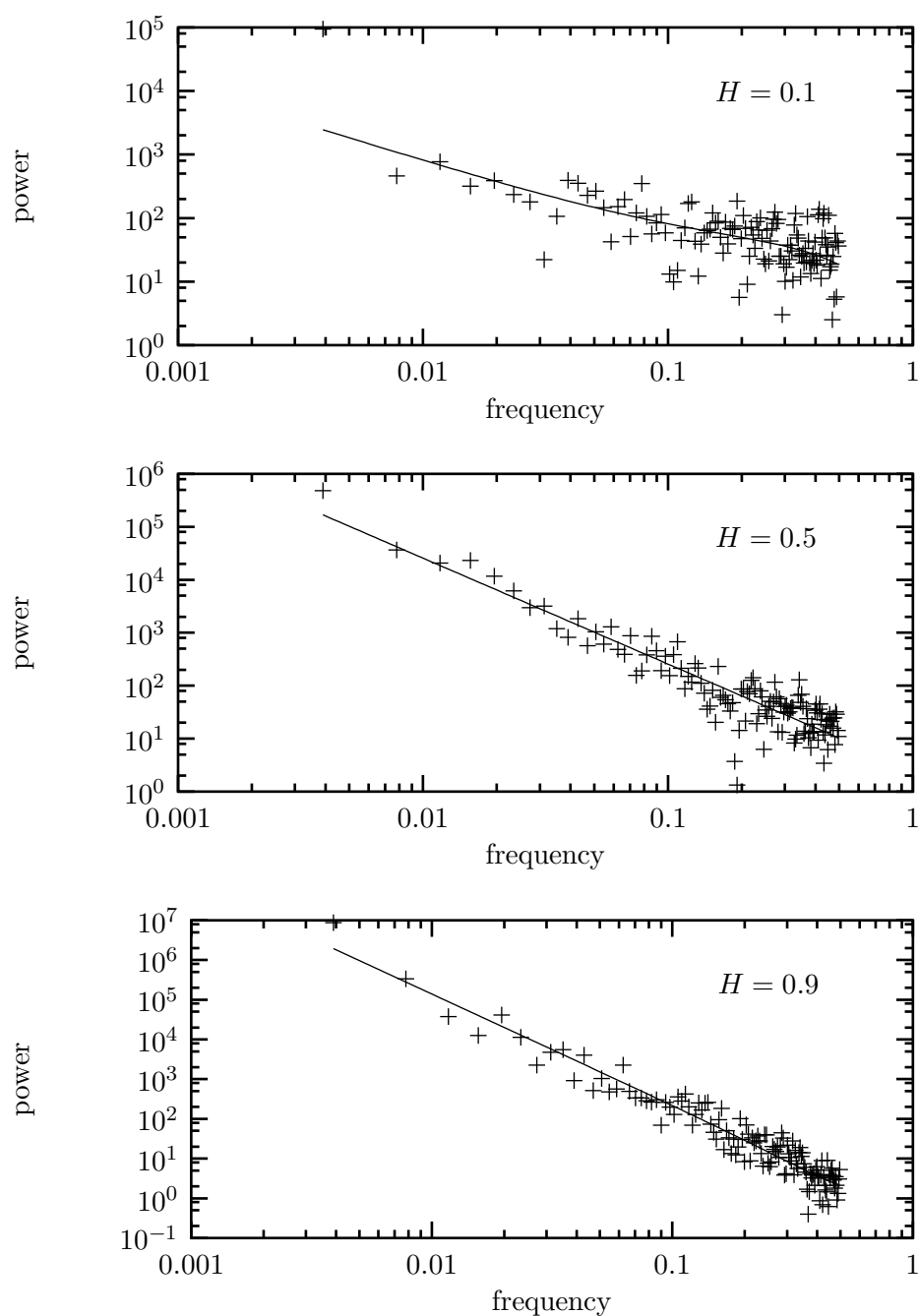
Figure 2.7: Power spectral densities for the fractional Brownian motion time series shown in Fig. 2.6. The points are from finite samples of 1000 points each and the line represents the theoretical spectrum. For low frequencies the power spectrum is well approximated by a power law $1/f^{2H+1}$.

infinite memory [65]. However, it is a mathematically elegant extension of standard
Brownian motion which introduces long-range memory effects and can be character-
ized by a single parameter H. Hence, it is an ideal experimental control for testing
procedures of measuring the Hurst coefficient in real data sets (see Sec. 2.7).

A very important question is what the distribution of a sum of random variables
will ultimately be. In other words, given N statistically independent and identically
distributed random variable x_i with zero mean and variance σ^2, one is interested on
knowning what is the probability distribution of:

$$S_N = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} x_n. \tag{2.29}$$

Calculating the characteristic function of the sum variable and using the moment-
generating propertiy of it, as presented in [10], one is able to show that the proba-
bility distribution is Gaussian (Central Limit Theorem):

$$p_N(z) \xrightarrow{N \to \infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{z^2}{2\sigma^2}\right]. \tag{2.30}$$

This result gives a justification of the ubiquitous appearance of the Gaussian distri-
bution in statistical phenomena. Indipendently from the exact form of the underly-
ing distribution of the individual random variables, the sum of them always obeys
a Gaussian distribution in the large N limit, provided that the first two moments
exist. For more details, included the deviations from the asymptotic behaviour when
the sum is truncated to a finite number of random variables, see [30].

The restrictions on the first and second moment involved in the central limit
theorem are so mild that almost all distributions belong to the domain of attraction
of the Gaussian. There are, however, some exceptions, the most famous being the
Cauchy distribution[3]:

$$p(x) = \frac{a}{\pi} \frac{1}{a^2 + x^2}, \tag{2.31}$$

whose second moment is infinite. A natural question is therefore to check what
happens when considering the sum of several random variables distributed accord-
ing to Eq.(2.31). The Cauchy distribution belongs to the following class of stable
distributions (*Lévy distributions*):

$$p(x) \sim \frac{1}{|x|^{1+\alpha}}, \qquad 0 < \alpha < 2, \qquad (|x| \to \infty), \tag{2.32}$$

---

[3]The Cauchy distribution occurs in many physical situations, like in the theory of critical opales-
cence and in the lifetime broadening of spectral lines.

which have long, inverse-power-law tails. A distribution is stable if it is invariant under convolution. The interesting thing about the broad tails is that they are responsible of the non-convergence to the Gaussian distribution, but at the same time they do not preclude the existence of a limiting distribution.

The distinguishing property of Lévy distributions is the presence of long-range power-law tails which may lead to a divergence of even the lowest-order moments. For instance, both the first and second moments are infinite if in Eq.(2.32) we have $\alpha < 1$. Physically, these lower-order moments have a very important meaning, since they set the pertinent scales. In Brownian motion, for instance, the time scale was determined by the first moment of the waiting-time distribution, while the second moment of the jump-length distribution defines the physically relevant length scale. The divergence of the corresponding moments for certain Lévy distributions implies the absence of underlying physical scales, provided that these distribution are realizable in nature. A very interesting paradox arised from this concept is the so-called *St. Petersburg Paradox*, sketched in Appendix: The absence of physical scales could be interpreted as scale invariance, which in turn invokes the notion of self-similarity and fractals. The intimate relation between Lévy distributions and self-similar behaviour is the link that ties the mathematical properties to physical applications.

From the viewpoint of practical applications, *stable distributions*, namely distributions which are invariant under convolution, have both appealing and non-appealing features. Certainly a very attractive property is the scaling behaviour, a typical feature of fractality, since the whole looks like its parts and no characteristic scale can be found. On the other hand, Lévy distributions possess divergent lower-order moments. The divergence of the variance, in particular, is a very disturbing feature if a time series analysis gives rise to a Lévy distribution, although the variance is a priori known to exist. A good example comes from an analysis of the human heart beat in [90], where the probability of the difference between the durations of successive beats was studied. This distribution could much better be fitted with a symmetric Lévy distribution (with $\alpha = 1.7$) than with a Gaussian, since large differences occur more frequently than what a normal distribution would predict. Physiologically, this is not unreasonable. The heart can respond more easily to changing external influences if successive beat durations may substantially deviate one from the other. However, the deviation cannot be arbitraly large, since
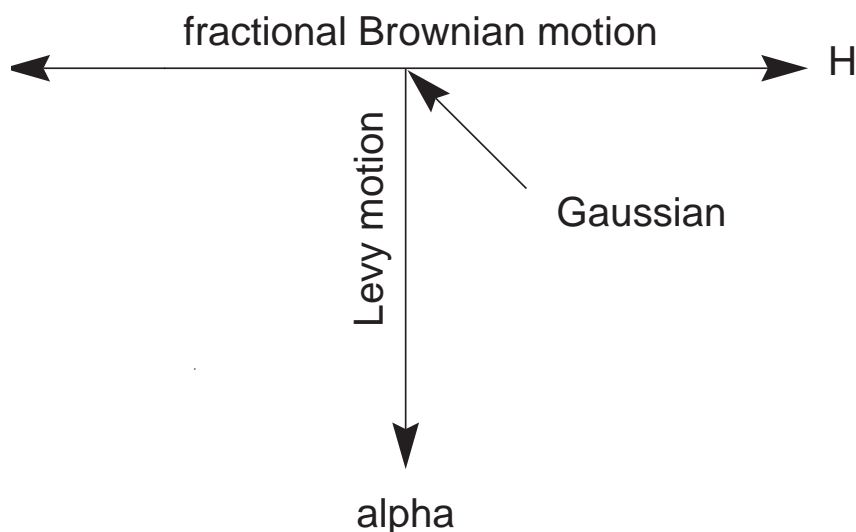
Figure 2.8: Schematic representation of relation between fractional Brownian motion and Lévy flight. Traditional Brownian motion sits at the intersection ($H = 1/2$, $\alpha = 2$). The natural extension into the two-space is fractional Lévy motion which has correlated, non-Gaussian increments.

the physiology of our body poses strong upper and lower limits to it, which in the end cut off the tails of the Lévy distribution and impose a finite variance of the time series.

Mantegna and Stanley have proposed a way [69] to reconcile Lévy distributions with a finite variance, truncating the tails in what they call a truncated Lévy flight. A Lévy flight is similar to a traditional Brownian motion in that it is a cumulated series of independent, identically-distributed increments, but in this case the increments are Lévy distributed instead of normally distributed. This yields a probability distribution $p(\Delta x) \propto \Delta x^{-1/H}$ with $0.5 \leq H \leq 1$. A Lévy flight creates enhanced diffusion. The idea of introducing the truncated Lévy flight was derived from the observation that financial time series show an exponential fall-off, implying that the second moment is finite. The central part of the price difference distribution shows a robust Lévy scaling over more than three orders of magnitude. In [69] it is shown that Lévy scaling may hold over a long period of time for the dynamics of quasi-stable stochastic processes having a finite variance. Fig. 2.8 shows schematically the relation between fractional Brownian motion (correlated with Gaussian increments)

and Lévy flight (uncorrelated with non-Gaussian increments). These two extensions of Brownian motion are not exclusive, but they can be combined to produce fractional Lévy motion with correlated, non-Gaussian increments. There is very little literature on the subject but it may be a useful model for some natural phenomena [49].

2.7 The Hurst exponent

The Hurst exponent gives a measure of whether a trend will persist or revert to some historical average. It also indicates the presence of cycles, although these are typically non-periodic. The interesting point of such an analysis is that no assumption on the frequency distribution of the data is required. We have already introduced the concept of Brownian motion, which provides the result that a particle suspended in a liquid would increase its covered distance with the square root of the time used to measure it. If R is the distance and N the time, we have the following relation:

$$R \simeq N^{0.5}. \tag{2.33}$$

This result is very used within the economist community in that the standard deviation of the price of a commodity over fifty days can be approximated with the standard deviation over five days multiplied by the square root of ten. In order to investigate the dynamical properties of a time series, we can proceed following the R/S analysis proposed by Mandelbrot and Wallis [66] and based on a previous hydrological analysis of Hurst [43]. When writing Eq.(2.33), we call H the unknown exponent which can vary between 0 and 1, measuring the intensity of long-range dependence in a time series. Defining $\{R/S\}_N$ as the range of cumulative deviations from the mean divided by the standard deviation, as discussed below, and using a constant c, we have:

$$\left\{\frac{R}{S}\right\}_N = (cN)^H, \tag{2.34}$$

where N is the length of the portion of the time series we are considering. Given the time series r_t, we define the average:

$$m(N, t_0) = \frac{1}{N} \sum_{t=t_0+1}^{t_0+N} r_t, \tag{2.35}$$

the biased standard deviation:

$$S(N, t_0) = \left\{ \frac{1}{N} \sum_{t=t_0+1}^{t_0+N} [r_t - m(N, t_0)]^2 \right\}^{\frac{1}{2}}, \qquad (2.36)$$

the cumulative deviation from the mean:

$$X(N, t_0, \tau) = \sum_{t=t_0+1}^{t_0+\tau} (r_t - m(N, t_0)), \qquad 1 \leq \tau \leq N, \qquad (2.37)$$

and the maximum exursion:

$$R(N, t_0) = max_\tau X(N, t_0, \tau) - min_\tau X(N, t_0, \tau). \qquad (2.38)$$

For each subset of length N, beginning at each t_0 in the time series:

$$\left\{ \frac{R}{S} \right\}_N = \frac{\sum_{t_0} R(N, t_0)}{\sum_{t_0} S(N, t_0)} \qquad (2.39)$$

and finally:

$$\left\{ \frac{R}{S} \right\}_N = (cN)^H. \qquad (2.40)$$

A linear regression on $ln(\{\frac{R}{S}\}_N)$ and $ln(N)$ is performed and the gradient of the regression line is the Hurst exponent. According to this definition, $H = 0.5$ corresponds to a truly random time series (center panel of Fig. 2.6), $0.5 < H \leq 1$ describes a dynamically persistent, or trend reinforcing series (bottom panel of Fig. 2.6). As H approaches 1 the certainty increases as to the direction of change that has been currently seen will continue. A straight line with non-zero gradient will provide $H = 1$. Alternatively, the series has memory that increases with H. Importantly, the values at the beginning of the time series are as important to the dynamics as the most recently observed (infinite memory, as discussed in Sec. 2.6). An exponent such that $0 \leq H < 0.5$ describes an anti-persistent, or mean-reverting system (top panel of Fig. 2.6). At the limit of zero the time series must change direction every sample, as in the case of flipping a coin. The correlation C between disjoint increments of the time series is given by [107]:

$$C = 2^{2H-1} - 1. \qquad (2.41)$$

Thus, once again, if $H = 0.5$ the disjoint intervals are uncorrelated. For $H > 0.5$ the segments are correlated, exhibiting a memory effect which tends to amplify patterns.

For $H < 0.5$ the time series is characterized by $C < 0$. A lack of robustness of the R/S statistic in the presence of short memory or heteroskedasticity (see Sec. 2.5) has been discussed by Lo [59], suggesting a modified rescaled range statistics, which replaces the denominator S by a consistent estimator of the square root of the variance of the partial sum of r. This modified rescaled range statistics is more robust to short-range dependence.

2.8 Detrended fluctuation analysis

As already noted, the presence of trends and/or seasonality is a quite common source of non-stationarity. We have furthermore seen that a straight line with non-zero gradient results in an Hurst exponent $H = 1$. If one assumes trends to be additive, the non-stationarity due to them can be removed and consequently the other (eventual) type of non-stationarity originating from the lack of recurrences can be studied. One method that is able to perform this task is the detrended fluctuation analysis (DFA) [41, 91], which provides a simple quantitative parameter (the scaling exponent α of Eq.(2.32)) to represent the correlation properties of a signal[4]. Thanks to DFA the detection of long-range correlations embedded in non-stationary time series is possible; furthermore the spurious detection of apparent long-range correlations, artifact of non-stationarity, is also avoided. DFA has been applied in several contexts, like the study of cardiac dynamics, bioinformatics, economics, meteorology, geology and ethology. A further advantage of DFA is its ability to identify different states of the same system according to its different scaling behaviours. In the case of heart interbeat intervals, for instance, there exists a case study where the authors have found significant differences in the scaling exponent for healthy and sick patients [90].

As usual, one aspect is the method itself, another is the correct interpretation of the obtained scaling results, which is of crucial importance for the understanding of the intrinsic dynamics of the system. Typically, in fact, the correlation exponent is not always a constant (independent of scale) and crossovers often exist: Different

---

[4]The applicability of this concept is wider than just for directly observed diffusive processes: Given a sequence of bounded, seemingly stationary state x_t, their accumulation $\sum^T x_t = s(T)$ follows a diffusive-like path. The characterization of this path in terms of Hurst exponents yields insight into the nature of the increments x_t.

ranges of scales may result in different α, depending on a change in the correlation properties of the signal at different scales, or from trends in the data. Trends are the most common feature generated by natural systems, almost unavoidable. For example, the number of particles emitted by a radiation source in a unit time has a decreasing trend, since the source becomes weaker and weaker; the density of air due to gravity has a trend at different altitude; the air temperature in different geographic locations and the water flow of rivers have a periodic modulation due to seasonal changes; the occurrence rate of earthquakes in certain area has trends in different time periods. Most interestingly, as we will see in the next chapters, the time evolution of every economic index, in the very large scales, gradually drifts to larger values. Since trends are almost ubiquitous, the question whether they depend on external conditions or they are intrinsic within the system arises. A possible approach is to first recognize and filter out the trends before attempting to quantify correlations. A drawback of such an approach is that careful justifications should be given in order to allow this filtering, since an intrinsic trend is of course related to local properties of signal fluctuations.

The idea of DFA consists in dividing the time series $u(i)$ of length N into N/n equal size non-overlapping segments, after that an integration has been performed:

$$y(j) = \sum_{i=1}^{j}(u(i) - \bar{u}), \qquad \bar{u} = \frac{1}{N}\sum_{j=1}^{N} u(i). \tag{2.42}$$

In each box, the integrated time series is fitted by means of a polynomial function, $y_f(i)$, which is called the local trend. For order-ℓ DFA (DFA-1 if $\ell = 1$, DFA-2 if $\ell = 2$ etc.), ℓ-th order polynomial functions should be applied for the fitting. The integrated time series $y(i)$ is then detrended by subtracting the local trend $y_f(i)$ in each box. This operation results in the following detrended fluctuation function:

$$Y(i) = y(i) - y_{fit}(i). \tag{2.43}$$

The computation of the root mean square fluctuation:

$$F(n) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}[Y(i)]^2} \tag{2.44}$$

is repeated for different box sizes n in order to cope with different scales and to get a relationship between $F(n)$ and n. A power-law relation:

$$F(n) \propto n^{\alpha} \tag{2.45}$$

implies the presence of scaling. In case Eq.(2.45) holds, the parameter α is called the scaling exponent or correlation exponent. It represents the correlation properties of the signal, since if $\alpha = 0.5$, the signal is uncorrelated like a white noise; if $\alpha < 0.5$, the signal is anti-correlated; if $\alpha > 0.5$, positive correlations are present. In [41] the effect of three different types of trends is studied in the framework of artificially generated correlated noise. Linear, sinusoidal and power-law trends are used. The crossover in the scaling of noisy signals with trends also follow scaling laws: There is a long-range power-law dependence of the position of the crossover on the parameters of the trends. The DFA result of noise with a trend can be exactly determined by the superposition of the separate results of the DFA on the noise and on the trend, assuming that the noise and the trend are not correlated. If this superposition rule is not satisfied, this is an indication that the noise and the superimposed trend are not independent. Removing the trend could lead to changes in the correlation properties of the noise. DFA can also be used when wishing to minimize the effects of trends, and in order to recognize if a crossover indicates a transition from one type to a different type of underlying correlation, or it is just due to a trend without any transition in the dynamical properties of the noise.

The DFA procedure leads to a local measurement of the degree of long-range correlations, and consequently to the claim of the presence of local persistence or not, in the sense already explained when introducing the Hurst exponent. In [3] this concept has been applied to the exchange rate between the german mark and the british pound. Very interestingly, a change in slope of the local α corresponds to changes in the *Bundesbank* interest rate. Another interesting observation involves the analysis of DNA: Coding and non-coding sequences can be sorted out by looking at the local α value. It is $\alpha > 0.5$ in non-coding regions, suggesting the interpretation that it could be considered as a measure of information, an entropy variation indicating how information is managed by the system [91].

2.9   Overview of the approaches

In this chapter, (very) different tools and concepts have been introduced. Which
one is applied to a particular time series depends on the nature of the data and
represents a quite delicate task. We can anyway suggest some general guidelines:

- For *"dynamical"* data one should try with the reconstructed phase space, recur-
  rence plot analysis and over- embedding in order to recover the instantaneous
  dynamics of the system.

- For *"stochastic"* data the best solution consists on the application of station-
  arity tests and then on the use of fluctuation analysis in order to establish the
  correlation properties of the process.

In the next chapters we will focus on human voice and financial markets as prominent
examples of these two different kinds of data. Thanks to the two-mass model of the
vocal folds (to be introduced in Sec. 3.6) and to a lot of other models quite successful
in describing human vocal apparatus, one can assert that the time series representing
a record of a voice registered with a microphone belongs to the first type depicted
above. The system generating the signal can be characterized by few degrees of
freedom and non-constant parameters.

About the second application, the situation is not so clear. It is commonly
accepted in financial theory that time series of asset prices are unpredictable and
hence price dynamics are usually described as stochastic processes. On the other
hand it is quite clear that unpredictable time series and stochastic processes are not
synonymous. As stressed in Sec. 1.1, chaos theory has shown that unpredictable time
series can arise from deterministic non-linear systems. The question of whether the
time evolution of asset prices in financial markets might be due to underlying non-
linear deterministic dynamics arises quite naturally. To establish that, one could try
to reconstruct the (hypothetical) strange attractor, as introduced in Sec. 1.2, present
in the chaotic time evolution. Of course the corresponding dimension D plays an
extremely important role, since for chaotic systems with $D > 3$ it is difficult to
distinguish between a chaotic time evolution and a random process. The more
reliable estimation of D, as reported in [71], is the inequality:

$$D > 6. \tag{2.46}$$

From an empirical viewpoint, it is therefore quite unlikely that financial markets follow chaotic dynamics with a limited number of variable. Furthermore, it is a widely accepted believe (see Sec. 4.1) that the time evolution of an asset price depends on all the information affecting the investigated asset and is seems quite difficult to describe this huge amount of data with a small number of non-linear deterministic equations. Therefore financial data are considered to belong to the second type depicted above.

Chapter 3

Human Voice Signals

All the simple systems are simple in the same way, each complex system has its
own complexity.
(Lev Tolstoy, Anna Karenina)

Analysis of acoustic signals provides a lot of interesting reasons to be carried out. First of all there is an ever-growing need to store, code, transmit and synthesize voice signals, such that the telecomunications industry has dichotomized transmission of information into either voice or data, suggesting that voice signals are a class of their own. Typical investigations have been performed through the microphone signal in order to understand speech production and perception, given that the acoustic signal is the common link between them. From a health science viewpoint, the human voice analysis is of fundamental importance, since it provides much information about the general health and well-being of an individual. Our voice reveals who we are and how we feel, giving considerable insight into the structure and function of certain parts of the body.

Having to do with measured signals one has to develop methods to treat the noise unavoidably present in the time series. Sometimes the noise reduction can be a preprocessing of the signal for applications like voice recognition or data classification, in some other circumstances it can be the final goal of the treatment. In the following we will distinguish between two kinds of noise, namely the additive noise due to a measurement process and the internal noise of the human body, reflected in the always present small fluctuations in frequency, amplitude and waveshape of voice signals [68]. Voice production can be thought of as the activation of an en-

tire system of coupled oscillators. The intent to vocalize activates motor commands
that are responsible for the neural inputs to an array of biomechanical, neural and
acoustic oscillators. The vocal folds are the primary oscillating system that produce
the glottal airflow, what we might call the carrier signal. All other oscillators act
as modulators of the glottal airflow. Respiratory and heart beat modulations are
nearly sinusoidal; action potentials of muscles, air vortices and mucus in motion
are hugh dimensional; tracheal resonator, supraglottal vocal tract and sinuses are
passive oscillators that can strongly influence the primary oscillating system [6, 40].

3.1 Feasibility of iNoise Reduction

Essentially every recorded signal is contaminated by noise, which reduces the de-
tailed knowledge about it. Although generally one associates with noise specific
properties such as broad band power spectrum or a certain distribution, here we
want to be more general and denote by noise every source which is different from
the signal of interest. So, noise contamination can mean the effects of discretization
in a conversion from analogue to digital data, it can mean distortions of the mea-
surements due to fluctuations in the electronic equipment, due to impact of other
signals into the measurement device (such as 50 Hz noise from the electric power
supply), or it can be mixed into the signal in the transmission channel. One elim-
inates as much noise as possible from the recorded mixture of signal and noise by
filtering techniques, called noise reduction. More generally, one could call this signal
separation or demixing.

Every scheme for noise reduction requires criteria for distinguishing the desired
signal component and the noise component inside the recorded mixture. Typical
and often employed characteristics are spectral properties: If noise and signal have
distinct, well defined spectral contents, filters in the frequency domain can be suc-
cessfully used for their separation. The power spectrum is particularly useful for
studying the oscillations of a system. There will be sharper or broader peaks at
the dominant frequencies and their integer multiples, the harmonics. Purely pe-
riodic or quasi-periodic signals show sharp spectral lines and measurement noise
adds a continuous floor to the spectrum. Thus in the spectrum, signal and noise
are distinguished. Deterministic chaotic signals may also have sharp spectral lines
but even in the absence of noise there will be a continuous part of the spectrum.

Without additional information it is impossible to infer from the spectrum whether the continuous part is due to noise on top of a qusi-periodic signal or to chaoticity.

Human voice is a non-stationary signal, characterized by changes in the istantaneous dynamics during the speaking that usually should constitute undesired complications of the analysis (see also Fig. 2.5). In this case, however, such changes represent the most interesting structure in the recording and filtering them out would produce a quite poor signal quality (the time average spectrum is very broad band). In Fig. 3.1 the amplitude of a speech signal is shown at high time resolution. One observes the oscillatory nature of the sound waves, but, more importantly, the irregular change of the enveloping amplitude due to the non-stationarity. We will later show that also the wave form changes considerably and that therefore the spectral contents is time dependent; the severe unharmonicity requires a strong phase relationality between different frequencies while moving from phoneme to phoneme. There exist vowels with a rather pure harmonic spectrum, and there exist fricatives which resemble quite closely white noise. Thus, simple filters such as low-pass filters distort the voice signal more than they reduce noise. More sophisticated methods have to be used, such as adaptive filters, still working in the frequency domain, but on moving windows in time. Due to the huge technical relevance of noise reduction for human speech (telecommunication, electronic hearing aids, computer-based speech recognition), extremely elaborate and sophisticated algorithms have been developed. Examples of state-of-the-art methods are [32, 101].

The algorithm we want to present here is called *Local Projections in Embedding Spaces* and it is based on a very different philosophy, namely on structures in the reconstructed phase space. Its performance will be compared to a modern spectral subtraction scheme which can be considered standard, the Ephraim-Malah filter [23], and shows that already the application of the raw concept yields comparable gains in the signal to noise ratio. The results can be improved by post-processing the denoised signal with other filters which rely on properties different from those exploited by our method to distinguish between signal and noise (e.g. spectral properties).

The local projections noise reduction scheme has been developed in the framework of *non-linear time series analysis*, as a set of techniques for the analysis, manipulation, and understanding of aperiodic signals relying on the hypothesis of deterministic chaos [46]. This means that the signals reflect the complex dynamics of a purely deterministic (often few-degree-of-freedom) system. Since such signals
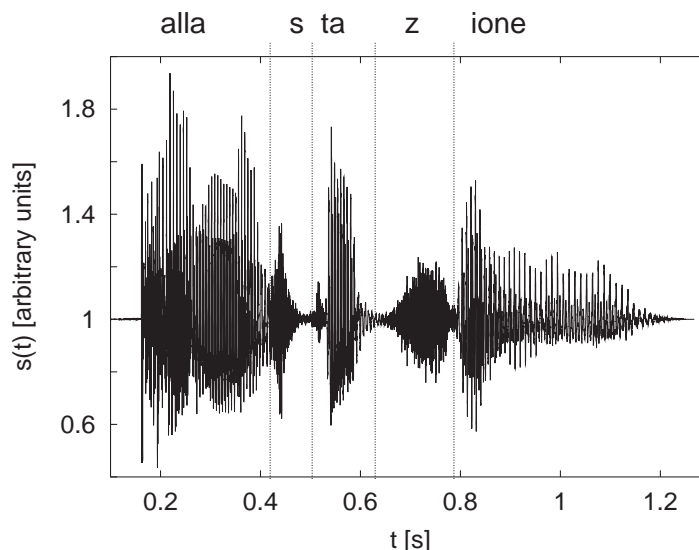
Figure 3.1: Time evolution of the italian sentence "alla stazione". A separation into the constituent phonemes is visualized.

represent a very limited class with most probably no relevance for time series data coming from outside the physicist's laboratory, it was of considerable interest to explore how far these concepts could be applied more generally to aperiodic signals with non-deterministic origin and strong non-stationarity like the human voice. Thus, one has to see the success of non-linear noise reduction of human voice not only as a new and alternative method for the solution of a prominent problem, but also as an outstanding example of the usefulness of phase-space methods for non-deterministic signals. This and other examples will eventually allow to compile a list of conditions when non-linear time series methods can be expected to be successful for signals that are clearly not stemming from deterministic dynamical systems of low dimension.

Although articulated human voice is, when considering full words, a highly non-stationary complex phenomenon, where the concatenation of phonemes is due to the arbitrariness of the speaker and thus largely unpredictable and in this sense non-deterministic, there is ample evidence that the sound generation by humans is a very deterministic dynamical phenomenon. Models of the sound generating mechanism have been designed; essentially the vocal folds as the mechanical parts modulating the sound generated by the vocal cords have been modelled. Both model simulations

and time series analysis of real data show that **stationary** articulated human voice such as the extended voweling ("aaa") have a low-dimensional dynamical origin. In [104], a two-mass model of vocal-fold vibrations is analyzed with methods from non-linear dynamics, showning that a sufficiently large tension imbalance of the left and right vocal fold induces bifurcations to subharmonic regimes, toroidal oscillations, and chaos. The reconstruction of attractors and the estimation of their properties indicate low dimensionality of the attractors generating the signal. Furthermore, it was shown by means of empirical orthogonal functions that normal phonation is well represented by only two eigenmodes. The simulation of disordered voice has shown that the three strongest modes contain 90% of the variance [7]. In contrast to stationary voice signals, the concatenation of different phonemes to full words or sentences does not represent a low-dimensional system, since there are frequent and arbitrary switches between different kinds of dynamical behaviour. Due to the transition regions from one phoneme to the next, a phoneme inside a word even differs from the same phoneme when it is spoken in an isolated and elongated way.

Hereby, the reconstruction of an embedding space allows us to efficiently cope with the problem of non-stationarity because different phonemes are identified implicitly if we provide a reconstructed embedding space of sufficiently large dimension, what we have called *over-embedding* in Sec. 2.2. The dynamics inside the single phonemes is very close to low-dimensional deterministic, being in fact almost periodic. The non-stationarity here plays the role of imposing different instantaneous dynamics to different phonemes throught a different set of parameters.

For an illustration of the noise reduction method let us consider the following task. We have to eliminate noise from a song stored on an old-fashioned vinyl disc, where the noise could be induced, for example, by scratches on the black LP. The task becomes almost trivial if we can retrieve several samples of this LP because if we play them synchronously, the signal part of the different discs is identical, whereas the noise part is independent, as depicted in Fig. 3.2. Already simple averaging will enhance the signal, and more sophisticated rules will allow us to completely remove the noise thanks to the availability of redundancy, a keyword in information theory. When coping with (almost) deterministic signals we do not need several copies of them. The deterministic dynamical constraints (as seen in Fig. 2.1) imply that the redundancy is stored in the past. Since determinism means that similar initial conditions will behave, at least for short periods, in a similar way, one solely has to
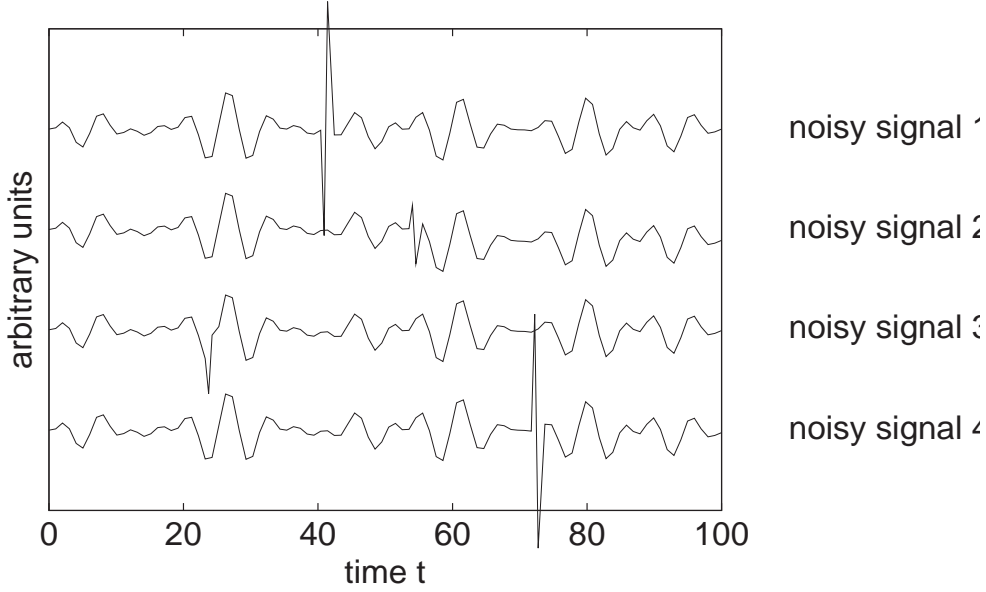
Figure 3.2: Sketch of noise reduction of music stored on LPs: Playing synchronously different samples of the identical piece of music, one can clearly identify noise due to its independence, whereas the aperiodic signals have a cross-correlation of unity.

look for near repetitions of the present signal in the past. Based on this idea, several approaches for noise reduction for deterministic chaotic data have been developed [31, 51].

