

SIXTH FRAMEWORK PROGRAMME

Project contract no. 043363

MANMADE

Diagnosing vulnerability, emergent phenomena and volatility in man-made networks**SPECIFIC TARGETED PROJECT***NEST PATHFINDER**Sub-Priority Tackling Complexity in Science***Work Package 5 – D5.3*****The study of energy spot prices and their correlation with faults in the Nordic region applying Recurrence Quantification and Cross Recurrence Analysis.***

Due date of delivery: Month 24

Actual submission date: 7th January 2009

Start date of project: 1st of January 2007

Duration: 36 months

Lead authors for this deliverable: [F. Strozzi (Università Cattaneo – LIUC), J.M. Zaldívar, (JRC-Ispra)]

	Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)	
	Dissemination Level	
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

D5.3:

The study of energy spot prices and their correlation with faults in the Nordic region applying Recurrence Quantification and Cross Recurrence Analysis.

Fernanda Strozzi, José Manuel Zaldívar*

LIUC University, Castellanza, VA, Italy

*JRC, Ispra, VA, Italy

PREFACE

This work is composed of two parts. The first consists in the application of non-linear time series analysis techniques to the Nordic spot electricity market and the second in the study of the correlation between disturbances and prices. Both studies are performed using the data respectively of spot prices, Total consumption and Disturbances in the Nordic Region publicly available from www.nordpool.com and www.nordel.org. In both parts we have applied, together with other techniques, Recurrence Quantification Analysis (RQA), that in the case of different time series becomes Cross Recurrence Plot Analysis (CRP).

In the first part of the work, we have studied the electricity spot prices Recurrence Plot that allows to a new representation of data in which new measures can be applied (Determinism and Laminarity) and they have demonstrated to be able to distinguish between real and surrogate (random Gaussian with the same FFT) data. Moreover they give a new measure of volatility that takes into account the dynamic properties and not only the statistical distribution of the data.

The second part studies the correlation between electricity prices and disturbances. In this case Cross recurrence plot allows, given two time series to identify a shift and a temporal window on which both series are linearly correlated.

Keywords: Recurrence Plot, RQA, CRP, linear correlation analysis.

Table of Contents

1. APPLICATION OF NON-LINEAR TIME SERIES ANALYSIS TECHNIQUES TO THE NORDIC SPOT ELECTRICITY MARKET DATA

- 1.1. R/S analysis confirms long range correlation and anti-persistence
- 1.2. Stable distribution fitting.
- 1.3. Recurrence Quantification Analysis
- 1.4. *Determinism* and *Laminarity* as volatility measures.

2. CORRELATION ANALYSIS BETWEEN FAULTS IN THE ELECTRICITY GRID AND ELECTRICITY PRICES IN THE NORDIC REGION

- 2.1. Linear correlation coefficient: correlation matrix.
- 2.2. Principal Component Analysis
- 2.3. Cross Correlation function
- 2.4. Cross Recurrence Plot

3. CONCLUSIONS

REFERENCES

- ANNEX I. *Application of non-linear time series analysis techniques to the Nordic spot electricity market*. LIUC Paper 200. October 2007.
- ANNEX III. *Measuring volatility in the Nordic spot electricity market using Recurrence Quantification Analysis*. Eur. Phys. J. Special Topics 164 (2008), 105-115.
- ANNEX III. *Time series analysis and long range correlations of Nordic spot electricity market data*. Physica A 387 (2008), 6567-6574
- ANNEX IV. *Correlation analysis between electricity spot prices and faults in the electricity grid in the Nordic region*. LIUC paper xxx (2009)
- .

1. APPLICATION OF NON-LINEAR TIME SERIES ANALYSIS TECHNIQUES TO THE NORDIC SPOT ELECTRICITY MARKET DATA

This is a summary of the extended work contained in Strozzi *et al* 2007, Strozzi *et al* 2008 and partially in Erzgräber *et al.*, 2008. (Annexes I -III). In these two works, we have applied non-linear time series techniques to the Nordic spot electricity market data. The time series are given in two periods, from May 1992 to December 1998 in Norwegian Kröne per MWh and from January 1999 to January 2007 in EUR per MWh. First, a preliminary study was carried out with the aim of characterising the time series in terms of long term memory (R/S analysis), and tails (stable distributions). Surrogate time series were also generated to test if the original time series were similar to a stationary Gaussian linear process. In a second step, state space reconstruction parameters: time delay and embedding dimension were used to carry out the analysis of these two series in the reconstructed state space. We applied Recurrence Quantification Analysis (RQA), which is based on the definition of several parameters that allows the quantification of the Recurrence Plots (RP). The RQA analysis of both time series and in particular *determinism* and *laminarity* has shown the ability to distinguish between real and surrogate data and to measure the financial volatility.

1.1. R/S analysis confirms long range correlation and anti-persistence

A tool for studying long-term memory and fractality of a time series is the Rescaled Range analysis (R/S analysis) first introduced by Hurst (1951) in hydrology. Mandelbrot (1983) argued that R/S analysis is a more powerful tool in detecting long range dependence compared to more conventional analysis like autocorrelation analysis, variance ratios and spectral analysis. In this method, one measures how the range of cumulative deviations from the mean of the series is changing with the time. It has been found that, for some time series, the dependence of R/S on the number of data points (or time) follows an empirical power law described as $(R/S)_n = (R/S)_0 n^H$, where $(R/S)_0$ is a constant, n is the time index for periods of different length, and H is the Hurst exponent. $(R/S)_n$ is defined as

$$\left(\frac{R}{S}\right)_n = \frac{\max_{1 \leq t \leq n} A(t, n) - \min_{1 \leq t \leq n} A(t, n)}{\sqrt{\frac{1}{n} \sum_{t=1}^n (s(t) - \langle s \rangle_n)^2}} \quad (1)$$

where $A(t, n)$ is the accumulated departure of the time series $s(t)$ from the time average over the time

$$\text{interval } n: \langle s \rangle_n \quad A(t, n) = \sum_{i=t}^{t+n} (s(i) - \langle s \rangle_n).$$

The Hurst exponent, $0 \leq H \leq 1$, is equal to 0.5 for random walk time series, <0.5 for anticorrelated series, and >0.5 for positively correlated series.

A long memory process is a process with a random component, where a past event has a decaying effect on future events. The process has some memory of past events, which is "forgotten" as time moves forward. The Hurst exponent has a relationship with the rate at which the correlation function decays. In this work we have calculated Hurst exponent for the given time series and a set of their surrogate (random Gaussian with the same FFT). All the time series show antipersistence i.e. $H < 0.5$. This has already been found by several authors. The results show that the H exponent of surrogates are slightly lower than the one of the correspondent real data but it is not possible to find a numerical H values that separate real from surrogate data. Different methods to calculate Hurst exponent these electricity prices are applied in Erzgräber et al. (2008).

1.2. Stable distribution fitting.

Stable distributions are a class of distributions that include Gaussian, Cauchy and Levy distributions. They allow skewness and heavy tails. The general stable distribution is described by four parameters the first two are $\alpha \in (0, 2]$, an index of stability and $\beta \in [-1, 1]$, a skewness parameter. α and β determine the shape of the distribution. The last parameters are $\gamma \in [0, \infty)$ a scale parameter and $\delta \in (-\infty, \infty)$ a location parameter. There are no closed formulas for density and distribution function with the exception of Gaussian, Levy and Cauchy.

Stable distributions have been proposed as a model for many types of physical and economic systems because many large data sets exhibit heavy tails and skewness. Anyway, while non-Gaussian stable distributions are heavy tailed, most heavy-tailed distributions are not stable. Stable distributions have the important property of stability: if a number of independent and identically distributed (iid) random variables have a stable distribution, then a linear combination of these variables will have the same distribution, except for possibly different shift and scale parameters.

A stable probability distribution is defined by the Fourier transform of its characteristic function $\varphi(t)$:

$$f(x; \alpha, \beta, \gamma, \delta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(t) e^{-itx} dt \quad (2)$$

where $\varphi(t)$ is given by

$$\varphi(t) = \exp[it\delta - |t|^\alpha (1 - i\beta \operatorname{sgn}(t)\Phi)] \quad (3)$$

and $\operatorname{sgn}(t)$ is just the sign of t and Φ is given by

$$\Phi = \tan(\pi\alpha/2) \quad (4)$$

for all α except $\alpha=1$ in which case:

$$\Phi = -(2/\pi)\log(t) \quad (5)$$

The heavy tail behaviour causes the variance of stable distribution to be infinite for $\alpha < 2$ (for $\alpha = 2$ is Gaussian).

There is no general analytic expression for a stable distribution. There are, however four special cases which can be analytically expressed:

a/ for $\alpha = 2$ the distribution becomes a Gaussian distribution with variance $\sigma^2 = 2\gamma^2$ and mean δ

b/ for $\alpha = 1$ and $\beta = 0$ the distribution reduces to a Cauchy distribution with scale parameter γ and shift parameter δ

c/ for $\alpha = 1/2$ and $\beta = 1$ the distribution reduces to a Levy distribution with scale parameter γ and shift parameter δ

d/ In the limit as $\gamma \rightarrow 0$ or as $\alpha \rightarrow 0$ the distribution will approach a Dirac delta function $\delta(x - \delta)$

In order to analyse these series we have fitted the histogram to the first normalized logarithmic return. A typical situation in these time series is the existence of a high number of zero values normally in correspondence with weekends or holidays. To compare the results, we have eliminated from the original series the points where the exchange rate was unchanged, i.e. the zero value. Table 6 summarizes the fitted parameters using the maximum likelihood estimation (Nolan, 1997 and 1999).

Table 1. Nord Pool data fitted parameters using STABLE (Nolan, 1999).

Data set	α	β	γ	δ
KRN	0.412	-0.365	0.035	-0.00018
KRN(0)	1.116	0.127	0.242	-0.0514
EUR	1.308	0.164	0.268	-0.068
EUR(0)	1.315	0.173	0.272	-0.069

Due to the high amount of zero in the price in Norwegian Kröne, it is difficult to find a good fit for this time series.

Afterwards, the surrogate time series for the Nord Pool in EUR have been compared with the original time series. We have found that they have a probability distribution function more similar to a Gaussian (α near 2) in comparison with original data ($\alpha = 1.308$) and they have β closer to 0 than original data which mean that their probability distribution functions are less skew.

1.3. Recurrence Quantification Analysis

Eckmann *et al.* (1987) introduced a new graphical tool, which they called a recurrence plot (RP). The recurrence plot is based on the computation of the distance matrix between the reconstructed points in the phase space, i.e. $s_i = \{s(t), s(t-\tau), s(t-2\tau), \dots, s(t+(d_E-1)\tau)\}$,

$$d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\| \quad (6)$$

This produces an array of distances in a $N \times N$ square matrix, \mathbf{D} , being N the number of points under study. Once this distance matrix is calculated, in the original paper of Eckmann *et al.* (1987), it was displayed by darkening the pixel located at specific (i,j) coordinates which corresponds to a distance value between i and j lower than a predetermined cutoff, i.e. a ball of radius ε centered at \mathbf{s}_i . Requiring $\varepsilon_i = \varepsilon_j$, the plot is symmetric and with a darkened main diagonal correspondent to the identity line. The darkened points individuate the recurrences of the dynamical systems and the recurrent plot provides insight into periodic structures and clustering properties that are not apparent in the original time series (Eckmann *et al.*, 1987).

To extend the original concept and made it more quantitative Zbilut and Webber (1992) developed a methodology called Recurrence Quantification Analysis (RQA) (Webber and Zbilut, 1994). For an excellent overview the reader is referred to Marwan *et al.* (2007). As a result, they defined several measures of complexity to quantify the small scale structures in RP. These measures are based on the recurrence point density and the diagonal and vertical line structures of the RP. A computation of these measures in small windows (sub-matrices) of the RP moving along the main diagonal yields the time dependent behaviour of these variables. Some studies based on RQA measures show that they are able to identify bifurcation points, especially chaos-order transitions (Trulla *et al.*, 1996). The vertical structures in the RP are related to intermittency and laminar states: those measures quantifying the vertical structures enable to detect chaos-chaos transitions (Marwan *et al.*, 2002). In these work we will use the measure of the percentage of diagonal and vertical lines (determinism and laminarity respectively).

To check if RQA measures are able to distinguish between real data and their surrogates (linear Gaussian processes) we calculated all of them for both. Using *%determinism*, *%laminarity* we obtain values which are always smaller for surrogate data in comparison with original data sets. The fact that these two parameters are able to distinguish between the original time series and the surrogate time series points toward the explanation that the original series have more diagonal and vertical lines, and therefore their state remain near or at the same place longer in time more often than for its surrogates linear Gaussian process and that they possess a different decaying of the autocorrelation function.

1.4. Determinism and Laminarity as volatility measures.

We have applied Recurrence Quantification Analysis (RQA) to data sets taken from the Nordic spot electricity market (Strozzi *et al.*, 2008) Our main interest was in trying to correlate their volatility with variables obtained from the quantification of recurrence plots (RP). For this reason we have

based our analysis on known historical events: the evolution of the Nord Pool market and climatic factors, i.e. dry and wet years, and we have compared several dispersion measures with RQA measures in correspondence of these events. The analysis suggests that two RQA measures: determinism (DET) and laminarity (LAM) can be used as a measure of the inverse of the volatility. The main advantage of using DET and LAM is that these measures provide also information about the underlying dynamics. This fact is shown using shuffled and linear Gaussian surrogates of the real time series.

Several measures of volatility has been used in literature (Simonsen, 2003, Hsu and Murray, 2007, Figueiredo, et al. 2005), between them we have considered:

$$V_1 = SD(s_t) \quad (7)$$

$$V_2 = SD(s_t - s_{t-1}) \quad (8)$$

$$V_3 = SD((s_t - s_{t-1}) / s_{t-1}) \quad (9)$$

where s_t and SD refer to the time series values and the standard deviation, respectively. To calculate the standard deviation the following formula was used:

$$SD(s_t) = \frac{1}{n-1} \sqrt{\sum_{i=1}^n (s_i - \bar{s})^2} \quad (10)$$

with $\bar{s} = \frac{1}{n} \sum_{i=1}^n s_i$ and n is the number of points considered. In V_3 the argument of SD is an

approximation of $\ln(s_t / s_{t-1})$ which is often used to measure financial volatility. In order to compare these quantities with RQA measures we have inverted and normalized them as follows:

$$IV_i = 1/V_i, i = 1, \dots, 3 \quad (11)$$

$$nIV_i = \frac{IV_i - \min(IV_i)}{\max(IV_i) - \min(IV_i)} \quad (12)$$

We have assumed that an increase of the dispersion measure corresponds to a decrease of RQA measures that account for the predictability of the underlying dynamical system.

As a first step, we have compared the RQA measures of the original time series with two types of surrogate series: shuffled and linear Gaussian with the same FFT. We have observed that RQA measures do not characterize the probability distribution of the data, because the shuffled and the real data have the same mean and variance, but different values of RQA measures. In addition, we have found that two RQA measures: DET (*%determinism*) and LAM (*%laminarity*) are able to distinguish between real and linear Gaussian surrogate with 95% of confidence. For this reason and because of the hypothesis that high volatility can imply small DET and LAM, we have compared them with the inverse of the normalized dispersion measures given by Eq. 12 on a one month

moving window translated of one month. We have found that these measures are correlated with the inverse of dispersion measures that are used to evaluate the volatility of financial time series.

We have found a qualitative agreement from the point of view of high and low values corresponding to wet and dry periods and a general decrease of the measures with the entrance of new countries in the Nord Pool. The linear correlation between these measures decreases for the linear Gaussian surrogates as well as the agreement with historical events.

To see if the RQA measures have some advantages in comparison with the other dispersion measures (Eq. 12), we have observed that DET and LAM show more pronounced jumps between the periods analyzed. This behaviour is lost when we apply the same treatment to surrogate data sets.

2. CORRELATION ANALYSIS BETWEEN FAULTS IN THE ELECTRICITY GRID AND ELECTRICITY PRICES IN THE NORDIC REGION

In this Section we have summarized the work contained in Strozzi and Zaldívar (2009), see Annex IV.

The deregulation has caused considerable changes in the electricity market. On one hand the increase in the competition has modified the prices volatility; on the other hand this competition has stressed electricity grids with the variation of the flow in the physical network. Thus it is natural to assume that some correlations between electricity prices and disturbances in the electricity grid should exist. The correlations, once detected, can help in the prevention of the disturbances acting on the electricity price or, at least, in the management of the contingency.