3.2   Local projective noise reduction scheme

Let a dynamical system be given by the map $\mathbf{F} : \Gamma \to \Gamma$ in a state space $\Gamma \subset \Re^d$. The equation of motion thus reads $\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n)$. Not knowing $\mathbf{F}$, one can determine it in linear approximation from a long time series $\{\mathbf{x}_k\}$, $k = 1, \ldots, N$, by determining a set of neighbouring points $\mathcal{U}_n$ of $\mathbf{x}_n$ and minimizing

$$\sigma_n^2 = \sum_{k:\mathbf{x}_k \in \mathcal{U}_n} \left(\mathbf{A}_n \mathbf{x}_k + \mathbf{b}_n - \mathbf{x}_{k+1}\right)^2, \tag{3.1}$$

the one-step prediction error, with respect to $\mathbf{A}_n$ and $\mathbf{b}_n$ (see [15, 26]). The implicit relation $\mathbf{A}_n \mathbf{x}_k + \mathbf{b}_n - \mathbf{x}_{k+1} = 0$ expresses that data are confined to a hyperplane in the extended phase space. When the signal $\mathbf{x}_k$ is superimposed by random noise, $\mathbf{y}_k = \mathbf{x}_k + \eta_k$, the set $\mathcal{U}_n$ will no longer be embedded in a manifold whose tangent space is the hyperplane defined by $\mathbf{A}_n$ and $\mathbf{b}_n$, but will form a cloud scattered around

it. Reducing noise now means to project the noisy $\mathbf{y}_n$ onto this hyperplane. Also in the noisy case, the unknowns $\mathbf{A}_n$ and $\mathbf{b}_n$ can be determined by the least squares problem (Eq.(3.1)), such we can gain all relevant information from the data.

As already noted, usually only scalar measurements are performed, i.e. not the full phase space vectors $\mathbf{x}$ are measured, but an observable $s = g(\mathbf{x})$, which is a non-linear projection of the phase space to the reals. A scalar time series $\{s_n = g(\mathbf{x}_n)\}$ is a sequence of measurements equidistant in time. The constancy of the time intervals in between successive measurements in the precondition of Takens theorem and generalizations thereof [37, 97, 106]: The Takens time delay embedding method allows us to reconstruct a sequence of vectors, from the sequence of scalar observables:

$$\mathbf{s}_n = (s_n, s_{n-\tau}, \ldots, s_{n-(m-1)\tau}), \tag{3.2}$$

which are equivalent to the unobserved state vectors, as already noted in Sec. 2.2.

The basic idea of the noise reduction, i.e. identification of the hyperplane containing the unperturbed data and projecting the noisy data onto this manifold, can be ported into the embedding space with very little modification [31, 38]. The local linear equations of motion in delay embedding space reduce to:

$$s_{n+1} = \mathbf{a}_n \mathbf{s}_n + b_n \,, \tag{3.3}$$

since all other components of the future delay vector $\mathbf{s}_{n+1}$ are copied from $\mathbf{s}_n$. Working in m-dimensional delay coordinates and assuming that the data are confined locally to an $(m - Q)$-dimensional hyperplane, $\mathbf{a}_n$ is not uniquely defined and we can find up to Q mutually independent linear subspaces $\mathbf{a}^q, q = 1, \ldots, Q$ fulfilling $s_{n+1} - \mathbf{a}_n \mathbf{s}_n - b_n = 0$. For obvious reasons we will call the linear space spanned by these Q vectors the *nullspace* at point $\mathbf{x}_n$. Since the noise free attractor does not extend to this space, the component of $\mathbf{s}_n$ we find in it must be due to noise. The locally projective noise reduction algorithm tries to identify this nullspace and then removes the corresponding component of $\mathbf{s}_n$.

When we assume that the nullspace has Q directions, we have to find Q orthonormal vectors $\mathbf{a}^q$ such that the local projection onto these vectors is minimal. If we use the notation $\mathbf{z}_n = (\mathbf{s}_n - \bar{\mathbf{s}})$ ($\bar{\mathbf{s}}$ is the average of $\mathbf{s}_n$ on $\mathcal{U}_n$), the projection of $\mathbf{z}_n$ onto the nullspace (assuming normalised vectors $\mathbf{a}^q$) is $\sum_{q=1}^{Q} \mathbf{a}^q \cdot (\mathbf{a}^q \cdot \mathbf{z}_n)$ and we require $\sum_{n' \in \mathcal{U}_n} [\sum_{q=1}^{Q} \mathbf{a}^q \cdot (\mathbf{a}^q \cdot \mathbf{z}_n)]^2$ to be minimal for the correct choice of the
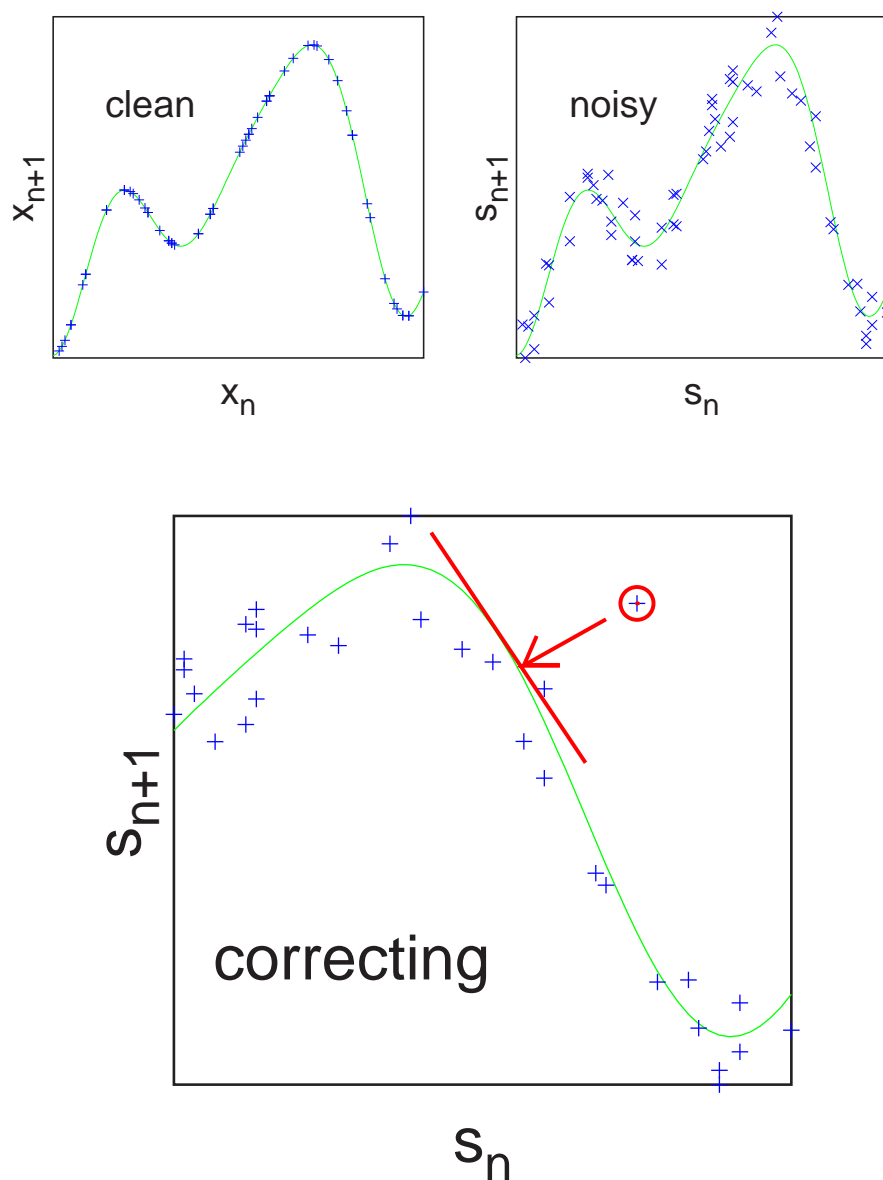
Figure 3.3: Schematic representation of the noise reduction method for a signal from a 1-dimensional deterministic map in the time delay embedding space. Neighbourhoods in embedding space exploit redundancy of the signal which is non-local in time. Upper left panel: Clean data align on a graph (1-dimensional manifold) $x_{n+1} = g(x_n)$. Upper right panel: Noisy data are scattered around it. Lower panel: Noise reduction can be performed by projection.

set of $\mathbf{a}^q$. If we introduce the constraint that the $\mathbf{a}^q$ have unit length by means of Lagrange multipliers λ^q and use that the $\mathbf{a}^q$ are orthogonal[1], $\mathbf{a}^q \cdot \mathbf{a}^{q'} = 0, \quad q \neq q'$, we have to minimize the Lagrangian:

$$L = \sum_{n' \in \mathcal{U}_n} \left[\sum_{q=1}^{Q} \mathbf{a}^q \cdot (\mathbf{a}^q \cdot \mathbf{z}_{n'}) \right]^2 - \sum_{q=1}^{Q} \lambda^q (\mathbf{a}^q \cdot \mathbf{a}^q - 1) \tag{3.4}$$

with respect to $\mathbf{a}^q$ and λ^q. This can be done separately for each q and yields:

$$\mathbf{C}\mathbf{a}^q - \lambda^q \mathbf{a}^q = 0, \qquad q = 1, \ldots, Q, \tag{3.5}$$

where $\mathbf{C}$ is the $m \times m$ covariance matrix of the vectors $\mathbf{z}_{n'}$ within the neighbourhood $\mathcal{U}_n$:

$$C_{ij} = \sum_{n' \in \mathcal{U}_n} (\mathbf{z}_{n'})_i (\mathbf{z}_{n'})_j . \tag{3.6}$$

Of course, the solutions of Eq.(3.5) are nothing but the orthogonal eigenvectors $\mathbf{a}^q$ and eigenvalues λ^q of $\mathbf{C}$. These can be readily determined with standard software. The global minimum of L is given by the eigenvectors to the Q smallest eigenvalues. The noise component of the vector $\mathbf{z_n}$ is thus removed by replacing it with:

$$\hat{\mathbf{z}}_n = \mathbf{z}_n - \sum_{q=1}^{Q} \mathbf{a}^q \cdot (\mathbf{a}^q \cdot \mathbf{z}_n) . \tag{3.7}$$

Finally, we write the result in terms of the original delay vectors $\mathbf{s}_n$:

$$\hat{\mathbf{s}}_n = \mathbf{s}_n - \Delta \mathbf{s}_n = \mathbf{s}_n - \sum_{q=1}^{Q} \mathbf{a}^q \cdot [\mathbf{a}^q \cdot (\mathbf{s}_n - \bar{\mathbf{s}})] . \tag{3.8}$$

The noise reduction scheme outlined above is called *local projective noise reduction*. It is illustrated in Fig. 3.3. The conceptual steps for the implementation can be riassumed in the following way:

- For every delay vector $\mathbf{s}_n$, all neighbours in the delay embedding space are collected (i.e. $\mathcal{U}_n$ is formed).

- The covariance matrix $C_{ij} = \sum_{\mathcal{U}_n} (\hat{\mathbf{z}}_k)_i (\hat{\mathbf{z}}_k)_j$ is computed, and its singular values are determined.

---

[1] We could also require orthogonality by additional Lagrange multipliers. This would complicate the algebra but leads to the same result.

- The vectors corresponding to the largest singular values are supposed to represent the directions spanning the hyperplane defined above by $\mathbf{a}_n$ and b_n.

- To reduce noise, $\mathbf{s}_n$ is projected onto these dominant directions.

- A suitable average over the m corrections of the same s_n is performed.

The latter point is the most difficult to treat complication arising from the fact that we are working in a time delay embedding space: The minimization problem given by Eq.(3.4) is local in the time index n, not global for the whole time series. Solving it, we gain vector valued corrections for every delay vector, but we intend to correct the scalar time series. Each time series element is member of m different delay vectors, hence there are m different corrections for it. We usually employ an arithmetic average over them. This cancels partly the corrections done and thus can require to iterate the method a few times for convergence. On the other hand, this enables us to reduce noise components which are *inside* the manifold at one time instant. The choice of the parameters entering the algorithm (embedding window $((m-1)\tau+1)$, embedding dimension m, time lag τ, diameter ϵ of the neighbourhoods $\mathcal{U}_n$, number Q of singular vectors to project on) is the crucial problem and has to respect the particular properties of the signal and of the noise. We will show in Sect. 3.5 a completely automated way to optimize ϵ, the most important parameter.

3.3 Structure and redundancy in human voice

The crucial aspect for the application of the non-linear noise reduction algorithm is the choice of the correct embedding parameters and the dimension of the linear subspace onto which we project in order to capture the structure we want to preserve. However, even more relevant than the choice of the parameters is the issue whether a given data set has at all enough deterministic structure to be used by the method. The surprising result of this study is that structure in embedding space does in fact exist in human voice signals and survive even for signal to noise ratios of unity or less (see again Fig. 3.7). Human voice forms an aperiodic and highly non-stationary signal. In Fig. 3.1 we show the trace of the italian words "alla stazione". It is composed of sub-units, called phonemes, which can be considered as different kinds of limit-cycle-like dynamics. Zooming into such signals confirms the claims made in the
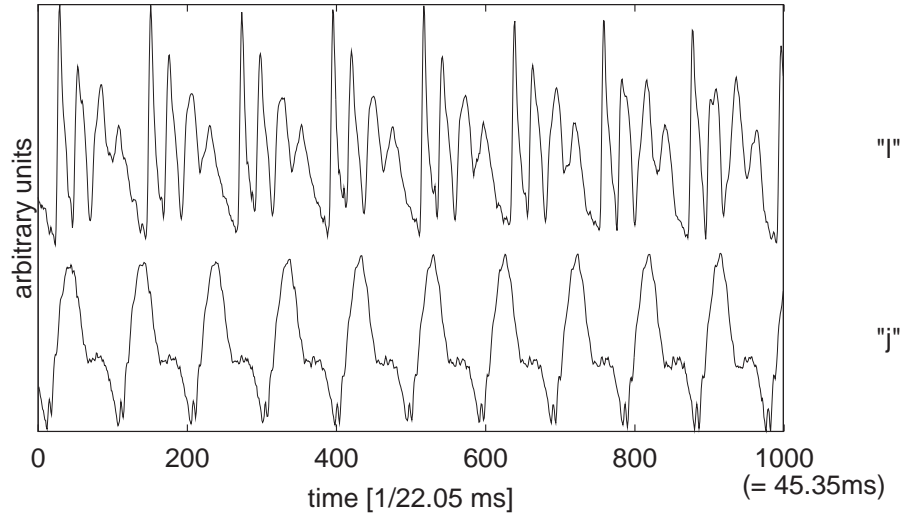
Figure 3.4: Two different phonemes in high time resolution (equal scale, different off-set).

introduction, namely that the dynamics underlying individual and purified phonemes is rather low-dimensional. Fig. 3.4 shows two different phonemes belonging to a word: Inside them, on time scales of about 5ms, human voice is almost periodic. Of course, the transitions from one phoneme to the next one and the fricatives are aperiodic. Inside the transitions, still clear but time dependent patterns are present. When we convert a finite sample of measurements of these waves into delay vectors, then vectors representing the same wave form and the same phase of this wave are close neighbours in the high dimensional space. Thus, searching for neighbours in the embedding space means to identify redundant parts of the wave trains, a task very close to pattern matching.

For a suitable delay reconstruction we now need the embedding dimension m and the time lag τ. Due to the non-stationarity related to the presence of the very many different instantaneous dynamical regimes we have to cover a full wave cycle by our delay vectors (called *over-embedding*, as introduced in Sec. 2.2, since a reasonable time delay embedding of a close-to-periodic signal would cover only one quarter of the wave cycle), i.e. the embedding window:

$$((m-1)\tau + 1) \tag{3.9}$$

should be about 5ms (or more, depending on the pitch of the voice). The large

variety of existing phonemes forces us to use about 25 measurements per cycle for the clean data to be able to distinguish between all of them. When noise is added, we even have to increase the embedding dimension, since we have to compensate the loss of neighbours by noise contamination (originally almost identical wave cycles look different after distortion) by larger neighbourhood diameters, but we do not want to accept false neighbours.

Recurrence plots [14, 18] are a suitable tool for the assessment of the quality of an embedding. In the time-time plane, one represents the couple of time indices (i, j) by a dot, if $|\mathbf{s}_i - \mathbf{s}_j| < \epsilon$, i.e. if the distance between the corresponding delay vectors is less than some tolerance ϵ. Fig. 3.5 shows a slice around the diagonal of a recurrence plot on clean data. As desired, there are recurrences for $|i - j| \approx 5$ms, i.e. we identify correctly the almost periodicity. More importantly, there are time intervals without recurrences, and these are the transition regions between phonemes. Our method thus implicitly identifies different phonemes and it locates the correct phases of the waves cycle inside the same phoneme. As larger sections of the time-time plane reveal (not shown in Fig. 3.5), there are almost no recurrences across phonemes in different words. Thus our algorithm essentially relies on intra-phoneme redundancy, computationally very fast to establish even if not always enough to garantee optimal performances.

When the local projective noise reduction scheme is applied to low-dimensional chaotic data, a quite critical problem consists on how to control the exponential instability of the dynamics, which, among other aspects, led us to introduce a non-trivial metric in the delay embedding space, as in [31]. For voice data, this is not an issue, since intra-phoneme dynamics is limit cycle like and thus marginally stable. The problem of noise reduction for voice lies in the fact that typical noise levels for which the method should be employed are much higher and of the order of the signal itself. Thus, it becomes a non-trivial problem to identify the correct neighbours (see Fig. 3.6).

The reasons why a local projections noise reduction scheme can be successfull have to found in the identification and exploration of quasi-deterministic structures in the voice signal. This process is very robust against noise contamination, as shown by Fig. 3.7: We consider a short word, lasting 0.5 s ($\approx$ 10000 points), noisefree, and we report with filled circles the number of neighbours along the time series. The 3 numbers in the legend of the figure have the following meaning: Embedding dimen-
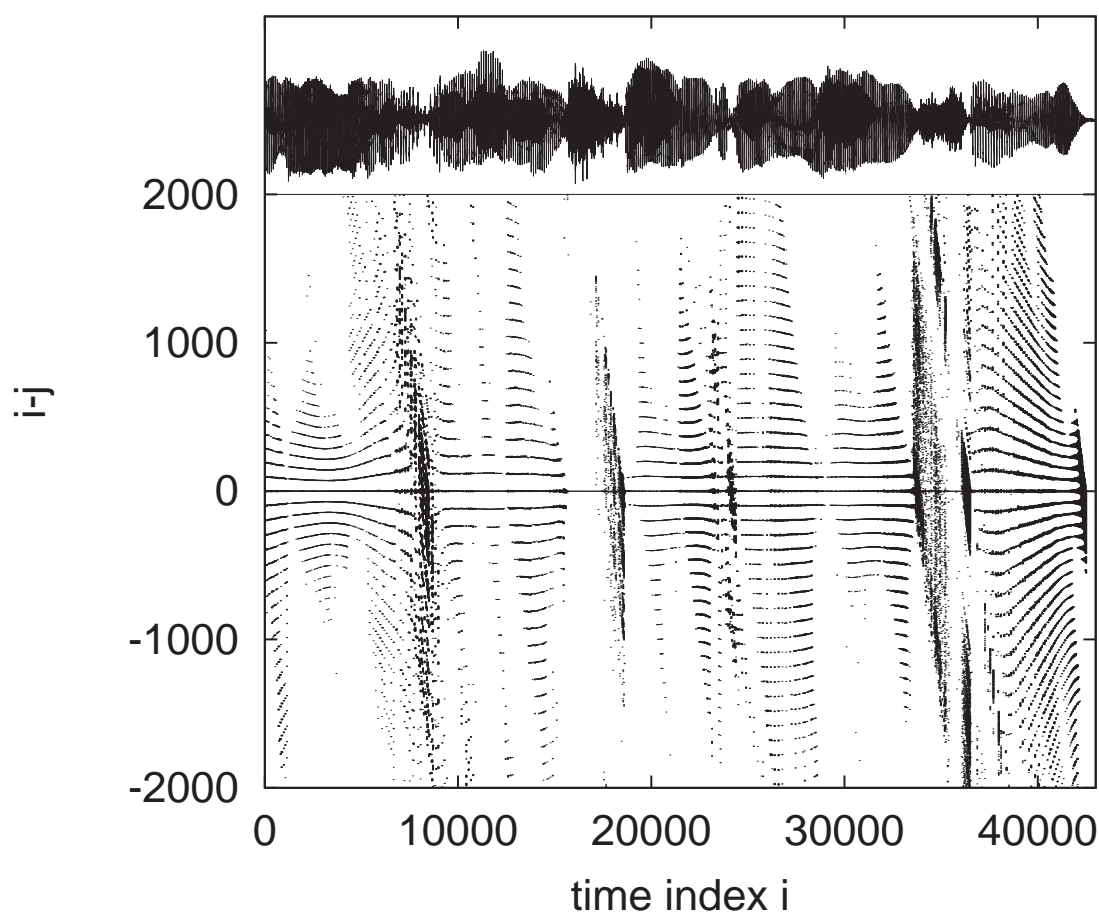
Figure 3.5: Main panel: Section of a recurrence plot: In the plane of indices i, j a dot is printed, whenever the delay vectors fulfill $|\mathbf{s}_i - \mathbf{s}_j| < \epsilon$. It proves that our delay vectors really represent meaningful states, where the line structure shows the approximate periodicity inside the phonemes and the number of intra-phoneme neighbours. There are almost no dots for $|i - j| > 2000$, reflecting the lack of inter-phoneme similarities (for this particular ϵ). Upper panel: The speech signal underlying the recurrence plot.
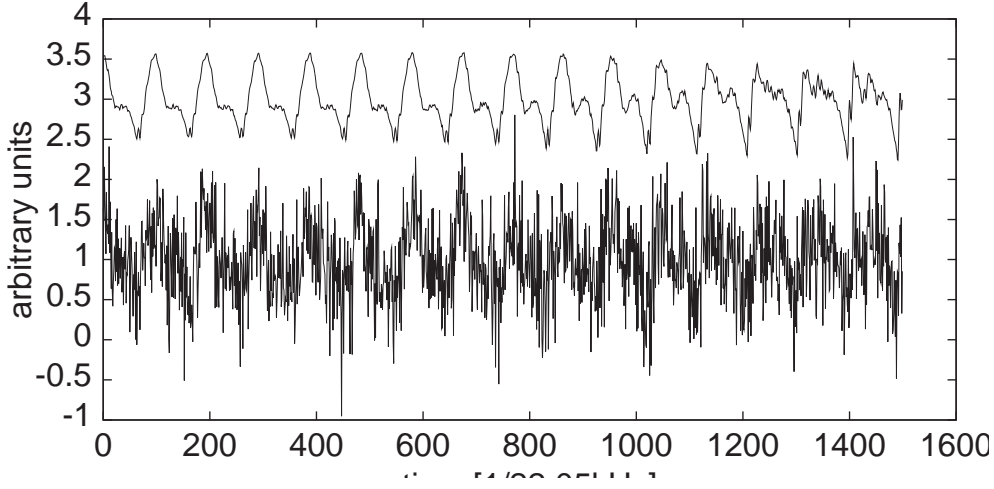
Figure 3.6: Part of the clean (top) and noisy (-4 dB SNR, bottom) signal "buon giorno". The difference in wave form of the left and the right part of this signal should be detected even after noise contamination. As shown in the text, euclidean distances on delay vectors with high dimensionality help to solve this problem.

sion m, delay τ and neighbourhood size ϵ. After the contamination with uncorrelated noise, we note that using the same set of parameters all the neighbourhood relations are destroyed. The reason, as explained in more details with the help of Fig. 3.14, is that ϵ is now too small and due to noise all the neighbourhoods we are considering are almost uniformly filled by points and therefore no precise identification of the underlying structure can be performed (filled squares). The solution is simple and powerful at the same time: We just have to consider a bigger ϵ (together with a bigger m for stability reasons) in order to recover all the neighbourhood structures (crosses in Fig. 3.7). We will address in Sect. 3.5 the trade-off between collecting as many neighbours as possible and avoiding false ones.

Here, the choice of a suitable norm, sampling rate and time lag is essential. If we denote the signal values by x_i, the additive noise by η_i, then the euclidean distance between two delay vectors $\mathbf{s}_n$ and $\mathbf{s}_k$ reads:

$$
\begin{aligned}
d^2_{\text{noisy}} &= \sum_i^m (x_{n+i} + \eta_{n+i} - x_{k+i} - \eta_{k+i})^2 \\
&= \sum_i^m (x_{n+i} - x_{k+i})^2 + \sum_i^m (\eta_{n+i} - \eta_{k+i})^2 \quad\quad (3.10)
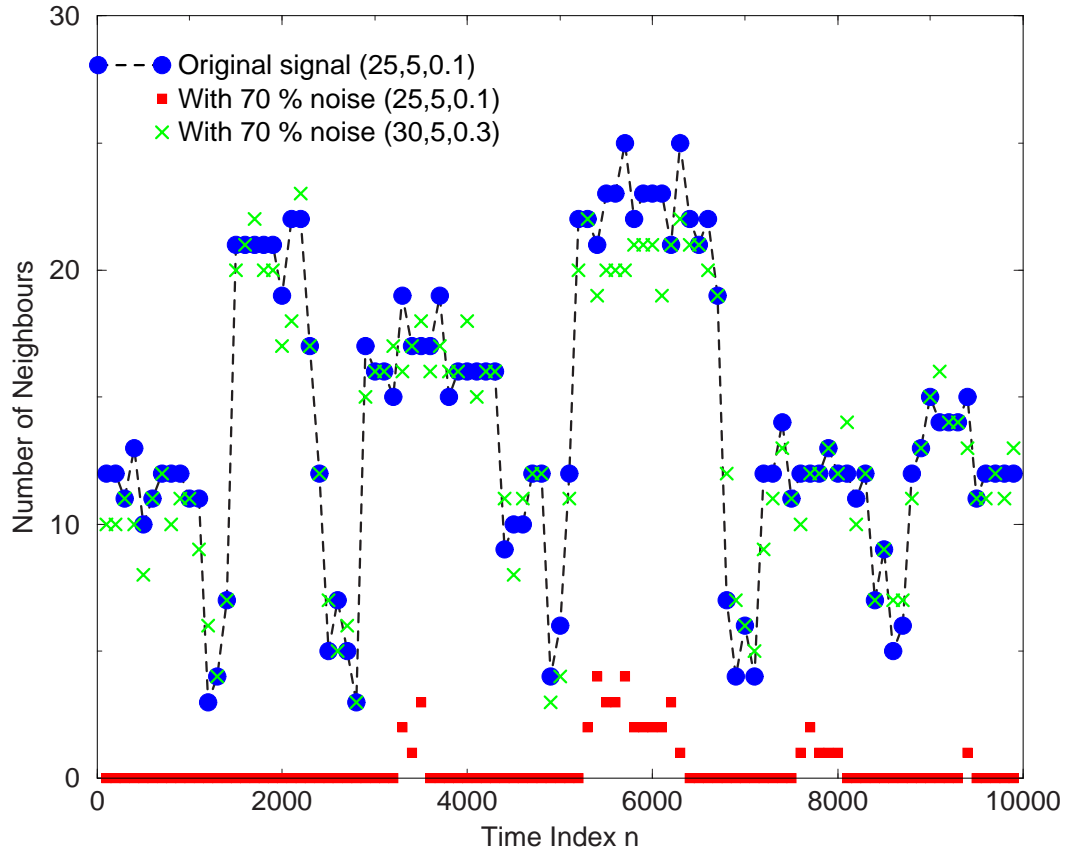\end{aligned}
$$

Figure 3.7: Number of neighbours as a function of the time index of a short word. Filled circles indicate the case without noise using $m = 25$, $\tau = 5$ and $\epsilon = 0.1$. In order to afford the contamination with noise, one has to consider a bigger ϵ (and also a bigger m for stability reasons), as reported by crosses. Otherwise if we want to use the same set of parameters as in the noisefree situation we do not recover any structure and therefore we do not get any significant neighbour, as reported by filled squares. The main message here is that the noise does not destroy completely the original redundancy and the filtering is then possible. Of course one has to optimize ϵ in order to get as many real neighbours and as few false neighbours as possible.

$$+2\sum_{i}^{m}(x_{n+i} - x_{k+i})(\eta_{n+i} - \eta_{k+i})$$

$$\approx \quad d_{\text{clean}}^2 + 2m\sigma^2 + \langle x_{n+i} - x_{k+i}\rangle\langle\eta_{n+i} - \eta_{k+i}\rangle$$

$$= \quad d_{\text{clean}}^2 + 2m\sigma^2, \tag{3.11}$$

where σ^2 is the variance of the noise. The rightmost term of the third line disappears due to the fact that the noise differences should have zero mean, and the splitting is due to the uncorrelatedness of signal and noise. The approximation becomes the better the larger the number of summands m, i.e. the higher the dimension of the delay vector. For given size of the embedding window (e.g. 5 ms) this requires a high sampling rate and a time lag of $\tau = 1$. For the 22.050kHz data this means that a delay vector of dimension $m = 100$ with $\tau = 1$ is reasonable. In summary, the squares of euclidean distances between high dimensional delay vectors suffer from a simple off-set by the variance of the (local) noise level. Numerically, euclidean distances are more costly than L^∞-distances (max-norm), and a lag of $\tau = 1$ is more costly than $\tau \approx 5$, which, in our example, would be completely sufficient for clean data.

Fig. 3.8 shows how far these considerations are valid in practice. We computed the distances between a large number of pairs of delay vectors from the time series of "buon giorno" on embedding windows of length 100, for 4 different situations: Euclidean norm for $\tau = 1$, $m = 100$ (upper left panel), $\tau = 5$, $m = 20$ (upper right panel), max-norm for $\tau = 1$, $m = 100$ (lower left panel), and $\tau = 5$, $m = 20$ (lower right panel). Plotted are the distances between every given pair of delay vectors, d_{ij}, computed on the clean data, versus the distances computed on the data after numerically adding white noise. The upper left panel (euclidean with maximal dimensionality) is close to ideal: The distances between noisy vectors and clean vectors fulfill in reasonable approximation the law:

$$d_{\text{noisy}}^2 \approx d_{\text{clean}}^2 + const. \tag{3.12}$$

The deviation from this law is essentially due to the fact that on a "sample" of $m = 100$ noise values, $\sum_{i=1}^m (\eta_{n+i} - \eta_{k+i})^2 \neq 2m\sigma^2$ and $\sum_{i=1}^m \eta_{n+i} - \eta_{k+i} \neq 0$, but that there are fluctuations of the order of $\sqrt{m}$.

Although the algorithm is usually implemented in a way such that the neighbourhood size ϵ is increased until the desired number of neighbours is found, the
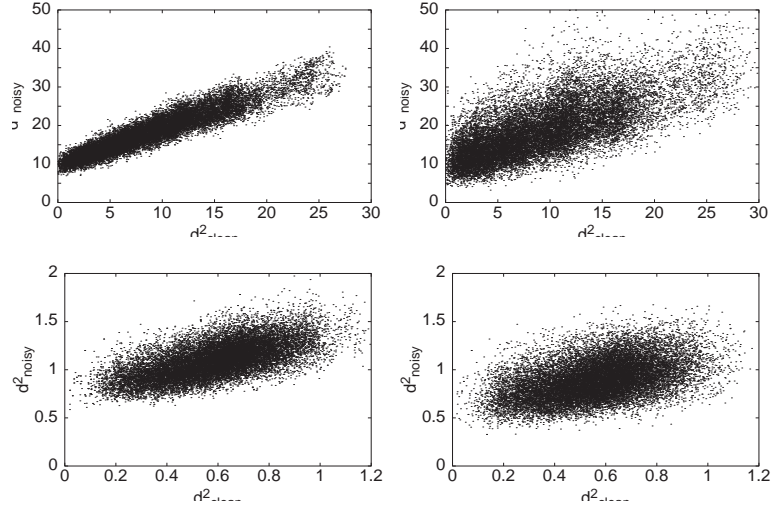
Figure 3.8: Distances between pairs of delay vectors of clean data versus the distances of the corresponding pairs after numerical contamination with additive white noise, for different norms and embedding spaces (for more details see text). The upper left panel (euclidean norm, $m = 100$, $\tau = 1$) is optimal. Noise level: 80% (SNR: 2dB).

knowledge of the noise level can be used as an initial value ϵ_0 to start with, and thus speed up the neighbour search. The (instantaneous) noise level can be easily estimated, if the noise is uncorrelated (otherwise it will be underestimated) by the following procedure: We know that the highly sampled clean voice signal is smooth. Thus, the distance between adjacent delay vectors of the clean signal, $|\mathbf{s}_n - \mathbf{s}_{n\pm1}|$, is close to zero. Together with Eq.(3.10) and Eq.(3.12), respectively, it follows that the distance between adjacent noisy delay vectors is in good approximation $|\mathbf{s}_n - \mathbf{s}_{n\pm1}| \approx \sqrt{2m}\sigma$.

Even for 100% of noise one can thus reasonably recover the neighbourhood relations of the clean signal. Typically, we choose ϵ such that about 20 neighbours are found. Based on these neighbourhoods, the noise reduction algorithm works with a projection onto the dominant ≈ 5 dimensions which can be identified with at least as many neighbours. This rather large subspace leaves enough degrees of freedom also for the transition regions between phonemes. Every phoneme populates a different region of the reconstructed space and there is almost no intersection since we are considering an appropriate *over-embedding* that is able to perform and automatic and implicit segmentation of the time series into constituent phonemes.

All values reported here cannot be given in a systematic way since everything depends slightly on the particular voice, the noise, the sampling rate and even the words. In Sec. 3.5 we will develop more powerful tools for their instantaneous adaptation [83]. Although the number of neighbours used is undesirably small, it cannot easily be increased. A single phoneme does not offer more than ≈ 20 almost repetitions of a given wave. Searching for neighbours in other words (presumably in identical phonemes) introduces large numerical effort, requires longer sentences and, most importantly, does not improve the situation much, since changed amplitudes and dilatation or compression of identical phonemes in different words destroy the closeness in embedding space. Thus all results presented here were gained from intra-phoneme neighbours, and the algorithm thus is semi-local in time: Only a time window of about 200ms (the maximal duration of a phoneme) has to be stored and is used as a data base for the construction of the neighbourhoods. In order to find maximal redundancy for the identification of the signal, the algorithm works in a non-causal way on the time span of a phoneme, i.e. a given time series element will be denoised only after already the future of about 200ms is recorded. Thus even if its processing speed is real-time, the output will be delayed with respect to the input by these 200ms.

3.4   Performance of the noise reduction of human voice

A first demonstration of the method is done on a stationary phoneme, a recording lasting 3 seconds of the vowel "a" (pronounced as in "far"). It is contaminated numerically by 30 percent white noise, and afterwards filtered by the noise reduction algorithm. The power spectrum before adding noise, after adding noise, and after noise reduction is shown in Fig. 3.9. Obviously, we were able to restore significant parts of the spectrum which are well below the noise level and thus invisible for any global band-pass filter.

The voice signals studied in the following are taken from a language course on CD ROM with a sampling rate of 22.050 kHz, sampled with 16 bit. The data were converted to real numbers and numerically contaminated by different types and amplitudes of noise and subjected to the noise reduction algorithm. Since we start
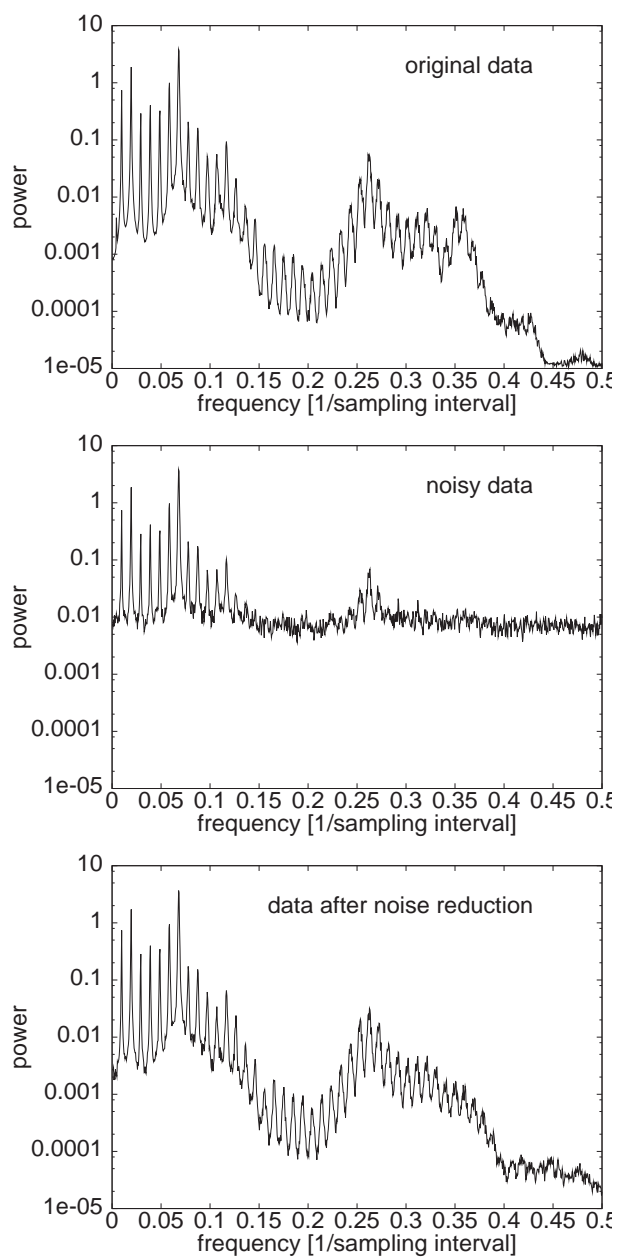
Figure 3.9: The power spectrum of 3 seconds of the vowel "a": Original recording, after adding noise numerically, and after non-linear noise reduction. Most of the structure of the original spectrum below the noise level could be reconstructed.
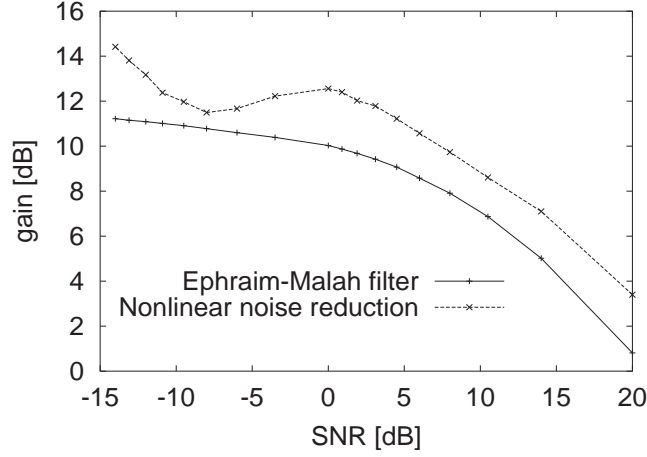
Figure 3.10: The gain of the local projective noise reduction scheme as a function of the noise level. The breakdown around SNR of 0 is due to the switching from the local linear to the locally constant model in our algorithm. Comparison with performance of the Ephraim-Malah adaptive filter [23].

from very clean signals, we can quantify the gain in dB, given by

$$\text{gain} = 10 \log_{10} \left(\frac{\sum (y_k - s_k)^2}{\sum (\hat{y}_k - s_k)^2} \right) , \tag{3.13}$$

where s_k is the clean, y_k the noisy and $\hat{y}_k$ the signal after noise reduction, as a measure of performance. Additionally, we reconvert noisy and denoised signals into the `wav`-audio format and inspect the results accoustically.

The projection is done onto subspaces of dimension in the range from 3 to 7. As will be discussed by the help of Fig. 3.10, noise-free signals are only marginally distorted by our non-linear filtering technique. The fact that the voice signal passes through this "dimension reduction filter" almost unperturbed confirms again that the wave dynamics inside every phoneme represents only few degrees of freedom, once it has been identified in the high-dimensional space. This result is in agreement with the study presented in [40] and explains also why as few as 20 neighbouring points are sufficient for the algorithm: Only the subspace onto which the projection is done has to be identified, all the remaining directions are irrelevant. Thus only the large singular values and the corresponding singular vectors of the covariance matrix C_{ij} in Eq.(3.6) are needed.

The high dimension of the embedding space helps to identify neighbours also for
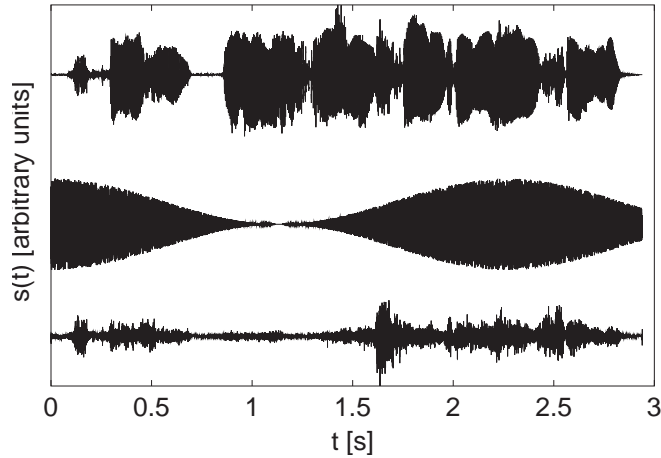
Figure 3.11: The clean signal of "Scusi, lei è Alessandra Janssen?", the white noise with time dependent amplitude added to the signal, and the remaining distortion after noise reduction (same scale, different offsets). The amplitude of the remaining noise varies systematically, i.e. the success of the noise reduction depends partly on the signal.

rather high noise levels: Usually, neighbourhoods merge if all data are contaminated by large amounts of noise. Here, due to the fact that the signal is rather sparsely filling the 20 to 30-dimensional space, this is not a problem, and we find reasonable results (Fig. 3.10). For noise levels bigger than $\approx 150\%$ the computation of the singular values of the covariance matrix gives unreliable results and the filtering is instead performed via an averaging over the identified neighbours rather than a projection into the approximated sub-manifold. However, this may be seen as a degenerate projection onto a zero dimensional space and is fully contained in the general algorithm.

An ideal filter would leave the unperturbed signal unaffected. Due to the violation of the basic assumptions about stationarity and determinism, this is not true for the voice signal, but only below 5% of noise these distortions become larger than the gain due to noise elimination. The results of our algorithm have been opposed to the performance of a particular spectral subtraction scheme which as to some extent become a benchmark filter, the Ephraim and Malah filter [23]. The Ephraim-Malah adaptive filter is particularly powerful when applied to speech signals, which are disturbed by additive noise with slowly varying spectral characteristics. Spectral subtraction is performed by subtracting a mean magnitude of the noise spectrum

from the disturbed spectrum to obtain an estimation of the magnitude of the noise-free spectrum.

The method may be interpreted as spectral equalization of the noisy-speech signal by applying spectral weights to the transformed signal; the spectral analysis (synthesis) is usually performed by a discrete Fourier (inverse) transformation with overlap-add techniques and by analysis and synthesis filterbanks with non-uniform frequency bands. The main difficulty involves the estimation of the mean magnitude of the noise spectrum; there exist methods based on spectral-minima tracking of the smoothed magnitude of the noisy-signal spectrum. The spectral minima are determined in a time window, the length of which is chosen such that in most practical cases a speech pause is present within the actual window, and an estimation within a sliding window is performed. Since we know the noise spectrum beforehand, we can operate the algorithm under optimal conditions, and the comparison of Fig. 3.10 is absolutely fair.

A detailed description of the spectral subtraction rule due to Ephraim and Malah is provided by [13]. Let us consider a signal with additive noise:

$$x(k) = s(k) + n(k), \tag{3.14}$$

being $x(k)$ the noisy signal, $s(k)$ the speech signal and $n(k)$ the noise. Let us define $X(e^{j\Omega})$ and $N(e^{j\Omega})$ the Fourier transformations of $x(k)$ and $n(k)$. The estimation of the spectrum of the filtered time series is obtained in the following way:

$$\hat{S}(e^{j\Omega}) = \left(|X(e^{j\Omega})| - \overline{|N(e^{j\Omega})|} \right) e^{j\Phi_x(\Omega)}, \tag{3.15}$$

where $\Phi_x(\Omega)$ is the phase of the disturbed speech signal. $\overline{|N(e^{j\Omega})|}$ is an estimated mean spectral magnitude of the noise from the spectral magnitude of the signal containing speech plus noise. One can show that Eq.(3.15) may be interpreted as spectral weighting of the noisy signal:

$$\hat{S}(e^{j\Omega}) = G(e^{j\Omega})X(e^{j\Omega}). \tag{3.16}$$

The core of the Ephraim and Malah filter is really the estimation of the spectral weights. It is performed after having divided the time series in blocks of a proper length. First the *a posteriori* signal-to-noise ratio:

$$R_{post,b}(e^{j\Omega}) = \frac{|X_b(e^{j\Omega})|^2}{|N_b(e^{j\Omega})|^2} - 1 \tag{3.17}$$

and the *a priori* signal-to-noise ratio:

$$R_{prio,b}(e^{j\Omega}) = (1 - \Theta)\max(R_{post,b}(e^{j\Omega}), 0) + \Theta\frac{|G_{b-1}(e^{j\Omega})X_b(e^{j\Omega})|^2}{|N_b(e^{j\Omega})|^2} \qquad (3.18)$$

are defined. G_{b-1} denotes the spectral weightings of the previous signal block, being b the counting index of the block. Θ is a parameter balancing the *a posteriori* signal-to-noise ratio with an estimate of the *a priori* signal-to-noise ratio. Its optimization is discussed in [13]. If we define the function:

$$M[u] = e^{-\frac{u}{2}}\left[(1+u)I_0(\frac{u}{2}) + uI_1(\frac{u}{2})\right], \qquad (3.19)$$

with I_0 and I_1 Bessel functions of the first and second order, then the spectral weights for the block b are:

$$
\begin{aligned}
G_b(e^{j\Omega}) &= \frac{\sqrt{\pi}}{2}\sqrt{\left(\frac{1}{1 + R_{post,b}(e^{j\Omega})}\right)\left(\frac{R_{prio,b}(e^{j\Omega})}{1 + R_{prio,b}(e^{j\Omega})}\right)} \\
&\quad \times M\left[(1 + R_{post,b}(e^{j\Omega}))\left(\frac{R_{prio,b}(e^{j\Omega})}{1 + R_{prio,b}(e^{j\Omega})}\right)\right]. \qquad (3.20)
\end{aligned}
$$

More details and suggestions about the optimal tuning of the Ephraim and Malah filter can be found in [13, 23]. As already said, the comparison between the local projective noise reduction scheme and the Ephraim and Malah filter is really fair, since we know the noise spectrum and therefore Eq.(3.18) gives the real *a priori* signal-to-noise ratio instead of just an estimate. In other words, we are using the Ephraim and Malah filter in its most favourable way.

Fig. 3.11 shows the result of our noise reduction scheme on white noise with variable amplitude, where the numerically generated initial noise and the remaining distortion after noise reduction are plotted together with the clean signal. As long as the variation takes place on time scales which are of the order of or larger than the duration of typical phonemes, the filter has no problem. However, this figure clearly shows that lack of redundancy or deterministic structure in the signal leads to reduced performance of the algorithm: Those instances where the remaining distortion is particularly high correspond to fricatives in the speech signal. The auditory impression of the denoised signal is much more homogeneus than the visual one, since exactly these parts where noise reduction is less successful are noise-like parts of articulated voice signals. Without any changes of the algorithm, also non-random noises such as the 50Hz component of electronic equipment have been succesfully suppressed.
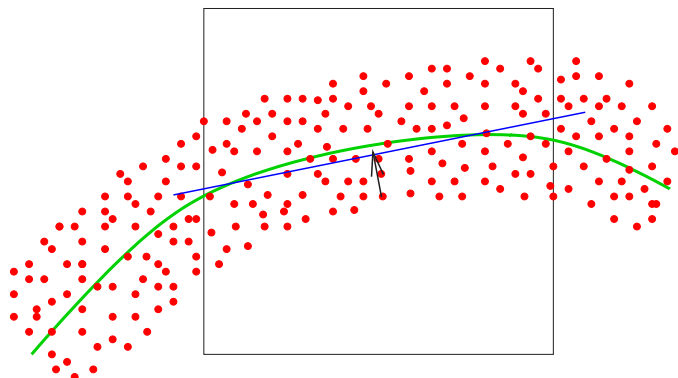
Figure 3.12: How the local projective noise reduction scheme works. Schematic representation. The original manifold (curved line) has been destroyed by the contamination with noise, getting the resulting cloud of points. In order to remove noise from the point at the base of the arrow, one has to collect neighbours of this point in a proper sub-region of the embedding space (square box), then identify a local approximation of the original attractor (straight line) and finally perform the projection onto this manifold (arrow).

3.5   Optimizing of Recurrence Plots for Noise Reduction

When looking for neighbours, we have to restrict our search to a subset of the embedding space. The size of it plays a crucial role. In this section we want to address an optimization problem, namely we want to provide a mechanism that is able to automatically identify the best neighbourhood size, one of the most important parameter of the filter. The two limit cases are trivial: *(i)* A very small value will provide no neighbours but the point itself and therefore the filter will produce no effect to the time series; *(ii)* A very big value will identify all the points as possible neighbours and the algorithm will perform just a global averaging, destroing completely the original voice.

From a computational point of view, the smaller the size of the neighbourhood is, the faster runs the program. But there is a lower limit for the size of the subspace, given by the noise level. As depicted in Fig. 3.12, the diameter of the neighbourhood has to be bigger of the size of the cloud of points contaminated by the noise. The point at the base of the arrow represents the actual point to be filtered. The square is the subset of the embedding space where we look for neighbours. The bold curve represents the original attractor and the cloud of points is the effect of the contamination with noise. The straight line is what the algorithm can identify as
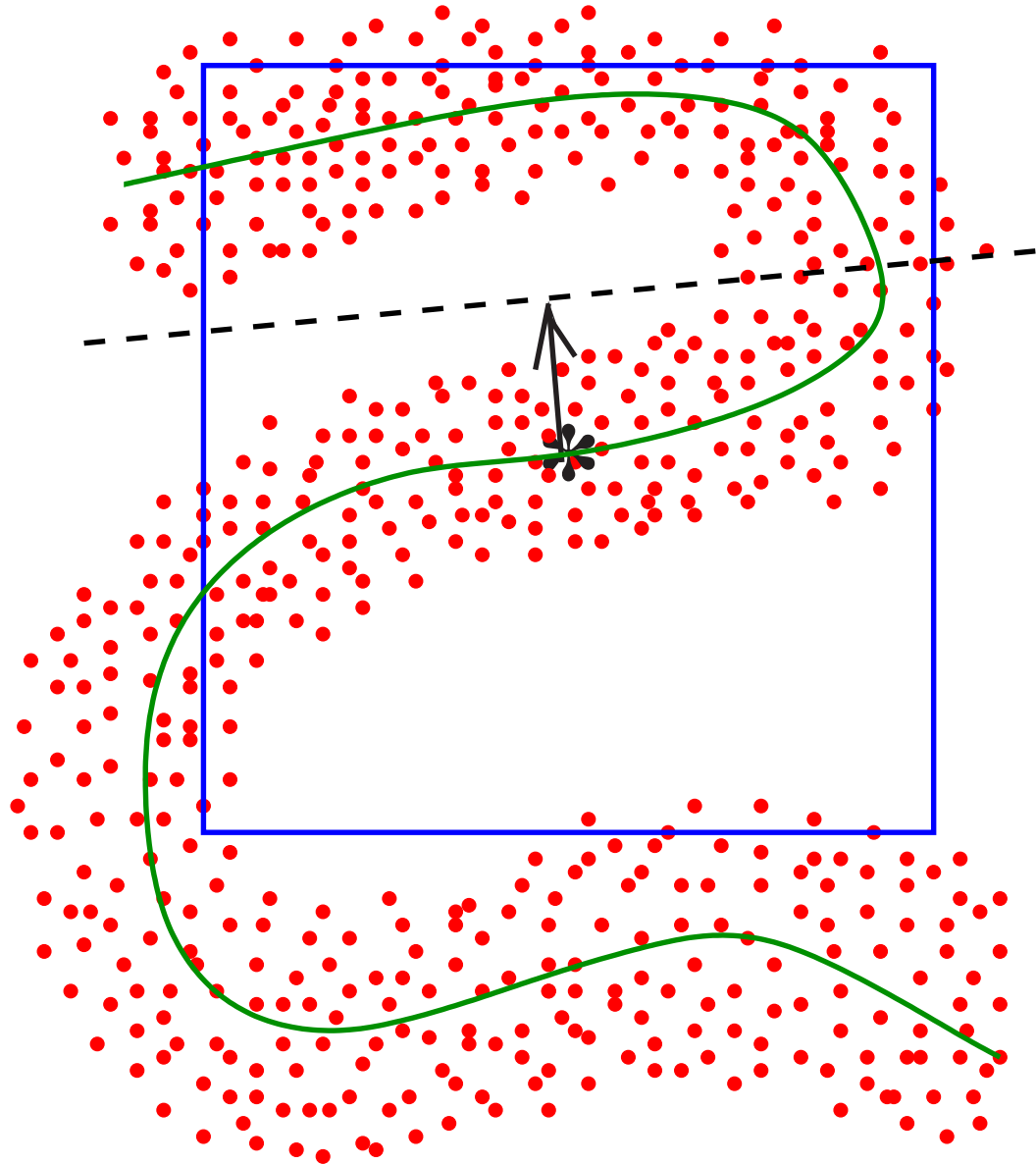
Figure 3.13: Effect of a too big neighbourhood. The algorithm is not able to correctly identify the manifold because two different branches of the original attractor are erroneously considered as neighbours. Therefore the quality of the filtering can only be bad because the local approximation (dashed line) can be considered as a low pass filtered version of the attractor. In this case one has collected too many neighbours, including false ones.
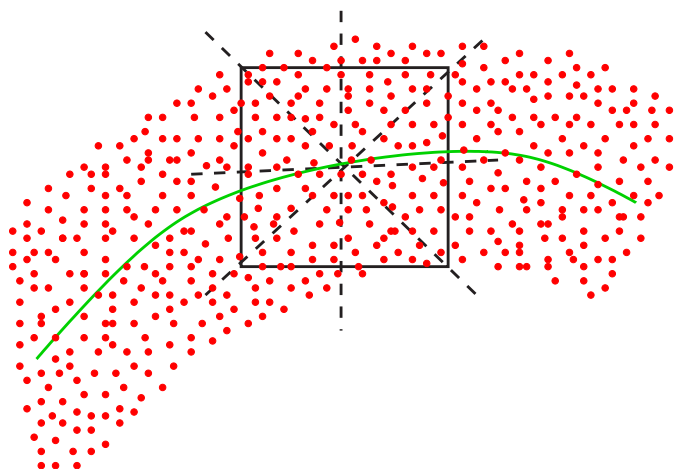
Figure 3.14: Effect of a too small neighbourhood. The algorithm is not able to correctly identify the original manifold because the distribution of points inside the considered subspace is almost uniform. Therefore the quality of the filtering is bad, again comparable to a low-pass filtering.

the original manifold starting from the knowledge of this cloud. The arrow indicates the effect of the filter, namely the projection of the actual point onto the local linear reconstruction of the attractor.

The effect of a too big neighbourhood is reported in Fig. 3.13, where the identification of the original manifold cannot be correctly performed and therefore the projection of the actual point does not act along the proper direction. This happens because of the minimization procedure: We have to find a local linear approximation of the attractor which minimizes the sum of distances from the noisy points. With such a neighbourhood we consider too many points, included false ones (because belonging to another branch of the attractor) and the resulting manifold is far away from the correct one. Also in the case depicted in Fig. 3.14 a clear identification of the original manifold is not possible. Here the solution of the problem is not unique, due to the fact that the size of the neighbourhood is too small and the points are distributed almost uniformly.

It is thus evident that one needs a mechanism to decide which is the best size of the neighbourhood to be taken into consideration. As a further example let us have a look at Fig. 3.15: Here we consider a whistle, one of the simplest acoustic signals that an human being can generate. The signal is almost sinusoidal and therefore,
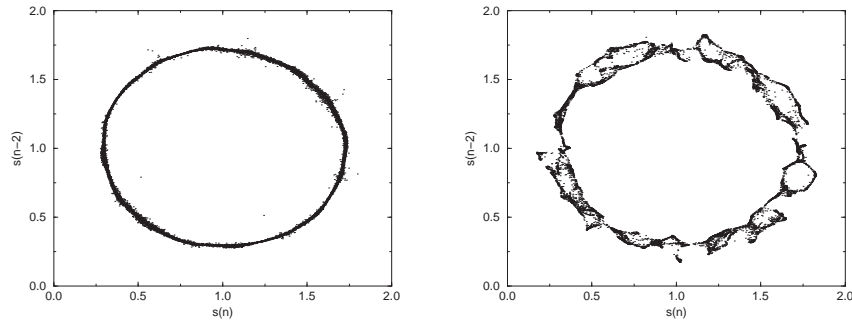
Figure 3.15: Effect of a too small neighbourhood. Left panel: Attractor of a clean whistle. Right panel: Attractor of the same whistle after the contamination with a 30% additive noise and the filtering with a wrong set of parameters, namely with a too small value of the neighbourhood size. The distribution of neighbours is almost uniform and therefore a clear identification of the local projection to be performed is not possible. The result is the creation of completely artificial structures due to averaging effects and low-pass filtering.

with the proper parameters, the attractor looks like a circle. The left panel of Fig. 3.15 refers to the reconstructed embedding space related to this signal. We proceed now adding a 30% noise to the whistle and filtering the new signal with a wrong set of parameters, namely with a too small value of the neighbourhood size. The right panel of Fig. 3.15 shows how the reconstructed attractor looks like. We do not report the picture of the attractor after the correct filtering, since within the resolution of this paper it would be almost indistinguishable from the original. The reason of the strange shape of the right panel is the following: Once considered the embedding space for the noisy signal, we look for neighbours in such small regions that the distributions of points within them is almost uniform; therefore we are not able to identify the original manifold and we perform local projections onto wrong (because essentially random) directions.

In order to develop the optimized scheme, we proceed by adding noise to a phoneme, as illustrated in Fig. 3.16. Starting from the upper panel we have the original time series plus 0%, 30% and 50% of noise. The voice was recorded with a 20 kHz sampling rate, so that the 2000 points correspond to 100 ms. Starting from the recurrence plot of the noise-free time series (reported in the middle part of Fig. 3.17), let us define the following quantities:
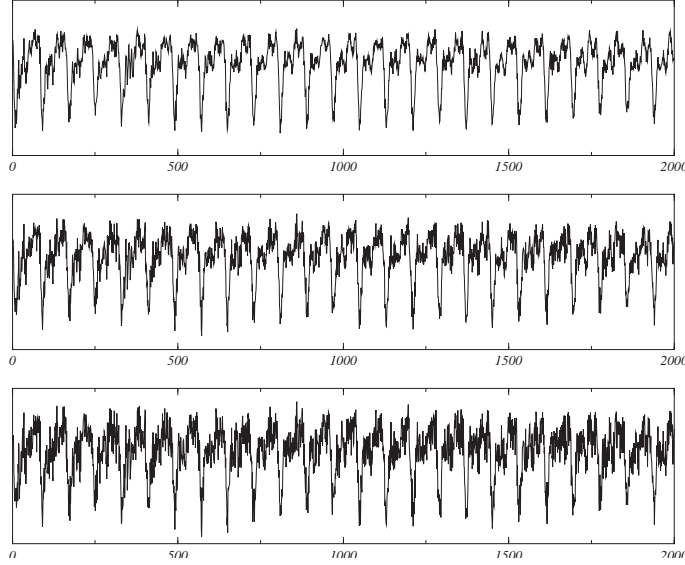
Figure 3.16: Example of a single phoneme (2000 points $\approx$ 100 ms). Uppen panel: Original time series, without noise. Middle panel: With 30% noise. Lower panel: With 50% noise.

$N_p(\epsilon)$: We compute the histogram along the main diagonal direction

$$h_i = \sum_{k-j=i} r_{jk}, \tag{3.21}$$

where r_{jk} is a point in the recurrence plot and h_i the histogram we get after this computation. We want to count the number of peaks and to be sure that they are sharp. For this reason we compute the average height in the histogram and we define a threshold in the histogram (dashed line in the lower panel of Fig. 3.17) as the average value plus three times the standard deviation. The number of peaks is then given by the number of non-consecutive h_i such that $h_i >$ threshold.

$N_\perp(\epsilon)$: We compute the histogram along the other diagonal, perpendicular to the previous one. Averaging over all the N points we get $N_\perp(\epsilon)$. This quantity can be very easily computed also through $\sum_{i,j} r_{ij}/N$ and in fact it is given by the correlation sum (Eq.(3.24)) times N. Just a technical note concerning the fact that in order to compute $N_\perp(\epsilon)$ we use a slightly different RP, namely a RP where the main diagonal is not present. After having introduced $\beta(\epsilon)$, the reason of such a choice will become clear.
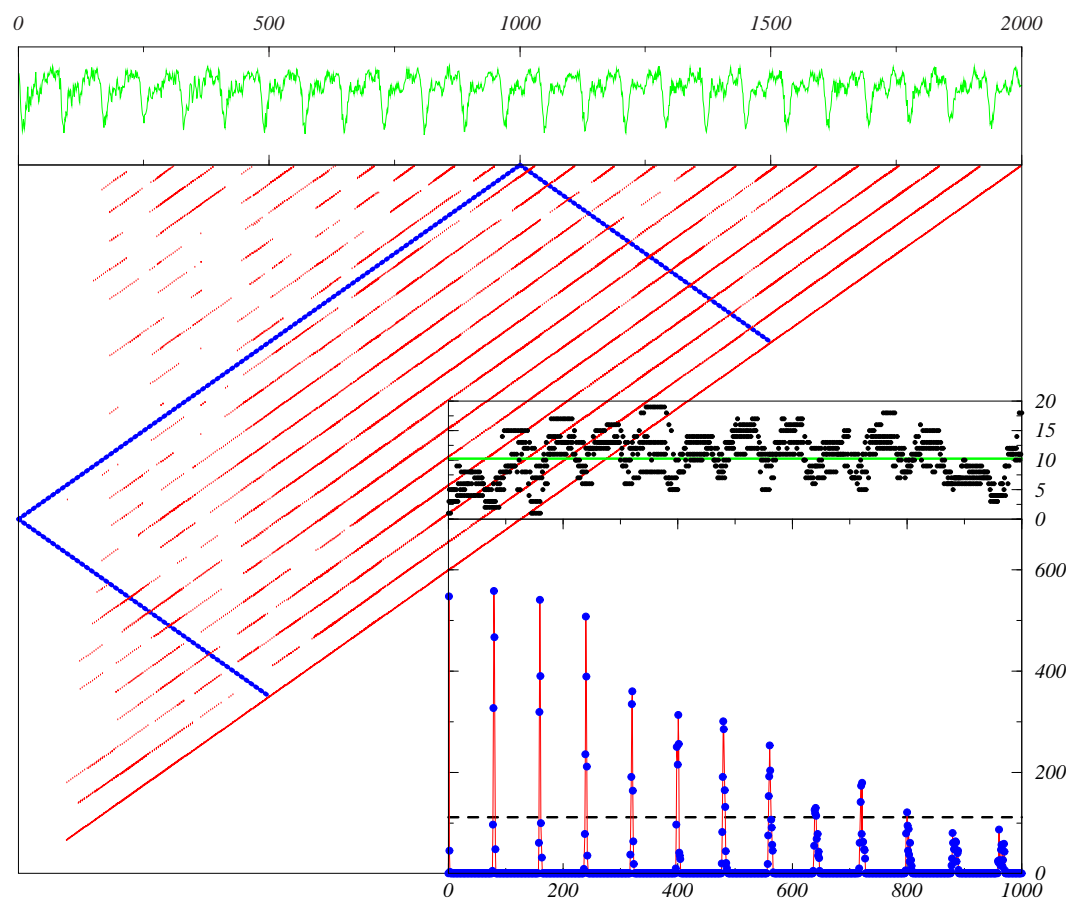
Figure 3.17: Estimation of the parameters entering the β index. Upper panel: Original time series. Middle panel: Recurrence plot with parameters m=25, d=4 and r=0.14 (optimal in this case). Upper part of the lower panel: Distribution of neighbours along the time series and average of them. Lower part of the lower panel: Histogram along the main diagonal of the recurrence plot. To avoid edge effects we have considered only the central part of the time series.
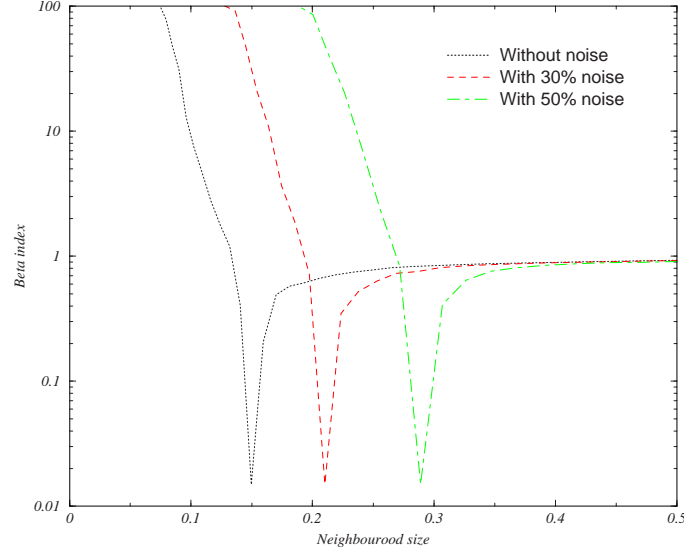
Figure 3.18: Identification of the best ϵ for the three cases of Fig. 3.16. It is very intuitive that increasing the noise level one has to consider bigger and bigger neighbourhoods.

They are both depicted in the bottom-right panel of Fig. 3.17: $N_p(\epsilon)$ is the number of peaks above the threshold (dashed line); $N_\perp(\epsilon)$ is the horizontal line in the upper sub-panel. Of course these two quantities depend on ϵ: The best value of it is the one that maximizes the number of peaks and produces a value of $N_\perp(\epsilon)$ as close as possible to $N_p(\epsilon)$. We want to maximize the length of the lines without making them fat, in order to avoid phase identification problems.