In Strozzi and Zaldívar (2009) we have analyzed possible correlations between electricity prices and disturbances using the data of the Nordic electricity market. We have used the monthly spot prices, disturbances and consumption from the beginning of January 2000 until the end of December 2006 in the Nordic region, i.e. Denmark, Finland, Norway and Sweden. The preliminary treatment of the data include the elimination of the trends applying the difference operator and subtracting the regression line. In addition, we have considered the price volatility and similarly the volatility of disturbances and of total consumption. The questions we were interested in addressing were the following: Are the monthly spot prices correlated with disturbances? Can we increase the correlation by shifting the time series and can we use the evolution of one time series to anticipate the behaviour of the other and/or to prevent adverse events? Can we detect windows of correlation and find a correspondence of the starting and ending point with some know events? To answer the mentioned questions we have proposed the following methodology. First, starting from prices, disturbances and consumption we have generated other 9 time series: the detrended ones, the first

differences and the volatilities. Then we have try to extract relevant correlations performing the mean (or standard deviation in the case of volatilities) on different time windows shifted by different time intervals and we have calculated all the correlation matrices and Cross Correlation Function to see if a relevant linear correlation exist or directly or after a shift of one series in respect to the other. To see if a linear correlation exists only on one portion of a time series in respect to another one and then disappears due to some external event, we have applied the Cross Recurrence Analysis that is a generalization of the Cross Correlation Function. The Principal Component Analysis is applied to understand if the set of 12 time series considered contain more information than the one contain only in Spot prices, Disturbance and consumption from which they have been generated.

2.1. Linear correlation coefficient: correlation matrix.

The correlation coefficient matrix represents the normalized measure of the strength of linear relationship between variables. To measure the significance of each correlation we have applied the t-test. In every correlation matrix R we have considered the correlation values $R(i,j)$ higher than 0.7071 (i.e. a determination coefficient $R^2 > 0.5$) with a significance level of 95% i.e. $P(i,j) < 0.05$. Each $P(i,j)$ value gives the probability of getting a correlation as large as the observed value by random chance, when the true correlation is zero. The results are presented in Table 2 and 3 In which we have underlined the correlation between different variables and in bold the correlations values between Disturbances and prices.

Table 2. Significant linear correlations coefficient $R(i,j)$ between data sets for different when w equal to sh .

w=1, sh=1	w=3 (seasonal); sh=3	w=6; sh=6	w=12; sh=12
S,Sdt (0.7317) Sfd,Vs(0.8607) Dfd,V _D (0.8761) Tfd,V _T (0.9896)	<u>D,T (-0.8154)</u>	Dfd,D(-0.8503) Tfd,T(-0.8686) V _D ,Dfd(0.7698) <u>D,T(-0.8594)</u> <u>D, Tfd(0.776)</u> <u>T,Dfd(0.7752)</u>	Tdt,T(0.9842) <u>V_D,T(-0.9057)</u> <u>V_D,Sdt(0.8138)</u> <u>V_D,Tdt(-0.9014)</u>

Table 3. Significant linear correlations coefficient $R(i,j)$ between data sets for different w and $sh=1$.

w=2; sh=1	w=3 (seasonal); sh=1	w=6; sh=1	w=12; sh=1
<u>T,D (-0.7354)</u> S, Sdt(0.7195)	<u>T,D (-0.8057)</u>	<u>T,D(-0.9044)</u> <u>Tfd,Dfd(-0.8010)</u>	<u>T,D(-0.7807)</u> <u>D,Tdt(-0.7586)</u> D,Ddt(0.8060) T,Tdt(0.9904) <u>V_D-Sdt (0.7567)</u>

Since we are interested mostly in the correlations between price and disturbances we can conclude that it exists only for $w=12$ and $sh=12$ or $sh=1$, particularly between the volatility of disturbances and the mean Spot prices de-trended.

2. 2. Principal Component Analysis

Principal component analysis (PCA) is a technique used to reduce multidimensional data sets (Jackson, 1991, Jolliffe, 2002). It is a way to identify patterns (linear) in data and then to compress them by reducing the number of dimensions without much loss of information. The eigenvector of the covariance matrix are the components. The eigenvector with the highest eigenvalue is the *principal component* of the data set. A subset of the eigenvectors is selected as basis vectors: the more significant and the others are cancelled. Usually those eigenvalues which sum is 90% of the sum of all eigenvalues are considered. The first principal component is that linear combination of the original variables which accounts for the maximum amount of variance in a single line. It is the line of best fit through the data, and the residual variance about this line is then a minimum for the data set. The second principal component is that line which is orthogonal to the first principal component and accounts for the maximum amount of the remaining variance in the data. The first two components therefore represent the plane of best fit through the data. The eigenvalues obtained from Principal Components Analysis are equal to the variance explained by each of the principal components, in decreasing order of importance. The summary of PCA analysis is presented in Table 4.

Table 4. Summary of PCA results.

w	sh	# points	#PC to explain at least 50% variance	% variance explained	#PC to explain at least 90% variance	% variance explained
1	1	83	3	63.39	6	91.03
2	1	82	3	53.08	8	91.08
3	1	81	3	56.35	8	92.67
6	1	78	3	62.62	7	92.85
12	1	72	2	54.80	6	93.26
3	3	27	3	56.55	7	91.69
6	6	13	2	61.44	5	94.38
12	12	6	2	65.63	4	96.75

In the first two columns of Table 4 there are the values of w and sh and, in the third, the number of points of each time series considered in calculating PCA. In the fourth column the number of principal components able to explain at least the 50% of variance is listed. It seems that an hyper plane of dimension three can fit the data. This is not so strange since we built the twelve time series

starting from three of them (S, D, T), but if we are interested in explaining at least 90% of variance we can see that we need always more than 3 principal components. Sometimes even 8 principal components are necessary i.e. the original time series and their first difference, for example, do not contain still all the independent information.

2.3. Cross Correlation function

Cross correlation is a generalization of the correlation coefficient and a standard method of estimating the degree to which two series are correlated when we shift them one in respect to the others (Orfanidis, 1996).

We have calculated the cross correlation function for every window, w , and every shift, sh . The maximum values obtained are listed in Table 5 together with the correlation coefficients without delay, $R(0)$, and the p values of the t-tests. V_D is correlated with price volatility, price first difference and price de-trended but only considering windows of six or twelve months.

Table 5. Results from the cross correlation analysis.

<i>Time series</i>	w	sh	$R(0)$	p	<i>delay (months)</i>	$R(delay)$	p
$V_S V_D$	12	1	0.5183	0.0000	-6	0.8906	0.0000
$V_S V_D$	6	1	0.1855	0.1040	-6	0.5959	0.0000
$Sdt V_D$	12	1	0.7567	0.0000	-3	0.8536	0.0000
$Sfd V_D$	6	1	-0.4273	0.0001	-8	0.7430	0.0000
$Sfd V_D$	6	6	-0.7778	0.0017	-1	0.8725	0.0002

The correlation function between D-Sfd and D-S which, that, even if it never reaches R-values higher than 0.4, it has a regular oscillating behaviour in respect to the delay (Fig. 1)

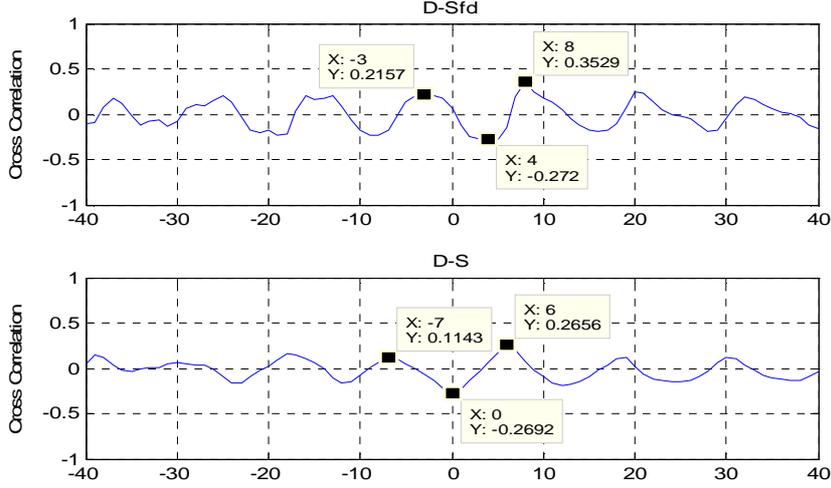


Figure 1. Cross Correlation functions for Disturbances with $w = 2$, $sh = 1$.

2.4. Cross Recurrence Plot

CRP is a bivariate extension of Recurrence Plot and was introduced to analyse the dependencies between two different time series by comparing their joint recurrence (Marwan and Kurths, 2002). It can be considered as a generalization of the linear cross-correlation function (Marwan *et al.* 2007), infact they introduced the Line Of Synchronization (LOS) which is a particular diagonal line in RP which local slope corresponds to the transformation of the time axes of the two considered trajectories. A time shift between the trajectories causes a dislocation of the LOS. Hence LOS allows finding the rescaling function between different time series. In the time window in which LOS has a slope 1 the two time series are linear correlated directly or after a shift of one in respect of the other, this is the case in which LOS is parallel to the main diagonal of RP but not coincident. An example of LOS obtained by the CRP of Disturbances (D), and first differences of Total Consumption (Tfd) is shown in Fig 2.

A disadvantage of using CRP is that in order to obtain a good LOS quality, which means that information given by LOS show real changes in the correlation properties, there is the need of a certain minimum amount of points. In this work we have been able to obtain good LOS quality using only data with $w=2$ and $sh=1$; in the other cases there were not enough points to perform this analysis.

To confirm the fact that LOS allows in detecting windows of higher linear correlation, we have compared the correlation of the entire time series with the one obtained using only the portion of the data in which the LOS is parallel to the main diagonal (R_{LOS}) and with the one suggested by the correlation function (R_{CCF}) i.e. obtained translating the entire time series. All the results are shown in Table 6. Moreover, looking to Table 6, we can observe that LOS allows identifying the time in

which Spot Prices changes at the beginning of the dry period (June-July 2002) and in which the prices increase due to the dependence from external sources.

Table 6. Correlation coefficient for different portion of the time series. *ns*: not significative. *R*: correlation Coefficient of the entire time series and without shift. *R_{CCF}*: max correlation obtained using Cross Correlation Function. *R_{LOS}*: Correlation coefficient of the portion of the time series suggested by LOS.

Time Series	R	R _{CCF}	R _{LOS}	Date correspondent to the points considered
-Disturbances -Prices	-0.2692	-0.2692	0.3979 (ns)	July 01- May 02
-Disturbances -Total Consumption	-0.7354	-0.7354	-0.8037	Feb 00- Sept 01
-Disturbances -Prices first differences of prices	0.0702	-0.3529	-0.3953	Feb 00- July 02
-Disturbances - Disturbances first differences	-0.4119	-0.6809	-0.7021	Feb 00- June 02 March 00 -July 01
-Disturbances - Total Cons. first differences	0.2429	0.6896	0.7455	Feb 00-July 06 May 00-Dec 06
-Disturbances volatility -Price detrended	0.1545	0.4418	-0.2248 (ns)	Feb 00-Dec 02

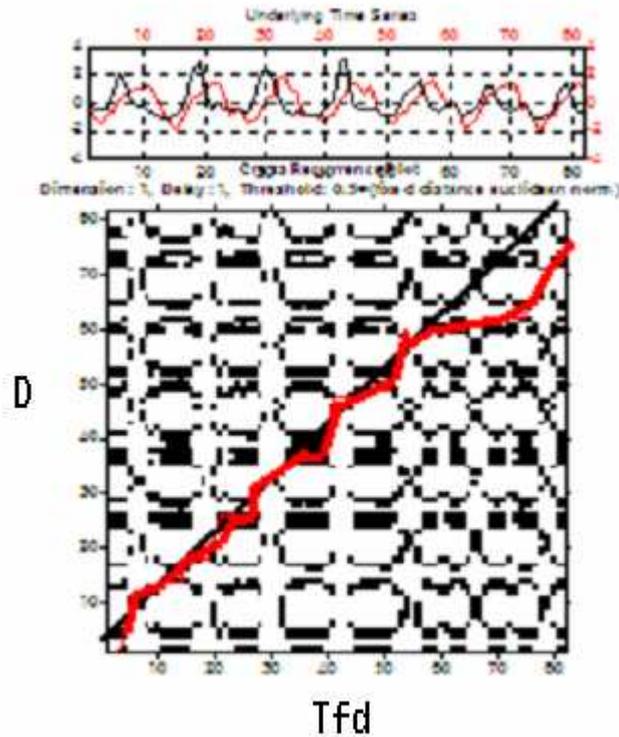


Figure 2. Example of LOS in the CRP obtained using Disturbances (D), and first differences of Total Consumption (Tfd).

3. CONCLUSIONS

The main conclusions of Section 1 of this report i.e the analysis of electricity spot prices in the Nordic region are that R/S analysis confirms their long range correlation and antipersistence. Stable distribution fitting has characterized the electricity spot price first difference from the statistical point of view confirming that the distribution is “fat tail” and that extreme events far from the mean value have higher probability to occur in comparison with a Gaussian distribution. Applying Recurrence Quantification Analysis and in particular two measures: DET and LAM we were able to distinguish between real and surrogate data sets. Moreover the same measure demonstrated to be able to detect the time windows of higher volatility and in this way they provide a bridge between the concept of volatility as dispersion and volatility as lack of predictability i.e. lack of determinism. The main conclusions of Section 2 of this report i.e. analysed possible correlations between electricity prices and disturbances in the Nordic Region, are that a strong linear correlation ($R > 0.7$) exists between the volatility of disturbances and the de-trended spot price if we consider mean of the time series on windows of six or twelve months. Using the Cross Correlation Function i.e. shifting the time series one in respect to the other some correlations increases but the one between

Disturbances and Prices never reaches values higher than 0.4, anyway it has a regular oscillating behaviour in respect to the delay and this can be a sign of similarity between the two dynamics.

Finally we have applied Cross Recurrence Plot analysis, which gives an extension of the Cross Correlation Function and it helps to detect portion of the time series that are linear correlated. We have demonstrated that some correlations increases. We found time window in which the linear correlation between disturbances and total consumption and disturbances and spot prices increases but the correlation values are not always significative if we apply a t-test. The only disadvantage of CRP is that we can apply it to extract reliable information only if we have a minimum amount of data

REFERENCES

- Eckmann, J. P., Kamphorst, S. O. and Ruelle, D., 1987, Recurrence plots of dynamical systems, *Europhys. Lett.* **4**, 973-977.
- Erzgräber, H., Strozzi, F., Zaldívar, J.M., Touchette, H., Gutiérrez, E. and Arrowsmith, D.K. 2008. Time series analysis and long range correlations of Nordic spot electricity market data. *Physica A*, 387, 6567-6574.
- Figueiredo, A., Gleria, I., Matsushita, R., Da Silva, S., (2005). Financial volatility and independent and identically distributed variables. *Physica A* 346, 484.
- Hsu, S.D.H. Murray, B.M. (2007). *Statistical Mechanics and its Applications*. *Physica A* 380, 366.
- Hurst, H. E., 1951, Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civ. Eng.* **116**, 770-779.
- Jackson, J. E., 1991. *A User's Guide to Principal Components*, John Wiley and Sons.
- Jolliffe, I. T. 2002. *Principal Component Analysis*, 2nd edition, Springer.
- Mandelbrot, B. B., *The Fractal Geometry of Nature*, 1983, W. H. Freeman. New York.
- Marwan, N., Romano, M. C., Thiel, M. and Kurths, J., 2007. Recurrence plots for the analysis of complex systems. *Physics Reports* 438, 237-329.
- Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., Kurths, J. 2002. Recurrence plot based measures of complexity and its application to heart rate variability data. *Phys. Rev. E* **66**(2), 026702.
- Nolan, J.P., 1997. Numerical computation of stable densities and distribution functions. *Commun. Stat.: Stochastic models* **13**, 759-774.
- Nolan, J.P., 1999. Fitting data and assessing goodness of fit with stable distributions. In *Proceedings of the Conference on Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, American University, Washington DC, June 3-5.
- Orfanidis, S.J., 1996. *Optimum Signal Processing. An Introduction*. 2nd Edition, Prentice-Hall, Englewood Cliffs, NJ.
- Simonsen, I. (2005) *Volatility of power markets*, *Physica A* 355, 10
- Strozzi, F., Gutiérrez, E., Noè, C., Rossi, T., Serati, M. and Zaldívar, J.M. 2008. Measuring volatility in the Nordic spot electricity market using Recurrence Quantification Analysis. *Eur. Phys. J. Special Topics* 164, 105-115.