So the task is finding the value of ϵ that minimizes:

$$\beta(\epsilon) = \frac{|N_\perp(\epsilon) - N_p(\epsilon)|}{N_\perp(\epsilon)}. \tag{3.22}$$

The purpose of $N_\perp(\epsilon)$ at the denominator of Eq.(3.22) is a normalization one and its introduction is reflected in a different slope of the function $\beta(\epsilon)$. The automatic identification of the best ϵ proceeds in the following way: We plot the index $\beta(\epsilon)$ for different values of ϵ and we select the ϵ corresponding to the minimum $\beta(\epsilon)$. In Fig. 3.18 we can see the result of such a computation for the three different noise levels depicted in Fig. 3.16, namely 0%, 30% and 50%. Not surprisingly, a bigger noise requires a bigger neighbourhood. For small values of ϵ we have almost no point outside the main diagonal; therefore $N_\perp(\epsilon)$ is close to zero (as already said, for the computation of $N_\perp(\epsilon)$ we use a RP without the main diagonal), but this is
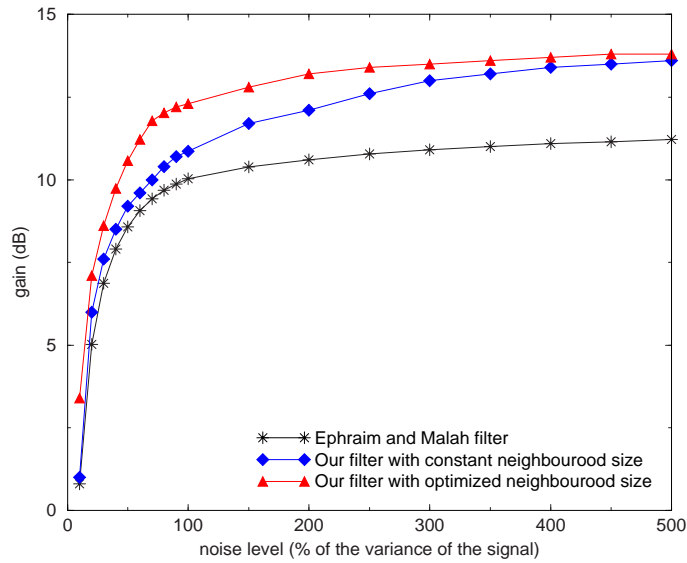
Figure 3.19: Comparison of the optimized local projective noise reduction scheme with the Ephraim and Malah filter, state of the art.

not the case for $N_p(\epsilon)$, since the few points are considered as isolated peaks: Hence $\beta(\epsilon)$ is very big. Big values of ϵ are such that the recurrence plot is almost full of points: $N_\perp(\epsilon)$ is close to N (the length of the time series under observation) and there is no isolated peak because the threshold becomes bigger than N and no line, obviously, can be longer than N. Consequently $\beta(\epsilon) \approx 1$. The optimal situation is when $N_\perp(\epsilon) \approx N_p(\epsilon)$, namely the lines are neither fragmented nor fat: All the recurrence points belong to line structures and the number of them is maximized.

Being applied to segments of the time series, this gives a local estimator of the optimal neighbourood diameters and hence the time dependent noise amplitude can be optimally identified. In Fig. 3.19 we show the performance of the filter compared to the Ephraim and Malah noise reduction scheme. Also the improvement of this optimized ϵ (compared to the case where ϵ is kept constant) is visible. As a measure of performance, we use again the gain in dB, given by Eq.(3.13). In this example the recorded voice is very short (≈ 1 s) and the noise almost stationary; as a consequence the improvement is not so surprising, provided one chooses the best ϵ for the standard algorithm too. The difference becomes more evident if one wants to filter a full sentence, where the noise may affect different words in a very different fashion. Furthermore, the attractor may have branches with different curvature and

where there is the risk of mixing distinct part of it, there the ϵ should be smaller than where the attractor is completely unfolded and regular. It is therefore clear that the use of a constant ϵ cannot be optimal even if the noise level is constant because of the different instantaneous dynamics associated with every phoneme. The new implemented tool contributes efficiently to the goal of getting the best possible quality whitin the shortest possible computational time.

3.6   Analysis of Vocal Disorders in a Feature Space

Leon Glass and Michael Mackey [29] at McGill University in Montreal were among the first to explore the possibility that many medical problems may have their roots in some underlying dynamical effect, the so-calles dynamical diseases. Mathematical models suggest that, when physiological parameters (such as CO_2 transport rates in the blood stream) are changed, processes that normally are rhythmical may be replaced by erratic or chaotic fluctuations. For instance, in some blood diseases the numbers of blood cells show large oscillations that are not normally present. Glass and Mackey demonstrated that simple, but realistic, mathematical models [29] for controlling blood cell production display similar periodic and chaotic oscillations as a particular parameter is varied. The changes in the parameters have themselves a physiological interpretation.

Neurophysiology also offers a wide range of phenomena that are candidates for dynamical diseases, like abnormal oscillations and complex rhythms posing clinical problems. Sometimes, there is a marked oscillation in a neurological control system that does not normally have a rhythm. Examples are ankle tremor in patients with corticospinal tract disease, various movement disorders like Parkinson's tremors and abnormal paroxysmal oscillations in the discharge of neurons that occur in many seizures. Alternatively, there can be a qualitative change in the oscillations within an already rhytmic process, as in abnormality in walking, altered sleep-wake cycles, or rapidly cycling manic depression. Yet again, clinical events may recur in seemingly random fashion, as in seizures in adult epileptics. Neural processes are however so complex that it is not easy to see how models for these dynamical diseases can be developed, tested and fully understood.

Here we want to restrict the analysis of dynamical diseases to vocal disorders exhibited by humans. The normal phonation of the voicing source is basically char-

acterized by almost periodic vibrations, but it can also produce a great variety of complex signals. This is due to the fact that many sources of non-linearity are involved in the air-flow production and in the laryngeal vibration processes, like the pressure-flow relation in the glottis, the stress-strain curves of vocal fold tissues and the vocal fold collisions. These features have been modelled successfully in the past years by an asymmetric two-mass model of the vocal folds [44]. Standard methods of voice analysis, such as the estimation of jitter and shimmer (to be defined later) and the harmonics-to-noise ratio, are valuable for the characterization of regular phonation. However, in the case of voice disorders abrupt changes to irregular regimes occur and these measurements are of limited relevance [48].

The transitions to qualitatively new oscillatory behaviour indicate the suitability of the methods from non-linear dynamics; e.g. in [40] it is shown that a sufficiently large tension imbalance of the left and right vocal fold induces bifurcations to chaos. Asymmetries due to paralysis, polyps, papilloma, cancer etc. produce the same effects and simulations of coupled oscillators [95] can provide insight into the sources of such vocal instabilities and are, therefore, of potential use for diagnosis and treatment of voice disorders. Furthermore, the reconstruction of attractors and the estimation of their properties indicate low dimensionality in the systems generating the voice signals. Herzel et al. [39] have shown that indications of secondary Hopf bifurcations can be found in pathological voices, as well as sudden jump from one limit cycle to another one with different period and amplitude. Different attractors may coexist in non-linear systems and, therefore, even extremely tiny changes of parameters, like muscle tension, may lead to abrupt jumps to other regimes. The occurrence of such a situation should be reflected in quantities like the entropy and the fractal dimension of the (global) attractor.

In the two-mass model of the vocal folds, each fold is approximated by two coupled oscillators of masses $m_{1\alpha}$ and $m_{2\alpha}$ respectively, arranged one upon the other. Springs $k_{1\alpha}$, $k_{2\alpha}$ and $k_{c\alpha}$ and dampers $r_{1\alpha}$ and $r_{2\alpha}$ represent the viscoelastic properties of vocal-fold tissue. The elongations $x_{1\alpha}$ and $x_{2\alpha}$ are the time-varying variables of the model ($i = 1, 2$):

$$m_{i\alpha}\ddot{x}_{i\alpha} + r_{i\alpha}\dot{x}_{i\alpha} + k_{i\alpha}x_{i\alpha} + \Theta(-a_i)c_{i\alpha}(a_i/2l) + k_{c\alpha}(x_{i\alpha} - x_{j\alpha}) = F_i. \qquad (3.23)$$

$\Theta(x)$ is such that $\Theta(x) = 1$ if $x > 0$ and $\Theta(x) = 0$ if $x \leq 0$. The forces F_i describe the action of the pressure in the glottis, whose length is indicated with l, the parameters

a_i indicate the rest areas and $c_{1\alpha}$ are additional spring constants during collision. With this model one is able to analyze bifurcations which are closely related to observations in voice pathology. The origin of the instabilities can be traced back to the desynchronization of two oscillators, the left and right fold. For sufficiently large effects of non-linearities related to large subglottal pressure and overcritical detuning of the eigenfrequencies, complex oscillation patterns are found.

Qualitatively, the origin of bifurcations and low-dimensional attractors can be understood as follows: Normal phonation corresponds to an essentially synchronized motion of all vibratory modes. A change of parameters such as muscle tension or localised vocal fold lesions may lead to a desynchronization of certain modes resulting in bifurcations and chaos. The following modes are of particular relevance: Motion of the left and right vocal fold, horizontal and vertical modes, interaction of the ventricular and vocal folds, interaction of vocal fold vibrations with sub- and supra-glottal acoustic resonances, and vortices generated at the glottis. Although normal phonation and voice disorders can be distinguished qualitatively by human very easily, a quantification scheme of the disease is highly desirable. This is motivated by clinical interests, which lie in objectively evaluating the effort made by cordectomised patients during an utterance, as it could be indicative of patient status, also as far as post-operatoty functional recovery is concerned [47, 68].

Before going into the details of geometric signal separation and feature space and therefore of the data classification performed on the voice, it is worth describing the general phenomenon of irregularity in the human voice as it is known in medicin. The following is just a descriptive terminology, without any precise mathematical formulation. In particular, no numbers or physical units of measurement have been attached to them, although in some cases they can be rated psychophysically.

- A **perturbation** is usually thought to be a minor disturbance, or a temporary change, from an expected behaviour. Perturbations are usually such that they do not alter the qualitative appearance of a visual or temporal pattern, at least not indefinitely. They are small irregularities that are for the most part overlooked. An expected circular orbit, which assumes a slightly elliptical shape, is said to be perturbed [92].

- A **fluctuation** suggests a more severe deviation from a pattern. It reflects an inherent instability in the system. Whereas a perturbed system usually

returns to normal since it is attracted to a stable state, a fluctuating system is somewhat out of control and cannot find a stable state. A vocal tremor or vibrato may be described as a fluctuation in fundamental frequency and amplitude; it is a pattern itself, rather than a small deviation from a pattern.

- **Variability** is the ability of someone or something to vary, intentionally or not. In other words, it is the amount of variation as determined by a statistical measure. The concept of variability is strictly related to the concept of reproducibility and therefore of stationarity, since it may cause the final result to be far from what expected.

- **Jitter** and **shimmer** refer to a short-term (cycle-to-cycle) perturbation in the fundamental frequency and amplitude of the voice, respectively. Here the problem of a precise definition arises, since there are many ways of quantifying a deviation from an expected pattern or trend. Unfortunately a solution of this problem has not yet been found, but a shimmering voice is quite easily recognized, since it is percived as a crackling or buzzing sound.

- **Tremor** is a low-frequency fluctuation in amplitude and/or frequency. Its origin is usually neurologic. Physiologic tremors in the body have fluctuation rates in the 0 to 15 Hz range, but not all are perceived the same way auditorily when they are part of the vocal signal. This term is also used by the recording industry to describe variability in the speed of the tape drive of an audio recorder. Anyway, without a small degree of tremor, steady vowel production has a buzzy quality.

- **Flutter** describes the variability associated with tape contact on the recording head. It has been used to describe neurologic fluctuations in the 9 to 15 Hz range. It appears to be associated with rapid onset and offset of phonation, reflecting the natural oscillating frequency of the control system in phonation. There is something about a low frequency fluctuation in the voice that makes it warm and acceptable.

- A **creaky voice** sounds like a creaking door, like two hard surfaces rubbing against each other. Acoustically a complex pattern of subharmonics and modulations is observed that reflect a complexity of modes of vibration of the vocal folds.

- **Biphonia** is a phonation with two independent pitches. There are essentially two non-commensurate fundamental frequencies, which can appear as non-parallel harmonic lines in a spectrogram as either or both pitches change. The lines may be parallel but not rationally dependent.

- **Roughness** is a kind of pathologic voice, which refers to an uneven and bumpy quality. It results from irregularity in the energy contained in a critical band of the auditory system. Periodic sounds like a vocal fry can have roughness, but more often there is a lack of periodicity. Acoustically the waveform is often aperiodic with the modes of vibration lacking synchrony, but voices with subharmonics can also be perceived as rough.

- **Breathiness** is a vocal quality that contains the sound of breathing during phonation, with particular reference to expiration. Acoustically there is a significant component of noise in the signal due to glottal air turbulence. A breathy voice has most of its energy in the fundamental. In hyperfunctional breathiness, air leakage may occur in various places along the glottis, whereas in normal voice air leakage is usually at the vocal processes.

- **Hoarseness** is a vocal pathology that combines the effect of roughness and breathiness together. It is not clearly distinguishable from the previous two cases, since the percentage of the two effects is variable.

After this small and incomplete list of some common irregularities in the human voice, it is not hasty to assert that our voice reveals a lot of information about us, who we are and how we feel, giving considerable insight into the structure and function of certain parts of the body. This is the reason why it is worth now addressing the problem of the classification of vocal disorders for clinical applications, introducing a proper feature space. The idea guiding us in this purpose is the following [98]:

- Extract a significant quantity from the whole time series, somehow compressing a very long vector into a scalar.

- Repeat it for N different quantities measuring different characteristics: From a time series we get therefore N scalars = 1 feature (small) vector.

- Define a space having these quantities as components. In such a space a feature vector is represented as a point.

- Repeat it for M different time series (M time series $\rightarrow M$ points $\in R^N$).

This concept has been used for several purposes, mainly in order to identify the dynamical state of a complex system, where for practical reasons the system itself is connected to a simple measurement device which records a scalar time series. Sufficiently long subsections of this series are transformed into feature vectors v in a feature space V. The entries of v are chosen to be quantities which contain the compressed information of the signal relevant for the task to be performed and which can be estimated directly from the time series. Neighbourhood relations in this space, based on the computation of distances between feature vectors, allow for various classification and diagnosis tasks. It can be necessary to introduce a local metric, since the variability of quantities may range over very different scales. During a training period pre-classified (e.g. by human experts) data sets are collected and each one is converted into a feature vector. This set of vectors is divided into clusters, where each cluster represents a dynamical state of the system. In the data classification period feature vectors are calculated and compared to the clusters in V. The distance of each test vector to the closest cluster is thresholded to yield a distinction between classes.

The basic idea of using a feature space for our purpose is therefore to eliminate the short-time variability of the time series, extracting characteristic features. As a first step for the classification of vocal disorders, one has to select suitable entries of the feature vectors v. These have to contain extremely condensed information from the time series, reflecting somehow a pseudo-state of the dynamical system that has produced the sentence. In other words, healthy patients and sick ones should be associated to feature vectors that populate different regions of the feature space.

We have already said about the local low dimensionality of the voice and the bifurcation scenario leading to chaos as a consequence of an asymmetry in the two mass model of vocal folds. The fractal dimension of the attractor is then a good entry of the feature space, since healthy people should produce smaller dimension-values than patients with some kind of disease. There are several ways to quantify the self-similarity of a geometrical object by a dimension. From a computational point of view it is convenient to proceed in the way illustrated in [46]. Let us define the correlation sum for a collection of points x_n in some vector space to be the fraction of all possible pairs of points which are closer than a given distance ϵ in a

particular norm. The basic formula is:

$$C(\epsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta(\epsilon - ||x_i - x_j||), \qquad (3.24)$$

where Θ is the Heaviside step function. The sum just counts the pairs (x_i, x_j) whose distance is smaller than ϵ. In the limit of an infinite amount of data $(N \to \infty)$ and for small ϵ, we expect C to scale like a power law, and we can define the correlation dimension D by:

$$D = \lim_{\epsilon \to 0} \lim_{N \to \infty} \frac{\partial \ln C(\epsilon, N)}{\partial \ln \epsilon}. \qquad (3.25)$$

Since in the latter there are two limits involved, and both limits are not computable in a closed form, one has to look very carefully to the results before claiming some numbers as the value of the correlation dimension. It should be clear that one needs a lot of points in order to estimate $C(\epsilon)$ over a large enough range of length scales. A few hundred are definitely not enough to yield a statistically significant result at the small length scales, where also the noise starts to play a noteworthy role. With our time series a clear identification of a dimension was not always possible; this is essentially the reason why we will introduce a pseudo-dimension or dimension-like quantity[2]. It is important to remember a result reported in [46]. In theory, the maximum dimension D_M that can be calculated for a data file of length N is:

$$D_M \approx 2 \log_{10}(N). \qquad (3.26)$$

Therefore, approximately eight dimensions is the maximum which can be calculated from a 22 KHz sampled sentence lasting one second.

One other candidate for the feature space is represented by the entropy, a fundamental concept in statistical mechanics and thermodynamics. Entropy describes the amount of disorder in the system, but one can generalise this concept to characterise the amount of information stored in more general probability distributions. Let us introduce a partition $\mathcal{P}_\epsilon$ on the dynamical range of the observable, and the joint probability $p_{i_1, i_2, \ldots, i_m}$ that at an arbitrary time n the observable falls into the interval I_{i_1}, at time $(n+1)$ it falls into interval I_{i_2} and so on. Then one defines

---

[2]In order to be correctly defined, the concept of dimension requires the stationarity of the data. It is of course not the case for a human voice time series.

block entropies of block size m and partition radius ϵ the quantity:

$$H_q(m, \epsilon) = \frac{1}{1-q} \ln \sum_{i_1, i_2, \ldots, i_m} p_{i_1, i_2, \ldots, i_m}^q.$$ (3.27)

The order-q entropies are then:

$$h_q = \sup_{\mathcal{P}_\epsilon} \lim_{m \to \infty} \frac{1}{m} H_q(m, \epsilon) = \sup_{\mathcal{P}_\epsilon} \lim_{m \to \infty} (H_q(m+1, \epsilon) - H_q(m, \epsilon)).$$ (3.28)

In the strict sense only h_1 is called the Kolmogorov-Sinai entropy [50, 100], but in fact all order-q entropies computed on the joint probabilities are entropies in the spirit of Kolmogorov and Sinai, who were the first to consider correlations in time in information theory. Due to the numerical problems encountered in the estimation of the entropy from real data, namely the finite length of the time series and the presence of noise, we will use a pseudo-entropy[3] in the following. The full feature vector contains the following quantities:

- **Spectral Factor:** It is the averaged ratio between the amplitude of frequencies under 1 KHz and frequencies between 4 and 6 KHz; it is motivated by the effort sick subjects have to face when they want to speak; this induces instabilities such that sick subjects should present a smaller value of this quantity.

- **Pseudo-Entropy:** The quantity h_2, as defined in Eq.(3.28), is averaged for ϵ-values of 5% to 10% of the variance of the data for embedding dimensions ranging between 2 and 8, upper limit suggested from Eq.(3.26). Sick subjects should present a bigger value than healthy people.

- **Pseudo-Correlation Dimension:** The quantity D, as defined in Eq.(3.25), is averaged for ϵ-value of 5% to 10% of the variance of data for embedding dimensions ranging between 2 and 8. Sick subjects should present a bigger value than healthy people.

- **First zero-crossing of the Autocorrelation Function:** This parameter is related to the ability of the subject in correctly pronouncing a word. In particular, dysphonic patients are not able to isolate every vowel, and the resulting time series is more correlated than for healthy subjects. The estimation method is the same as in [46].

---

[3]The concept of entropy requires the stationarity of the data in order to be correctly defined, similarly to the correlation dimension.

- **First Lyapunov Exponent:** This is a convenient indicator of the sensitivity to small orbit perturbations characteristic of chaotic attractors, as it gives the average exponential rate of divergence of infinitesimally nearby initial conditions (see Eq.(2.17)). Some sicknesses can induce sudden jumps from the limit cycle (to which a zero maximum Lyapunov exponent is associated) to another one with different period and amplitude (but again with zero maximum Lyapunov exponent); if the jump is due to a bifurcation one can see a positive value. Estimating this quantity, anyway, one has to be careful because Lyapunov exponents for speech data are subject of a great debate[4].

- **Prediction Error:** We apply the idea presented in [26], namely we expect to get a small error when an attractor is present. Normal phonations were found indeed to lie with a good approximation on a limit cycle, while sick people commonly produce more disordered time series that are therefore more difficult to predict.

- **Jitter:** As in [58], to take into account the short-term (cycle-to-cycle) variation in the fundamental frequency of the signal. It reflects the internal noises of the human body. Sources of fundamental frequency micro-perturbation are biomechanical, aerodynamic and neurological [108]. Commonly, for healthy voices, the jitter is lower than 1%, while higher values indicate disphonic phonation. In order to compute it, the time series is divided into overlapping segments (usually 2048 points with an overlap of 1024 points, so that the $(k+2)$-th segment starts exactly at the end of segment k) and for each segment the fundamental frequency is estimated. The jitter is the standard deviation of the fundamental frequency distrubution.

- **Shimmer:** As in [40], to take into account the short-term (cycle-to-cycle) variation in the amplitude of the signal. The entity of sickness is somehow

---

[4]Kumar and Mullik [53] found the maximum exponent to be positive, characteristic of chaos, for normal vowel and consonant productions. In contrast, Narayanan and Alwan [87] calculated a maximum exponent of zero, indicating a limit cycle, for normal vowel phonation and a positive value for voiced and voiceless fricative productions. Herzel [40] found a maximum exponent of zero for healthy vowel phonation, while a dysphonic vowel sample yielded a positive value. All authors cautioned, however, that calculation of the exponent is highly sensitive to short data sets and non-stationarity. In fact, critical sensitivity to noise [99] makes calculation of a Lyapunov exponent highly suspect and inconclusive for voice data.

proportional to the shimmer. The computation of the shimmer involves the same segmentation used for the jitter, but instead of estimating the fundamental frequency, here the power carried out by the fundamental frequency is computed. The shimmer is then estimated as standard deviation of this data distribution.

- **Peak in the Phoneme Transition:** The transition between one phoneme and the following shows a much longer transient for sick people, reflecting the problems they encounter when having to switch from a dynamical regime to a different one. Every phoneme contains a pitch that is repeated a number of time variable between 10 and 20 (see [36, 37]). One has then to identify the time length of such a pitch and to compare the distance between this pattern and one of the same length coming from the same phoneme. Moving the second pattern along the full phoneme, one gets a distribution of distances (with a zero when the reference pitch is considered and saturation values close to the end of the actual phoneme); the maximum peak of this distribution before the saturation, i.e. before the end of the current phoneme, gives what we call Phoneme Trasition. Although comparable to the harmonics-to-noise ratio, we prefer to estimate the Peak in the Phoneme Transition because it looks more reliable and stable. Sick subjects present bigger values than healthy people.

- **Residual Noise:** The algorithm presented in [36] is applied to the time series. Noise can be easily removed if the series contains redundancy and, roughly speaking, redundancy is a synonym of health[5]. After applying the noise reduction algorithm, we look at the variance of the difference between the original and the processed signal. Healthy subjects are related with small values, sentences spoken by sick people look very noisy and therefore present a bigger value of the residual noise after the filtering.

This feature vector is redundant, but this is something somehow unavoidable, since all the entries have to come from the same time series and in this meaning they must be more or less correlated. Furthermore, working in a larger dimensional

---

[5]In [47, 68] the Normalized Noise Energy is introduced to measure the dysphonic component of the voice spectrum related to the total signal energy. The deviation from periodicity in the sub-phoneme structures, due to the dysphonia, can be interpreted as noise and then a way to quantify this additive component is preposed.

feature space helps to better identify the pathology, since the distances between points become larger. The *Principal Component Analysis* method is very helpful in detecting the trade-off between redundancy and robustness [42, 110].

We now proceed to collect male voice samples from three categories of subjects[6]:

- People suffering from dysphonia[7] (12).

- Healthy people (17).

- People with pathologies under medical treatment[8] (4).

All these subjects have been asked to say the italian word *aiuole* (flower-beds), as it is made up with the five main Italian vowel sounds: 'a', 'e', 'i', 'o', 'u'. Every time the word was uttered in isolation. The use of a complete word instead of sustained vowels is due to the clinical interest in evaluating the effort made by the patient during the entire vocal emission, also as far as the glide between vowels ('ai', 'iu', 'uo') is concerned. The sentences have been recorded in a **.wav** mono, 16-bits linear, 22050 Hz format converted into sequences of real numbers[9] and given to the feature vector building algorithm, which always treats them as whole words.

A two-dimensional projection of the feature space is visible in Fig. 3.20, namely onto the pseudo-correlation dimension and the first zero-crossing of the autocorrelation function. A principal component analysis has revealed it to be one of the clearest bidimensional projections. Others provide slightly better separations, but since they have combinations of features as axes we prefer to present the results as in Fig. 3.20. As expected, a normal voice does not exceed the third dimension, while up to 5 degrees of freedom are necessary to represent a pathologic phonation. A similar sharp distinction is given by the autocorrelation function: Dysphonic subjects are not able to well isolate every vowel and the resulting time series is more correlated, i.e. the first zero-crossing is reached after 1.2 ms (typical values for healthy subjects range between 0.2 ms and 0.4 ms). The benefit of the feature space analysis with re-

---

[6]These voice samples were recorded in a quiet room at the Phoniatric Section of the Otholaryngoiatric Institute, Careggi Hospital, Firenze.

[7]Adult subjects affected by T1A glottis cancer, a tumour confined to the glottis region with mobility of the vocal cords.

[8]Subjects operated via endoscopic laser or traditional lancet technique.

[9]Every voice sample is slightly shorter than 1s; this means that the corresponding time series contains about 20000 points
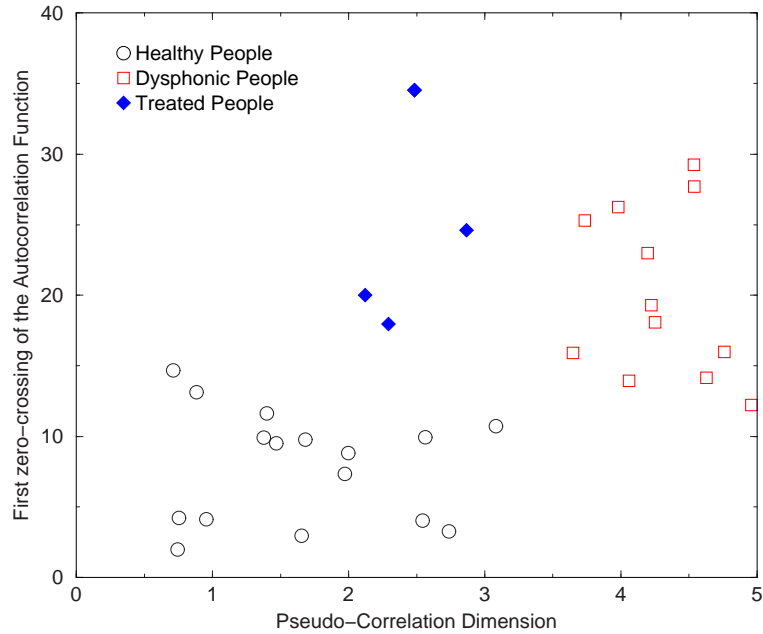
Figure 3.20: Two-dimensional projection of the feature space onto the third and fourth components of the feature vector.

spect to a simple classification by, say, the first zero of the autocorrelation function, is also evident in Fig. 3.20: The chance for misclassification is much lower.

Similar results are readable through the entropy as far as sick people produce more disordered series; the prediction error is smaller for normal phonations because in the phase space they lie with a good approximation in a limit cycle. Also the filtering of the time series is easier when no disease is there; the absence of bifurcations facilitates the search of neighbours and improves the quality of the noise reduction [36]. In order to check the classification ability of the method, we have collected two more sets of data:

- Healthy people simulating a disease (17).

- Artificial voices (12).

In the first set, people have tried to say *aiuole* in the strangest possible way (getting sometimes very impressive records!), simulating hoarseness. They could listen as many time as they wanted to sentences spoken by diseased speakers and they were given the instruction to imitate them as close as possible. In the second
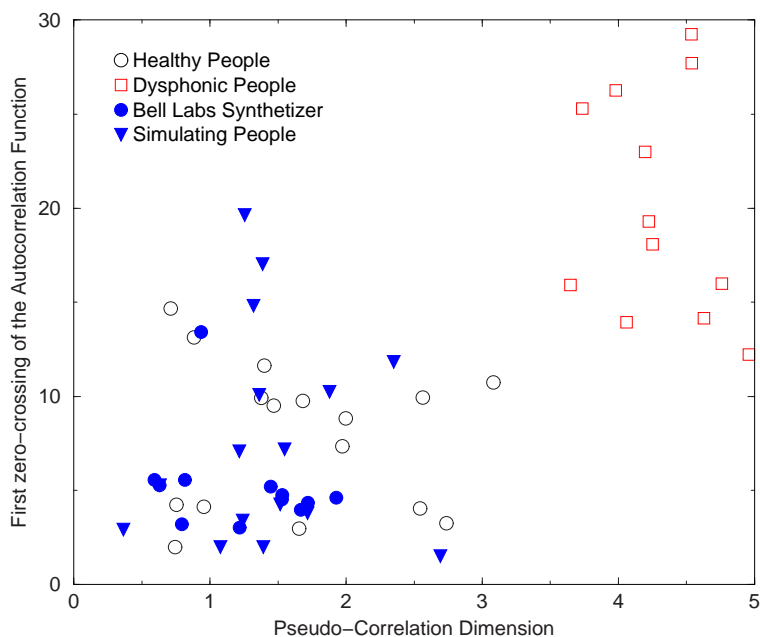
Figure 3.21: Two-dimensional projection of the feature space onto the pseudo-correlation dimension and the first zero-crossing of the autocorrelation function.

set we have used a speech synthetizer with different languages and several voices (man, gnat, raspy, woman, coffee drinker, ridiculous, child, big man). Fig. 3.21 shows the results. It is very interesting to note that, although the simulated and the diseased sentences sound quite similar, they are correctly classified by the algorithm. In particular nobody was able to exceed the dimension three, since the dysphonia is something that one cannot directly control: Just think about the two mass model and the impossibility to directly control and impose the asymmetry. Less amazingly, artificial voices lie inside the healthy zone, but they also sound somehow less rich and, so to say, artificial.

Fig. 3.20 shows also the results concerning the set of voices coming from cordectomised patients. They lie between the healthy and the sick region, with small pseudo-correlation dimension values but large first zero-crossing of the autocorrelation ones. This result, though preliminary, shows that cordectomised subjects still require more effort in speaking with respect to healthy people, due to the limited extension of the produced scar fold. More results could be obtained if the long-time path followed by patients after treatment were available.
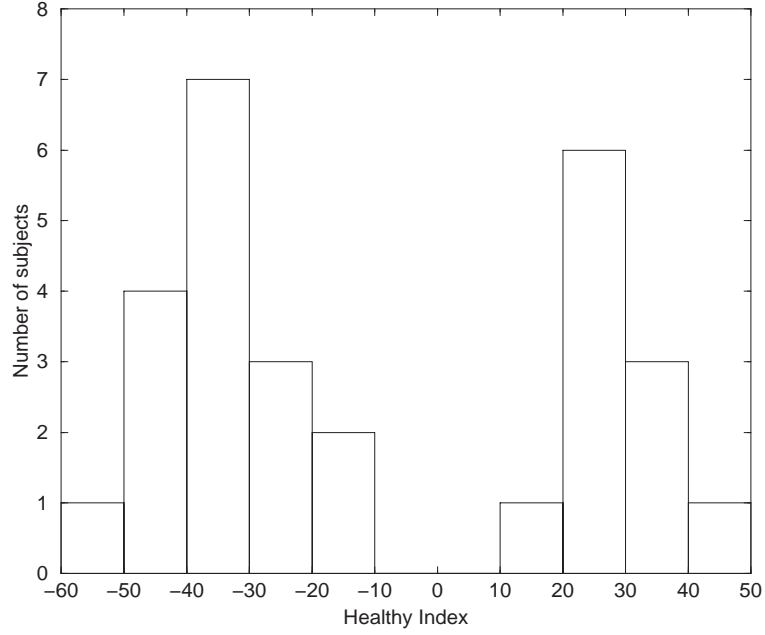
Figure 3.22: Distribution of healthy and sick people according to the healthy index.

In order to quantify the degree of illness, we define the following index. First, the centre of mass of both the healthy ($\overline{healthy}$) and pathologic clusters ($\overline{sick}$) is computed. Then the distances $d(new, \overline{healthy})$ and $d(new, \overline{sick})$ of the voice sample under test from $\overline{healthy}$ and $\overline{sick}$ respectively are evaluated. Due to the different range covered by the different components of the feature vector, a weighted distance is considered, where the weights are the inverse of the standard deviations of the distribution of the entries. The **healthy index** $\mathcal{H}$ is defined as:

$$\mathcal{H}(new) = 20 \log_{10} \frac{d(new, \overline{healthy})}{d(new, \overline{sick})}. \qquad (3.29)$$

A strong negative value of $\mathcal{H}$ indicates a healthy voice, while a big positive one reveals the presence of some kind of pathology. Of course it is necessary to introduce some thresholds to get a good classification, according to clinical considerations.

We have computed the healthy index $\mathcal{H}$ for the full set of data, getting the distribution for healthy and sick people shown in Fig. 3.22. The peak on the left is relative to normal phonation, while the right one regards only disphonic voices. Cordectomised patients got an index value ranging between -7 and 4, artificially generated voices between -77 and -45, healthy people simulating a disease are situated

between -53 and -5. It is again interesting to note how difficult (if not impossible) it is for healthy people to simulate a real disease, since it operates directly into the physiologic level.

Notice that the healthy index is based on all the entries of the feature vector. Reducing the number of features causes a smoothed distribution of voices, with a worse classification. Since in this application the computation time was not a problem (the word *aiuole* is very short and the total number of samples is very limited), we did not put too much effort in determining up to which extent we can neglect some features. As already discussed, some of them play a more important role than some other and this aspect needs to be considered in more details if the speed of the classification becomes a crucial point.

3.7   Software Corrections of Vocal Disorders

We discuss now how vocal disorders can be post-corrected thanks to the local projective noise reduction scheme, with a twofold advantage: *(i)* Physicians can take advantage when perform surgical interventions from the availability of a tool showing graphically the result of their work; *(ii)* Researchers can try to build up devices that can help to improve voice quality avoiding any surgigal treatment. The first simple difference between dysphonic and healthy voices is shown in Fig. 3.23, where the time evolution of the amplitude of a microphone-registered sound is represented. The upper panel could be interpreted as a highly noisy time series, but careful investigations reveal that this is not the case [2]. Therefore the task is very sensitive and delicate and of course applying a simple low-pass filter would only introduce a distortion bigger than the original noise level. Some of the noise-like structures belong to the time series and one has to be able to correctly identify what is worth keeping and what has to be eliminated during the correction procedure [82].

We use the local projective noise reduction algorithm and tune the parameters according to the recipes introduced in Sec. 3.5, the filtered signal sounds more *normal* than the original, even if some characteristic aspect of the voice has been lost. The desired corrections are reported in Fig. 3.24, where the dotted signal is the dysphonic sample and the bold curve represents the time series after noise reduction (that was able to perform what one would expect from a filter, namely correctly remove the noise and keep the signal). In Fig. 3.25 we see a projection of the feature space onto
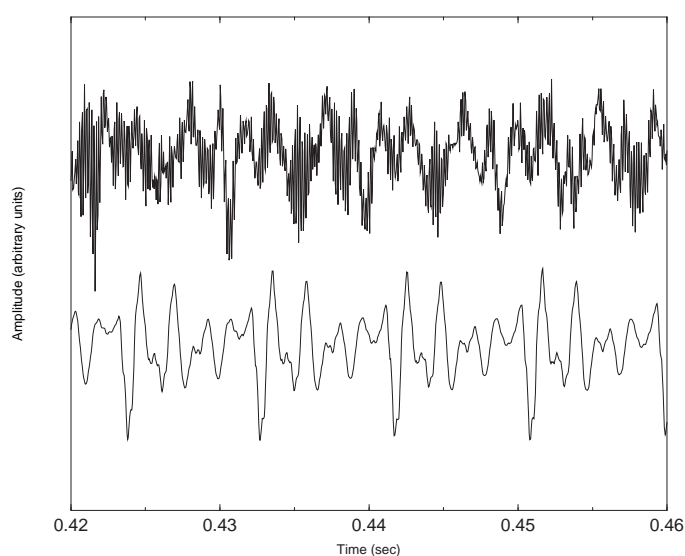
Figure 3.23: Typical shapes of the investigated time series. Upper panel: Dysphonic voice. Lower panel: Healthy sample. The difference looks like noise [2], but removing it completely one deteriorates the signal quality and the voice sample sounds artificial. Part of the depicted structures are due to the unicity of the speaker, other comes from the vocal pathology, one further contribution is just noise.
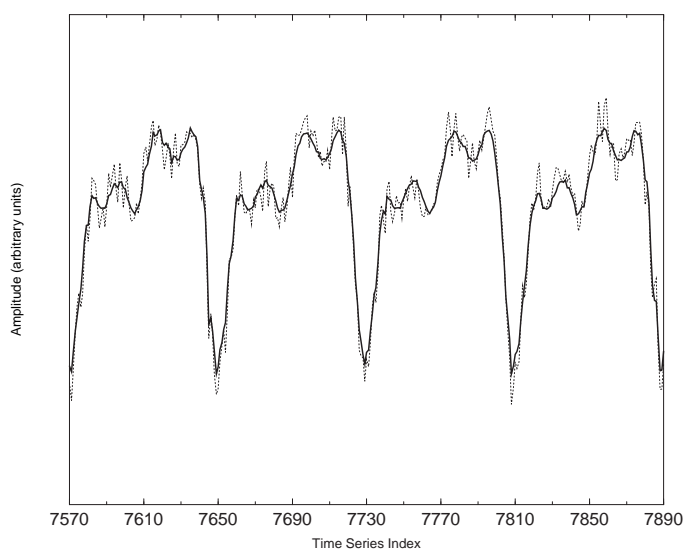

Figure 3.24: Dotted line: Original dysphonic time series. Bold line: After noise reduction. Although it may look like, the task cannot be performed with a low-pass filter, since it would not be able to keep abrupt peaks that are part of the signal. This task is quite sensitive and delicate.
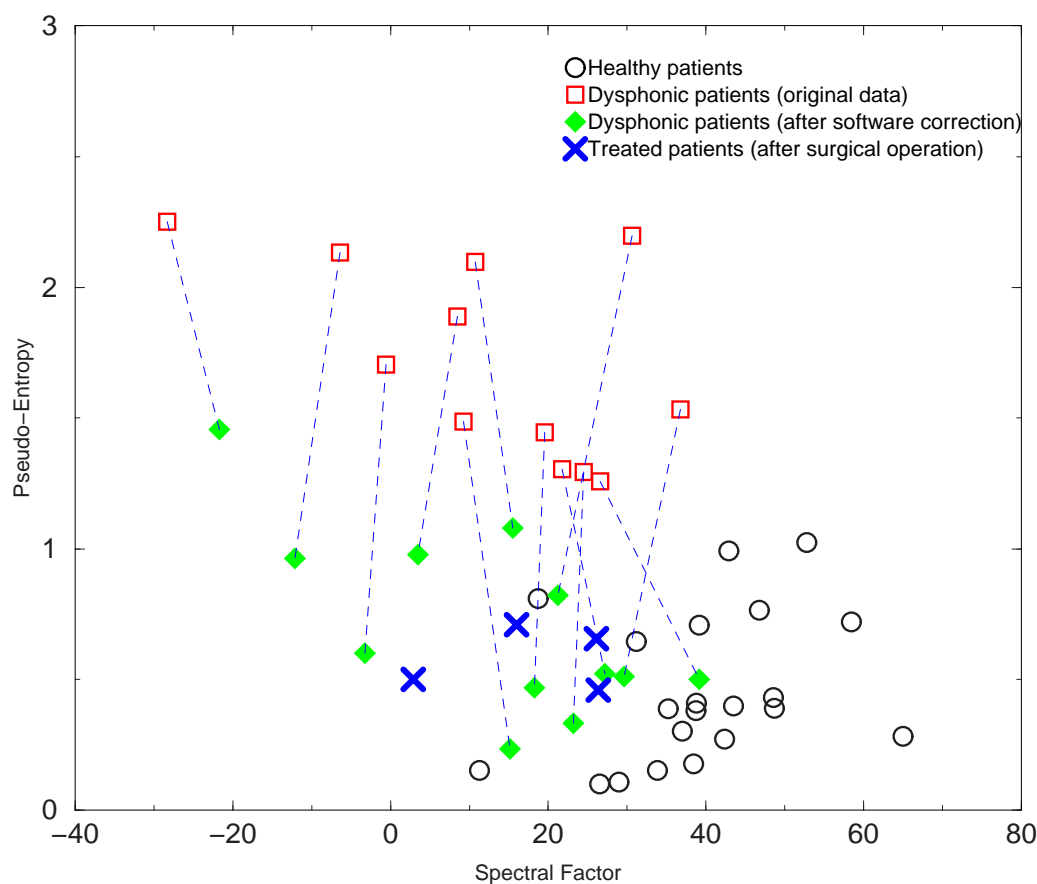
Figure 3.25: Two-dimensional projection of the feature space. Healthy patients are characterized by a pseudo-entropy close to 0.5 and a spectral factor within the 10 to 60 range. Sick patients populate a region of the feature space where the pseudo-entropy has a value close to 2 and the spectral factor ranges between -30 and 30. A bold cross refers to a patient after a surgical operation. The effect of this treatment is a drastic decrease of the pseudo-entropy and a moderate increase in the value of the spectral factor. Dotted lines link the points before the filtering to the same samples after the proposed software corrections. The algorithm produces similar effects of a surgical treatment (indicated with crosses).
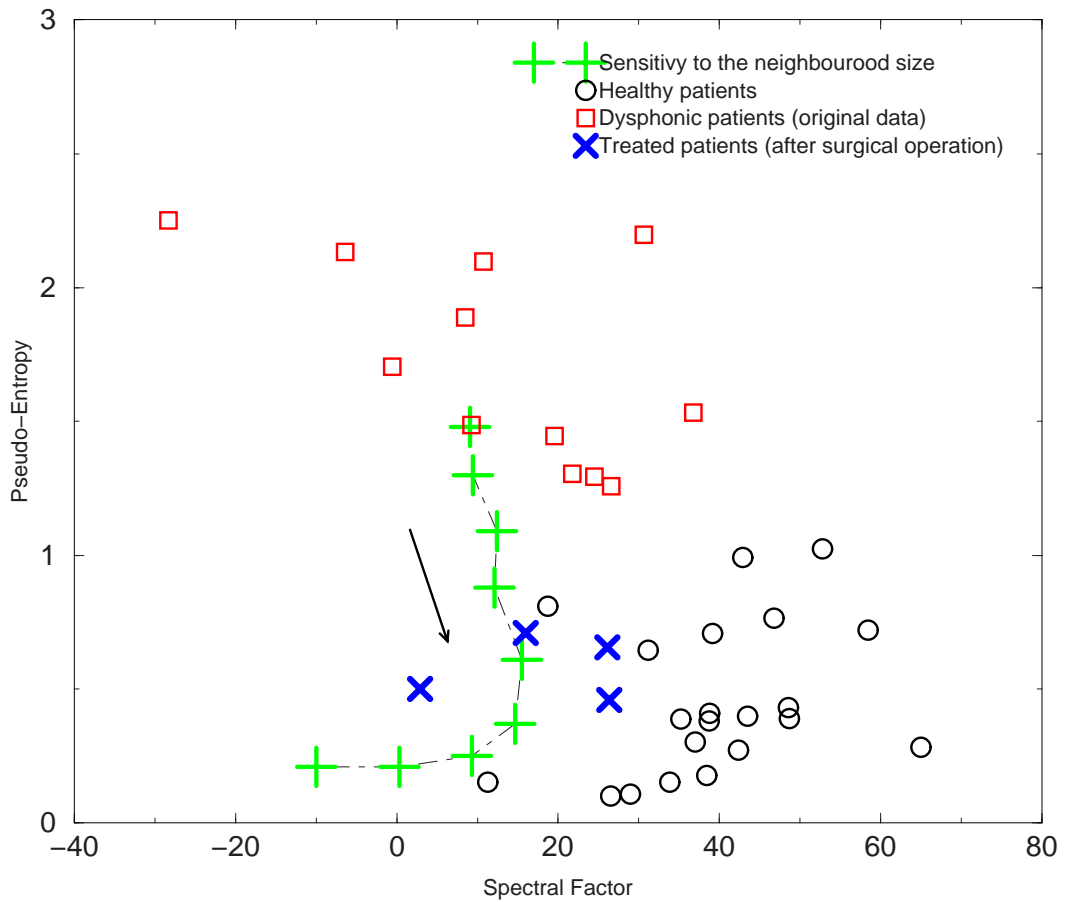
Figure 3.26: Two-dimensional projection of the feature space. The path followed by the bold crosses is the effect of corrections with increasing ϵ (ϵ increases along the direction indicated by the arrow). $\epsilon = 0$ coincides with the original sample, since no neighbour can be identified and therefore no correction is performed. The last three points are the result of a filtering with a too big neighbourhood size. Only for central values of ϵ the corrected time series lies in the healthy region.

the spectral factor and the pseudo-entropy dimension; dotted lines link the points before the filtering to the points after the attenuation of the noise-like features, using the following set of parameters: $m = 30$, $\nu = 4$ and ϵ optimized according to the previously introduced scheme. Dysphonic patients, in the (spectral factor, pseudo-entropy) plane, are spread around the average point (5,1.8); the centre of mass of the healthy cluster is (35,0.5). The surgical treatment produces an average correction located in (20,0.6). Our software moves the centre of mass of sick patients to (15,0.7). In getting these results the value of m and ν is not so crucial, provided that $m\nu > l_p$, being l_p the extension of a pitch. The program is more sensitive to the choice of ϵ, in the following way: The length of the dotted line of Fig. 3.25 is somehow proportional to the value of ϵ. Unfortunately the angle is not constant: This means that up to a certain value of ϵ all the corrections act along the same direction, beyond that threshold they start to deteriorate the voice. In the extreme case of ϵ as big as the full embedding space, the averaging procedure performed by the filter destroyes almost completely the signal, producing a pseudo-entropy and a spectral factor close to zero.

The sensitivity of the program to the choice of ϵ is illustrated in Fig. 3.26. There we have filtered one sample with 8 different values of the neighbourhood size. The original position is the one with the biggest value of the pseudo-entropy. ϵ is increasing along the direction indicated by the arrow and only for two values of it the corrected point lies in the healthy region. For the last three corrections the neighbourhood size was absolutely too big. The other samples behave in a qualitatively similar way.

Chapter 4

Financial Markets

In physics, you're playing against God; in finance, you're playing against people.

(Emanuel Derman)

The stock market is from a physicist's point of view the largest, most well tuned, efficient and well maintained emergence laboratory in the world. With the most dense and precise measurementes performed, recorded, transmitted, stored and documented flawlessly on extremely reliable data bases. Add to this the potential relevance to the most money-saturated human activity in the world and obtain a very promising vast area to exercise our drives for understanding.

As opposed to field theory in which the microscopic "bare" interactions are to be inferred from the emerging dynamics, and to the cosmology where the emerging macroscopic features are unknown at the largest scales, in financial markets, both the microscopic operations and the macroscopic trends are fully documented. The old dream of Boltzmann and Maxwell of following in detail the emergence of macroscopic irreversibility, generic universal laws and collective robust features from microscopic simple elementary interactions can now be fully realized with the help of this immense thermodynamic machine where the Maxwell demons are human size and the (Adam Smith's) invisible hand is more transparent then ever.

The approach carried out by physicists is not superior, just complementary to the one performed by economists. Most of the economists treated the short time fluctuations as noise. In fact when the Olsen Associates started to accumulate their high density market data, some of the people in the finance field expressed surprise at this unexpected interest in recording all that noise. For physicists, the tick-by-

tick individual trade data are the very engine of the entire system, similarly to the way in which the molecular collisions are ultimately the engine for all statistical mechanics or thermodynamic phenomena[1]. The stock market is the ideal space for the strategical opening to a new kind of science, since it offers a perfectly rigorous experimental and theoretical research framework.

The possibility of accessing and processing rather easily huge quantities of data on financial markets opens the path to new methodologies where systematic comparison between theories and real data not only becomes possible, but mandatory. This perspective has spurred the interest of the statistical physics community, with the hope that methods and ideas developed in the past decades to deal with complex systems could also be relevant in Finance. However the existing literature roughly falls into two categories: Either rather abstract books from the mathematical finance community, which are quite difficult to read for people trained in natural science, or more professional books, where the scientific level is usually quite poor. In particular, only few books in this context are discussing the physicists' way of approaching scientific problems, in particular a systematic comparison between theory and empirical results, the art of approximations and the use of intuition [10, 71].

The perception of the financial market as a complex many-body system offers an interesting challenge for testing well-established physical concepts and methods in a new field. If some of these tools can be successfully transferred, improved insight into the underlying mechanisms of the market should be gained. The supporters of this idea expect that this young branch of physics will soon mature into an independent field, which they call econophysics.

4.1   Efficient Market Hypothesis

The market is the place where buyers and sellers meet in order to exchange products, whose price at time t is called the spot price $S(t)$. It is determined by the interplay beetween buyers and sellers in a free-market economy. A special kind of market is the financial one, where money is lent, borrowed and invested in commodities and securities. Commodities are physical things like metal or corn, securities are more virtual objects, like bonds or stocks. Traders are willing to accept a certain amount

---

[1]This point of view is obviously very optimistic: In statistical physics the microscopic collisions can surely be assumed to have no hidden structure, but human decisions do have.

of risk in order to get a greater return than the interest rate granted by banks. The expected gain is the main driving force for the market participants to trade at all and to deliberately tolerate a certain exposure to risk.

A market is called efficient if:

- The participants quickly and comprehensively obtain all information relevant to trading.

- It is liquid. This means that an investor can easily buy or sell a financial product at any time. The more liquid a market is, the more secure it is to invest. The investor knows that he/she can always cash-in the assets. This easy exchange between money and financial products raises the attractivity of the market. On a mature liquid market, the myriad transactions efficiently balance the decision of a single investor so that individual purchases or sales are possible at any time without affecting substantially the asset prices.

- There is low market friction. Market friction is a collective expression for all kinds of trading costs. These include trader provisions, transaction costs, taxes and bid-ask spreads. The sum of these costs is negligible compared with the transaction volume if the market friction is low.

The efficient market hypothesis states that a market with these properties digests the new information so efficiently that all the current information about the market development is at all times completely contained in the present prices. No advantage is gained by taking into account all or part of the previous price evolution. This amounts to a Markov assumption. The rationale behind this hypothesis rests on the following argument: Imagine that a time series exhibits a structure from which the rise of an asset price could be predicted in the near future. Certainly, investors would buy the asset now and sell it later in order to pocket the difference. However, an efficient market immediately responds to the increased demand by increasing the price. The profitable opportunity vanishes due to competition between the many active traders. This argument limits any correlations to a very short time range and advocates the random nature of the time series. One could say even more stating that an efficient financial market is unpredictable by construction and therefore there is no periodically working financial process which generates a risk-free profit from nothing.

It is very interesting to follow the debate of whether a competitive market were efficient or not, since academics, practitioners and general public are divided roughly into two categories: Believers and non-believers of the efficient market hypothesis [114]. Efficient markets are a natural consequence of neoclassical economics, whose founding doctrine is that economics is about optimally allocating scarce resources among competing uses. According to it every hypothetical market inefficiency would be arbitraged away by competitive players whose greed assures that markets stay at least closely to the perfect efficiency. On the other hand, the market, no matter how big, is governed by a small group of people called market-makers, whose behaviour influences the distribution laws of the entire market. According to this alternative point of view, the so-called *free* economy is in fact predefined by properties of a small group of people who are ready to do their best to support its spirit and first principles. In social terminology they are called *bureaucracy* or *government* [5].

4.2 The Liquid Analogy

One of most famous conjecture about the behaviour of financial markets is the Samuelson's paradigm [96], according to which a financial market exhibits a random character and consequently it is unpredictable. It has recently been proved, nevertheless, that a certain degree of correlation is still present on extremely short time scales [60]. Despite that, the intermediate scales are dominated by random behaviour with Lévy stable statistics of asset returns [24, 64]. The possibility to extract information on the future evolution of a single asset by knowing a big enough ensemble of its past values matters indeed institutional traders, who can generally intervene on the market in real time, with delays smaller than few seconds. Their presence reduces at minimum time correlations and consequently speculation possibilities in a way such that the efficient market hypothesis almost holds.

Time dependence is however only one possible domain for surveying similar patterns inside financial signals, another one being the *spatial* domain. In fact, albeit much efforts are spent in studying correlations in the time dynamics of a *single* asset (see [12] and [71] for a digest of the recent economist and physicist approach, respectively), there are many applicative and fundamental reasons for understanding deeply spatial, commonly referred as *multivariate*, correlations. A financial market is not simply the juxtaposition of different prices which are organized on an inde-

pendent basis, but rather a complex system of interacting constituents [25]. The latter are then monitored by sampling single prices with respect to an arbitrary currency. Hence the study of correlations among different asset time signals is of peculiar importance. By the way, this is also the case in many problems involved in the modern risk management theory, where the composition of a certain portfolio strongly depends on the movements of different underlying assets. On a more fundamental level, the interesting issue is the comprehension of how price changes can be separated, with a sufficient degree of confidence, in *single asset–* and *collective–* behaviour [17].

Since the Markowitz's work on the theory of optimal portfolio [75], much effort has been spent to characterize correlation matrices of financial assets [19]. In recent contributions, different physics concepts have been adopted to endeavor this type of problem, mainly because the study of correlations represents a paradigm of a wide class of physical problems for which powerful tools have been developed. A bivariate analysis of the futures on the German and Italian bonds showed that despite the perfect uncorrelation of the single tracks, the crosscorrelation of the two signals was significantly non-zero: The signals considered described two random, but similar, processes [16]. This behaviour emerges quite generally in the stock market, where certain asset clusters *move* in a particularly correlated way with respect to remaining titles. Using equal time cross–correlation matrices and several physics–borrowed tools such as the random matrix theory, these conjectures have been quantified [54]. In a recent study, the structure of a N stock market has been investigated as regarding the multivariate structure in a global window period [72].

We will investigate here asset correlations by interpreting asset growth rates as observables of a particle system scenario. This idea is carried out by introducing a formal map between the logarithmic returns and the distances among gas particles. The strength of this analogy resides in the possibility to separate collective motion from the single asset dynamics through the investigation of mutual interactions among titles. We can study the thermodynamics of the system and interpret its temperature as a measure of spatial volatility, as compared with the more familiar (temporal) volatility. The 2–asset interacting potential is then calculated on the isothermal (isovolatile) market. Finally a time dependent asset–distance and a moving frame model are introduced. The implementation of this scheme is performed on daily stock market data taken among the 30 most capitalized titles forming the

Deutscher **A**ktien inde**X** (**DAX30**) in the period 30 Dec 1987 to 7 Mar 1995 (1800 trading days). To maintain a continuity of quotation, we have selected the maximal subset of 23 assets which, in the above mentioned period, remained in the DAX30 basket and did not perfom any split[2].

As a general starting point, we consider a collection of asset, which is a suitable subpart of titles in a stock market (better if one representative for every economic sector), a collection of currency prices, or any combination of them. The value of the *asset* Ω_i at time t, is expressed in unity of asset Ω_j by means of conversion factors $P_{ij}(t)$:

$$\Omega_i(t) = P_{ij}(t)\Omega_j(t). \tag{4.1}$$

The indices i and j span all N considered assets forming the *market*. By writing Eq.(4.1) for another couple of indices, a no–arbitrage equation for a liquid market is obtained $P_{ij} = P_{ik}P_{kj}$. Its multiplicative symmetry is reflected in a corresponding additive symmetry of the logarithmic returns:

$$d_{ij}^\alpha(t) = \frac{1}{\tau_\alpha} \log\left(\frac{P_{ij}(t)}{P_{ij}(t - \tau_\alpha)}\right), \tag{4.2}$$

where $\tau_{\alpha \leq H}$ is a collection of H time horizons. The rescaling of the log–returns to the considered time horizon is solicited by its interpretation; in the idealized limit of prices with (deterministic) growth laws, we get $P_{ij}(t) \propto \exp(d_{ij}t)$, so that the quantity defined in Eq. (4.2) turns out to be the growth rate between asset i and j, independently on the time horizon. The latter can be considered as a long term limit when one refers —for example— to prices of stocks with respect to currencies. In the opposite limit of extremely small returns (which eventually corresponds to short time lags), d^α is the rate of the absolute return, $d = \Delta P/(P\Delta t)$, obtained by logarithmic expansion.

As Eq. (4.2) points out, the display of the time series P_{ij} by arranging them in the H dimensional variable $\vec{d}_{ij}$, gives a natural *embedding* for a dynamical system oriented analysis [97]. This is not difficult to understand when thinking that the log–return on a certain time horizon τ^* is proportional to the average of log–returns on sub–multiples of τ^*. Thus the component d^{α^*} can be written as a linear combination

---

[2]When a company performs a split, the price X of the emitted shares becomes X/α and the number of them changes from N to αN. Typical values of α range between 2 and 10.

of delayed components $d^{\alpha < \alpha^*}$. The no–arbitrage symmetry of the log–returns hints for the further identification of $\vec{d}_{ij}$ as an (oriented) distance vector between asset i and j; in fact (*a.*) $\vec{d}_{ii} \equiv \vec{0}$, (*b.*) $\vec{d}_{ij} = -\vec{d}_{ji}$, (*c.*) $\vec{d}_{ij} = \vec{d}_{ik} + \vec{d}_{kj}$ for all k.

It is easy to see that any norm in a H dimensional euclidean space induces a well defined distance $\|\vec{d}_{ij}\|$ between asset i and asset j [28]. As an intrinsic character of financial markets no asset can be regarded a priori as an absolute quantity, that is why we ended up only with mutual distances among asset. Nevertheless some truly single asset property can be extracted by the symmetry of the problem and interpreted consequently. If we define:

$$\vec{x}_i \equiv \frac{1}{N} \sum_{j=1}^{N} \vec{d}_{ij}, \tag{4.3}$$

we can observe that $\vec{x}_i - \vec{x}_j = \vec{d}_{ij}$. We have introduced a frame in which every single asset is assigned to an absolute position: The problem of the behaviour of the N assets of the market is now translated to a physical problem of N interacting particles (a liquid) in H dimensions, with coordinates $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N$. At time t, $\vec{x}_i(t)$ is the H dimensional position of particle i. Note that the distance between two assets, expressed by Eq. (4.2), is zero when the price of one with respect to the other remains constant. Furthermore, it is easy to check that the $\vec{x}$'s vectors are centered, hence the positions $\vec{x}_i$ are referred to a coordinate frame which attributes to the center of mass of our liquid a trivial dynamics. From the financial point of view, it states the *closure* of our system: The N assets are watched as complementary, with zero overall return. This does not mean that the applicability of the present construction is restricted to those markets where this property is nearly fulfilled (as an example in the foreign exchange). In stock markets, which experience escape and retention events that is positive and negative return periods, the $\vec{x}$ are automatically selected within a neutral frame which keeps track of the particle cloud. Of course nothing prevents from starting the analysis of an extended market with a huge number of constituent assets. Some of them would follow similar dynamics by evolving in a closer cluster with respect to others.

Coming back to the map construction, it is easy to show that as a consequence of the centered character of the $\vec{x}$'s coordinates, the following quantity σ:

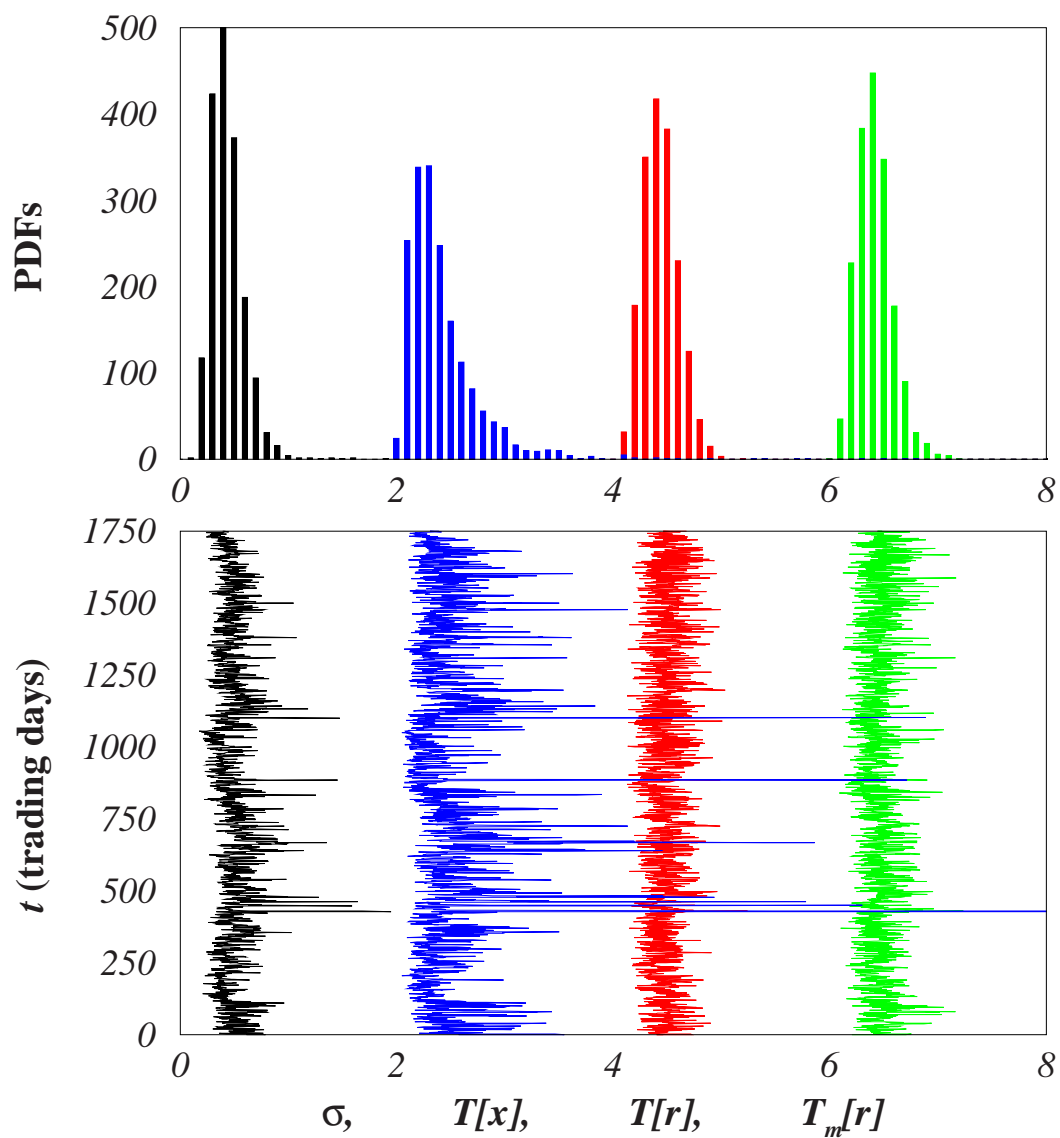$$\sigma \equiv \frac{1}{N} \sqrt{\sum_{1 \leq i < j \leq N} \left\| \vec{d}_{ij} \right\|^2} \tag{4.4}$$

Figure 4.1: Lower panel: Time dependence of the correlated volatility σ, and the temperatures T (shifted as a visual aid) relative to the x– and r–coordinates. Upper panel: The corresponding PDFs. All the calculations refer to four horizons ($H = 4$) of 1, 5, 20, and 250 market days.
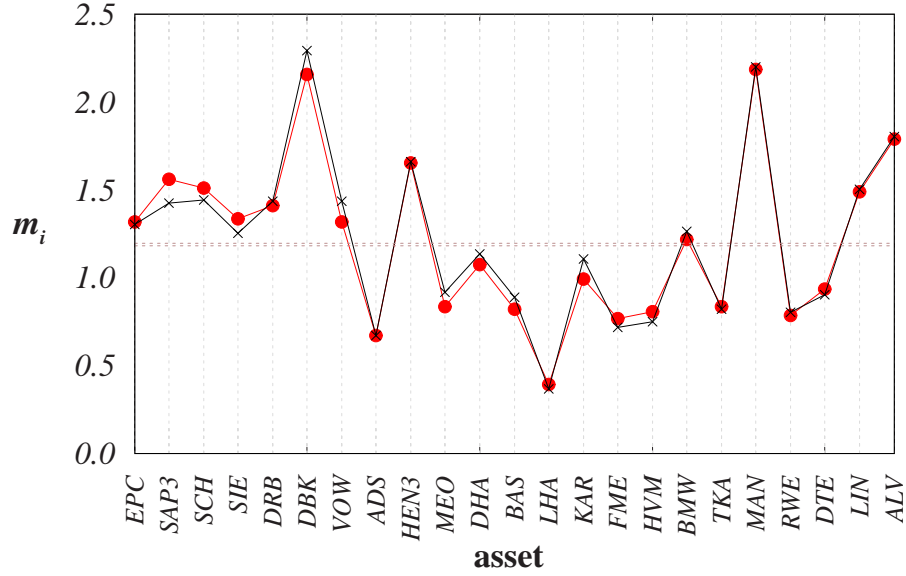
Figure 4.2: Masses m_i as calculated after Eq.(4.8) versus the asset label. Crosses indicate x–frame calculations with mean 1.195, circles the r–frames with mean 1.184.

is exactly their standard deviation. Its Hth power is a measure of the *volume* of our system. The financial counterpart of it is what we call *correlated volatility*, so to stress that it is a quantity merely connected to the spatial interactions of the particles at a certain time. As the usual volatility takes into account the *temporal* variability of an analyzed fixed asset, we are here referring to a measure of a *spatial* variability of a group of interacting assets at a fixed time. Moreover, even after the compensation of the split discontinuities, the correlated volatility shows clusterization around bubble and crash periods [85].

We proceed now rescaling the x coordinates to volume renormalized ones:

$$\vec{r}_i \equiv \frac{\vec{x}_i}{\sigma}. \tag{4.5}$$

Their difference is obviously $\vec{r}_i - \vec{r}_j = \vec{d}_{ij}/\sigma$. The r–frame is a volume preserving frame. Once the volume of the system is stabilized, one may wonder which is the dependence of the liquid temperature on time. Thus, by analyzing the empirical behaviour of the ensembled averaged square (finite difference) velocities:

$$\vec{v}_i(t) = (\vec{r}_i(t) - \vec{r}_i(t - \tau_1))/\tau_1, \tag{4.6}$$

we found that the r–system is thermostated at a fix temperature:

$$T = \frac{\langle\langle v_i^2(t)\rangle_i\rangle_t}{H}.$$

(4.7)

The *correlated volatility* is therefore a measure of the *temperature* of our system. A similar concept has been also introduced by Lillo and Mantegna [57], addressing the question whether the complexity of a financial market was limited to the statistical behaviour of each financial time series or rather a complexity of the overall market did exist.