- Strozzi, F., Gutiérrez, E., Noè, C., Rossi, T., Serati, M. and Zaldívar, J.M., October 2007. *Application of non-linear time series analysis techniques to the Nordic spot electricity market*. LIUC Paper 200.
- Strozzi, F. and Zaldívar, J.M., January 2009. *Correlation analysis between electricity spot prices and faults in the electricity grid in the Nordic region*. LIUC paper xxx.
- Trulla, L.L, A. Giuliani, J.P. Zbilut, and C.L. Webber, Jr. (1996). Recurrence quantification analysis of the logistic equation with transients. *Phys. Lett. A* **223**, 255-26.
- Webber Jr. C. L. and Zbilut, J. P., 1994, Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **76**, 965-973.
- Zbilut, J. P. and Webber Jr. C. L., 1992, Embeddings and delays as derived from quantification of recurrence plots. *Phys. Lett. A* **171**, 199-203.

ANNEX I. *Application of non-linear time series analysis techniques to the Nordic spot electricity market.* LIUC Paper 200. October 2007.

APPLICATION OF NON-LINEAR TIME SERIES ANALYSIS TECHNIQUES TO THE NORDIC SPOT ELECTRICITY MARKET DATA

Fernanda Strozzi, Eugénio Gutiérrez Tenreiro, Carlo Noè, Tommaso Rossi, Massimiliano Serati, José-Manuel Zaldivar Comenges

Contents

1. Introduction
2. Data provision and treatment
 - 2.1. Data treatment
 - 2.2. Historical background
 - 2.3. Materials and methods
3. Embedding theory
 - 3.1. Embedding parameters
4. Chaotic time series analysis
 - 4.1. Preliminary Analysis
 - 4.1.1. Surrogate time series generation*
 - 4.1.2. R/S Analysis*
 - 4.1.3. Power Spectral density*
 - 4.1.4. Fitting Nord Pool data with stable distributions*
 - 4.2. Finding the time delay and embedding dimension
 - 4.2.1. Time delay*
 - 4.2.2. Embedding dimension*
 - 4.3. Detecting non-stationarity
 - 4.4. Testing for non-linearity
 - 4.5. Recurrence quantification analysis (RQA)
 - 4.5.1. Selection of the threshold or cutoff value ε*
 - 4.5.2. Quantification of recurrence plots*
 - 4.5.3. Analysing the complete time series*
 - 4.5.4. RQE analysis*
5. Conclusions
- References

1. Introduction

The complex behaviour of financial time series, which linear stochastic models are not able to account for (Mantegna & Stanley, 2000; Johnson et al., 2003), has been attributed to the fact that financial markets are nonlinear stochastic, chaotic or a combination of both. Specifically, in the last decades there have been a considerable amount of discussion about the characterization of financial time series using the theory of Brownian motion (Osborne, 1959; Malkiel, 1990), fractional Brownian motion (Mandelbrot, 1998), non-linearity (Brock *et al.*, 1991), chaos and fractals (Hsieh, 1991; Lorenz, 1993; Peters, 1996), scaling behaviour (Mantegna and Stanley, 1995 and 1996), and self organized criticality (Bak and Chen, 1991; Shlesinger *et al.*, 1993). The problem of characterizing financial time series is still an open question. Most of the test developed in the area of economic theory, provide evidence of nonlinear dynamics, which is a necessary but not sufficient condition for chaos. This nonlinearity may be deterministic or not deterministic. In fact, there is no convincing evidence of deterministic low-dimensionality in price series (Scheinkman and LeBaron, 1989; Papaioannou and Karytinos, 1995) and the claims of low-dimensional chaos have never been well-justified. For example, Andreadis (2000) analysing the S&P 500 index time series favours the stochastic hypothesis, whereas Friederich *et al.* (2000), using the high frequency price changes of the US dollar-German Mark support the analogy of turbulence and financial data (Mantegna and Stanley, 1996). Therefore, even though there is no conclusive evidence of low dimension deterministic (chaotic) structure, in the last few years, nonlinear time series analysis has expanded rapidly in the fields of Economics and Finance. This is also due to the fact that economic and financial time series seem to provide a promising area for the development, testing and application of nonlinear techniques (Soofi and Cao, 2002) and the fact that high frequency financial time series are readily available.

Between these time series, energy spot prices have also been analysed with several nonlinear techniques. Weron and Przybyłowicz (2000) studied the electricity prices using Hurst R/S analysis and showed that they are anti-persistent with a Hurst exponent lower than 0.5. Using another technique, the Average Wavelet coefficient method, Simonsen (2003) calculated also the Hurst exponent and obtained a value of $H \approx 0.41$ in agreement also with another energy spot prices time series. In a recent study, Bask et al. (2007) estimated the Lyapunov exponents and concluded that the dynamic system that generates these prices appeared to be chaotic for the period July 1, 1999 to September 30, 2000. The question of modelling spot electricity prices has also been addressed by several researchers. Because of the high volatility in Nord Pool electricity prices, Byström (2005) applied extreme value theory (EVT) to investigate the tails of the price change distribution and then used the peaks-over-threshold (POT) method to deal with the data that exceed the threshold. Then he used a combined AR and GARCH model to fit the filtered time series to estimate as well as to forecast the time series. Along the same lines, Perelló *et al.* (2007) proposed a GARCH model for the spot price. Weron *et al.* (2004) fit a jump diffusion and regime switching model to Nordic Pool spot prices. Vehviläinen and Pyykkönen (2005) developed a stochastic factor based approach to mid-term modelling of spot prices taking into account climate data, hydro-balance, base load supply and the underlying mechanisms in spot price generation. The model was able to provide simulated values for the fundamental data, demand and supply information, and pricing strategies.

In this work we have applied non-linear time series techniques the Nordic spot electricity market data. The time series are given in two periods, from May 1992 to December 1998 and from January 1999 to January 2007. Our main interest was on trying to classify these series and analysing if their dynamical behaviour were in some way correlated with known events, e.g. the evolution of the Nord Pool and the climatic factors. This work is a first step in the direction of finding correlation of some features of the time series with the frequency and intensity of blackouts.

First, a preliminary study was carried out with the aim of characterising the time series in terms of power spectral distribution, long term memory (R/S analysis), stationarity (space-time separation plots) and tails (stable distributions). Surrogate time series were also generated to test if the original time series were similar to a stationary Gaussian linear process. In a second step, state space reconstruction parameters: time delay and embedding dimension were used to carry out the analysis of these two series in the reconstructed state space. We applied Recurrence Quantification Analysis (RQA) (Webber and Zbilut, 1994), which is based on the definition of several parameters that allows the quantification of the Recurrence Plots (RP) introduced by Eckmann *et al.* (1987). The RQA analysis of both time series has shown a certain coherent structure with a regime shift in the first time series. Moreover, the RQA analysis was repeatedly performed on 720-point epochs (approx. one month) in order to analyse the dynamic information obtained. Neighbouring epochs were shifted also by 720 points and the nonlinear variables: *%recurrence*, *%determinism*, *%laminarity* and *trapping time* obtained for the time series analysed. A similar analysis has also been performed with the surrogate time series. As discussed in the report, it is possible to correlate certain events with changes in *%recurrence*, *%determinism*, *%laminarity* and *trap time*. Furthermore, the RQA method allows distinguishing the original time from the surrogate the time series, indicating a certain nonlinear behaviour in the original series. The preliminary results following the analysis of these series have shown that there are some similarities in terms of certain statistical characteristics, but also differences with other high frequency financial time series (Strozzi *et al.*, 2002; Strozzi *et al.*, 2007). Finally, we used two RQA measures, *%determinism* and *%laminarity*, for developing a new measure of volatility which is able of detecting important historical and meteorological events with better resolution than by measuring the time series standard deviation.

2. Data provision and treatment

We have analyzed hourly data from the Nord Pool system spot prices. The series is divided into two parts. In the first part, that goes from 4th May 1992 until 31st December 1998 and comprises 58,392 data points (fig.1), the prices are indicated in Norwegian Krone (NOK)/MWh, whereas in the second time series that goes from 1st January 1999 until 26th January 2007 and comprises 70,752 data points (fig.2), the prices are expressed in EUR/MWh.

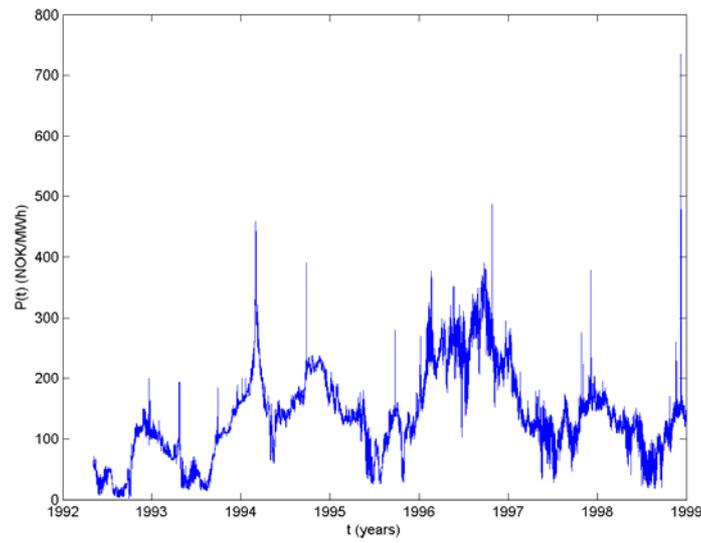


Figure 1. Spot prices in the Nordic electricity market (Nord Pool) from May 1992 until December 1998.

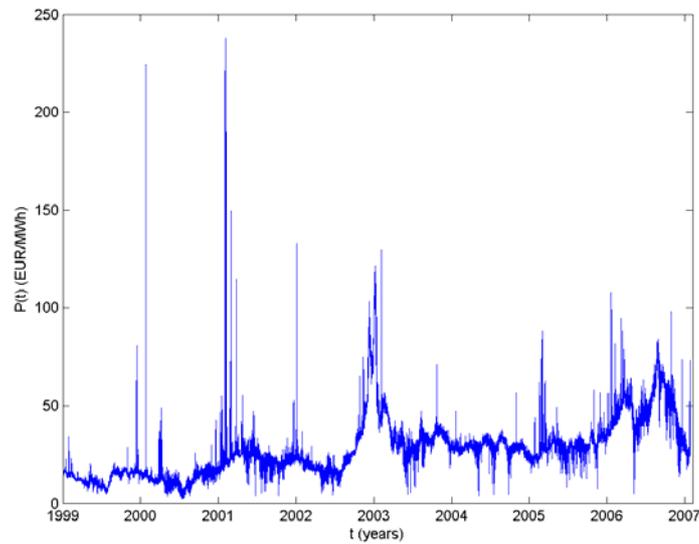


Figure 2. Spot prices in the Nordic electricity market (Nord Pool) from January 1997 until January 2007.

2.1. Data treatment

We have considered the prices time series as well as the corresponding logarithmic returns over the time horizon Δt , defined as:

$$r_{\Delta t}(t) = \ln\left(\frac{P(t)}{P(t - \Delta t)}\right) \quad (1)$$

Figures 3 and 4 show the hourly returns for the two prices time series considered.

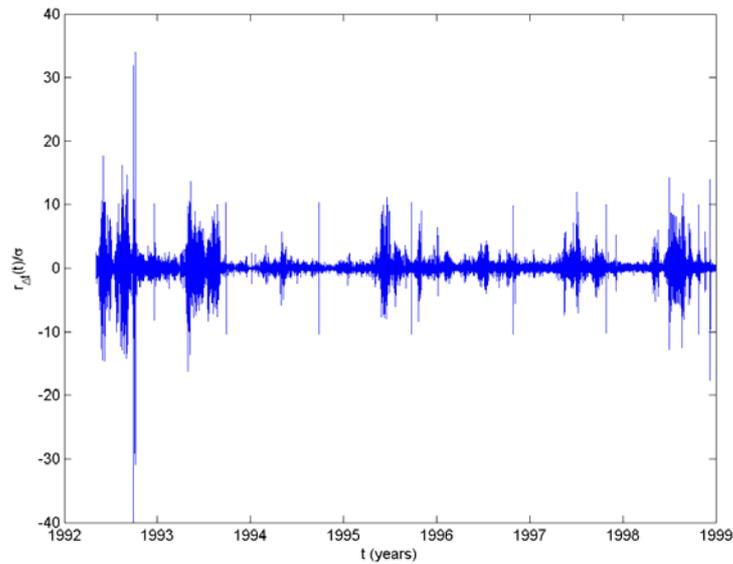


Figure 3. Hourly logarithmic return (Eq. 1) for the spot prices in the Nordic electricity market (Nord Pool) from May 1992 until December 1998.

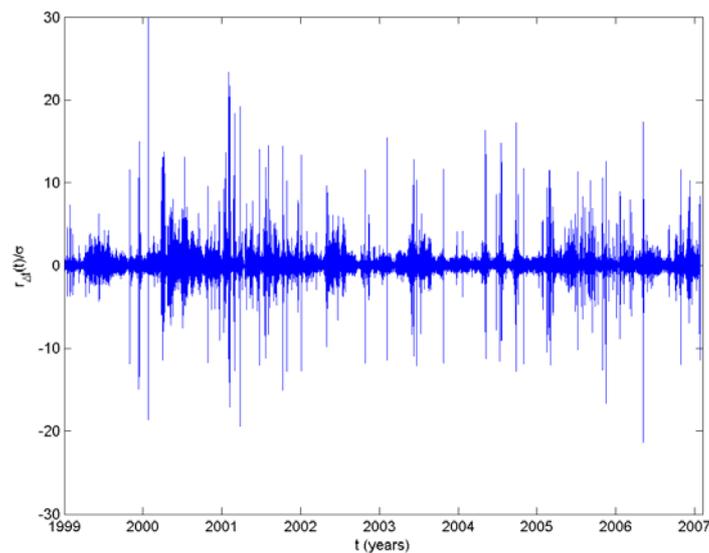


Figure 4. Hourly logarithmic return (Eq. 1) for the spot prices in the Nordic electricity market (Nord Pool) from January 1997 until January 2007.

2.2. Historical background

Electricity deregulation started in individual countries, notably United Kingdom (1990) and Norway (1991), and the Norwegian effort spread to the rest of the Nordic region before the European Union's 1996 Electricity Directive started to have real impact. This directive required that all EU countries opened up their electricity markets to competition to consumers of more than 9 GWh by 2003. The various countries are free to choose their own methods of deregulation in accordance to the criteria of the Directive. There were no provisions in the Directive for a power pool or the establishment of financial markets (Mork, 2001). The Nordic electricity market, known as Nord Pool (<http://www.nordpool.no>) was created in 1993 and it is owned by the two national grid companies, Statnett SF in Norway (50%) and Affärverket Svenska Kraftnät in Sweden (50%). It was

established as a consequence of the decision in 1991 by the Norwegian Parliament's to deregulate the market for power trading.

Therefore, between 1992 and 1995 only Norway contributed to the market, in 1996 a joint Norwegian-Swedish power exchange was started-up and the power exchange was renamed Nod Pool ASA. Finland started a power exchange market of its own, EL-EX, in 1996, and joined Nord Pool in 1997. Beginning of 15th June 1998, Finland became an independent price area on the Nord Pool Exchange. The western part of Denmark (Jutland and Funen) has been part of the Nordic electric power market since 1 July 1999, whereas the eastern part of Denmark entered after 1st October 2000. On 5th October 2005 also the German area KONTEK was added in the Nord Pool exchange market. Table 1 summarises the historical evolution of the Nord Pool, whereas in Table 2 the deregulation process is also indicated.

Table 1. Nord Pool participating countries and dates of entry.

Countries	Date of entry of new country (dd/mm/yy)
Norway	1/1/93
Norway and Sweden	1/1/96
Norway, Sweden and Finland	29/12/97
Norway, Sweden, Finland and western Denmark	1/7/99
Norway, Sweden, Finland, western and eastern Denmark	1/10/00
KONTEK (Germany)	5/10/05

Table 2. Summary of the deregulation process in Nord Pool members.