Fig. 4.1 shows this fact: In the lower panel we have plotted the time dependence of the correlated volatility σ and of the temperatures $T[x]$ and $T[r]$ calculated by averaging the square velocities in the x– and r–frame, respectively. In order to contrast the results, the time averages of σ and $T[x]$ are rescaled to T (the time average of $T[r]$). The scale of T is in fact fixed by the underlying assumption of an unitary Boltzmann's constant. To check possible ergodicity properties of the system we have also analyzed the time averaged square velocities of the single assets and extracted from them mass terms:

$$m_i = \frac{HT}{\langle v_i^2(t)\rangle_t}.$$

(4.8)

Fig. 4.2 shows that the masses are only slightly affected by the reference frame used to calculate them. This indicates that they are an intrinsic property of the asset regardless of the kinetics details. To prove this statement, we have plotted, in Fig. 4.1, the correction to the temperature due to the asset masses:

$$T_m[r] = \frac{\langle m_i v_i^2(t)\rangle_i}{H}.$$

(4.9)

In order to investigate the nature of the interaction of the particle system under study, we have calculated the two point correlation function as expressed by Eq. (3.24):

$$C(r) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta(r - \|\vec{r}_i(t) - \vec{r}_j(t)\|)$$

(4.10)

and the related pair potential:
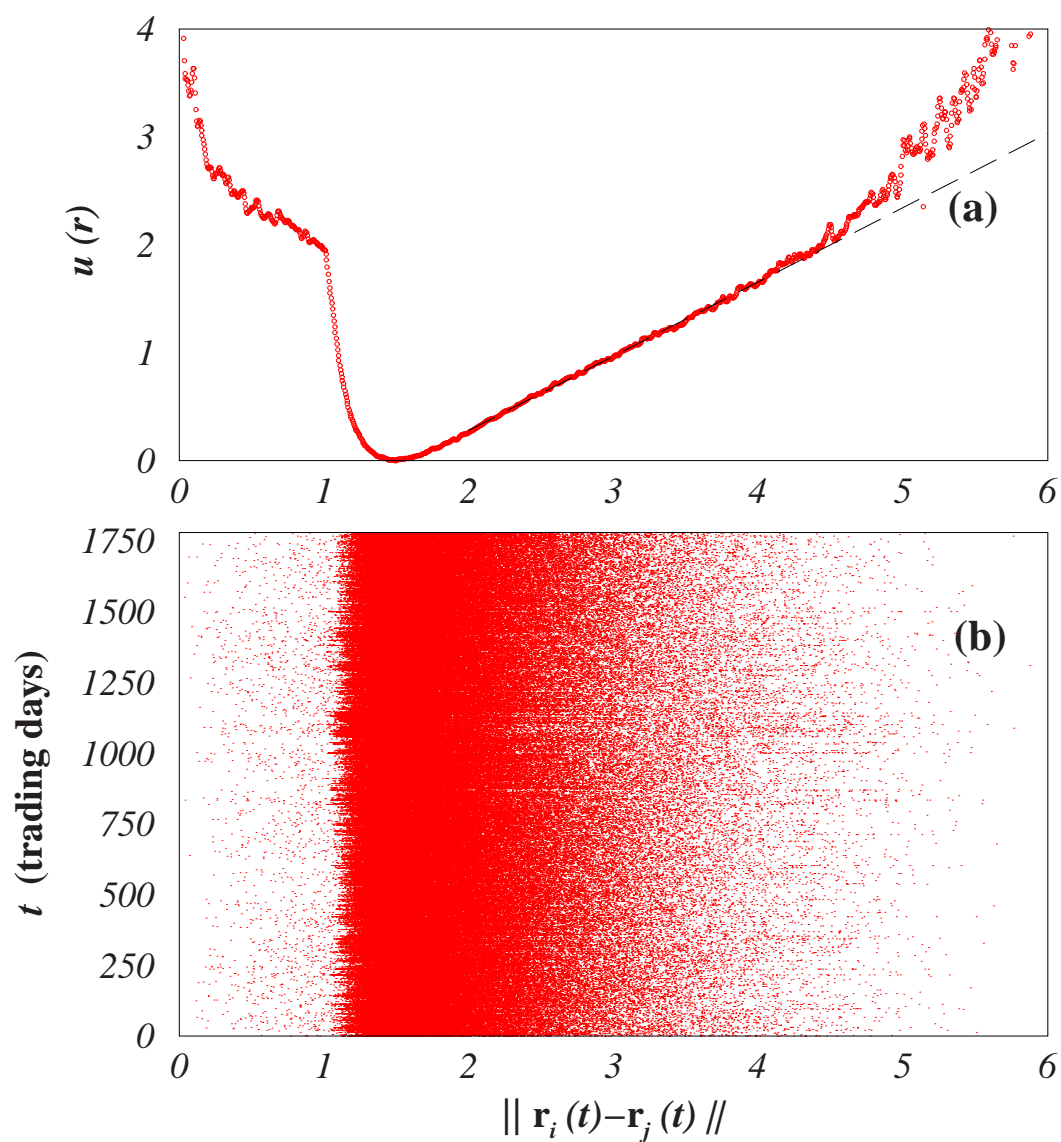
$$u(r) \propto -\log C(r).$$

(4.11)

Figure 4.3: (a) Plot of the pair potential $u(r)$ for the whole data set over four horizons ($H = 4$) of 1, 5, 20, and 250 market days, and (b) the time distribution of the inter–asset distances.

In Fig. 4.3., the potential $u(r)$ is shown. The great distance tail of $u(r)$ is linear (correlation coefficient= 0.9994, for a regression in the region $2 < r < 4$ over 446 points giving the line $u = ar + b$, with $a = 0.689 \pm 0.001$ and $b = -1.101 \pm 0.004$) indicating the strong long range attraction of the market liquid. On the other hand at small distances two different behaviours emerge. By decreasing the asset–asset distance an equilibrium point is reached. At smaller distances a barrier is presented, followed by a region corresponding to less intense repulsive forces. We interpret it as a signature of the inhomogeneity of the system, which allow at small distances the formation of privileged pairs (clusters). As a consequence, we expect that in a wider market (here we consider the quite diversified but small pool of the DAX30 assets) this tendency could even be more pronounced [57, 72].

The validity of a no–arbitrage condition is guaranteed by the assumed liquid character of the market. By implementing the embedding, naturally prompted by the structure of the returns, we have been able to map the financial signals in positions of particles of an interacting gas (a liquid). One of the strength points of this method is its easy generalizability to the case of great N, albeit here we have restricted our analysis to a relative small asset market. On the other hand a word of caution is needed in a great N market. The results presented here share the plain assumptions of isotropy and homogeneity of the market liquid. Indeed they should become weaker for very large and differentiated markets. There, the pair potential introduced here is supposed to maintain the same great–distance properties (linearity). At the low–distances (where clustering emerges), in analogy to what is done in the study of ionic liquids [77], a generalized pair potential could be introduced in order to include both anisotropy, cluster formation, and specie diversification [85]. Besides, this approach is straightforwardly employable for *time dependent clustering*. A procedure similar to the one adopted to organize static distances between assets in hierarchy trees, given in [72], could be generalized to the time dependent distance matrix expressed by Eq. (4.2).

4.3 Predictability of Foreign Exchange Markets

We want to test now a possible further application of the liquid analogy presented in the previous section, combined with a Markov model, studying the foreign exchange market data related to the three following currencies: American Dollar (USD), German Marc (DEM) and Japanese Yen (YEN). The data we use cover the period from 9 Apr 1973 to 15 Nov 1997 with a daily tick and are shown in Fig. 4.4. In a foreign exchange market the closure property (Eq.(4.3)) is by definition fulfilled. As clearly observable in Fig. 4.4, it is in fact obvious that:

$$\frac{USD}{DEM} = \frac{YEN}{DEM} * \frac{USD}{YEN}. \tag{4.12}$$

Let us consider the quantity $s_t^i = sign(x_i^1(t))$ and compute the statistic:

$$P(s_t^i | s_{t-1}^{USD}, s_{t-1}^{DEM}, s_{t-1}^{YEN}), \tag{4.13}$$

where $i = USD, DEM, YEN$ for all the three currencies. For $i = USD$, the quantity $P(+| + -+)$ indicates the probability to get a positive value for $x_i^{USD}(t)$, given a positive value for the USD at time $(t-1)$, a negative value for the DEM at time $(t-1)$ and a positive value for the YEN at time $(t-1)$. Doing that we implicitly assume the validity of a Markow model for the foreign exchange market system, namely a model where $P(s_t^i)$ only depends on 3 other values. Of course this may not be the case, but we think it is worth trying this analysis if we end up with a satisfactory statistical significance. According to the size of the data set we are using, we are not allowed to consider a Markov model of bigger order. The different combinations related to Eq. (4.13) are 16, but 2 of them, namely $P(\pm_t | +_{t-1}, +_{t-1}, +_{t-1})$, never occur since $\sum_{i=1}^{N} \vec{x}_i = \vec{0}$. The two $P(\pm_t | -_{t-1}, -_{t-1}, -_{t-1})$ are positive because in the simbol "$-$" also the zero value is included.

The 14 probabilities are given from $P(+| + +-)$, $P(+| + -+)$, to $P(-| - --)$. As we can see in Fig. 4.5, the slight difference from the overall constant value of 0.5 reveals the presence of structures and redundancy. They behave in a synchronous way that could not otherwise be explained only with the closure property. This allows us to perform some kind of prediction of future exchange rates. The most accentuated situation is the one related to the YEN (right panel): One of the probabilities goes down to 22.73 (with the corresponding going up, obviously, to 77.27). Altough this result clearly tells us that in principle one could get advantage
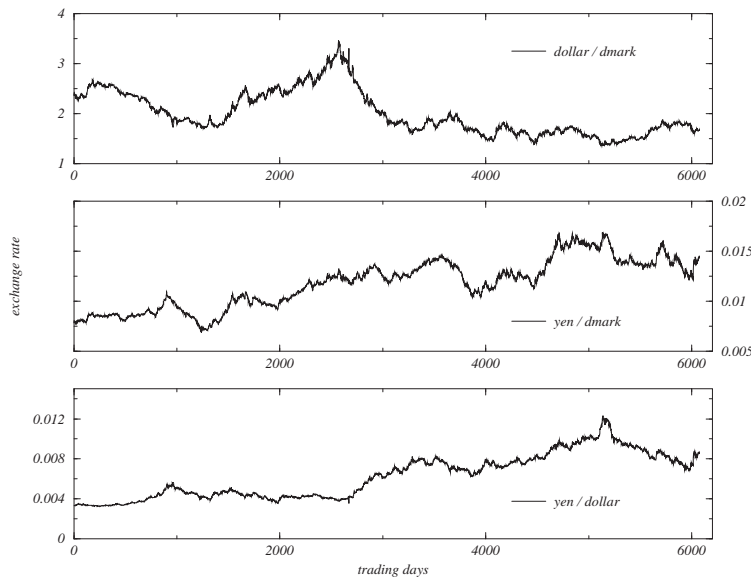
Figure 4.4: Daily exchange rate of USD, DEM and YEN for the period of time 9 Apr 1973 to 15 Nov 1997.

speculating with the foreign exchange market, we get no quantification of the gain since the quantity $s_t^i = sign(x_i^1(t))$ takes into account only the direction of a currency but no the amplitude of the step. So the good result of Fig. 4.5 could just mean that we can predict with probability almost 80% that the YEN will become more *powerful* tomorrow, but we cannot quantify this improvement. Maybe the YEN will gain only a small fraction of a percent and if we consider transaction costs and commissions this information becomes absolutely useless.

A better approach, therefore, consists in introducing some threshold in order to differentiate between strong increase, moderate increase, almost no change, moderate decrease and strong decrease situations. Once computed the quantity in Eq.(4.3), we compare it with 4 values in order to get a better classification. The output of this comparison can be one of the following: *(i)* $x_i^1(t) > s$, namely we have a strong increase (4), *(ii)* $s_e < x_i^1(t) < s$, moderate increase (3), *(iii)* $-s_e < x_i^1(t) < s_e$, almost unchanged (2), *(iv)* $-s < x_i^1(t) < -s_e$, moderate decrease (1) and *(v)* $x_i^1(t) < -s$, strong decrease (0). The numbers in brackets indicate the label associated to the sistuation. Of course the condition $0 < s_e < s$ is fulfilled. In order to have an idea of the proper value of s and s_e let us consider Fig. 4.6. There we plot the percentage of values of $x_i^1(t)$ that falls into the interval $[-s_e, s_e]$ with respect to s_e. We see that
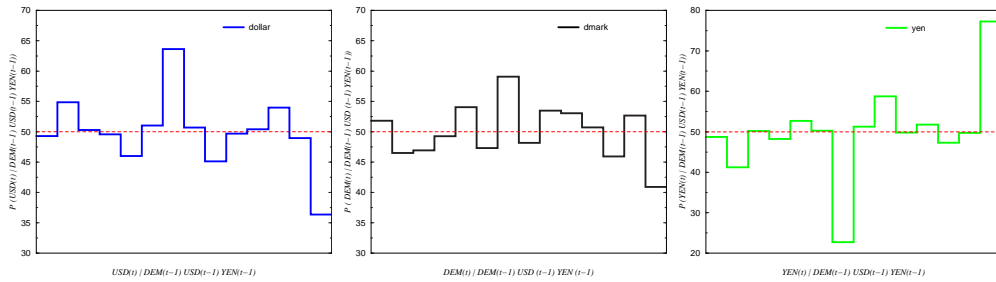
Figure 4.5: Statistic on the $s_t^i = sign(x_i^1(t))$. Left panel: USD. Middle panel: DEM. Right panel: YEN. Note a quite far behavior from constancy, leaving space for an attempt to prediction. The sample size is big enough to guarantee us that everything outside the 50 ± 1 % is statistically significant.

there is no value smaller than 0.1% and also no one bigger than 2%. The data set we are using has a daily sampling rate and for an exchange market it is very uncommon to get variations bigger than some percent.

Given the distribution of $x_i^1(t)$s, we consider the following thresholds: $s_e = 0.2\%$ and $s = 0.8\%$. Now we have 5 possibilities for each currency at each time step. Therefore we have to cope with 625 cases when evaluating the statistics of $P(X_{t+1}|DEM_t, USD_t, YEN_t)$, where X stays for DEM, USD and YEN.

In Fig. 4.7 we show the situation we get for the DEM. The upper part of it refers to the forecasted increase for the next time step, the middle part deals with the almost unchanged case, the lower part reports on decreasing situations. Along the X axes all the 125 possible combinations of DEM, USD and YEN at a given time are reported, using a 5-based code: 0 refers to 000_5 and 124 to 444_5. Of course not all the possible combinations are allowed, since $\sum_{i=1}^N \vec{x}_i = \vec{0}$, as already discussed. This conditions excludes almost half of the possibilities. In order to really take profit from this speculation, one cannot use the middle part because the gain is so small (smaller than $s_e = 0.2\%$) that it would be completely compensated from commission costs. Looking at the upper part, however, we note that in 4 situations the value of the probability is bigger than 50%. The first peak, in position $19_{10} = 034_5$ means that: $P(s_e < x_{DEM}(t+1) < s | x_{DEM}(t) < -s, s_e < x_{USD}(t+1) < s, x_{YEN}(t) > s) = 67\%$, namely it is very probable that the day after a bad situation for the DEM (0) together with a strong USD (3) and a very strong YEN (4), the DEM will recover part of its losses (3). In other words, buying some DEM after such a situation has happened,
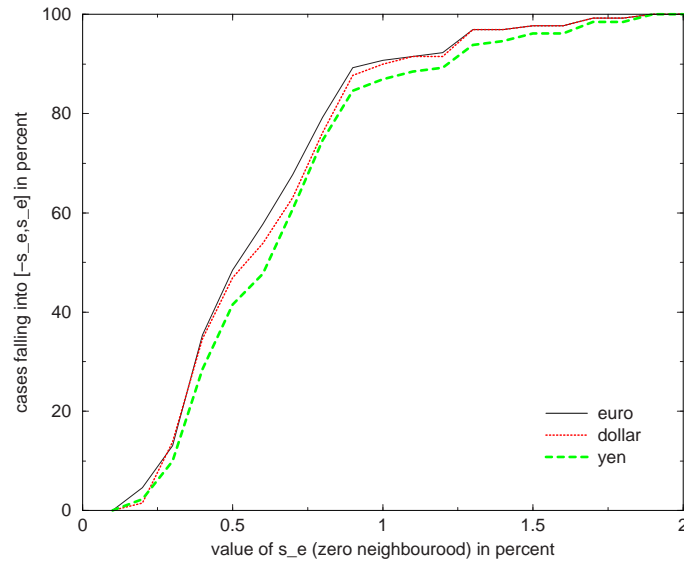
Figure 4.6: Percentage of values of $x_i^1(t)$ that falls into the interval $[-s_e, s_e]$ with respect to s_e. All the values are contained within the interval $[0.1\%, 2\%]$.
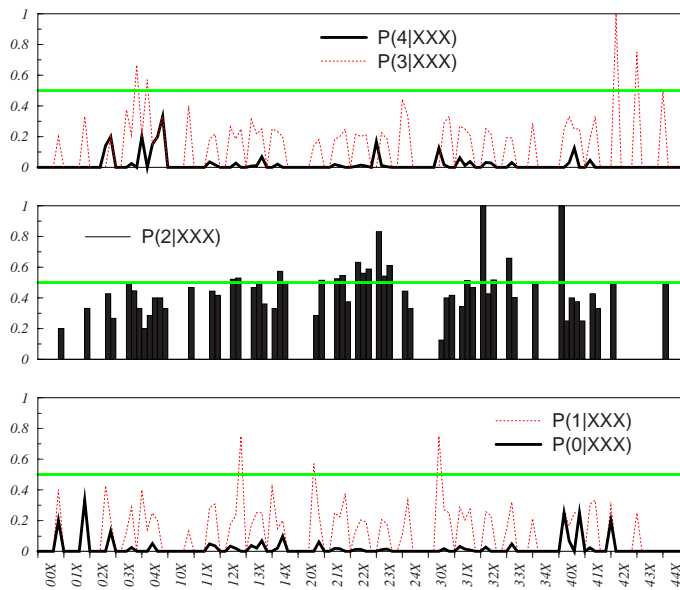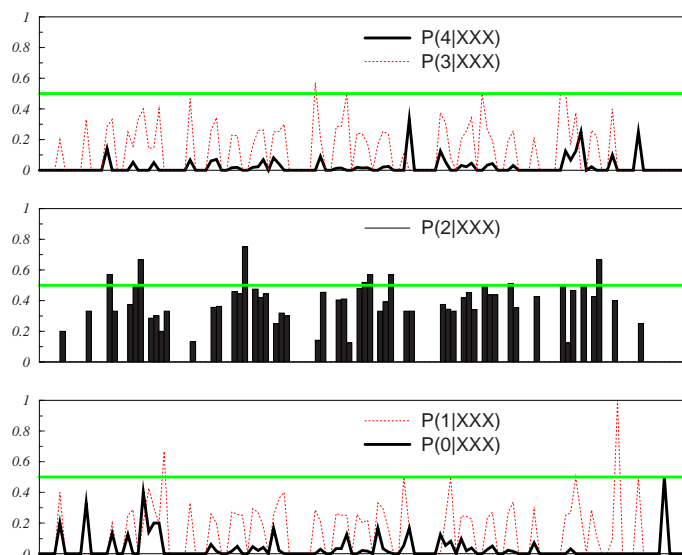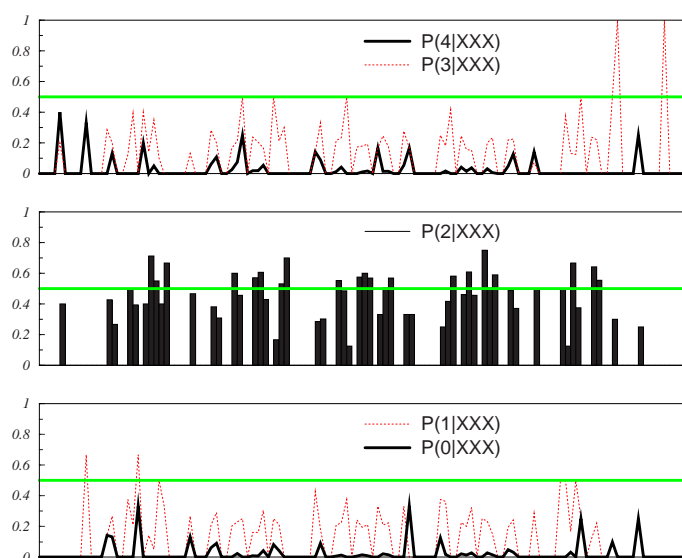

Figure 4.7: Statistic of $x_{DEM}^1(t)$ after the introduction of thresholds. The upper part refers to the probability of having an improvement at time $(t+1)$, the middle graph to stationarity and the lower one to a dicrease in the relative value. Note the presence of peaks above the 50% line.

Figure 4.8: Statistic for $x^1_{USD}(t)$, same considerations of Fig. 4.7.

Figure 4.9: Statistic for $x^1_{YEN}(t)$, same considerations of Fig. 4.7.

should pay with some gain. Conceptually similar cases are depicted in Fig. 4.8 and Fig. 4.9, dealing respectively with USD and YEN.

If willing to use the present approach in order to speculate in the foreign exchange market, one has to keep in mind that it is convenient to operate only when falling in a situation related to a next step probability bigger than 50%. This means that the applicability of the present method is limited to some favourable cases: It is better to do nothing when we do not get proper indications and to be active only when suggested from Fig. 4.7, Fig. 4.8 and Fig. 4.9. We have already said that only one half of all the combinations of labels for DEM, USD, YEN is effectively possible. Counting the number of occurrence of the probability being bigger than 50% we get 14 (as previously mentioned we cannot use the middle part of the last three illustrations because they are related to the neighbourhood of zero case): Therefore this speculation scheme is really applicable only in rougly the 20% of the situations (that maybe do not happen the 20% of the trading days but much more rarely).

4.4   Financial Markets in a Book

Parallel to the analysis of empirical features exhibited by financial market, also the development of models able to reproduce part of them is interesting, especially if the parameters of the model have a clear meaning. Since stock markets resemble the scaling laws characterizing physical systems in which large numbers of units interact [34, 71], we introduce here a frustrated and disordered many-body system. Frustration enters in that not all the individual inclinations can be satisfied simultaneously, whereas the model is disordered because interaction couplings are random and traders have randomly chosen expectations and resources but they share the same strategy. In contrast with previuos works we do introduce only one kind of investor: In our opinion the usual distinction between fundamentalists and noise traders is not necessary. Looking at both strategies, in fact, we note that *fundamentalists* follow the premise of the efficient market hypothesis in that they expect the price to follow the fundamental value of the asset: A fundamentalist trading strategy consists of buying when the actual market price is believed to be below the fundamental value and selling in the opposite case. *Noise traders*, on the other hand, do not believe in an immediate tendency of the price to revert to its underlying fundamental value: They try to identify price trends and consider the behaviour

of other traders as a source of information. This gives rise to the tendency towards *herding behaviour.*

BUY ORDERS					SELL ORDERS			
time	trader	shares	price		price	shares	trader	time
21005	240	4	11122		11123	4	576	19802
25008	207	70	11121		11124	4	876	14706
24506	647	3	11118		11125	2	806	12150
19002	820	2	11108		11130	49	201	17203
20148	100	12	11106		11130	4	792	20101

TABLE 1. Example of the first five levels of the book. No transaction can take place because the highest buying price is smaller than the cheapest selling order. Entries are ordered according to price and occurrence time in case of equal prices.

The model presented in [62, 63] relies on (artificial) movements of individuals from one group to another together with the exogenous changes of the fundamental value. However, as clearly stated in [4], a 20% of fundamentalists is enough to confine the price within the range of the rational traders. We want to raise serious doubts on the blind use of the fundamental price, expecially when assumed that the relative changes are Gaussian random variables: Why should agents risk money just believing in a random walk behaviour? A more realistic assumption is based on the following wish: Gaining the maximum, taking the smallest possible risk. Agents make decisions to buy or to sell, adjust prices, and so on according to the information available at the time, as well as individual preferences such as tolerance for risk and time deadlines. We therefore simulate the book for the ask and for the bid. Every trader, when buying a share, has to identify a fair price and then put an order keeping in mind a target price and a stop-loss price. These quantities are the result of the interaction with other agents, the study of past values of the price and the influence of incoming news. The decision to sell some shares is made on the basis of the current price (to be compared with the personal target and stop-loss price) and of the *age* of the shares: After a sufficently long period of time agents start to ask themselves whether it is worth keeping the money invested in that way.

The model leads to a kind of *self-organized criticality* which is responsible of the alternation between *laminar* and *turbulent* trading [89].

We consider N traders and one stock with M shares on the market. Each trader is characterized through the following information: *(i)* Initial amount of money. *(ii)* Inclination towards investment: Usually traders tend to keep cash a part of their resources, in order to be able to have money to exploit the market at special time. *(iii)* Number of shares owned. *(iv)* List of friends with which he is sharing information, to model the herding effect. *(v)* Invested money, to keep trace of the average buying price. *(vi)* Desired gain. *(vii)* Maximum loss. *(viii)* Threshold: Amount of time after which the trader may start to change idea about his/her investment.

Every order is stored in the corresponding list, according to the type of it (buy or sell), to the requested price and to the time at which it was submitted. A transaction occurs whenever the cheapest price among the sell list matches with the most expensive offer in the buyers' list: This value defines the market price of the stock at that particular instant and it will of course affect the future behaviour of the traders. We provide a mechanism to produce *news*, whose purpose is to let all the agents know some information about the overall behaviour of the market, namely the unbalance of the two books and the actual volatility. It is very important to note that these signals are endogenous: The information they provide was already present in the system. In this way our model takes into account both a *local and global coupling*, via shared information with neighbours and generation of news and advertisements, respectively.

The simulation consists of two parts. At the beginning we assign all the shares to one trader and we broadcast advertisements to induce people to put a buy order. The purpose of this first step is to simulate the **I**nitial **P**ublic **O**ffer (**IPO**) and the selected trader can be thought as the bank responsible of the initial selling of the stock. This part ends when the number of shares owned by the IPO trader vanishes. We then reset his amount of money and enter the second phase. At each update step the algorithm performs the following operations:

- Select randomly a trader.

- If the trader has no pending order and no share then he/she is probably willing to buy, formulating a limit price, a target price and a stop-loss price and

inserting a buying order.

- In case of owning some shares, the trader may decide to sell, according to the market price and the threshold.

- If the trader has a pending order, he/she may change some parameter because the conditions that led to that decision may have changed.

We suppose that every trader can afford only one pending order. Every time a buyer has got all the desired shares, he/she is immediately asked whether he/she wants to place a selling order.

When formulating the prices, every trader makes the decision in a *deterministic* way, computing a weighted average among the opinion of some acquaintance, the indication of the news and some past values of the stock price. Every time that two complementary orders match, we define it as the market price for that particular instant: Usually only one of the two orders disappears, namely the one with the smallest amount of shares involved. The other cannot be removed from the corresponding list, since only a part of the desired transaction has taken place.

So the main ingredients of the model are: *(i) Disorder*, since the interaction couplings are random and some parameters characterizing the agents are randomly chosen: The overall dynamic follows deterministic rules triggered by stochastic events. *(ii) Frustration*, because not all the individual wishes can be satisfied at the same time and traders have to change idea, i.e. insert a modified order in the book, quite often, behaviour that resemble the typical multiplicity of states of frustrated systems. *(iii) Delayed feedback*, involving the use of past values of the market price and volatility during the decision formation. *(iv) Phase transition* between the excess-demand and the excess-supply phases [74]: The model presents a symmetry breaking reflected in a discontinuity in such a global quantity like the market price.

Fig. 4.10 shows the result of a typical simulation on the market price and the corresponding amount of exchanged shares. The model is able to reproduce all the typical features observed in empirical data. At the beginning the price remains constant due to the *IPO's phase*, namely the bank offers the shares to the traders at the fixed price, the *IPO price*. After that, one can see the typical pressure made by agents who did not get enough shares during the initial public offer: The volumes are high and the price tends to raise. Then, after a normal settlement, the price
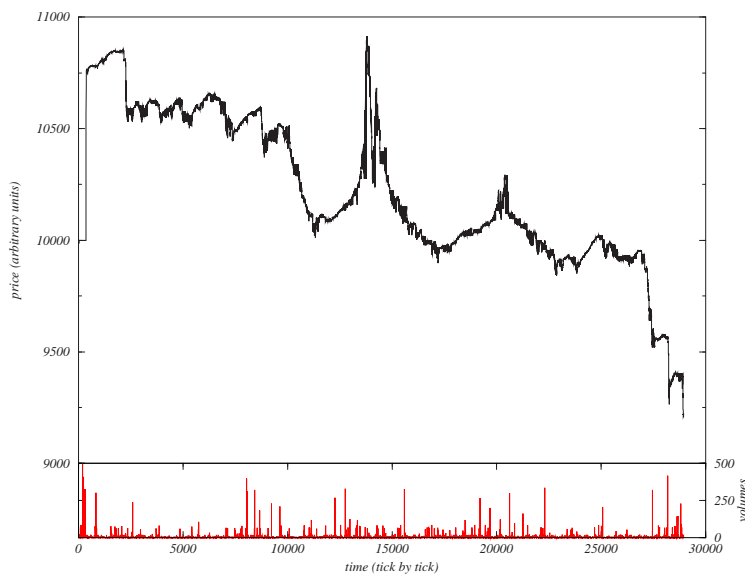
Figure 4.10: Typical time series segment from a simulation run. Upper panel: Development with time of the market price. Lower panel: Development with time of the corresponding volumes of exchange.
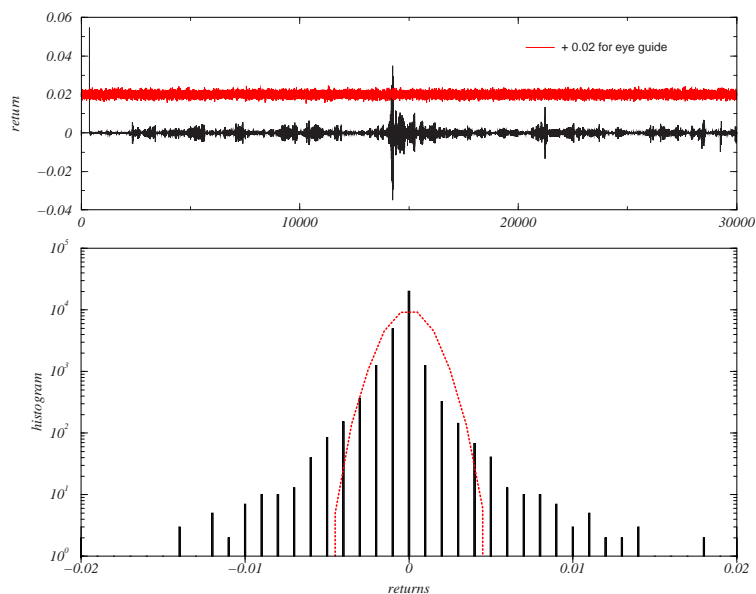

Figure 4.11: Upper panel: Price returns, with the random series shifted for eye guide. Lower panel: Returns distribution. The comparison with a best-fitted normal distribution reveals the presence of *fat tails*.
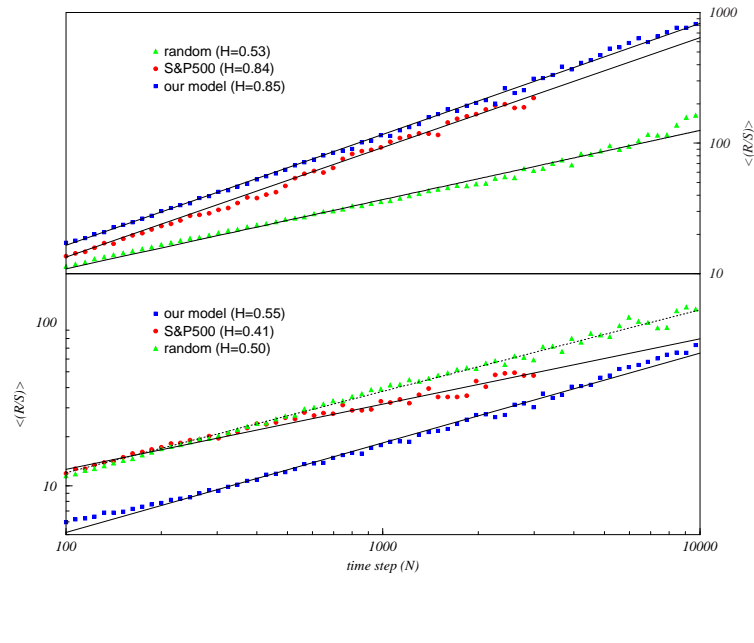
Figure 4.12: Estimation of the self-similarity parameter H. Upper panel: Absolute returns. Lower panel: Raw returns. Average of many runs.

starts to oscillate with very low volumes: Traders with shares do not want to sell because they hope to get more money if they wait a little bit more, agents without shares do not buy because the price is too high and there is no evidence of a trend on it. Then oscillations become bigger and bigger and when the volumes are big as well, then a small crash occurs and the price comes back to a more interesting value for potential buyers. As a consequence, volumes remain high and the price follows a so called rally period, followed again by a crash, maybe due to the fact that the bubble phase has been too optimistic.