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2005	2004	2005
Norway	Green	Green	Blue	Blue											
Sweden	Green	Green	Green	Green	Green	Blue	Blue								
Finland						Green	Green	Blue	Blue						
West Denmark								Green	Blue	Blue	Blue	Blue	Blue	Blue	Blue
East Denmark										Blue	Blue	Blue	Blue	Blue	Blue
Kontek								Green	Blue						

green= deregulation process; blue= NordPool member

The new bidding area named KT offered geographic access to the Vattenfall Europe Transmission control area from East Denmark and allowed Nord Pool to compete directly with European Energy Exchange (EEX). Kontek cable connects Zealand and Germany. Nord Pool owns 17.39% of the shares of EEX and proposed a common market with EEX, but EEX did not agree (Kristiansen, 2006; 2007). Nevertheless the existence of a common electricity market, there are still national transmission system operators and some differences with respect to transmission pricing.

The spot market operated by Nord Pool is an exchange market where participants' trade power contracts for physical delivery the next day and is thus referred to as a day-ahead market. The spot market is based on an auction with bids for purchase and sale of power contracts of 1-h duration covering the 24 h of the following day. At the deadline for the collection of all buy and sell orders the information is gathered into aggregate supply and demand curves for each power-delivery hour. From these supply and demand curves the equilibrium spot prices-referred to as the system prices-are calculated.

When no grid congestion exists there will be a single identical price across the area with no congestions. However, when there is insufficient transmission capacity in a sector of the grid, grid congestion will arise and the market system will establish different "price areas". This is because the Nordic market is partitioned into separate bidding areas which become separate price areas when the contractual flow between bidding areas exceeds the capacity allocated by the transmission system operators for spot contracts. In the case of congestion the transmission system operators ask generators to reduce (increase) production or large buyer to increase (decrease) demand until excess of supply or demand are eliminated. The fact that separate prices may coexist depending upon regional supply and demand causes the relevant market definition to vary with time. Sometimes the prices are of the entire Nordic region. Sometimes more than one price area exists (Haldrup and Nielsen, 2006). Thus, whenever the relevant interconnector capacity is insufficient the Nord Pool area is divided into two or several "price areas". Sweden is always one single price area, and the same applies to Finland. In Denmark the transmission system is divided into two parts, West and East, and consequently there are two price areas. In Norway the congestion charges effectively divides the country into five price areas. In addition to the "area prices" there is a "system price". This price is determined under the assumption that no transmission constraint is binding. The system price is the reference price in the financial contracts (Amundsen and Bergman, 2007). Haldrup and Nielsen (2006) found that looking to hourly data from 3.1.2000 to 25.10.2003, 34.24% of time all the prices for the entire Nordic region were identical. Two price areas existed in 34.55% and three in 20.86 % of the time. In only 11 hours there was complete congestion and six different price areas existed i.e. one for each geographical market. Despite these differences, in this work we will only consider "system price".

The variation of the prices in the Nord pool system is well correlated with the variations in precipitation in Norway and Sweden because of its strong dependence of the hydropower generation. Table 3 summarises the climatic conditions during the last years. The 1996 was a "dry" year, while 1997-2000 was a series of "wet" years. The 2000 was not very "wet" and the first part of 2001 was quite "dry" but the autumn was very rainy and 2001 started well with a water reservoir above the normal. Very special hydrological conditions appeared during the autumn and winter season of 2002-2003 with a sharp decline of precipitation. This was a rare event that could happen only every 100-200 years (Weron *et al.*, 2004). The result was the increasing of spot prices in 2003.

Table 3. Summary of meteorological conditions: Dry and wet years.

<i>year</i>	<i>state</i>	<i>Period considered</i>
1996	dry	1.1.96-31.12.96
1997-2000	wet	1.1.97-31.12.99
2000	not very wet	1.1.2000-31.12.2000
First part 2001	dry	1.1.2001-31.8.2001
Autumn 2001	very wet	1.9.2001-31.12.2001
2002-2003	very dry (rare event)	1.1.2002-31.12.2003

By looking into figs. 1-2 and comparing with Table 3, we can observe these correlations in the electricity price. However, weather conditions are not able to explain all the features in the time series. For example, the relative sharp price increase between 2000 and 2001 could be explained by a combination of the market power exercised by the mayor generators, the increased demand and higher fuel prices (Weron *et al.*, 2004). Moreover spot prices can increase tenfold during a single hour. Jumps in the spot prices are an effect of extreme load fluctuations, caused by severe weather conditions often in combination with generation outages or transmission failures. These spikes are normally quite short lived, and as soon as the weather phenomenon or outage is over, prices fall back to a normal level. Jumps tend to be more severe during high price periods and a positive jump may be followed by a negative jump to capture the rapid decline of electricity prices (Weron *et al.*, 2004).

2.3. Material and methods

There are different freely available software packages on the Internet that may be used to perform nonlinear time series analysis. In this work, we have used several of them for different purposes as indicated bellow.

One of the most complete is the TISEAN software package (<http://www.mpiyks-dresden.mpg.de/~tisean>) which has incorporated an impressive quantity of algorithms developed in the nonlinear time series analysis field (Kantz and Schreiber, 1997). There is a version for MATLAB[®] users developed at Göttingen University, called TSTOOL, that can be download at <http://www.physik3.gwdg.de/tstool/>. Furthermore, a commercially available software package developed by Abarbanel and co-workers (Abarbanel, 1996) and commercialised by Randle Inc., called Csp, can be found at <http://www.chaotic.com/>.

Concerning Recurrence Quantification Analysis, the original programs developed by Weber and Zbilut (1994) can be download at <http://homepages.luc.edu/~cwebber>., whereas a MATLAB[®] version of RQA developed at the University of Postdam called CRP toolbox can be found at <http://tocsy.agnld.uni-postdam.de> (Marwan *et al.*, 2007). In addition, there is a commercially available version called VRA (Visual Recurrence Analysis) that can be obtained at <http://home.netcom.com/~eugenek/download.html>

Finally, the analysis of stable distributions has been carried out using the program STABLE for univariate data (<http://www.cas.american.edu/~jpnolan>).

3. Embedding theory

The mathematical basis of continuous dynamical modelling is formed by differential equations of the following type:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, \alpha) \quad (2)$$

where the real variable t denotes time, $\mathbf{x} = (x_1, x_2, \dots, x_n)$ represents the state variables of the system, depending on time t and on the initial conditions, and α_j are parameters of the system, while $\mathbf{F} = (F_1, F_2, \dots, F_n)$ is a nonlinear function of these variables and parameters. Actual states of these systems are described by the vector variable \mathbf{x} consisting of n independent components. Each state of the system corresponds to a definite point in phase space, which is called phase point. The time variation of the state of the system is represented as a motion along some curve called phase trajectory.

Experimentally, it is not always possible to measure the complete state of a system and, normally, when analysing a dynamical system, we have access to few observable quantities which, in the absence of noise, are related to the state space coordinates by:

$$s(t) = \mathbf{h}(\mathbf{x}(t)) \quad (3)$$

where \mathbf{h} is normally an unknown nonlinear function called measurement function. The theory of embedding is a way to move from a temporal time series of measurements to a state space "similar" -in a topological sense- to that of the underlying dynamical system we are interested in analysing. Techniques of state space reconstruction were introduced by Packard *et al.* (1981) and Takens (1981), which showed that it is possible to address this problem using measurements of a sufficient long time series, $s(t)$, of the dynamical system of interest. Takens proved that, under certain conditions, the dynamics on the attractor of the underlying original system has a one-to-one correspondence with measurements of a limited number of variables. This observation opened a new field of research. In fact, if the equations defining the underlying dynamical system are not known, and we are not able to measure all the state space variables, the state space of the original system is not directly accessible to us. However, if by measuring few variables we are able to reconstruct a one-to-one correspondence between the reconstructed state space and the original, this means that it is possible to identify unambiguously the original state space from measurements. Embedding theory has opened a new field of research: nonlinear time series analysis (Tong, 1990; Abarbanel, 1996; Kantz and Schreiber, 1997; Diks, 1999, amongst others).

In order to explain the relationship that occurs between the reconstructed and the real state space, let us consider the following dynamical system

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}); \quad \mathbf{x} = (x_1, x_2, x_3) \quad (4)$$

We can define $\mathbf{y} = (y_1, y_2, y_3)$ as follows: $\mathbf{y} = (x_1, dx_1/dt, d^2x_1/dt^2)$, then the equations of motion take the form

$$\frac{dy_1}{dt} = y_2$$

$$\frac{dy_2}{dt} = y_3 \tag{5}$$

$$\frac{dy_3}{dt} = \mathbf{G}(y_1, y_2, y_3)$$

for some function \mathbf{G} . In this coordinate system, modelling the dynamics reduces to constructing the single function \mathbf{G} of three variables, rather than three separate functions, each of three variables.

In this way we may proceed from the state space (x_1, x_2, x_3) to the space of derivatives $(x_1, dx_1/dt, d^2x_1/dt^2)$. The dynamics in this new space will be related to the dynamics of the original space by a nonlinear transformation which is called the reconstruction map. The extension of this approach to higher-dimensional dynamical systems is straightforward by considering higher derivatives.

The advantage in considering the space of derivatives is that we can approximate them from measurements of x_j . But what kind of information about the original space is preserved in the new one?

There are two types of preserved information: qualitative and quantitative. Qualitative information is that which allows a qualitative description of the dynamics described by topological invariants, such as for instance, singularity of the field, closeness of an orbit, stability of a fixed point, etc. (Gilmore, 1998) Quantitative information can be of two different types: geometrical and dynamical. Geometrical properties (Grassberger, 1983) consist on fractal dimensions or scaling functions. Dynamical methods (Wolf *et al.*, 1985) rely on the estimation of local and global Lyapunov exponents and Lyapunov dimensions. In order to guarantee that the quantities computed for the reconstructed attractor are identical to those in the original state space, we require that the structure of the tangent space, i.e. the linearization of the dynamics at any point in the state space, is preserved by the reconstruction process. The problem is to see under what conditions this can happen. Embedding theorems try to shed some light on this problem.

Let $s(t)$ be the measure of some variable of our system, see Eq. (3). Takens (1981) shown that instead of derivatives, $\{s(t), \dot{s}(t), \ddot{s}(t), \dots\}$, one can use delay coordinates, $\{s(t), s(t + \Delta t), s(t + 2\Delta t), \dots\}$, where Δt is a suitably chosen time delay. In fact, looking at the following approximation of the derivative of $s(t)$:

$$\frac{ds(t)}{dt} \cong \frac{s(t + \Delta t) - s(t)}{\Delta t} \tag{6}$$

$$\frac{d^2s(t)}{dt^2} \cong \frac{s(t + 2\Delta t) - 2s(t + \Delta t) + s(t)}{2\Delta t^2} \tag{7}$$

it is clear that the new information brought from every new derivative is contained in the series of the delay coordinates. The advantage of using delay coordinates instead of derivatives is that in case of high dimensions high order derivatives will tend to amplify considerably the noise in the measurements.

Another generally used method, for state space reconstruction, is singular value decomposition (SVD), otherwise known as Karhunen-Loève decomposition, which was proposed by Broomhead and King (1986) in this context. The simplest way to implement this procedure is to compute the covariance matrix of the signal with itself and then to compute the eigenvalues, i.e. if $s(t)$ is the signal at time t , the elements of the covariance matrix \mathbf{Cov} are:

$$c_{ij} = \langle s(t)s(t + (i - j)t) \rangle^T \quad (8)$$

where i and j go from 1 to n where n is bigger or equal to the dimension of the system in this new space. The eigenvectors of \mathbf{Cov} define a new coordinate system. Typically, one calculates the dimension of the reconstructed phase space by considering only eigenvectors whose eigenvalues are “large”.

Then, from the space of derivatives, time lags or eigenvectors, it is possible to extract information about the underlying system, which was generating the measured data.

In order to preserve the structure of tangent space and then the dynamic characteristic of it, the relation between the reconstructed space and the original one has to be an embedding of a compact smooth manifold into R^{2n+1} , which means a one-to-one immersion i.e. a one-to-one C^1 map with Jacobian which has full rank everywhere. The point now is to show under what conditions the reconstruction forms an embedding.

A general existence theorem for embedding in Euclidean spaces was given by Whitney (1936) who proved that a smooth (C^2) n -dimensional manifold may be embedded in R^{2n+1} . This theorem is the basis of the time delay reconstruction (or embedding) techniques for phase space portraits from time series measurements proposed by Takens (1981), who proved that, under certain circumstances, if d_E -the dimension of the reconstructed state vector, normally called the embedding dimension- is greater or equal to $2n+1$, where n is the dimension of the original state space, then the reconstructed states fill out a reconstructed state space which is diffeomorphic, i.e. a one-to-one differentiable mapping with a differentiable inverse, to the original system. Generally speaking, the embedding dimension is the minimal number of dynamical variables with which we can describe the attractor when we know only one of its state variables or a function related to them.

Apart from the methods mentioned above, there are several other methods of reconstructing state space from the observed quantity $s(t)$ that have appeared in the literature -for a critical review see Breeden and Packard (1994). Although the method of reconstruction can make a big difference in the quality of the resulting coordinates, it is not clear in general which method is the best. The lack of a unique solution for all cases is due in part to the presence of noise and to the finite length of the available data sets.

For Takens' theorem to be valid we need to assume that the underlying dynamics is deterministic and that both the dynamics and the observations are autonomous, i.e. \mathbf{F} and \mathbf{h} in Eqs. (2) and (3) depend only on \mathbf{x} and not on t . Unfortunately, this is not the case of many systems in the field of control and communications which are designed to process some arbitrary input and hence, cannot be treated as autonomous. The extension of Takens' theorem to deterministically forced stochastic systems has been recently developed by Stark *et al.* (1997). In particular they proved that such an extension is possible for deterministically forced systems even when the forcing function is unknown, for input-output systems (which are just deterministic systems forced by an arbitrary input sequence) and for irregular sampled systems.

Another problem in embedding theory is that Takens' theorem has been proven for noise-free systems. Unfortunately, there is always a certain amount of noise, $\sigma(t)$, in real data. Such noise can appear in both the measurements and the dynamics (Diks, 1999). Observational noise, i.e. $s(t)=h(\mathbf{x}(t))+\sigma(t)$, does not affect the evolution of the dynamical system, whereas dynamical noise acts directly on the state of the dynamical system influencing its evolution, for example: $d\mathbf{x}/dt=\mathbf{F}(\mathbf{x}, \alpha)+\sigma(t)$.

The effects of relatively small amount of observational noise may put severe restrictions on the characterisation and estimation of the properties of the underlying dynamical system. In order to remove the observational noise different possibilities are available which can be broadly divided into two categories: linear filters (Badii *et al.*, 1988) and special nonlinear noise reduction methods that make use of the deterministic origin of the signal we are interested in (for a recent survey see: Kostelich and Schreiber, 1993; Davies, 1994). However, in the case of dynamical noise, the reconstruction theorem does not apply and it may even be impossible to reconstruct the state of the system (Takens, 1996). In this situation, systems must be examined case by case before analysis. In particular, Stark *et al.* (1997) showed that the extension of Takens' theorem is possible for deterministic systems driven by some stochastic process.

3.1. Embedding parameters

The embedding theorem is important because it gives a rigorous justification for the state space reconstruction. However, Takens' theorem is true for the unrealistic case of an infinite, noise-free, number of points. Takens showed that, in this case, the choice of the time delay is not relevant, and gave indications only on the choice of the embedding dimension.

Nevertheless, in real applications, the proper choice of the time delay τ and the calculation of an embedding dimension, d_E , are fundamental for starting to analyse the data. As a matter of fact, a lot of research on state space reconstruction has centred on the problems of choosing the time delay and the embedding dimension which we can call the parameters of the reconstruction for delay coordinates.

If the time delay chosen is too small, there is almost no difference between the elements of the delay vectors, since that all points are accumulated around the bisectrix of the embedding space: this is called redundancy (Casdagli *et al.*, 1991). However, when τ is very large, the different co-ordinates may be almost uncorrelated. In this case the reconstructed trajectory may become very complicated, even if the underlying "true" trajectory is simple: this is called irrelevance. Unfortunately no rigorous way exists of determining the optimal value of τ . Moreover, similar problems are encountered for the embedding dimension. Working in a dimension larger than the minimum required by the data will lead to excessive requirements in terms of the number of data points and computation times necessary when investigating different questions such as, for example invariants calculation, prediction, etc. Furthermore, noise by definition has an infinite embedding dimension, so it will tend to occupy the additional dimensions of the embedding space where no real dynamics is operating and, hence, it will increase the error in the subsequent calculations. On the other hand, by selecting an embedding dimension lower than required, we would not be able to unfold the underlying dynamics, i.e. the calculations would be wrong since we do not have an embedding.

When derivatives, $\{s(t), \dot{s}(t), \ddot{s}(t), \dots\}$, or SVD are employed there is no need to determine an optimum time delay. Nevertheless, for the case of derivatives, the reconstruction will depend on the way they are numerically calculated (which turns out to depend on different parameters, see for example (Burden and Faires, 1996) for a review of numerical calculation of derivatives). In practice for each method we will carry out a slightly different state space reconstruction. For the case of SVD, the time delay chosen is unitary, but there is still the problem of choosing the time scale or window in which the calculations are performed. Broomhead and

King (1986) in fact, concluded that the effects of window length should be carefully investigated each time a state space reconstruction is carried out.

4. Chaotic time series analysis

Nonlinear analysis of experimental time series has, among its goals, the separation of high-dimensional and stochastic dynamics from low-dimensional deterministic signals, estimation of system parameters or invariants (characterisation), and, finally, prediction, modelling and control.

Unfortunately it seems very difficult to tell whether a series is stochastic or deterministically chaotic or some combination of these categories. More generally, the extent to which a non-linear deterministic process retains its properties when corrupted by noise is also unclear. The noise can affect a system in different way, either in an additive way or as a measurement error, even though the equations of the system remain deterministic.

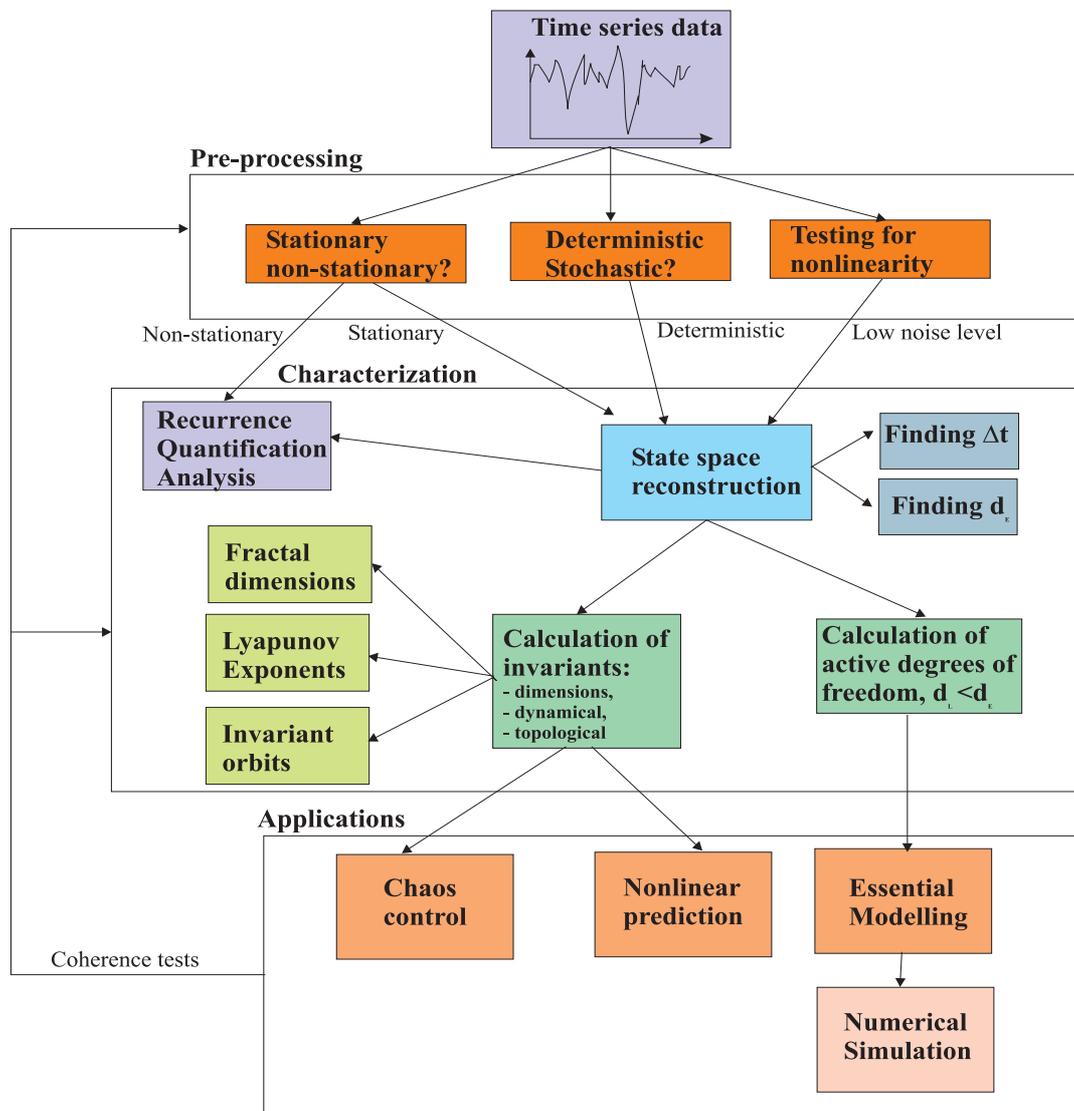


Figure 5. Schematic representation of nonlinear time series analysis using delay coordinate embedding (Strozzi and Zaldívar, 2002).

A schematic representation of the different steps is given in Fig. 5. Since a single reliable statistical test for chaoticity is not available, combining multiple tests is a crucial aspect, specially when one is dealing with limited and noisy data sets like in economic and financial time series.

There are different aspects that should be carefully studied before attempting to go further using nonlinear time series analysis methods. A long and exhaustive discussion can be found in Schreiber (1998) and the basic methodologies will be presented during the analysis part. Here, we are briefly going to indicate the main problems one should be aware of. These can be summarized as follows:

- has the phenomenon been sufficiently sampled?;
- is the data set stationary or can one remove the nonstationary part?;
- is the level of noise sufficiently low so that one can obtain useful information using nonlinear time series techniques?

Some tests to study these questions have been recently implemented in the TISEAN software package (Hegger *et al.*, 1999), which has incorporated a substantial quantity of algorithms developed for nonlinear time series analysis.

The problem of the number of samples needed to carry out state space reconstruction is related to the dimensionality of the problem we are dealing with. In order to characterize properly the underlying dynamics from the observed time series, we need to sample properly the phase space in which our dynamical system lies. As the dimension of the underlying system increases, a higher number of samples is needed. Ruelle (1990) discussed this problem, and based on simple geometrical considerations, he arrived at the following conclusion: if the calculated dimension of our system is well below $2\log_{10}m$, where m is the total number of points in the original time series, then we are using a sufficient number of data points. Of course having a sufficient number of data points is a necessary but not a sufficient condition for reliable nonlinear time series analysis.

Another related problem is the sampling rate. Consider the case when we are sampling data from a, presumably, chaotic system. Chaotic systems, like stochastic ones, are unpredictable in the long run. This long run is related to the speed at which nearby trajectories diverge in phase space, which turns out to be related to the Lyapunov exponents of the system under study. Hence, if we are sampling at a rate slower than our predictability window, even though the underlying system is chaotic, we will find that our system behaves as a stochastic one. In this situation, if one suspects that the underlying system is deterministic, the best thing to do is to repeat the experiment by increasing the sampling rate. Interpolating between data points would be of no use as no new information is introduced.

A time series is said to be strictly stationary if its statistical distribution does not change across time. More specifically, suppose we have a set of m samples of the series $s(t)$ made at times t_1 through t_m , these need not be contiguous times. Strict stationarity implies that the joint probability density function of those m samples is identical to the joint probability distribution of another m samples taken at times t_{1+k} through t_{m+k} . This must be true for all the choices of m and k , as well for the m relative sample times. Why is stationarity so important? Because almost all methods developed by linear and nonlinear time series analysis assume that the time series we are analysing is stationary, which implies that the parameters of the system that has generated the time series, remain constant. For this reason time-series analysis often requires one to turn a nonstationary series into a stationary one so as to use these theories. Unfortunately, nonstationary signals are very common in particular

when observing natural or man-made phenomena, and in some cases the nonstationary components, such as the trend, may sometimes be of more interest than that of the stationary part obtained by removing the trend or the seasonal variation from the signal.

Even though a precise definition of stationarity exists, there is no magic formula for deciding whether a series is stationary or not. However, strong violations of the basic requirements that the dynamical properties of the system must not change, beyond their statistical fluctuations, can be checked simply by measuring such properties, i.e. mean, variance, spectral components, correlations, etc., for several segments of the data set. Nonlinear time series analysis has also developed its own techniques to study nonstationarity as we will see bellow.

4.1. Preliminary Analysis

4.1.1. Surrogate time series generation

If the dynamics that has generated the time series is not known or if the data are noisy, it is important to investigate whether the amount of nonlinear deterministic dependencies is worth analyzing further or whether the series can be considered as stochastic. Hence, one of the first steps before applying nonlinear techniques to the Nord Pool data is to investigate if the use of such advanced techniques is justified by the data. The main reason behind this reasoning is that linear stochastic processes can create very complicated looking signals and that not all the structures that we find in a data set are likely to be due to nonlinear dynamics going on within the system. The method of surrogate data, see for example Schreiber and Schmitz (2000) for a review, has become a useful tool to address the question if the irregularity of the data is most likely due to nonlinear deterministic structure or rather due to random inputs to the system or fluctuations in the parameters.

The method of surrogate data, which was first introduced by Theiler *et al.* (1992) in nonlinear time series analysis, consists of generating an ensemble of “surrogate” data sets similar to the original time series, but consistent with the null hypothesis, usually that the data have been created by a stationary Gaussian linear process, and of computing a discriminating statistic for the original and for each of the surrogate data sets

In general a linear stochastic process can be described by

$$x_n = a_0 + \sum_{i=1}^{M_1} a_i x_{n-i} + \sum_{j=1}^{M_2} b_j \eta_{n-j} \quad (9)$$

where η_n are independent Gaussian random numbers with zero mean and unit variance and a_i , b_i , M_1 and M_2 are constants. This is called an ARMA(M_1 , M_2) process. Now we want to test the hypothesis that the data could be explained by a linear model. A statistical significance test consists on the following steps: a/ we compute some nonlinear observable λ_0 from the data; b/ we observe if the value obtained suggests that the data are nonlinear and we calculate the same quantity from a number of comparable linear models. If the results are completely different the data might be nonlinear. If we have any theory for the distribution of the values of λ_i for linear stochastic process, we can estimate their distribution using the method of *surrogate data*. The null hypothesis that we want to test is that the data results from a Gaussian linear stochastic process. Then we should specify the *level of significance*. If we allow for a 5% chance that we reject the null hypothesis although

it is in fact true (valid at a 95% significance level) then more than one wrong result out of 19 is usually not considered acceptable (Schreiber and Schmitz, 2000). How to make surrogate data sets? Let us suppose that the data came from a stationary linear stochastic process with Gaussian inputs. We consider the mean, the variance and the autocorrelation function of the real data or equivalently the mean and the power spectrum.

We can create surrogate data by taking their fast (discrete) Fourier transform (FFT) and multiplying it by a random phase parameter uniformly distributed in $[0, 2\pi[$, then it is possible to compute the inverse of FFT and we have a time series with the prescribed spectrum. Different realization of the random phase gives new surrogate data. This process of phase randomisation preserves the Gaussian distribution.

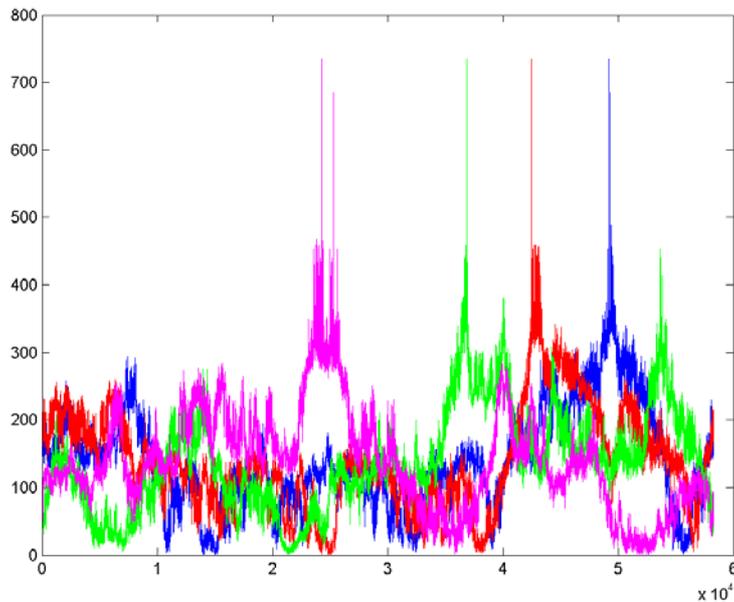


Figure 6. Four surrogate time series generated for the Nord pool spot price in Norwegian Krone (Fig. 1) using the TISEAN software package (Hegger *et al.*, 1999) with *surrogates* program.

In this work, we have created 19 surrogate data sets (same mean, variance and Fourier power spectrum) for each Nord Pool spot prices time series, for example see Fig. 6, these time series data comes from a stationary linear stochastic process with Gaussian inputs.

In addition, we have also considered another null hypothesis: the data are simply *temporally uncorrelated noise* i.e. the null hypothesis is any correlation at all. Surrogate data in this case are generated by a random shuffling of the original time series. Also, in this case, we have created 19 surrogate data sets from the original time series.

4.1.2. R/S Analysis

A tool for studying long-term memory and fractality of a time series is the Rescaled Range analysis (R/S analysis) first introduced by Hurst (1951) in hydrology. Mandelbrot (1983) argued that R/S analysis is a more powerful tool in detecting long range dependence when compared to more conventional analysis like autocorrelation analysis, variance ratios and spectral analysis. In this method, one measures how the range of cumulative deviations from the mean of the series is changing with the time. It has been found that, for some time series, the dependence of R/S on the number of data points (or time) follows an empirical power law

described as $(R/S)_n = (R/S)_0 n^H$, where $(R/S)_0$ is a constant, n is the time index for periods of different length, and H is the Hurst exponent. $(R/S)_n$ is defined as

$$\left(\frac{R}{S}\right)_n = \frac{\max_{1 \leq t \leq n} A(t, n) - \min_{1 \leq t \leq n} A(t, n)}{\sqrt{\frac{1}{n} \sum_{t=1}^n (s(t) - \langle s \rangle_n)^2}} \quad (10)$$

where $A(t, n)$ is the accumulated departure of the time series $s(t)$ from the time average over the time interval n : $\langle s \rangle_n$. $A(t, n) = \sum_{i=t}^{t+n} (s(i) - \langle s \rangle_n)$.

The Hurst exponent, $0 \leq H \leq 1$, is equal to 0.5 for random walk time series, < 0.5 for anticorrelated series, and > 0.5 for positively correlated series. The Hurst exponent is directly related to the "fractal dimension", which gives a measure of the roughness of a surface. The relationship between the fractal dimension, D , and the Hurst exponent, H , is:

$$D = 2 - H \quad (11)$$

Hurst exponents quantify the correlation of a fractional Brownian motion. A fractional Brownian motion (fBm) is a random walk with a Hurst exponent different from 0.5 and then with a memory. The decaying of spectral density of a fBm has a relationship with the Hurst exponent as follow:

$$spectral\ density \propto \frac{1}{f^\alpha} \quad (12)$$

where $\alpha = 2H + 1$.

Financial time series have been found to exhibit some universal characteristics that resemble the scaling laws typical of natural systems in which large numbers of units interact. For instance, the Hurst exponent has been extensively applied by Peters (1996) to various capital markets and in most of the cases he has found persistent memory.