The model makes use of the following parameters: *(i)* Number of traders **N**. *(ii)* Threshold **T**: The critical age of the shares. It ranges among a decade in order to take into account the differences between intraday speculators and long-time agents. *(iii)* Number of shares **S** and IPO's price **I**: Their product defines the initial value of the company. *(iv)* Amount of money **M** initially distributed among the traders. Inspired by [76, 88] we have decided to distribute the richness according to a Zipf's law: The 20% of the traders posses around the 80% of **M** and among the two groups this rule is applied again, recursively. In this way we are able to model the difference

between a normal agent and an institutional investor and the different effect they produce when they decide to enter the market. A minimal value of money **m** is provided to all the traders and added to the previous distribution. *(v)* Length of the past values' list **MEM**: Chartists look for trends and patterns in the historical time series of the market price. *(vi)* Threshold **B** for the unbalance of the book. We compute the ratio between the sum of all the shares people want to sell and the sum of all the shares people want to buy. If this value does not belong to the interval $[1/B, B]$, then news and advertisements are generated.

As shown in Fig. 4.11, the probability density function of the returns of our simulated stock shows a strong leptokurtic nature. For comparison, the Gaussian with the same measured standard deviation is also reported. The time series of returns exhibits a higher frequency of extreme events and clustering of volatility. This aspect becomes clear thanks to Fig. 4.12: When considering absolute returns as a measure of volatility, we see that the transformed price data behave differently from their counterpart derived from the Gaussian distribution. The estimation of the self-similarity parameter H, as introduced in Sec. 2.7, reveals a strong persistence in volatility ($H = 0.85$).

One comment about demand and supply. It had been a common sense in economics for a long time that demand and supply balance automatically, however, it becomes evident that in reality such balances are hardly be realized for most of popular commodities in our daily life [105]. The important point is that demand is essentially a stochastic variable because human action can never be predicted perfectly, hence the balance of demand and supply should also be viewed in a probabilistic way. If demand and supply are balanced on average the probability of finding an arbitrarily chosen commodity on the shelves of a store should be 1/2, namely about half of the shelves should be empty. Contrary to this theoretical estimation shelves in any department store or supermarket is nearly always full of commodities. This clearly demonstrates that supply is much in excess in such stores. In general the stochastic properties of demand and supply can be well characterized by a phase transition view which consists of two phases: The excess-demand and excess-supply phases. It is a general property of a phase transition system that fluctuations are largest at the phase transition point, and this property also holds in this demand-supply system. In contrast to that, in the case of markets of ordinary commodities, consumers and providers are independent and the averaged supply and demand are
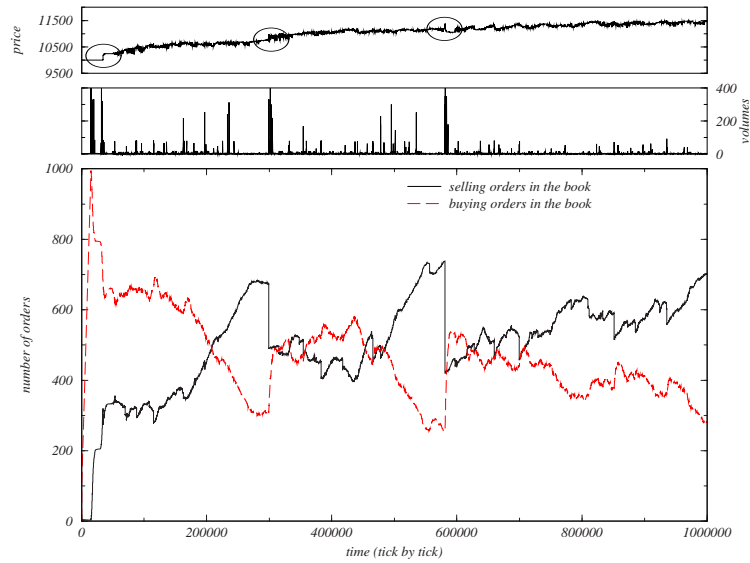
Figure 4.13: Simulation run with N = 1000. Upper panel: Market price evolution on time. Middle panel: Volumes of exchanged shares on time. Lower panel: Evolution of the book. Almost all the traders have placed an order and are waiting. Note the symmetry of the two paths with respect to the half of the number of agents (N/2 = 500).

generally not equal. The resulting price fluctuactions are generally slow and small in such market because the system is out of the critical point.

On the contrary in an open market of stocks or foreign exchanges, market is governed by speculative dealers who frequently change their positions between buyers and sellers. It is shown that such speculative actions make demand and supply balance automatically on average by changing the market price, as Fig. 4.13 clearly shows. Contrary to [4] we do not need to impose that the number of the shares has to be half of the number of traders in order to get a balance between demand and supply. The three circles in the upper panel indicate the most extreme events taking place in the price evolution: There are corresponding movements in the book and in the volumes, since they are the reason for such a sudden variation. As the system is always at the critical point the resulting price fluctuations are generally quick and large [105]. This result is in agreement with [102], where the authors present an analogy between large stock market crashes and critical points with log-periodic correction to scaling: Complex systems often reveal more of their structure and organization in highly stressed situations than in equilibrium.

Performing a correlation analysis on our simulations and comparing it to the results presented in [71], we can associate a temporal scale to our tick by tick time: Since the autocorrelation function vanishes after approximately 20 ticks and the typical correlation length in financial time series is supposed to be around 20 minutes, we can speculate that one of our tick corresponds to one real-life minute. Therefore the involved time scales are the following: *(i)* Total number of iterations, namely total number of ticks = 10^6 (**10 years**), *(ii)* threshold = 10000 (**1 month**), *(iii)* threshold variability in the range between 0.1 and 1. This gives a time variable from a minimum of **1 month** to a maximum of **1 year** to have second thoughts.

4.5   Spread-Volatility correlation

The key ideas here are the interplay between time and money and the risk aversion represented by the stop loss mechanism. Every order is stored in the corresponding list, according to its type (buy or sell), the requested price and the time at which it was submitted. A transaction occurs whenever the lowest price in the sell list matches with the highest offer in the buyers' list: This value is defined as the market price of the stock at that particular instant (tick).

When randomly selected, a trader usually tends to perform some operation if he/she has not yet inserted an order. Since, as already mentioned, we want to avoid any use of fundamental rules, there is no real recipe in our model to decide when to enter the market. It is much more important to identify the right moment to sell, because it is only when you sell that you get the extra money you have won or you realize your loss. Let us have a look at Fig. 4.14 for a better understanding of this concept. Suppose that a trader has bought shares at the price and the time marked by the filled circle. The basic strategy is represented by the *trading rectangle* [84], defined by the three following quantities: Target price (upper horizontal line), stop-loss price (lower horizontal line) and threshold in time (rightmost vertical line).

As long as the market price is confined within the trading rectangle, the agent does not feel the need to trade, but once this condition has been violated, it is very likely for him/her to perform an operation. If the price goes beyond one of the two horizontal lines, a market order to sell the shares is very probable (either to cash the win or to limit the loss). If the price remains almost constant within the trading rectangle, and therefore the time series ultimately crosses the rectangle at
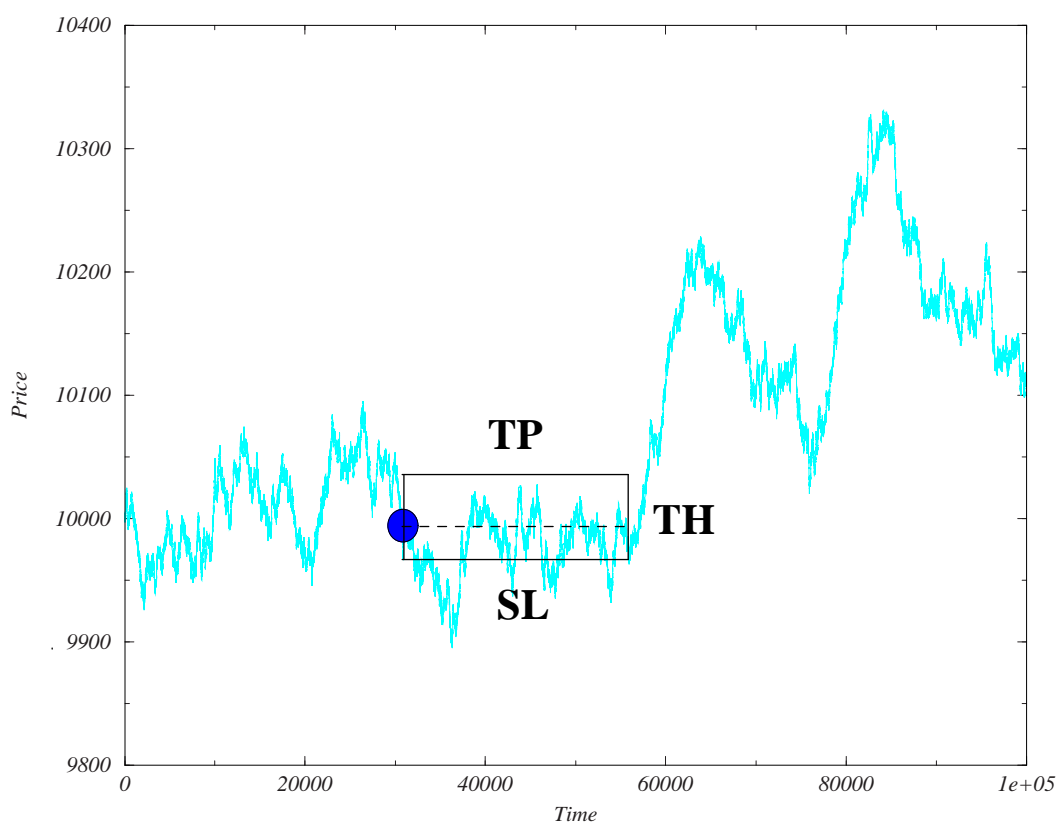
Figure 4.14: **The trading rectangle**. Reported is the market price versus trading time. The filled circle indicates the moment in which the trader has bought shares. The dotted line, constant at the buying price, is plotted only for eye guide. The upper line is the target price (**TP**), the lower line refers to the stop loss price (**SL**) and the threshold in time (**TH**) defines the right end of the trading rectangle.
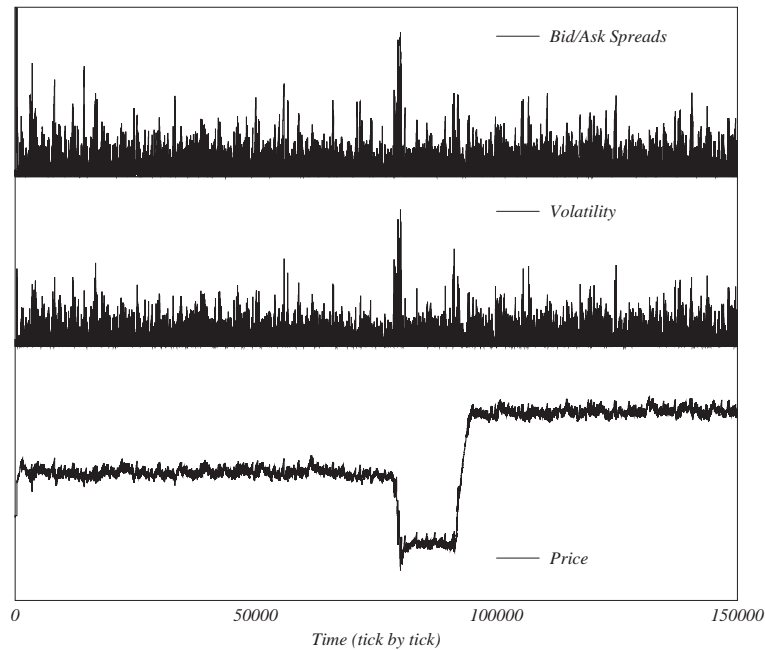
Figure 4.15: Temporal evolution of bid/ask spreads, volatility and market price. Bid/ask spreads and volatility are strongly correlated.

its rightmost vertical line, the decision of the trader depends on a global condition, which is given by the imbalance of the book, namely by the ratio between selling and buying orders. If too many people want to sell, this is a good reason to leave the market as soon as possible (therefore with a market order). If a lot of agents are willing to buy, then it can be better to keep the shares because their value could appreciate substantially in a near future. The cross at the right vertical line is related to the constraints of real life and a sort of practical considerations which must be considered in making a real gambling decision (see the Appendix).

The last consideration gives rise to a comment about minority games. It is almost evident that the financial market is not in favour of minorities. The best strategy is not acting as extravagant as possible, but doing exactly what the majority of people wants to do although within (and this is of fundamental importance) a shorter time, in advance. At the beginning of a crash, for example, the majority of traders wants to sell: Being with the minority and buying in that moment would not be a good idea. The same during a rally: The price is strongly increasing because there is a big request to buy and if you sell at the beginning of such an event you lose an
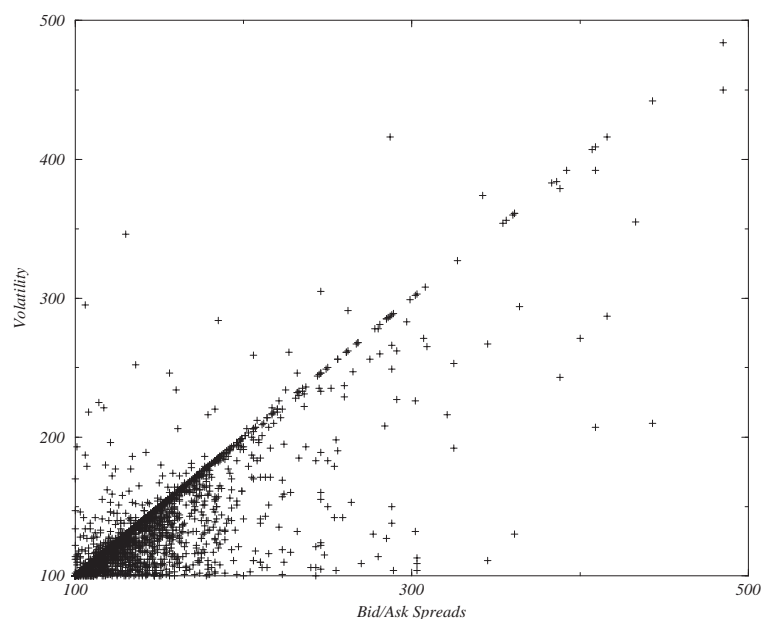
Figure 4.16: Scatter-plot of bid/ask spreads and volatility. Significant correlations between these two quantities can clearly be observed.

opportunity to make money. Furthermore, minority games assume two outcomes and it is unclear how to distinguish between buying and selling.

A relationship between dealers' bid/ask spreads and price volatility is presented in [86]. Dealers widen their spreads when volatility is higher in order to reduce their risk of loss. Due to the high degree of realism in our model, we are able to test the same relationship from a simulation run. Fig. 4.15 shows the temporal evolution of the market price, the volatility and the bid/ask spreads, defined as the difference between the lowest selling price and the highest buying request. A significant positive correlation between bid/ask spreads and volatility is apparent from the temporal evolution and from the scatter-plot of Fig. 4.16. Numerical analysis of the linear cross-correlation (when defining volatility as the local average of the absolute value of the price change) indicates a coefficient that ranges from 0.2 to 0.3, depending on the window size. In fact, considering our simulated data at lower frequencies leads to a higher cross-correlation coefficient, in agreement with the results reported in [86].

Up to our knowledge, this is the first attempt to reproduce this correlation within an artificial financial market. The availability of such a model is very fruitful when

dealing with risk management and option pricing and a similar behaviour cannot be shown by simpler models (like [78]). The main contribution to this result comes from the threshold, since it allows the establishment of a strong relationship between time and money: Traders do want to make money within a specific time when they trade stocks.

4.6   Deriving the Optimal Strategy

We want now to check whether the common basic strategy of the trading rectangle leaves place for the establishment of a kind of *meta-strategy*, a way to profit from mistakes and good intuitions of agents [27]. As a first step, we have then to evaluate the performance of every single trader. In order to simplify the task, we discard as many of the last operations as necessary in order to come back to a point where the agent is completely liquid (no invested money, just cash). We then consider the difference between the money realized after having sold the shares and the originally invested capital. This quantity, related to the initial amount of money, gives an accurate estimation of the quality of the trading. In more details the absolute performance $(AP)_i$ is defined as:

$$(AP)_i = \sum_{j=1}^{j=m_{is}} p_{isj} s_{isj} - \sum_{j=1}^{j=m_{ib}} p_{ibj} s_{ibj}, \qquad (4.14)$$

where m_{is} is the total number of selling orders, m_{ib} the total number of buying orders, p_{isj} and s_{isj} the selling price and the number of shares involved in the selling orders, p_{ibj} and s_{ibj} the same quantities related to buying orders. The relative performance $(RP)_i$ is obtained as:

$$(RP)_i = \frac{(AP)_i}{R_i}, \qquad (4.15)$$

with R_i indicating the initial resources of trader i.

The Fig. 4.17 shows clearly that not all the traders have been equally lucky during the simulation. The straight line refers to the zero gain and it represents of course the average amount of money won from the traders, since the market we are simulating is a closed one, we have introduced only one stock, the trading activity has no extra cost and no dividend is paid. To define a transaction one needs a buyer and a seller, for every investor gaining money there must be one other who is losing
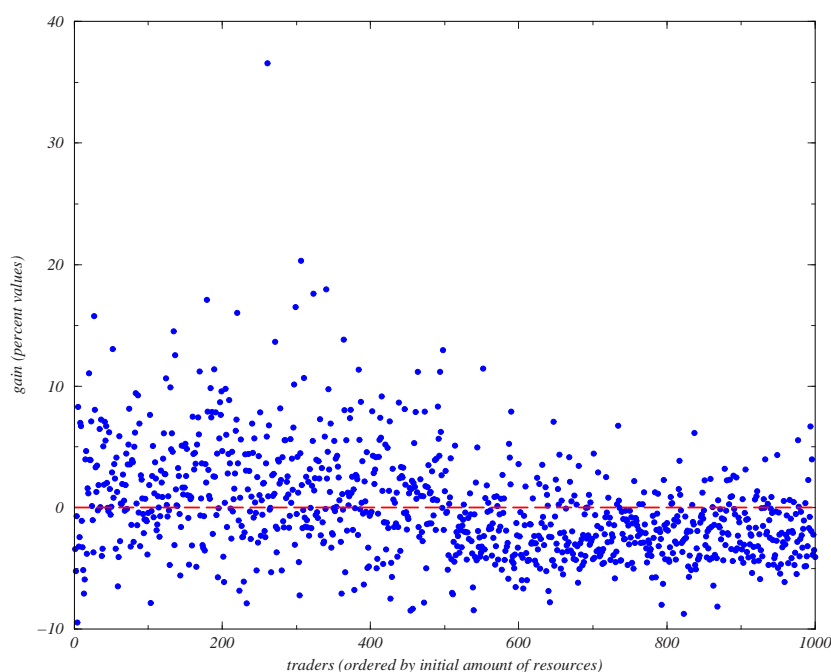
Figure 4.17: Distribution of the gain obtained by each trader.

it. But the distribution depicted in Fig. 4.17 indicates the possibility to derive an optimal strategy, since only few people are really getting money, while the majority of the traders is in the red.

This fact suggests to take a closer look at the best performer in order to learn something. In particular it is interesting to analyze the book and derive the strategy[3] adopted by the best investor. In our model there is only one kind of trader, since we do not make any distinction between chartists and fundamentalists, optimists and pessimists, and so on (see [56] for a good insight on the effect of microscopic diversity). All the investors want to get rich with the smallest possible risk and as fast as possible. They have access to the price history, they can communicate with a limited number of friends, they receive news and advertisements from the market. Then they decide whether to join or not the market and define a price, since the

---

[3]As stressed in [81] all the traders in this model perform according to the same deterministic rules. What makes the difference is the weight given to the opinion of media and acquaintances, the interpretation of the trend in the market price, personal targets and so on. In other words, every investor has a different set of parameters. Deriving the optimal strategy means therefore tuning in a proper way all the degrees of freedom that a trader has still free in the model.

only way to buy and sell some shares is through the book: One has to communicate the quantity and the price of the involved commodity.

Of course in a stockmarket one cannot decide both the price and the time at which to trade. If you specify the moment to buy or sell (performing a so-called *market order*) you do not know exactly at which price you will trade the shares. On the other hand, if you have a given price in mind and you insert a so-called *limit order* you have to wait an amount of time that is not trivial to evaluate. Usually in the first case the price is not so different from the last transaction and in the second situation the time one has to wait grows with the difference between the limit and the market price, namely the requested price and the price involved in the last exchange of shares.

Let us consider the example reported in Fig. 4.18. Here we can see the evolution of the price and the corresponding operations performed by the winner. Upper triangles indicate a buy, lower triangles a sell. A dotted line connects two consecutive operations and has no meaning but visual guide. Since in this example a buy is always followed by a sell, the slope of the above mentioned line indicates the variation of money: If the slope is positive the trader has got money and viceversa. The lower panel reports on the involved volumes: It seems that this agent had a preferred number of shares to be exchanged, namely all the orders involved 6 to 8 pieces.

The trader has performed 24 operations, 12 to buy shares and 12 to sell them. 10 times he has won money, twice he has lost something. It is interesting to comment the two losing cases, since they indicate two different aspects of the adopted strategy. The situation indicated with the letter **A** is a typical stop loss: The agent has bought at a relatively high price in a moment characterized by a high volatility. After a short period of time he has decided to cautionally sell the shares because the market price went down and he wanted to limit the losses. In fact this was a good idea because after that the trader could buy the shares again at a cheaper price. On the other hand, the situation under the letter **B** indicates the intervention of the threshold. Again the trader loses money (although this time a really small amount), but it was a good choice once more. After a while, in fact, he could buy the shares with a strong discount and sell immediately after, getting much more money than he had previuosly lost due to the threshold.

In order to derive the optimal strategy we proceed in the following way. We extract from the book all the transactions where the winner is involved and we
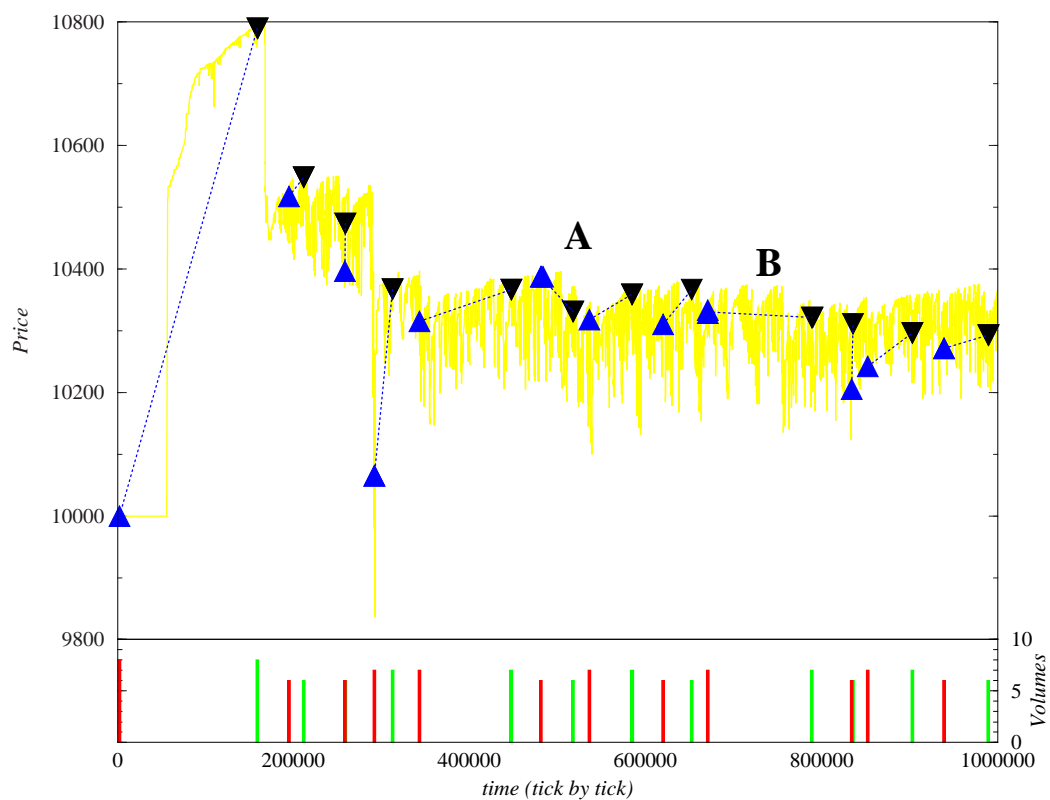
Figure 4.18: Upper panel: Evolution of the market price and the corresponding operations performed by the best trader. Upper triangles indicate buying shares, lower triangles selling shares. Lines connect a buy with the relative sell. The two cases depicted by the letters **A** and **B** are the situations when the agent has lost money. The former is due to a stop loss order, the latter is due to the threshold in time. Lower panel: Number of shares involved in the transactions.

define a trading space. Coordinates of this space are price returns, age of the shares, broadcasted information, remaining cash and so on. We need tens of iterations in order to clearly identify the best performer and hundreds of runs in order to properly populate the trading space with winners from several simulations. After this training period we proceed with a clusterization as in [80] of the collected data in order to get a lookup table to be used in the next step, when the new trader starts to play, too.

A two-dimensional projection of the trading space is reported in Fig. 4.19. Here only the price increment along the x-axis and the age of the share along the y-axis are shown. The first quantity is defined as the logarithmic difference of the price at time t_1 and the price at time t_2, where t_1 and t_2 refer to two consecutive transactions involving the trader under observation. The quantity along the y-axis is just time, here expressed in months due to the considerations reported in [81] about correlation analysis. In Fig. 4.19, a circle is a suggestion to buy shares, a rectangle to sell them. Apart the trivial idea to sell after a price increase and to buy in the complementary case, we can observe some interesting situations. According to the upper big circle (1), one should buy shares after a moderate price increment over a very long time. The second upper ellipse (2) gives similar indications but for the opposite case: If the market price is slowly falling down then it is better to sell before it is too late (anyway between region 2 and region 4 one should buy as by default). The circle labelled by 3 comes from the identification of an upper trend, since the price is raising within a quite short amount of time, a good moment to catch the train and buy shares. The fourth region is a consequence of the introduction of the stop loss price and it suggests to sell before it is too late. The four ellipses are not simmetric with respect to the zero price return and also the dimension and the shape are quite different. This reflects typical asymmetries found in empirical data between positive and negative returns (see [10, 11]).

The lookup table one derives from Fig. 4.19 contains very valuable information beyond the trivial behaviour to buy when cheap, sell when expensive or just follow trends. It is of fundamental importance during the tuning of the parameters characterizing every trader. To show that, let us see the performance of the new agent, whose strategy in nothing but consulting Fig. 4.19 when trading.

The results are shown in Fig. 4.20, where 4 distributions are reported. They have been derived running 200 times the program and individuating, for each simulation,
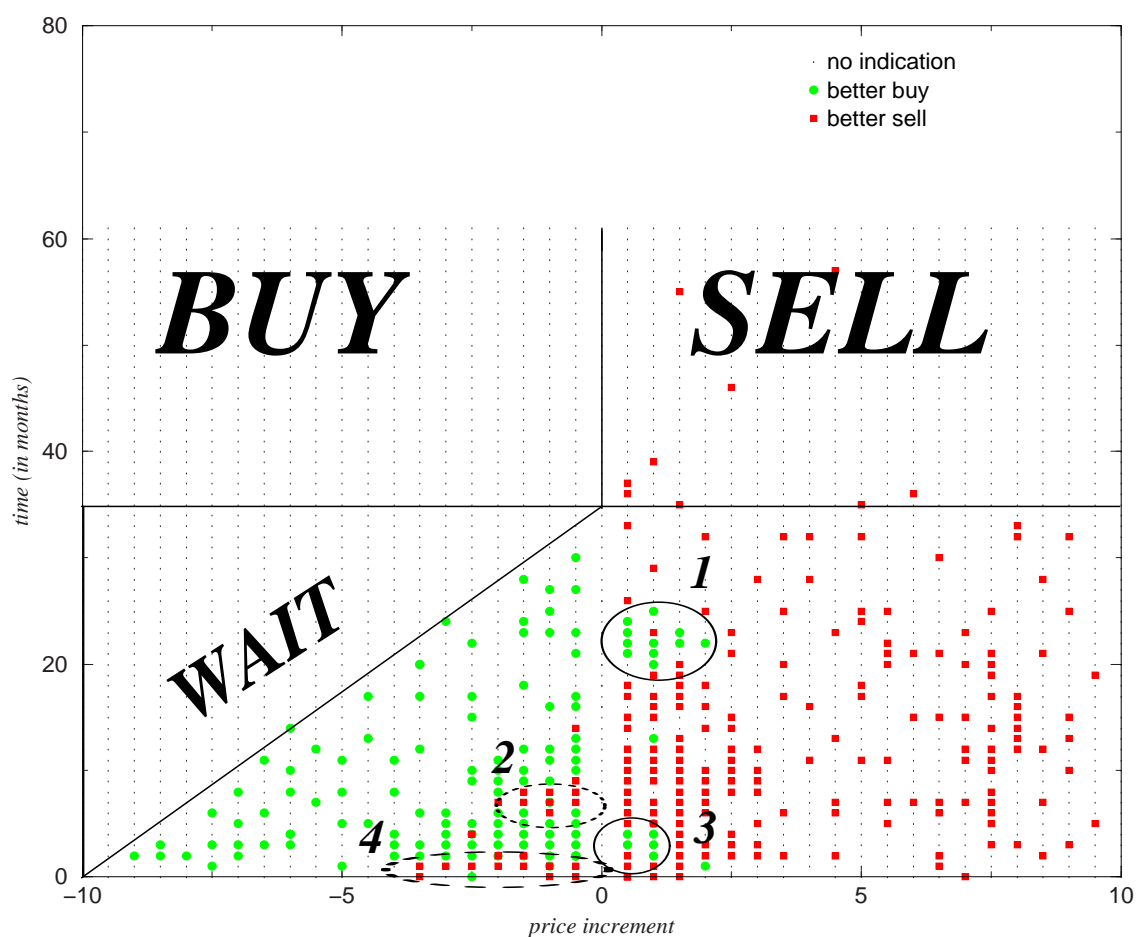
Figure 4.19: Behaviour of the best traders. Projection of the trading space over the (time, return) plane. The trivial recipe *"buy when low, sell when high"* is not enough in order to get good results. One has also to identify slope and duration of trends. Small circles suggest to buy shares, rectangles to sell them. When no clear indication is possible a small dot is reported. There are regions where it is obvious what to do: They are labelled with **BUY**, **SELL** and **WAIT**.
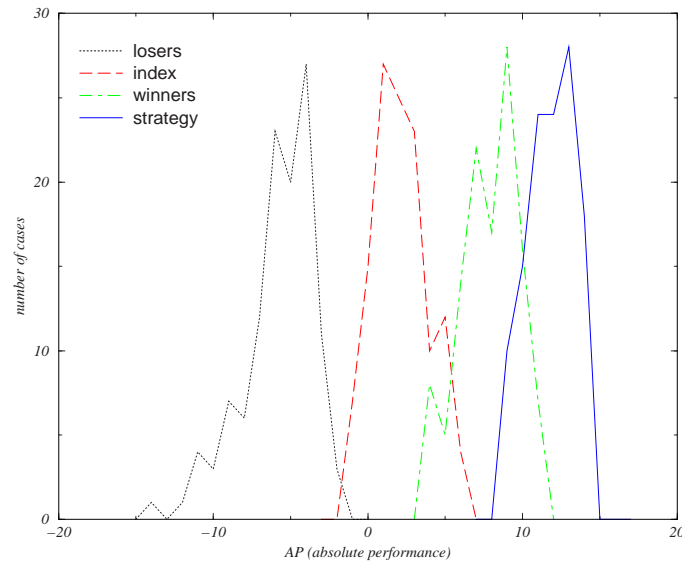
Figure 4.20: Distribution of losers and winners, together with the index (difference between the final and the initial market price) and the performance of the trader adopting the newly developed strategy.

the RP of the worst and the best trader, together with the index performance and the RP of the new agent trading with the help of the lookup table. It is necessary to introduce the index because usually the final market price is bigger than the initial value and one could ask wether it would be more profitable to buy at the beginning a certain number of shares and just keep them till the end. It is clear from Fig. 4.20 that the strategy we have introduced and developed is really useful, since the new trader is systematically able to get more money than the winner, namely than all the other agents.

As a further proof of the efficiency of such a strategy, we consider now a real time series. We have used one company belonging to the *S&P500*, whose temporal evolution of the quoted value is reported in Fig. 4.21. Before being able to make use of the lookup table, one has to tune it on the new time series. More precisely, one has to correctly identify the time scales, namely to be sure that the time reported along the y-axis of Fig. 4.19 corresponds to the time involved on the x-axis of Fig. 4.21. This point is very crucial and unfortunately we do not have any garantee that we manage to do it because of the non-stationarity of such a time series. In any case, in order to apply our strategy to Fig. 4.21, we split the time series in two parts: The
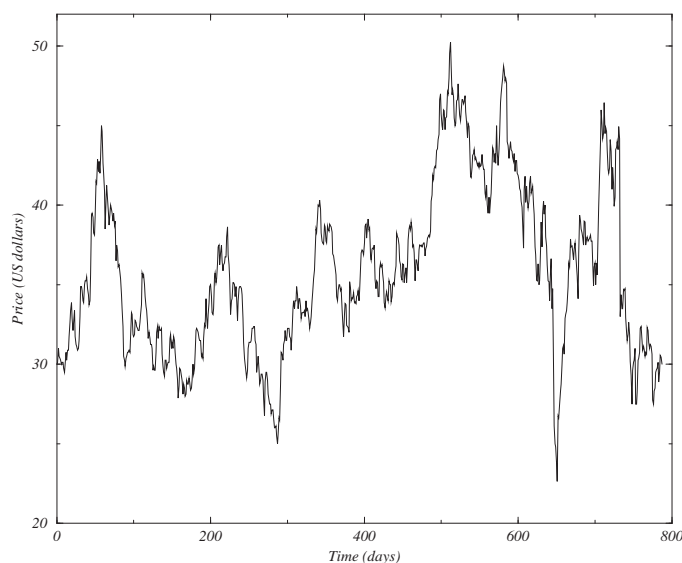
Figure 4.21: Evolution with time of the market price of the shares of a company belonging to the *S&P500* index.

first is used to tune the time with the lookup table, the second is given as input to the lookup table itself. In other words, after the calibration we buy a certain amount of shares (related to the initial resources in the model) and we scroll over the time series to get indications on what to do. If the price increment dp after a time t is such that (dp,t) on Fig. 4.19 belongs to a sell zone then we sell all the shares, if (dp,t) is on a buy region we buy shares with respect to the resources, otherwise we consider the next point of the time series.