A long memory process is a process with a random component, where a past event has a decaying effect on future events. The process has some memory of past events, which is "forgotten" as time moves forward. The mathematical definition of long memory processes is given in terms of autocorrelation. When a data set exhibits autocorrelation, a value x_i at time t_i is correlated with a value x_{i+d} at time t_{i+d} , where d is some time increment in the future. In a long memory process autocorrelation decays over time and the decay follows a power law, i.e

$$p(k) = Ck^{-\beta} \quad (13)$$

where, C is a constant and $p(k)$ is the autocorrelation function with lag k . The Hurst exponent is related to the exponent β by

$$H = 1 - \frac{\beta}{2} \quad (14)$$

In this work we have used the standard scaled windowed variance method (Cannon *et al.*, 1997) to estimate H by linear regression of $\log(R/S)$ versus $\log(Windowsize)$. The results for the two original time series and the surrogate series are shown in Tables 4-5. As it can be seen both time series show antipersistence, $H < 0.5$. This

has already been found by several researchers (Weron and Przybylowicz, 2000; Simonsen 2003; Perelló *et al.* 2007, amongst others). In all cases, the Hurst exponents of the original time series are slightly higher than those of the linear surrogate time series but this does not mean that the value of H helps us to distinguish between the original time series and their surrogates, because H for the linear surrogate of NOK is higher than H for real EUR (Table 4). For the shuffled surrogate time series we can observe that H for surrogates is nearer to 0.5 independently if we consider the surrogate of NOK or of EUR (Table 5).

Table 4. Hurst exponents for the Nord pool and the surrogate linearly correlated time series.

Data set	H	Data set	H
<i>NOK</i>	0.4406	<i>EUR</i>	0.2673
<i>Surr01 nl</i>	0.3632	<i>Surr01 el</i>	0.1231
<i>Surr02 nl</i>	0.3824	<i>Surr02 el</i>	0.0899
<i>Surr03 nl</i>	0.3399	<i>Surr03 el</i>	0.1402
<i>Surr04 nl</i>	0.3646	<i>Surr04 el</i>	0.1631
<i>Surr05 nl</i>	0.3276	<i>Surr05 el</i>	0.1597
<i>Surr06 nl</i>	0.4151	<i>Surr06 el</i>	0.1325
<i>Surr07 nl</i>	0.3480	<i>Surr07 el</i>	0.0914
<i>Surr08 nl</i>	0.3497	<i>Surr08 el</i>	0.2262
<i>Surr09n l</i>	0.3125	<i>Surr09 el</i>	0.1063
<i>Surr10 nl</i>	0.3024	<i>Surr10 el</i>	0.1673
<i>Surr11 nl</i>	0.3396	<i>Surr11 el</i>	0.1434
<i>Surr12 nl</i>	0.3624	<i>Surr12 el</i>	0.1612
<i>Surr13 nl</i>	0.3795	<i>Surr13 el</i>	0.1990
<i>Surr14 nl</i>	0.3602	<i>Surr14 el</i>	0.1604
<i>Surr15 nl</i>	0.3574	<i>Surr15 el</i>	0.1368
<i>Surr16 nl</i>	0.3406	<i>Surr16 el</i>	0.2096
<i>Surr17 nl</i>	0.3874	<i>Surr17 el</i>	0.1814
<i>Surr18 nl</i>	0.3369	<i>Surr18 el</i>	0.1089
<i>Surr19 nl</i>	0.3774	<i>Surr19 el</i>	0.0905

Table 5. Hurst exponents for the Nord pool and the surrogate shuffled time series.

Data set	H	Data set	H
<i>NOK</i>	0.4406	<i>EUR</i>	0.2673
<i>Surr01 ns</i>	0.4886	<i>Surr01 es</i>	0.4293
<i>Surr02 ns</i>	0.4842	<i>Surr02 es</i>	0.4293
<i>Surr03 ns</i>	0.4847	<i>Surr03 es</i>	0.4423
<i>Surr04 ns</i>	0.4773	<i>Surr04 es</i>	0.4136
<i>Surr05 ns</i>	0.4857	<i>Surr05 es</i>	0.4267
<i>Surr06 ns</i>	0.5036	<i>Surr06 es</i>	0.4437
<i>Surr07 ns</i>	0.4877	<i>Surr07 es</i>	0.4349
<i>Surr08 ns</i>	0.5026	<i>Surr08 es</i>	0.4238
<i>Surr09 ns</i>	0.4888	<i>Surr09 es</i>	0.4366
<i>Surr10 ns</i>	0.4846	<i>Surr10 es</i>	0.4274
<i>Surr11 ns</i>	0.4874	<i>Surr11 es</i>	0.4261

<i>Surr12 ns</i>	0.4948	<i>Surr12 es</i>	0.4266
<i>Surr13 ns</i>	0.4848	<i>Surr13 es</i>	0.4315
<i>Surr14 ns</i>	0.4824	<i>Surr14 es</i>	0.4245
<i>Surr15 ns</i>	0.4690	<i>Surr15 es</i>	0.4279
<i>Surr16 ns</i>	0.4882	<i>Surr16 es</i>	0.4253
<i>Surr17 ns</i>	0.4780	<i>Surr17 es</i>	0.4292
<i>Surr18 ns</i>	0.4795	<i>Surr18 es</i>	0.4299
<i>Surr19 ns</i>	0.4798	<i>Surr19 es</i>	0.4270

4.1.3. Power Spectral Density

The Fourier transform of a function $s(t)$ is given by:

$$\tilde{s}(f) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} s(t) e^{2\pi i f t} dt \quad (15)$$

and that of a finite, discrete time series by

$$\tilde{s}_k = \frac{1}{\sqrt{N}} \sum_{j=1}^N s_j e^{2\pi i k j / N} \quad (16)$$

Here, the frequencies in physical units are $f_k = k/(N\Delta t)$, where $k = -N/2, \dots, N/2$ and Δt is the sampling interval (1 hour in our case). The power spectrum of a process is defined to be the squared modulus of the continuous Fourier transform, $P(f) = |\tilde{s}(f)|^2$. The power spectrum is particularly useful for studying the main frequencies in a system, since there will be sharper or broader peaks at the dominant frequencies and their integer multiples, the harmonics.

In Figures 7-8 we observe the power spectral density of Nord Pool time series. For both of them, we have found behaviour of the type

$$P(f) \propto \frac{1}{f^\alpha} \quad (17)$$

where α is a positive real number. The values of α for Nord Pool time series and their surrogates are listed in Tables 6-7.

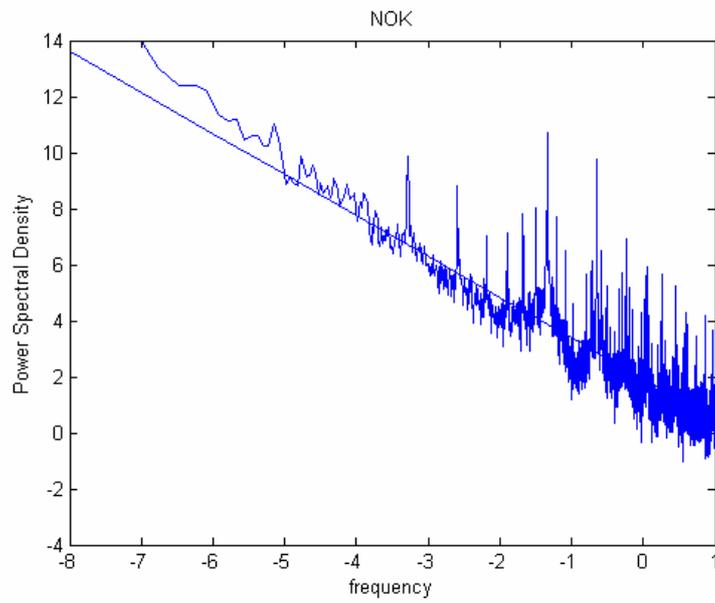


Figure 7. The power spectrum (log-log scale) , NOK ($\alpha=1.4612$).

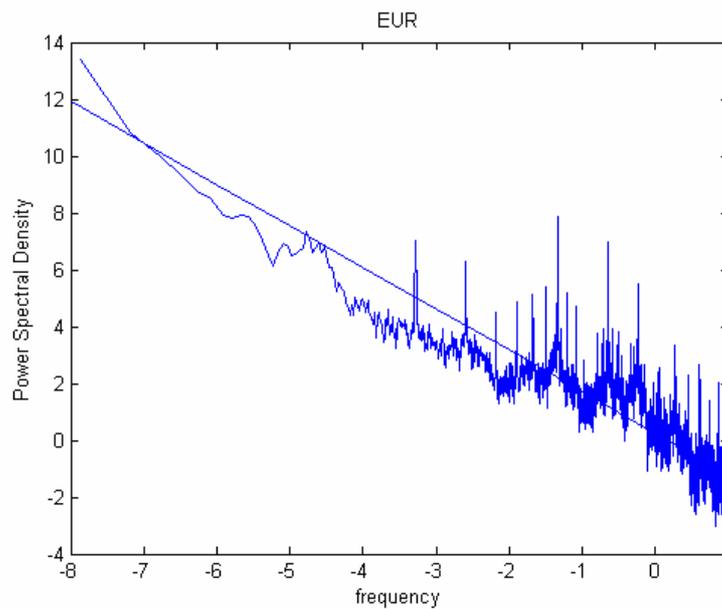


Figure 8. The power spectrum (log-log scale) EUR ($\alpha=1.4562$).

Table 6. Power spectra trend α calculated using linear regression (LR) and Hurst exponent H. Real and linearly correlated data

Data set	α (LR)	$\alpha=2H+1$	Data set	α(PS)	$\alpha=2H+1$
<i>NOK</i>	1.4612	1.8812	<i>EUR</i>	1.4562	1.5346
<i>Surr01 nl</i>	1.3626	1.7264	<i>Surr01 el</i>	1.4281	1.2462
<i>Surr02 nl</i>	1.3577	1.7648	<i>Surr02 el</i>	1.4453	1.1798
<i>Surr03 nl</i>	1.3527	1.6798	<i>Surr03 el</i>	1.4296	1.2804
<i>Surr04 nl</i>	1.3440	1.7292	<i>Surr04 el</i>	1.4546	1.3262
<i>Surr05 nl</i>	1.4719	1.6552	<i>Surr05 el</i>	1.3914	1.3194
<i>Surr06 nl</i>	1.4085	1.8302	<i>Surr06 el</i>	1.4300	1.2650
<i>Surr07 nl</i>	1.3383	1.6960	<i>Surr07 el</i>	1.4140	1.1828
<i>Surr08 nl</i>	1.3628	1.6994	<i>Surr08 el</i>	1.4131	1.4524
<i>Surr09 nl</i>	1.3406	1.6250	<i>Surr09 el</i>	1.4558	1.2126
<i>Surr10 nl</i>	1.3566	1.6048	<i>Surr10 el</i>	1.4289	1.3346
<i>Surr11 nl</i>	1.3564	1.6792	<i>Surr11 el</i>	1.4466	1.2868
<i>Surr12 nl</i>	1.3301	1.7248	<i>Surr12 el</i>	1.4404	1.3224
<i>Surr13 nl</i>	1.3652	1.7590	<i>Surr13 el</i>	1.4276	1.3980
<i>Surr14 nl</i>	1.3767	1.7204	<i>Surr14 el</i>	1.4330	1.3208
<i>Surr15 nl</i>	1.3690	1.7148	<i>Surr15 el</i>	1.4341	1.2736
<i>Surr16 nl</i>	1.3701	1.6812	<i>Surr16 el</i>	1.4506	1.4192
<i>Surr17 nl</i>	1.3589	1.7748	<i>Surr17 el</i>	1.4455	1.3628
<i>Surr18 nl</i>	1.3483	1.6738	<i>Surr18 el</i>	1.4554	1.2178
<i>Surr19 nl</i>	1.3646	1.7548	<i>Surr19 el</i>	1.4666	1.1810

Table 7. Power spectra trend α calculated using linear regression (LR) and Hurst exponent H. Real and shuffled data

Data set	α (LR)	$\alpha=2H+1$	Data set	α(PS)	$\alpha=2H+1$
<i>NOK</i>	1.4612	1.8812	<i>EUR</i>	1.4562	1.5346
<i>Surr01 ns</i>	0.0100	1.0200	<i>Surr001 es</i>	0.0139	1.0278
<i>Surr02 ns</i>	0.0106	1.0212	<i>Surr002 es</i>	0.0148	1.0296
<i>Surr03 ns</i>	0.0150	1.0300	<i>Surr003 es</i>	0.0084	1.0168
<i>Surr04 ns</i>	0.0294	1.0588	<i>Surr004 es</i>	0.0237	1.0474
<i>Surr05 ns</i>	0.0200	1.0400	<i>Surr005 es</i>	0.0097	1.0194
<i>Surr06 ns</i>	0.0151	1.0302	<i>Surr006 es</i>	0.0019	1.0038
<i>Surr07 ns</i>	0.0212	1.0424	<i>Surr007 es</i>	0.0153	1.0306
<i>Surr08 ns</i>	0.0099	1.0198	<i>Surr008 es</i>	0.0126	1.0252
<i>Surr09 ns</i>	0.0243	1.0486	<i>Surr009 es</i>	0.0147	1.0294
<i>Surr10 ns</i>	0.0208	1.0416	<i>Surr010 es</i>	0.0138	1.0276
<i>Surr11 ns</i>	0.0216	1.0432	<i>Surr011 es</i>	0.0208	1.0416
<i>Surr12 ns</i>	0.0183	1.0366	<i>Surr012 es</i>	0.0137	1.0274
<i>Surr13 ns</i>	0.0218	1.0436	<i>Surr013 es</i>	0.0114	1.0228
<i>Surr14 ns</i>	0.0265	1.0530	<i>Surr014 es</i>	0.0154	1.0308
<i>Surr15 ns</i>	0.0261	1.0522	<i>Surr015 es</i>	0.0182	1.0364
<i>Surr16 ns</i>	0.0124	1.0248	<i>Surr016 es</i>	0.0247	1.0494
<i>Surr17 ns</i>	0.0192	1.0384	<i>Surr017 es</i>	0.0177	1.0354
<i>Surr18 ns</i>	0.0160	1.0320	<i>Surr018 es</i>	0.0106	1.0212
<i>Surr19 ns</i>	0.0133	1.0266	<i>Surr019 es</i>	0.0201	1.0402

It has been observed experimentally (Shuster, 1995) that the power spectra of a large variety of physical systems diverge at low frequencies with a power law $1/f^\alpha$ ($0.8 < \alpha < 1.4$).

The appearance of a scaling behaviour in the power spectrum of economic time series support further, according to Theiler (1991), the existence of a self-organisation with many degree of freedom for these series.

If the motion is a fractional Brownian motion (fBm) a relationship exists between the Hurst exponent and the scaling factor of the power law α , see Eq. (11). We have calculated α from Hurst and directly from the spectrum. The results are presented in Tables 6 and 7. In all the time series considered, real and surrogate, the values are significantly different. However, this is not conclusive since there is a certain amount of variability calculating the Hurst exponents as well as α that may be responsible for these differences in particular for the linear surrogate time series.

4.1.4. Fitting Nord Pool data with stable distributions

Stable distributions are a class of distributions that have the property of stability: if a number of independent and identically distributed (iid) random variables have a stable distribution, then a linear combination of these variables will have the same distribution, except for possibly different shift and scale parameters. Special cases of stable distributions include Gaussian, Cauchy and Levy distributions. They are described by four parameters the first two are $\alpha \in (0, 2]$, an index of stability and $\beta \in [-1, 1]$, a skewness parameter. α and β determine the shape of the distribution. The last parameters are $\gamma \in [0, \infty)$ a scale parameter and $\delta \in (-\infty, \infty)$ a location parameter.

A stable probability distribution is defined by the Fourier transform of its characteristic function $\varphi(t)$:

$$f(x; \alpha, \beta, \gamma, \delta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(t) e^{-itx} dt \quad (18)$$

where $\varphi(t)$ is given by

$$\varphi(t) = \exp\left[it\delta - | \gamma t |^\alpha (1 - i\beta \operatorname{sgn}(t)\Phi) \right] \quad (19)$$

and $\operatorname{sgn}(t)$ is just the sign of t and Φ is given by

$$\Phi = \tan(\pi\alpha/2) \quad (20)$$

for all α except $\alpha=1$ in which case:

$$\Phi = -(2/\pi) \log(t) \quad (21)$$

There is no general analytic expression for a stable distribution. There are, however four special cases which can be analytically expressed:

a/ for $\alpha=2$ the distribution becomes a Gaussian distribution with variance $\sigma^2 = 2\gamma^2$ and mean δ

b/ for $\alpha=1$ and $\beta=0$ the distribution reduces to a Cauchy distribution with scale parameter γ and shift parameter δ

c/ for $\alpha=1/2$ and $\beta=1$ the distribution reduces to a Levy distribution with scale parameter γ and shift parameter δ

d/ In the limit as $\gamma \rightarrow 0$ or as $\alpha \rightarrow 0$ the distribution will approach a Dirac delta function $\delta(x - \delta)$

The heavy tail behaviour causes the variance of stable distribution to be infinite for $\alpha < 2$ (if $\alpha = 2$ the distribution is Gaussian).

Stable distributions have been proposed as a model for many types of physical and economic systems because many large data sets exhibit heavy tails and skewness. Anyway, while non-Gaussian stable distributions are heavy tailed, most heavy-tailed distributions are not stable.

In order to analyse these series we have fitted the histogram to the first difference, of each series with a stable distribution (Nolan, 1999), $X \sim S(\alpha, \beta, \gamma, \delta; A)$, using the program STABLE for univariate data (<http://www.cas.american.edu/~jpnolan>). The last parameter A can be 0 or 1 respectively if the characteristic function is continuous in all four parameters or not. We will consider the first case $A=0$. A typical situation in these time series is the existence of a high number of zero values normally in correspondence with weekends or holidays. To compare the results, we have eliminated from the original series the points where the exchange rate was unchanged, i.e. the zero value. Table 8 summarizes the fitted parameters using the maximum likelihood estimation (Nolan, 1997 and 1999); whereas in figs. 9-10 the fit obtained using both approaches is shown. The other methods implemented in STABLE, i.e. quantile/fractile method and sample characteristic function, gave similar results.

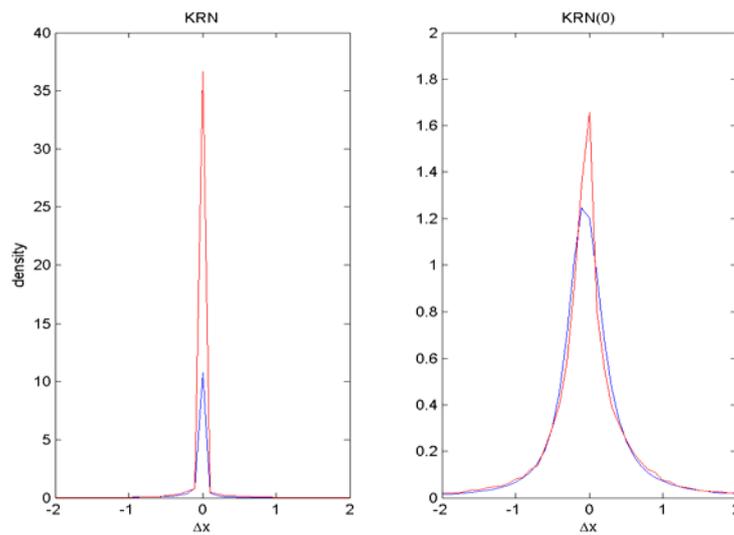


Figure 9. Fitted density plot for the Nord Pool Norwegian Krone time series data (blue line): a/Original time series, first difference; b/ without zero values (23962 values).

As can be observed in figs. 9 and 10, the first Nord Pool time series had a considerable amount of first differences equal to zero, i.e. no change between one spot price and the successive. This high value makes it difficult to fit a stable probability distribution (see fig. 9a). On the contrary, in the EUR Nord Pool time series this problem is not so evident and the stable parameters are quite similar with or without the zero values (see Table 8 and Fig. 10).

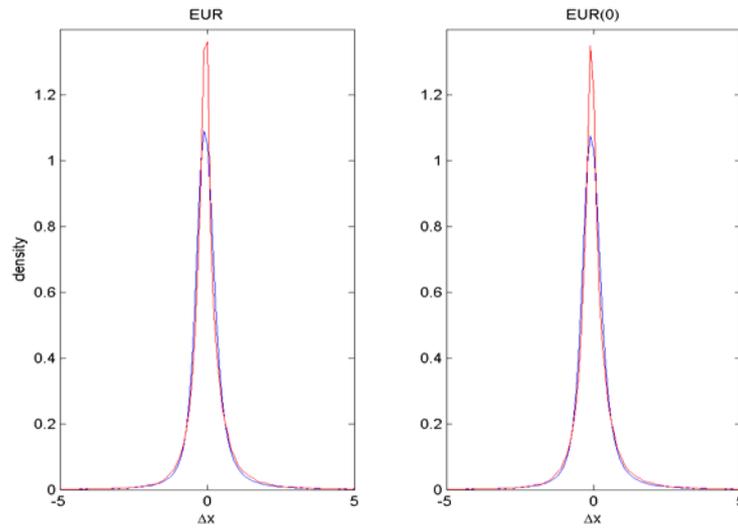


Figure 10. Fitted density plot for the Nord Pool Norwegian Euro time series data (blue line): a/Original time series, first difference; b/ without zero values (730 values).

Table 8. Nord Pool data fitted parameters using STABLE (Nolan, 1999).

Data set	α	β	γ	δ
KRN	0.412	-0.365	0.035	-0.00018
KRN(0)	1.116	0.127	0.242	-0.0514
EUR	1.308	0.164	0.268	-0.068
EUR(0)	1.315	0.173	0.272	-0.069

Afterwards, the surrogate time series for the Nord Pool in EUR have been compared with the original time series. The results are summarized in Table 9. The shuffled surrogate time series have all the same stable parameters as the original time series.

Table 9. Parameters of stable distribution that fit Nor Pool data in EUR/kwh and its surrogates linearly correlated.

Data set	α	β	γ	δ
<i>EUR</i>	1.308	0.164	0.268	-0.068
<i>Surr01 nl</i>	1.799	0.0397	1.16264	-0.6388E-2
<i>Surr02 nl</i>	1.7831	0.0264	1.13151	-0.61447E-2
<i>Surr03 nl</i>	1.6640	0.0378	1.02906	-0.0210
<i>Surr04 nl</i>	1.7762	0.0036	1.01902	0.1162E-2
<i>Surr05 nl</i>	1.8062	0.0205	1.11834	-0.3311e-2
<i>Surr06 nl</i>	1.6059	0.0011	0.903214	-0.66719E-2
<i>Surr07 nl</i>	1.7837	-0.0093	1.16422	0.30468E-2
<i>Surr08 nl</i>	1.7693	-0.0553	0.998577	0.15710E-1
<i>Surr09 nl</i>	1.7481	0.0135	1.18132	-0.22515E-2
<i>Surr10 nl</i>	1.7621	-0.0061	1.09915	0.35807E-2
<i>Surr11 nl</i>	1.7604	0.0624	1.13694	-0.27875E-1
<i>Surr12 nl</i>	1.7505	0.0450	1.07224	-0.15495E-1
<i>Surr13 nl</i>	1.7156	0.0161	1.05419	-0.827819E-2

<i>Surr14 nl</i>	1.7392	-0.0293	1.18483	0.67472E-2
<i>Surr15 nl</i>	1.7039	0.0619	1.00392	-0.15986E-1
<i>Surr16 nl</i>	1.7894	0.0155	1.10006	-0.69631E-2
<i>Surr17 nl</i>	1.8059	-0.0368	1.17309	0.87207E-2
<i>Surr18 nl</i>	1.7869	-0.0213	1.11810	0.60565E-2
<i>Surr19 nl</i>	1.7777	0.0038	1.19410	0.75471E-2

By comparing the surrogate data sets it is possible to observe that they have a probability distribution function (pdf) more similar to a Gaussian (α close to 2) in comparison with original data ($\alpha = 1.308$) and they have β closer to 0 which mean their pdfs have less skewness.

4.2. Finding the time delay and embedding dimension

Determining the time delay and the embedding dimension is considered as one of the most important steps in nonlinear time series modelling and prediction. A number of methods have been developed in determining the time delay and the minimum embedding dimension since the early beginning of nonlinear time series study. Here we will describe and apply several of them to the foreign exchange time series data sets.

4.2.1. Time delay

The first step in phase space reconstruction is to choose an optimum delay parameter τ . Different prescriptions have appeared in the literature to choose τ but they are all empirical in nature and do not necessarily provide appropriate estimates:

- *First passes through zero of the autocorrelation function*: In earlier works (Mees *et al.*, 1987) it was suggested to use the value of τ for which the autocorrelation function

$$C(\tau) = \sum_n [s(n) - \bar{s}][s(n + \tau) - \bar{s}] \quad (22)$$

first passes through zero which is equivalent to requiring linear independence.

The application of the zero crossing of the autocorrelation function gives quite high values for both time series, see Fig. 11.

- *First minimum of the Average mutual information*: Fraser and Swinney (1986) suggested to use the average mutual information (AMI) function, $I(\tau)$, as a kind of nonlinear correlation function to determine when the values of $s(n)$ and $s(n + \tau)$ are independent enough of each other to be useful as coordinates in a time delay vector but not so independent as to have no connection which each other at all. For a discrete time series, $I(\tau)$ can be calculated as,

$$I(\tau) = \sum_{n, n+\tau} P(s(n), s(n + \tau)) \log_2 \left[\frac{P(s(n), s(n + \tau))}{P(s(n))P(s(n + \tau))} \right] \quad (23)$$

where $P(s(n))$ refers to individual probability and $P(s(n), s(n + \tau))$ is the joint probability density. Following the method developed by Abarbanel (1996), to determine $P(s(n))$ we simply project the values taken from $s(n)$ versus n back onto the $s(n)$ axis and form an histogram of the values. Once normalised, this gives us $P(s(n))$. For the join distribution of $s(n)$ and $s(n + \tau)$ we form the two-dimensional histogram in the same way.

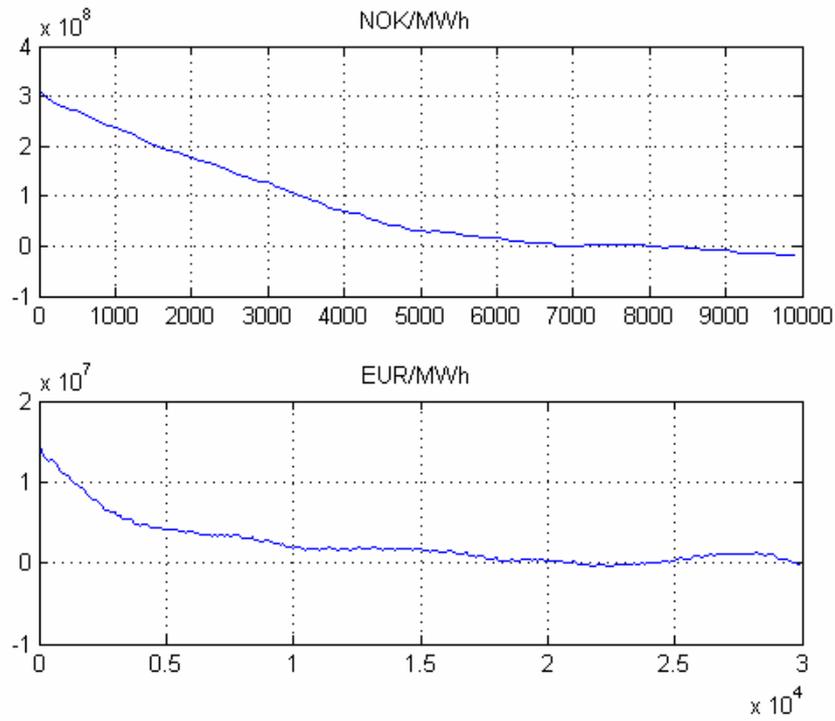


Figure 11. Correlation f of the Nord Pool time series.

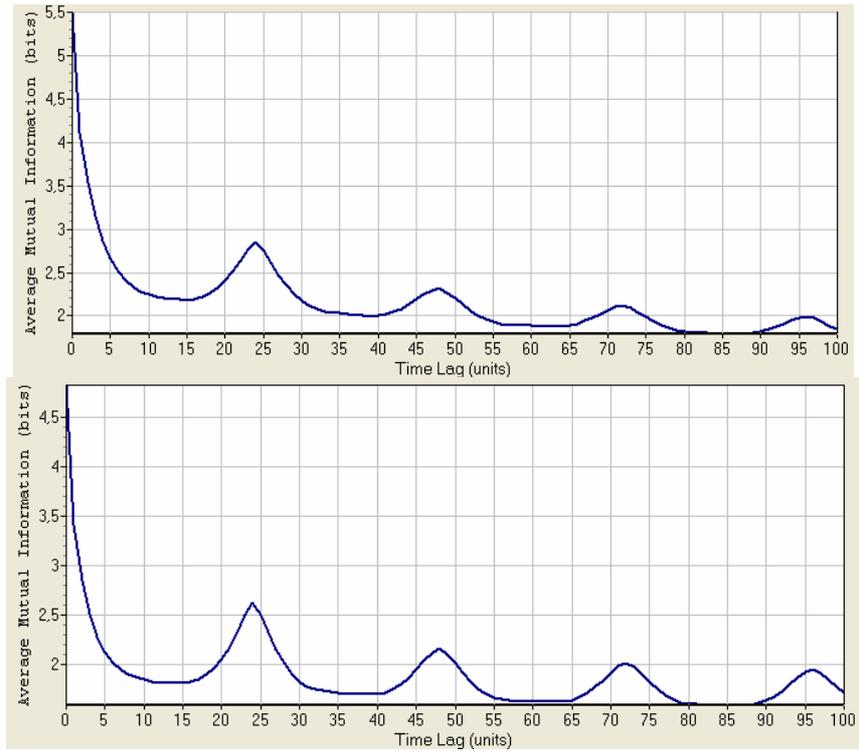


Figure 12. Average Mutual information function. First minimum for NOK/MWh time series (top) occurs at $\tau=15$, whereas for the EUR/MWh time series (bottom) for $\tau=13$.

In general, the time lag provided by $I(\tau)$ is normally lower than the one calculated with the $C(\tau)$, $\tau_{AMI} \geq \tau_{Correl}$, and provides the appropriate characteristic time scales for the motion. Even though $C(\tau)$ is the optimum linear choice from the point of view of predictability in a least square sense of $s(n + \tau)$ from knowledge of $s(n)$, it is not clear why it should work for nonlinear systems and it has been shown that in some cases it does not work at all.

4.2.2. Embedding dimension

The dimension, where a time delay reconstruction of the system phase space provides a necessary number of coordinates to unfold the dynamics from overlaps on itself caused by projection, is called the embedding dimension, d_E . This is a global dimension, which can be different from the real dimension. Furthermore, this dimension depends on the time series measurement, and hence, if we measure two different variables of the system, there is no guarantee that the d_E from time delay reconstruction will be the same from each of them.

The usual method for choosing the minimum embedding dimension is to compute some invariant of the attractor. By increasing the embedding dimension used for the computations, one notes when the value of the invariant stops changing. Since these invariants are geometric properties of the dynamics, they become independent of d for $d \geq d_E$, i.e. after the geometry is unfolded.

In this work, we have used three methods:

- Saturation of the correlation dimension: The correlation dimension is a measure of the dimension obtained considering correlations between points. If N is the number of points in the time serie, τ is a fixed increment of time and $\{x_i\}_{i=1}^T \equiv \{x(t + i\tau)\}_{i=1}^T$ the *correlation integral* is defined as:

$$C(\varepsilon) = \lim_{T \rightarrow \infty} \frac{1}{T^2} \sum_{\substack{i,j=1 \\ i \neq j}}^T \Theta(\varepsilon - \|x_i - x_j\|) \quad (24)$$

where

$$\Theta(x) = \begin{cases} 1 & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (25)$$

is the Heaviside function and $\|\cdot\|$ denote the Euclidean norm. The function $C(\varepsilon)$ behaves as a power of ε for small ε :

$$C(\varepsilon) \propto \varepsilon^{\nu} \quad (26)$$

the exponent ν is called *correlation dimension*.

The correlation dimension is frequently used to distinguish between chaotic and random behaviour. The idea behind it is to construct a function $C(\varepsilon)$ that is the probability that two arbitrary points on the orbit are closer together than ε . The correlation dimension is given by $\log(C)/\log(\varepsilon)$ in the limit $\varepsilon \rightarrow 0$, and $N \rightarrow \infty$. The correlation dimension is defined as the slope of the curve $C(\varepsilon)$ versus ε . $C(\varepsilon)$ is the correlation of the data set, or the probability that any two points in the set are separated by a distance ε . A noninteger result for the correlation dimension indicates that the data is probably fractal. In VRA, $C(\varepsilon)$ is calculated for every embedding dimension specified in the range and plotted against that range. For the truly random signals, the

correlation dimension graph will look like a 45-degree straight line, indicating that no matter how you embed the noise, it will evenly fill that space. Chaotic (and periodic) signals, on the other hand, have a distinct spatial structure, and their correlation dimension will saturate as some point, as embedding dimension is increased.