The portion of the time series reported in Fig. 4.21 is the one really used for the simulation, having performed the calibration of the lookup table on a previous part of it. We have chosen this portion with the constraint that the final value is identical to the initial because we want to avoid spurious effects due to the presence of a trend. Our strategy, once applied on Fig. 4.21, pays with an RP of +7%. It is not so much if we consider that the involved period of time is longer than three years, but it is without any doubt better than the zero-performance of the index.

There are of course several limitations if one wants to apply such a strategy to empirical situations. Some of them are due to the assumptions of the model (no transaction costs, no arbitrage possibilities, zero execution time) and could be overcome. But at least two others are quite crucial, namely the correct time tuning

between the empirical series and the modelled time and the perturbing effect due to a real trade on a real market. The first task has to be performed very accurately because the use of the lookup table strictly depends on it: A mistake during the *real time - simulated tick* translation reduces the potentially optimal strategy to a nonsense. Unfortunately this point is quite far from being considered solved. The second problem is inherent and it is a common limitation for all the attempts to get a model of the stockmarket.

In conclusion one could say that if people lose money in a stock market, usually they have sold the shares too late. You can buy whenever you want, but once you have bought you should have a clear strategy in mind. Something like the *trading rectangle*.

Chapter 5

Conclusions

We have presented several aspects related to the issue of how stochasticity and non-stationarity can be included into the theoretical framework of non-linear time series analysis and system modelling. On the basis of the classical deterministic viewpoint of Laplace, it is in principle possible to predict the state of a system, at any time, once the evolution laws and the initial conditions are known. However, they are never known with arbitrary precision and if the system is chaotic the prediction of its state is only possible up to a severly bounded time. The study of predictability can also be seen as a way to characterize the complexity of a dynamical system, at least in the sense of information theory, data compression and algorithmic complexity theory.

There are essentially three different kinds of non-stationarity, related to different situations and different approaches to the problem. We have investigated them and addressed related and challenging tasks from real situations.

- The control parameters of a dynamical system may be non-constant during the measurement time. We have seen what happens to a recurrence plot when the time series is generated from a deterministic system, the Hénon map, with several sets of different parameters; how to use the powerful idea of Takens' theorem and build an over-embedding in order to reconstruct an extended phase space for a D-dimensional deterministic dynamical system depending on P parameters with slow time dependence; and finally how to recover the instantaneous dynamics of the system and how to use its redundancy to perform very challenging tasks like noise reduction, data classification and software corrections of vocal pathologies.

- The characteristic joint probabilities of a stochastic process can be time dependent, although all underlying transition probabilities are fixed. This is a typical feature of Brownian motion and its generalizations, processes characterized by a lack of recurrence. We have introduced several tools, like the analysis of the Hurst exponent, to cope with this situation, and analyzed the relation between fractional Brownian motion and Lévy flight. The interactions between stable distributions, scaling behaviour and the divergence of low-order moments play a fundamental role in this context.

- Data may be characterized by trends and/or seasonality, which affect drastically the estimation of self-similarity parameters. Assuming the additivity of trends, the detrended fluctuation analysis is able to remove them and consequently to discover the other source of non-stationarity originating from the lack of recurrence. This tool is very useful when studying financial time series, where the seasonality provided by the high volatility at the begin and at the end of a trading day contributes to the long range correlations expressed, say, by the Hurst exponent.

We have also discussed some models reproducing further essential properties of financial markets. These models (e.g. ARCH and GARCH models) are, by their setting, equipped with time independent parameters, but nevertheless they show a kind of non-stationarity by the fact that large time scales are involved and (anomalous) diffusion drifts of some of their variables are possible. We have investigated asset correlations by interpreting their growth rates as observables of a particle system scenario, being able to separate the collective motion from the single dynamics. We have also proposed a model of stock markets based on the notion of the book, namely the data-base where all buying and selling orders are stored. Such a realistic model may help to better understand the *micro-macro* relation in economics, in particular the translation from single expectations of market participants to a global market price.

Non-stationarity is almost ubiquitous in nature and introduces many complications in the analysis and modelling of real systems; on the other hand if nature were stationary, it would not be as beautiful as it is.

Appendix A

The St. Petersburg Paradox

Risk is not knowing what you are doing.

The St. Petersburg game is played by flipping a fair coin until it comes up tails, and the total number of flips n determines the prize, which equals $\$2^n$. Thus if the coin comes up tails the first time, the prize is $\$2^1 = \$2$, and the game ends. If the coin comes up heads the first time, it is flipped again. If it comes up tails the second time, the prize is $\$2^2 = \$4$, and the game ends. If it comes up heads the second time, it is flipped again. And so on. There are an infinite number of possible sequences (runs of heads followed by one tail) possible. The probability of a sequence of n flips is $P(n) = 1/2^n$, and the expected payoff of each consequence is the prize times its probability, namely it always equals $\$1$.

The *expected value* of the game is the sum of the expected payoffs of all the sequences. Since the expected payoff of all possible sequence is $\$1$, and there are an infinite number of them, this sum is an infinite number of dollars. A rational gambler would enter a game if and only if the price of entry was less than the expected value. In the St. Petersburg game, any finite price of entry is smaller than the expected value of the game. Thus, the rational gambler would play no matter how large the finite entry price was. But it seems obvious that some prices are too high for a rational agent to pay to play. Many commentators agree with the estimation in [33] that few people would pay even $\$25$ to enter such a game. If this is correct, then something has gone wrong with the standard decision-theory calculations of expected value above.

Daniel Bernoulli answered to this problem with the observation that the calcu-

lations make a mistake by adding expected payoffs in dollars, whereas what should be added are the expected utilities of each consequence. He proposed the widely-accepted principle that money has a decreasing marginal utility, and suggested that a realistic measure of the utility of money might be given by the logarithm of the money amount. In such a case the sum of expected utilities is not infinite: It reaches a limit of about 0.60206 utiles (worth $4.00). The rational gambler, therefore, would not pay more than $4 to play and the system would crash, because 25% of the players should get more than $4 out.

This response to the paradox is, however, unsatisfactory. Let us agree that money has a decreasing marginal utility, and accept that a reasonable calculation of the utility of any dollar amount takes the logarithm of the amount in dollars. The St. Petersburg game as proposed, then, presents no paradox, but it is easy to construct another St. Petersburg game which is paradoxical, merely by altering the dollar prizes. Suppose, for example, that instead of paying 2^n for a run of n, the prize were $\$10^{2^n}$. The expected utility would be again 1. This version contains much larger prizes than the original version, and one would presumably be willing to pay more to play this version than the original. But the expected value of this game is infinite, and the paradox returns. Of course, it is not clear how in fact dollar values relate to utility, but we can imagine a generalized paradoxical St. Petersburg game which offers prizes in utiles at the rate of 2^n utiles for a run of n. This game would have infinite expected value, and the rational gambler should pay any amount, however large, to play. For simplicity, we continue to discuss the game in terms of the original dollar prizes, recognizing, however, that the diminishing marginal utility of dollars may make some revision of the prizes necessary to produce the paradoxical result.

Consider the following argument. The St. Petersburg game offers the possibility of huge prizes. A run of forty would, for example, pay a whopping $1.1 trillion. Of course, this prize happens rarely: Only once in about 1.1 trillion times. Half the time, the game pays only $2, and you're 75% likely to wind up with a payment of $4 or less. Your chances of getting more than the entry price of $25 are less than 1/25. Very low payments are very probable, and very high ones very rare. It's a foolish risk to invest more than $25 to play. Many of us are risk-averse, and unwilling to gamble for a very small chance of a very large prize, because the chance is so small. This sort of consideration could solve the St. Petersburg paradox.

But there are objections to this approach. For one thing, a factor for risk-aversion is not a generally applicable consideration in making rational decisions, because some people are not risk averse. In fact, some people may enjoy risk. What should we make, for example, of those people who routinely play state lotteries, or who gamble at pure games of chance in casinos[1]? It's possible to dismiss such behaviour as merely irrational, but sometimes these players offer the explanation that they enjoy the excitement of risk. In any case, it's not at all clear that risk-aversion can explain why the St. Petersburg game would be widely intuited to have a fairly small maximum rational entry fee, while so many people at the same time are not averse to the huge risk entailed by the very small expected probability of large prizes in lotteries.

The St. Petersburg game is sometimes dismissed because it has infinite expected value, which is thought not merely practically impossible, but theoretically objectionable - beyond the reach even of thought-experiment. But is it? Imagine you were offered the following deal. For a price to be negotiated, you will be given permanent possession of a cash machine with the following property: Every time you punch in a dollar amount, that amount is extruded. This is not a withdrawal from your account; neither will you later be billed for it. You can do this as often as you care to. Now, how much would you offer to pay for this machine? Do you find it impossible to perform this thought-experiment, or to come up with an answer? Perhaps you don't, and your answer is: Any price at all. Provided that you can defer payment for a suitable time after receiving the machine, you can collect whatever you need to pay for it from the machine itself.

Of course, there are practical considerations: How long would it take you to collect, say, a trillion dollars from the machine, if this were its price? Would you be worn out or dead by then? Any bank would be crazy to offer to sell you an infinite cash machine, and unfortunately the address of the crazy bank which has made this offer has been lost. Anyway, there appears to be nothing wrong with this thought experiment: It imagines an action - buying the machine - with no upper limit on expected value. We easily ignore practical considerations when calculating the expected value (in this case, merely potential withdrawals minus purchase price), which is infinite. Do your intuitions tell you to offer, say, $25 at most for this

---

[1]In these games, the entry fee is greater than the expected utility.

machine? But the only difference between this machine and a single-play St. Peters-
burg game is that this machine guarantees an indefinitely large number of payouts,
while the game offers a one-time lottery from among an indefinitely large number of
possible payouts, each with a certain probability. *The only difference between them
is the probability factor*: The same difference that exists between a game which gives
you a guaranteed prize of $5, and one which gives you half a chance of $10, and half
a chance of $0. The expected value of both the St. Petersburg game and the infinite
cash machine are both indefinitely large. You should offer any price at all for either.
It appears, then, that the notion of infinite expected value is perfectly reasonable.

There are all sorts of practical considerations which must be considered in making
a real gambling decision. For example, in deciding whether to raise, see, fold, or cash
in and go home, in a particular poker game, you must consider not only probability
and expected value, but also the facts that it's 5 A.M. and you are cross-eyed from
fatigue and drink; but it's not expected that classical decision theory has to deal
with these. The St. Petersburg game commits participants to doing what we know
they will not. The casino may have to pay out more than it has. The player may
have to flip a coin longer than physically possible. But this may not show a defect
with choice theory. Classical unrestricted theory is still serving its purpose, which
is *modelling the abstract ideal rational agent.* It tells us that no amount is too great
to pay as a ideally rationally acceptable entrance fee, and this may be right. What
it's reasonable for real agents, limited in time, patience, bankroll, and imaginative
capacity to do, given the constraints of the real casino, the real economy, and the
real earth, is another matter, one that the theoretical core of decision theory can
be forgiven for not specifying. From this point of view, the St. Petersburg paradox
does not point out any defect with classical decision theory, and it is not a paradox
after all[2].

Relating this game to Lévy distributions introduced in Sec. 2.6, we can observe
that here the rather unprobable tails are so *long* that even the mean is infinite. The
Lévy solution to the paradox arises from the idea of *finite size regularization* and
asserts that players should flip the coin up to n_M times. In this case the fair entry
price would be exactly $\$n_M$ and gamblers could even choose what the maximal gain
$\$2^{n_M}$ should be.

---

[2]One could say that the attempt to determine the entry price for the rational gambler is equiv-
alent to determine a characteristic scale for a problem that has no characteristic scale.

Bibliography

[1] H.D.I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer, New York, 1996.

[2] E.R.M. Abberton, D.M. Howard and A.J. Fourcin, *Laryngographic assessment of normal voice: A tutorial*, Clin. Ling. Phon. 281-296 (1989).

[3] M. Ausloos, *Statistical physics in foreign exchange currency and stock markets*, Physica A **285**, 48-65 (2000).

[4] P. Bak, M. Paczuski and M. Shubik, *Price variations in a stock market with many agents*, Physica A **246**, 430-453 (1997).

[5] A. Belenkiy, *Inner markets as a "Black Box"*, cond-mat/0106401 (2001).

[6] J. van den Berg, *Myoelastic-aerodynamic theory of voice production*, J. Speech Hearing Res. **1**, 227-244 (1958).

[7] D.A. Berry, I.R. Titze, H. Herzel and K. Krischer, *Interpretation of biomechanical simulations of normal and chaotic vocal fold oscillations with empirical eigenfunctions*, J. Acoust. Soc. Am. **95**, 3595-3604 (1994).

[8] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics **31**, 307-327 (1986).

[9] T. Bollerslev, R.Y. Chou and K.F. Kroner, *ARCH modelling in finance: A review of the theory and empirical evidence*, Journal of Econometrics **52**, 5-59 (1992).

[10] J.P. Bouchaud and M. Potters, *Theory of financial risks. From statistical physics to risk management*, Cambridge University Press, 2000.

[11] J.P. Bouchaud, A. Matacz and M. Potters, *The leverage effect in financial markets: Retarded volatility and market panic*, cond-mat/0101120 (2001).

[12] W. A. Brock, W. D. Dechert, B. D. LeBaron, and J. A. Scheinkman, Econometric Reviews **15**, 197 (1996).

[13] O. Cappé, *Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor*, Trans. Speech Audio Process., 345-349 (1994).

[14] M. Casdagli, *Recurrence plot revisited*, Physica D **108**, 12-44 (1997).

[15] J.P. Crutchfield and B.S. McNamara, *Equations of motion from data series*, Complex Systems **1**, 417 (1987).

[16] G. Cuniberti, M. Raberto, and E. Scalas, Physica A **269**, 90 (1999).

[17] G. Cuniberti and L. Matassini, *Liquid Markets and Market Liquids*, The European Physical Journal B **20**, 561 (2001).

[18] J.P. Eckmann, S.O. Kamphorst and D. Ruelle, *Recurrence plots of dynamical systems*, Europhys. Lett. **4**, 973 (1987).

[19] E. J. Elton and M. J. Gruber, *Modern Portfolio Theory and Investment Analysis*, Wiley, 1995.

[20] R.F. Engle, *Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation*, Econometrica **50**, 987-1008 (1982).

[21] R.F. Engle and G. Gonzalez-Riviera, *Semiparametric ARCH models*, Journal of Business and Economic Statistics **9**, 345-360 (1991).

[22] R.F. Engle and C. Mustafa, *Implied ARCH models from option prices*, Journal of Econometrics **52**, 289-311 (1992).

[23] Y. Ephraim and D. Malah, *Speech enhancement using a minimum mean square error short time spectral amplitude estimator*, IEEE Trans. Acoust. Speech Signal Process. **32**, 1109-1121 (1984).

[24] E. F. Fama, The Journal of Business **38**, 34 (1965).

[25] E. F. Fama, Journal of Financial Economics **49**, 283 (1998).

[26] J.D. Farmer and J.J. Sidorowich, *Predicting chaotic time series*, Phys. Rev. Lett. **59**, 845 (1987).

[27] F. Franci, R. Marschinski and L. Matassini, *Learning the optimal trading strategy*, Physica A **294**, 213 (2001).

[28] F. R. Gantmacher, *The Theory of Matrices*, Chelsea, 1990.

[29] L. Glass, M.R. Guevara, A. Shrier and R. Perez, *Bifurcation and chaos in periodically stimulated cardiac oscillator*, Physica D **7**, 89-101 (1983).

[30] B.W. Gnedenko and A.N. Kolmogorov. *Limit distributions for sums of independent random variables*, Addison Wesley, Redwood City, 1954.

[31] P. Grassberger, R. Hegger, H. Kantz, C. Schaffrath and T. Schreiber, *On noise reduction methods for chaotic data*, CHAOS **3**, 127 (1993).

[32] T. Gülzow, A. Engelsberg, U. Heute, *Comparison of a discrete wavelet transformation and a non-uniform polyphase filterbank applied to spectral-subtraction speech enhancement*, Signal Processing **64**, 5-19 (1998).

[33] J. Hacking, *Strange Expectations*, Phylosophy of Science **47**, 562-567 (1980).

[34] H. Haken, *Synergetics: An introduction*, Springer, Berlin, 1983.

[35] R. Hegger, H. Kantz, F. Schmüser, M. Diestelhorst, R.P. Kapsch and H. Beige, *Dynamical properties of a ferroelettric capacitor observed through non-linear time series analysis*, CHAOS **8**, 727 (1998).

[36] R. Hegger, H. Kantz, and L. Matassini, *Denoising human speech signals using chaoslike features*, Phys. Rev. Lett. **84**, 3197 (2000).

[37] R. Hegger, H. Kantz, L. Matassini, T. Schreiber, *Coping with non-stationarity by over-embedding*, Phys. Rev. Lett. **84**, 4092 (2000).

[38] R. Hegger, H. Kantz and L. Matassini, *Noise Reduction for Human Speech Signals by Local Projections in Embedding Spaces*, IEEE TCAS-I, accepted for publication (2001).

[39] H. Herzel, D. Berry, I.R. Titze and M. Saleh, *Analysis of Vocal Disorders with Methods from Non-linear Dynamics*, Journal of Speech and Hearing Research **37**, 1008-1019 (1994).

[40] H. Herzel, D. Berry, I. Titze and I. Steinecke, *Non-linear dynamics of the voice: Signal analysis and biomechanical modelling*, CHAOS **5**, 30-34 (1995).

[41] K. Hu, P.C. Ivanov, Z. Chen, P. Carpena and H.E. Stanley, *Effect of trends on detrended fluctuation analysis*, arXiv:physics/0103018 v2 (2001).

[42] W. Huadong, M. Siegel and P. Khosla, *Vehicle sound signature recognition by frequency vector principal component analysis*, IEEE Transactions on Instrumentation and Measurement **48**, 1005-1009 (1999).

[43] H.E. Hurst, Trans. Amer. Soc. Civil Eng. **116**, 770 (1951).

[44] K. Ishizaka and J.L. Flanagan, *Synthesis of voiced sounds from a two-mass model of the vocal cords*, Bell. Syst. Tech. J. **51**, 1233-1268 (1972).

[45] H. Isliker and J. Kurths, *Int. J. Bifurcation Chaos Appl. Sci. Eng.* **3**, 157 (1993).

[46] H. Kantz and T. Schreiber, *Non-linear Time Series Analysis*, Cambridge University Press, Cambridge, 1997.

[47] H. Kasuya, S. Ogawa, K. Mashima and S. Ebihara, *Normalized noise energy as an acoustic measure to evaluate pathologic voice*, J. Acoust. Soc. Am. **80**, 1329-1334 (1986).

[48] A.W. Kelman, *Vibratory pattern of the local folds*, Folia Phoniatr. **33**, 973-991 (1981).

[49] S.M. Kogon and D.G. Manolakis, *Signal modelling with self-similar α-stable processes: The fractional Lévy stable motion model*, IEEE trans. on Signal Processing **44**, 1006-1010 (1996).

[50] A.N. Kolmogorov, *A new invariant of transitive dynamical systems*, Dokl. Akad. Nauk. SSSR **119**, 861 (1958).

[51] E.J. Kostelich and T. Schreiber, *Noise reduction in dynamical systems*, Phys. Rev. E **48**, 1752 (1993).

[52] D. Kugiumtzis, *State space reconstruction parameters in the analysis of chaotic time series - the role of the time window length.* Physica D **95**, 13 (1995).

[53] A. Kumar and S.K. Mullick, *Non-linear dynamical analysis of speech*, J. Acoust. Soc. Am. **100**, 615-629 (1996).

[54] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, Physical Review Letters **83**, 1467 (1999).

[55] C.G. Lamoureux and W.D. Lastrapes, *Persistence in variance, structural change and the GARCH model*, Journal of Business and Economic Statistics **8**, 225-234 (1990).

[56] M. Levy, H. Levy and S. Solomon, *Microscopic simulation of the stock market: The effect of microscopic diversity*, J. Phys. I France **5**, 1087-1107 (1995).

[57] F. Lillo and R.N. Mantegna, *Variety and volatility in financial markets*, cond-mat/0006065 (2000).

[58] B. Linders, G.G. Massa, B. Boersma and P.H. Dejonckere, *Fundamental voice frequency and jitter in girls and boys measured with electroglottography: Influence of age and height*, International Journal of Pediatric Otorhinolaryngology **33**, 61-65 (1995).

[59] A.W. Lo, Econometrica **59**, 451-474 (1991).

[60] A.W. Lo, Econometrica **59**, 1279 (1991).

[61] E.N. Lorenz, *Deterministic non-periodic flow*, J. Atmos. Sci. **20**, 130-141 (1963).

[62] T. Lux, *The socio-economic dynamics of speculative markets: Interacting agents, chaos, and the fat tails of return distributions*, Journal of Economic Behaviour & Organization **33** 143-165 (1998).

[63] T. Lux and M. Marchesi, *Scaling and criticality in a stochastic multi-agent model of a financial market*, Nature **397**, 498-500 (1999).

[64] B.B. Mandelbrot, The Journal of Business **36**, 394 (1963).

[65] B.B. Mandelbrot and J.W. Van Ness, *Fractional Brownian motions, fractional noises and applications*, SIAM Review **10**, 422-437 (1968).

[66] B.B. Mandelbrot and J.R. Wallis, Water Resour. Res. **4**, 967 (1968).

[67] C. Manetti, M.A. Ceruso, A. Giuliani, C.L. Webber and J.P. Zbilut, *Recurrence quantification analysis as a tool for characterization of molecular dynamics simulations*, Physical Review E **59**, 992-998 (1999).

[68] C. Manfredi, P. Bruscaglioni, M. D'Aniello, L. Pierazzi and A. Ismaelli, *Pitch and noise estimation in hoarse voices*, Proc. Int. Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications, Firenze, 42-47 (1999).

[69] R.N. Mantegna and H.E. Stanley, *Stochastic process with ultraslow convergence to a gaussian: The truncated Lévy flight*, Physical Review Letters **73**, 2946 (1994).

[70] R.N. Mantegna and H.E. Stanley, *Scaling behaviour in the dynamics of an economix index*, Nature **376**, 46-49 (1995).

[71] R.N. Mantegna and H. E. Stanley, *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, 1999.

[72] R.N. Mantegna, The European Physical Journal B **11**, 193 (1999).

[73] R. Manuca and R. Savit, *Stationarity and non-stationarity in time series analysis*, Physica D **99**, 134 (1996).

[74] R. Manuca, L. Yi and R. Savit, *The structure of adaptive competition in minority games*, Physica A **282**, 559-608 (2000).

[75] H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, Wiley, 1959.

[76] M. Marsili and Y.C. Zhang, *Interacting individuals leading to Zipf's law*, Physical Review Letters **80**, 2741-2744 (1998).

[77] G.A. Martynov, *Fundamental Theory of Liquids: Method of Distribution Functions*, Adam Hilge, 1992.

[78] S. Maslov, Physica A **278**, 571-578 (2000).

[79] L. Matassini, R. Hegger and H. Kantz, *Filtering of Speech Signals by Over-Embedding*, in D.S. Broomhead, E.A. Luchinskaya, P.V.E. McClintock and T. Mullin, ed., Stochastic and Chaotic Dynamics in the Lakes, American Institute of Physics, Melville, NY, USA, 642, (2000).

[80] L. Matassini, C. Manfredi, R. Hegger and H. Kantz, *Analysis of Vocal Disorders in a Feature Space*, Medical Engineering and Physics **22**, 413-418 (2000).

[81] L. Matassini and F. Franci, *On financial markets trading*, Physica A **289**, 526-542 (2001).

[82] L. Matassini and C. Manfredi, *Software corrections of vocal disorders*, Computer Methods and Programs in Biomedicine, to appear (2001).

[83] L. Matassini, H. Kantz, J. Hołyst and R. Hegger, *Optimizing Recurrence Plots for Noise Reduction*, submitted to Physical Review E (2001).

[84] L. Matassini, *The Trading Rectangle Strategy within Book Models*, submitted to Physica A (2001).

[85] L. Matassini and G. Cuniberti, in preparation (2001).

[86] J. Moody and W. Lizhong, *High frequency foreign exchange rates: Price behaviour analysis and true price models*, in Non-linear modelling of high frequency financial time series, edited by C. Dunis and B. Zhou, Wiley, 1998.

[87] S.S. Narayanan and A. Alwan, *A non-linear dynamical systems analysis of fricative consonants*, J. Acoust. Soc. Am. **97**, 2511-2524 (1995).

[88] K. Okuyama, M. Takayasu and H. Takayasu, *Zipf's law in income distribution of companies*, Physica A **269**, 125-131 (1999).

[89] M. Paczuski, K.E. Bassler and A. Corral, *Self-organized networks of competing boolean agents*, Physical Review Letters **84**, 3185-3188 (2000).

[90] C.K. Peng, J. Mietus, J.M. Hausdorff, S. Havlin, H.E. Stanley and A.L. Goldberger, *Long-range anticorrelations and non-gaussian behaviour of the heartbeat*, Physical Review Letters **70**, 1343 (1993).

[91] C.K. Peng, S. Buldyrev, S. Havlin, M. Simons, H.E. Stanley and A.L. Goldberger, *Mosaic organization of DNA nucleotides*, Physical Review E **49**, 1685-1689 (1994).

[92] N.B. Pinto and I.R. Titze, *Unification of perturbation measures in speech signals*, Journal of the Acoustical Society of America **87**, 1278-1289 (1990).

[93] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 1992.

[94] G.M. Raymond and J.B. Bassingthwaighte, *Deriving dispersional and scaled windowed variance analyses using the correlation function of discrete fractional Gaussian noise*, Physica A **265**, 85-96 (1999).

[95] R. Reuter, H. Herzel and R. Orglmeister, *Simulations of Vocal Fold Vibrations with an Analog Circuit*, Int. J. Bifurcation and Chaos **9** (1999).

[96] P. A. Samuelson, Industrial Management Review **6**, 41 (1965).

[97] T. Sauer, J. Yorke, and M. Casdagli, *Embedology*, J. Stat. Phys. **65**, 579 (1991).

[98] T. Schreiber and D. Kaplan, *Signal separation by non-linear projections: The fetal electrocardiogram*, Phys. Rev. E **53**, 4326 (1995).

[99] T. Schreiber, and H. Kantz, *Dimension estimates and physiological data*, Chaos **5**, 143-154 (1995).

[100] A.G. Sinai, *On the concept of entropy of a dynamical system*, Dokl. Akad. Nauk. SSSR **124**, 768 (1959).

[101] I.Y. Soon, S.N. Koh and C.K. Yeo, *Noisy speech enhancement using discrete cosine transform*, Speech communication **24**, 249-257 (1998).

[102] D. Sornette and A. Johansen, *Large financial crashes*, Physica A **245**, 411-422 (1997).

[103] J. Stark, D.S. Broomhead, M.E. Davies and J. Huke, *Takens embedding theorems for forced and stochastic systems*, Non-linear Analysis **30**, 5303-5314 (1997).

[104] I. Steinecke and H. Herzel, *Bifurcations in an asymmetric vocal-fold model*, J. Acoust. Soc. Am. **97**, 1874-1884 (1995).

[105] H. Takayasu and M. Takayasu, *Critical fluctuations of demand and supply*, Physica A **269**, 24-29 (1999).

[106] F. Takens, *Detecting strange attractors in turbulence*, Lecture Notes in Math. **898**, Springer, New York 1981.

[107] C.S. Tapiero and P. Vallois, *Run Length Statistics and the Hurst Exponent in Random and Birth-Death Random Walks*, Chaos, Solitons & Fractals **7**, 1333-1341 (1996).

[108] I. Titze, *Principles of voice production*, Prentice-Hall Inc., Englewood Cliffs, 1994.

[109] L.L. Trulla, A. Giuliani, J.P. Zbilut and C.L. Webber, *Recurrence quantification analysis of logistic equation with transients*, Physics Letters A **223**, 43-47 (1996).

[110] D. Tzovaras and M.G. Strintzis, *Use of non-linear principal component analysis and vector quantization for image coding*, IEEE Transactions on Image Processing **7**, 1218-1223 (1998).

[111] A. Witt, J. Kurths and A. Pikovsky. *Testing stationarity in time series*, Physical Review E **58**, 1800-1810 (1998).

[112] Z.M. Yin, *New methods for simulation of fractional Brownian motion*, Journal of Computational Physics **127**, 66-72 (1996).

[113] J.P. Zbilut, A. Giuliani, and C.L. Webber, *Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification*, Physics Letters A **246**, 122-128 (1998).

[114] Y.C. Zhang, *Why financial markets will remain marginally inefficient*, cond-mat/0105373 (2001).

Acknowledgements

> *There is nothing more to say.*
> *(Madonna, The power of goodbye)*

me several ways to lose money with stock markets. Tommaso Agnoloni and Dimitris Kugiumtzis are a precious source of constructive criticisms. Robert Marschinski, Angelo Valleriani and Luis Craco are sharing with me the beautiful experience of putting together the latin and the germanic culture.

A special thought to my (new) family, growing up together with this thesis: Maud, with whom I share the most beautiful moments of my life and with whom I afford the sad side of it; Leonardo, born at the beginning of my PhD studies and now contributing to the non-stationary and fashinating aspects of my fatherhood; Marvin, born few months ago when this work was finished in almost all its parts and still too young to understand the concepts developed here, but able to provide a lot of empirical data.

Lorenzo Matassini

Persönliches

- Geboren am 9. April 1973 in Montevarchi, Italien
- Staatsangehörigkeit: italienisch
- Familienstand: ledig, 2 Kinder

Ausbildung

September 1979 – Juni 1984

    Scuola Elementare "Pestello", Montevarchi (Grundschule I)

September 1984 – Juli 1987

    Scuola Media "Francesco Petrarca", Montevarchi (Grundschule II)

September 1987 – Juli 1992

    Istituto Tecnico Industriale "Galileo Ferraris", San Giovanni Valdarno

    Studienrichtung Informatik. Abschlussnote 60/60 mit Auszeichnung

    Abitur, mit Auszeichnung

Oktober 1992 – November 1998

    Università degli Studi di Firenze, Facoltà di Ingegneria, Indirizzo Automazione Industriale

    Studium der Ingenieurwissenschaft, TU Florenz. Abschlussnote 110/110 mit Auszeichnung

    Diplom-Ingenieur, mit Auszeichnung

seit Dezember 1998

    Doktorand in der Arbeitsgruppe "nichtlineare Zeitreihenanalyse" bei PD Dr. Hol-

ger Kantz (Max-Planck-Institut für Physik komplexer Systeme, Dresden). Abschluss "Dr. rer. nat." voraussichtlich August 2001

Wettbewerbe und Auszeichnungen

- Rotary Award, Bester Schüler der Oberschule "Galileo Ferraris" (1992)

- Aufgrund guter Leistungen Stipendium zur Erlassung der Studiengebühren (1992 - 1997)

- Stipendium "Socrates-Erasmus" für die Diplomarbeit, TU Florenz und Max-Planck-Institut für Physik komplexer Systeme Dresden (März - Oktober 1998)

- Stipendium "Young Scientist" für die Teilnahme an der Konferenz "EUROAT-TRACTOR2000", Warschau (2000)

- Stipendium der NATO für die Teilnahme am Advanced Research Workshop in "Application of Physics in Economic Modelling", Prag (2001)

Berufserfahrung

November 1996 – Juni 1997

  Autohändler bei Citroën, eigenverantwortliche Kundenbetreuung und Verkauf

August 1997 – Februar 1998

  Kundenberater für Omnitel-Vodafone, Beratung und Aquisierung von neuen Kunden, Aktualisierung und Pflege der Intranet-Datenbank der Firma

April 2000 – Juli 2000

  Teaching and Research Assistant 'Non-linear Dynamics', TU Florenz

Leistungen in Forschung und Lehre

- Deutsches Patent Nr. 198 59 174.8, 'Verfahren und Vorrichtung zur Verarbeitung rauschbehafteter Schallsignale' (1999)

- U.S. Patent 14815, 'Method and Apparatus for Processing Noisy Sound Signals' (1999)

- Staatliche Prüfung zur Erlangung der Lehrbefugnis in Physik für Sekundarstufe II in Italien (2000)

- 12 Veröffentlichungen in Fachzeitschriften und Sammelbänden

- 24 Vorträge auf internationalen Kongressen und an Forschungseinrichtungen in 11 Ländern
- Fachgutachter für *Physica A* und *American Institute of Physics*

Spezialkenntnisse

- Erfahrung mit diversen Betriebssystemen (Unix, DOS/Windows, Linux, QNX), Programmiersprachen (C/C++, Fortran, Pascal), wissenschaftlichen Textsatzsystemen (LaTeX), Datenbanksystemen (SQL, DBIII, Clipper) und MS-Office
- Dreisprachig (Italienisch, Englisch, Deutsch) plus Grundkenntnisse Französisch

Hobbies

- Argentinischer Tango (Mitglied "Gare de la Lune" Dresden)
- Marathon (Bestzeit von 3h25 beim Lauf von Florenz 1998), Triathlon und MTB
- Autos und Motorräder
- Online Börsenhandel