For our two time series the saturation does not occur at least until of an embedding dimension of 20, but this can be due to the presence of noise in the signal.

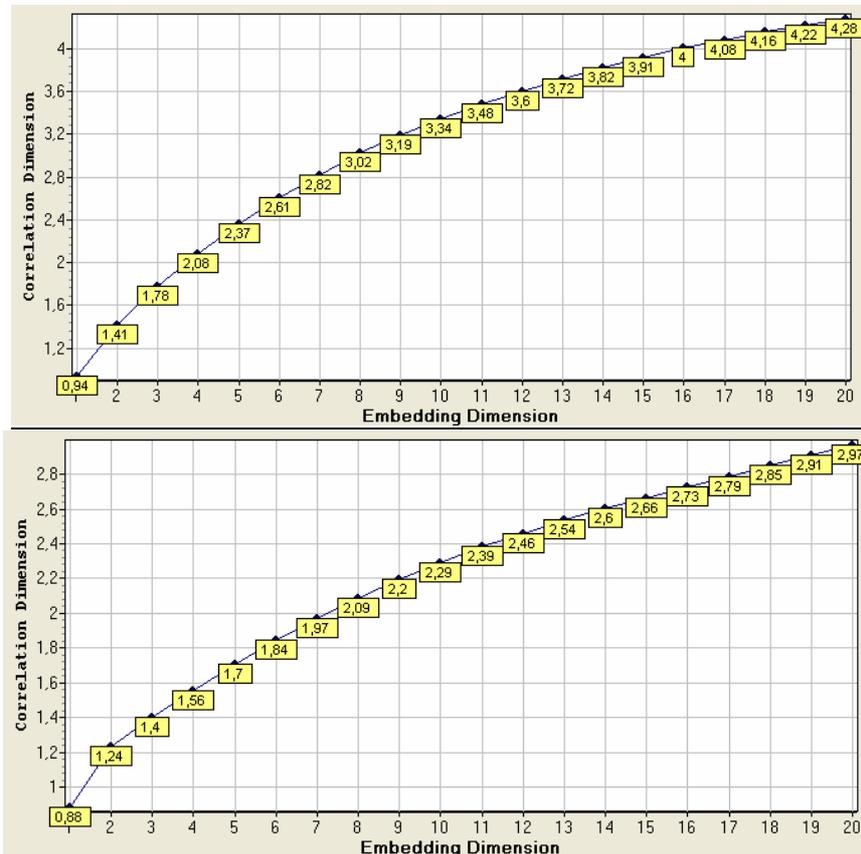


Figure 13. Correlation dimension for NOK/MWh time series (top), and EUR time series (bottom).

- *False Nearest Neighbours*: The method of False Nearest Neighbours (FNN) was developed by Kennel *et al.* (1992). In this case, the condition of no self-intersection states that if the dynamics is to be reconstructed successfully in R^d , then all the neighbour points in R^d should be also neighbours in R^{d+1} . The method checks the neighbours in successively higher embedding dimensions until it finds only a negligible number of false neighbours when increasing dimension from d to $d+1$. This d is chosen as the embedding dimension.

It was found by Kennel *et al.* (1992) that if the data set is clean from noise, the percentage of false nearest neighbours will drop from nearly 100% in dimension one to strictly zero when d_E is reached. Further, it will remain zero from then on since the dynamics is unfolded. If the signal is contaminated with noise (infinite dimension signal) we may not see the percentage of false nearest neighbours drop to near zero in any dimension. In this case, depending on the signal to noise ratio the determination of d_E will degrade.

For both time series, the FNN method suggest an embedding dimension of 6, see fig 14. The increase of the number of FNN after a certain d_E is an indication of the presence of noise in the signal.

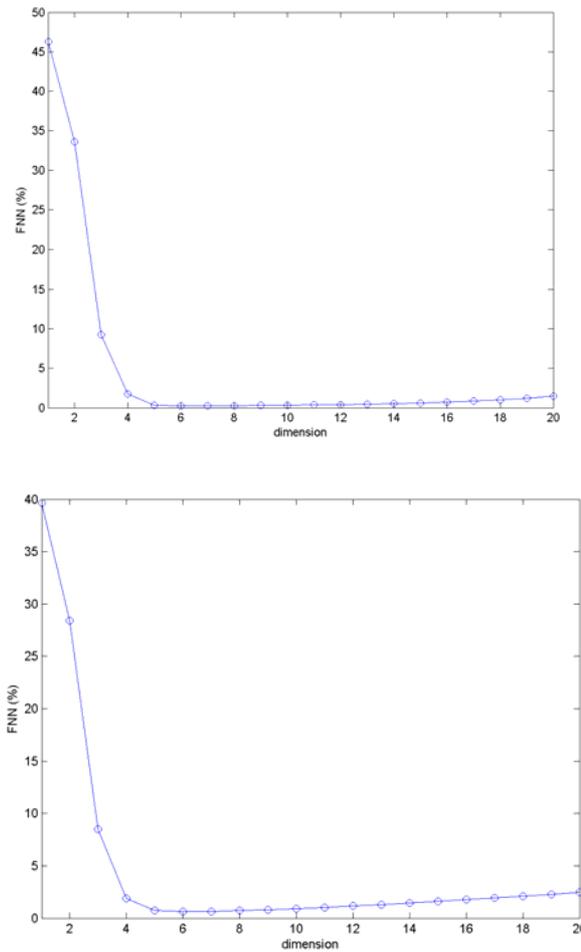


Figure 14. Embedding dimension using the FNN method. a/NOK/MWh time series; b/ EUR/MWh time series.

- *E1 & E2 Method* : The method of FNN has some subjectivity in defining that a neighbour is false since the values of two threshold parameters have to be defined, Kennel *et. al* (1992). To improve this situation, Cao (1997) developed a similar method, which is based on evaluating the mean value of the distance between time-delay vectors, $E1(d)$. However, if we look only to the quantity $E1(d)$ we can obtain wrong results in the case of random signals. For time series data from a random set of numbers $E1(d)$, in principle, will never reach a saturation value as d increase. But in practical computations, it is difficult to resolve whether the $E1(d)$ is slowly increasing or has stopped changing if d is sufficiently large. In Fact, since available observed data samples are limited, it may happen that the $E1(d)$ stop changing at some d although the time series is random. To solve this problem Cao (1997) suggested to consider the quantity $E2(d)$. Let $y_i(d) = \{s(i), s(i + \tau), \dots, s(i + (d - 1)\tau)\}$ and $y_{n(i,d)}$ the nearest neighbour of $y_i(d)$ in the d -dimensional reconstructed state space, then it is possible to define:

$$E^*(d) = \frac{1}{N - d\tau} \sum_{i=1}^{N-d\tau} |s_{i+d\tau} - s_{n(i,d)+d\tau}| \quad (27)$$

$$E2(d) = \frac{E^*(d+1)}{E^*(d)} \tag{28}$$

Since the future values are independent of the past values, $E2(d)$, for random data, will be equal to 1 for any d . However, for deterministic data, $E2(d)$ is certainly related to d , and it cannot be a constant for all d . In other words, there must exist some d 's such that $E2(d) \neq 1$. The E1&E2 method depends only on the time delay, and the embedding dimension is calculated, as in the other methods, when the values of E1 and E2 reach saturation. Cao (1997) showed that the method does not strongly depend on how many points are available, provided there are enough and it can clearly distinguish between deterministic and stochastic.

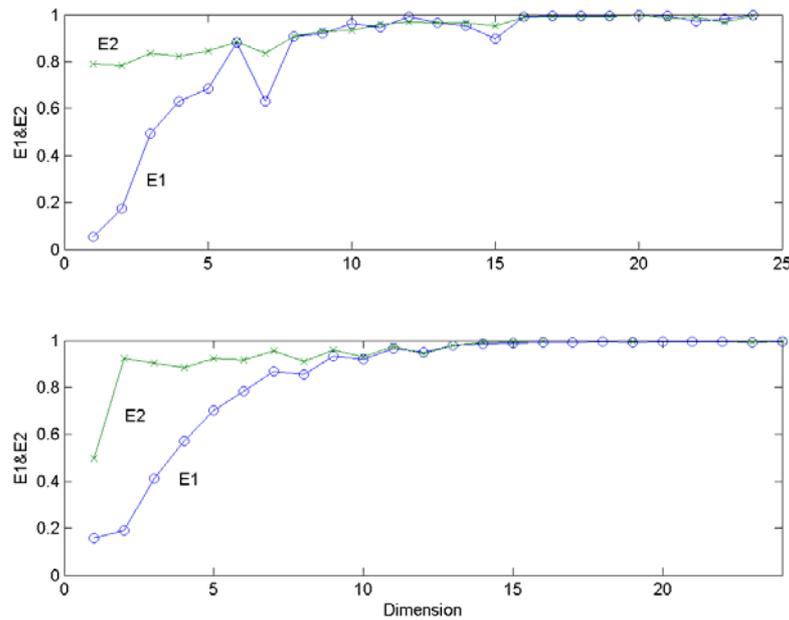


Figure 15. Embedding dimension calculation for the Nord pool time series. NOK/MWh (top), EUR/MWh (bottom).

Table 10 summarises the results obtained analysing Nord Pool time series. The time delay has been obtained using the first minimum of the AMI, Eq. (23). The embedding dimension has been computed using the methods of FNN (Kennel *et al.*, 1992) and the E1&E2 method (Cao, 1997). The results of this last method can be seen by looking at the value of $E2$ (fig 15), furthermore it can be also observed that the time series analysed does not behave as stochastic signals, i.e. $E2 \approx 1$ for all d . Furthermore, both time series have high dimensionality, $d_E \geq 7$. This high values are in agreement with similar analysis carried out by Cao (2002) for other economic time series, i.e. daily variations in the British Pound and Japanese Yen/US dollar.

Table 10. Time delay, τ , and embedding dimension, d_E , found for the Nord Pool data sets.

Data set	τ	d_E (FNN)	d_E (E1&E2)
NOK/MWh	15	7	10
EUR/MWh	13	6	10

4.3. Detecting non-stationarity

Broadly speaking a time series is said to be stationary if there is no systematic change in mean (no trend), in variance, and, if strictly periodic variations have been removed. Most of the probability theory of time series is concerned with stationary time series, and for this reason time series analysis often requires one to turn a non-stationary series into a stationary one so as to use this theory. However, it is also worth stressing that the nonstationary components, such as the trend, may sometimes be of more interest than the stationary residual.

We only report here a relatively simple stationarity test, called *space time separation plot (stp)*, introduced by Provenzale *et al.* (1992). The idea below is that in the presence of temporal correlations the probability that a given pair of state points in the reconstructed state space, $\{s(t_i), s(t_i-\Delta t), s(t_i-2\Delta t), \dots\}$, has a distance smaller than r , i.e. $\|s_i - s_j\| < r$, does not depend only on the position of the state but also on the time that has elapsed between them. This dependence can be detected by plotting the number of neighbour points as a function of two variables, the time separation and the spatial distance. In principle, one can create for each time separation an accumulated histogram of spatial distances. In the case of power-law noises the only points with small spatial separation are dynamically near neighbours, i.e. the series is non-recurrent in phase space. In this case the contour curves do not saturate. In the case of stationarity, we will find saturation in the plot.

Figures 16 show the results of the test to the analysed time series. In those graphics the separation time is represented in the horizontal axis whereas the base 2 logarithm of the separation in space is represented in the vertical axis. For small Δt points are always near neighbours in space, as their time separation increases so does their separation in space, in principle (Provenzale *et al.* 1992). Technically we have to create, for each time separation Δt an accumulate histogram of spatial distance ε . We have used the program *stp* of Tisean (Kantz and Schreiber, 1997) which returns level lines for 10%, 20%, ... of the pairs with a given temporal separation Δt .

As can be observe the Nord Pool time series saturate, Fig. 16a, 16c but the high frequency exchange rates do not (Strozzi *et al.*, 2002), fig. 16b and 16d, which gives the indication that Nord Pool time series are more stationary than other financial high frequency series. The non saturation, a part from the non-stationarity, is an indication that the data we are analysing has significant power in the low frequency, such as $1/f$ noise or Brownian motion. In this case, all points in the data set are temporally correlated and there is no way of determining an attractor dimension from the sample. A similar situation arises if the data set is too short. Then there are no pair left after removing temporally correlated pairs. If we regard the problem from a different point of view, correlation times of the order of the length of the sample (nonsaturating curves) mean that the data does not sample the observed phenomenon sufficiently (Kantz and Schreiber, 1997).

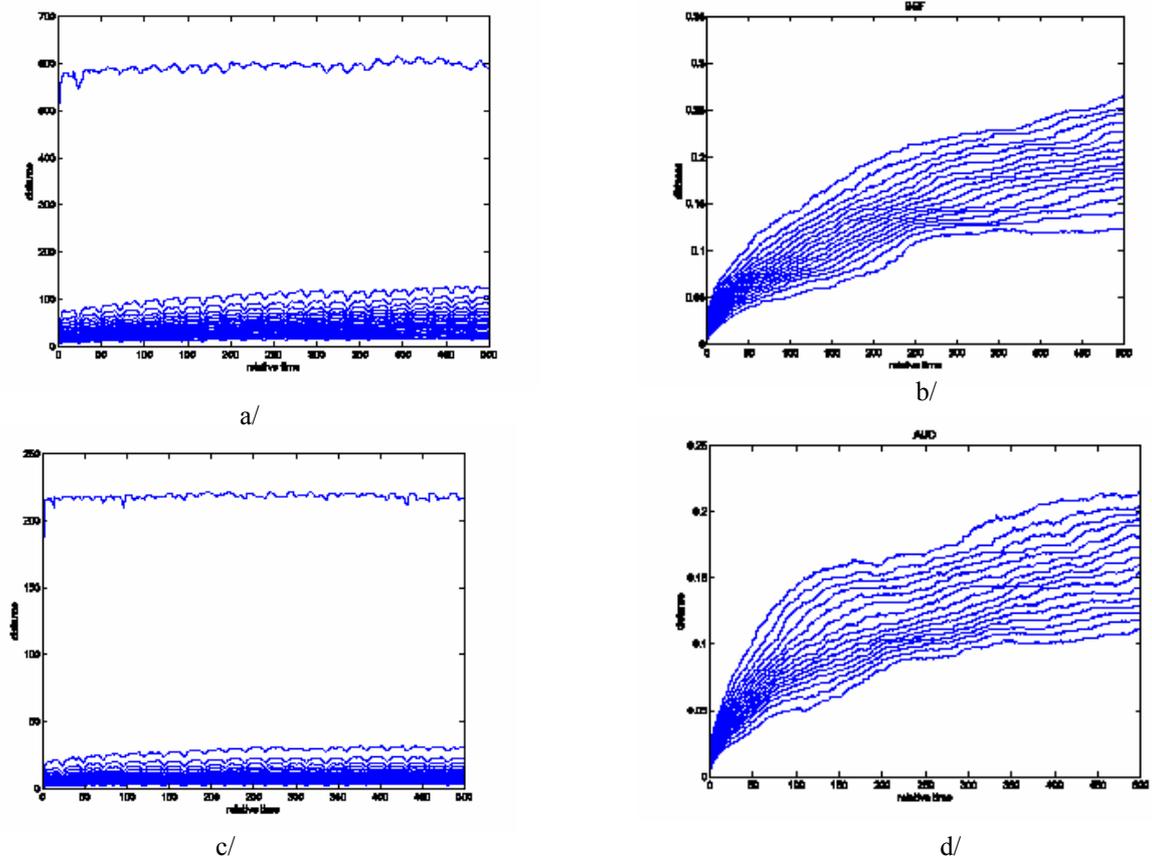


Figure 16. a/ Space-time separation plot (stp) of the Nord Pool spot prices (NOK/MWh); b/ Space-time separation plot of Australian-US dollar foreign exchange time series; c/ Space-time separation plot of the Nord Pool spot prices (EUR/MWh); d/ Space-time separation plot Belgium Franc-US dollar foreign exchange time series .

In Figure 17 we have plotted the space-time separation plot for several of the surrogates time series. As it is possible to observe, in the case of linear surrogates, the results are very similar to the ones obtained for the real time series. In addition, the space-time separation plots finds that the series are stationary.

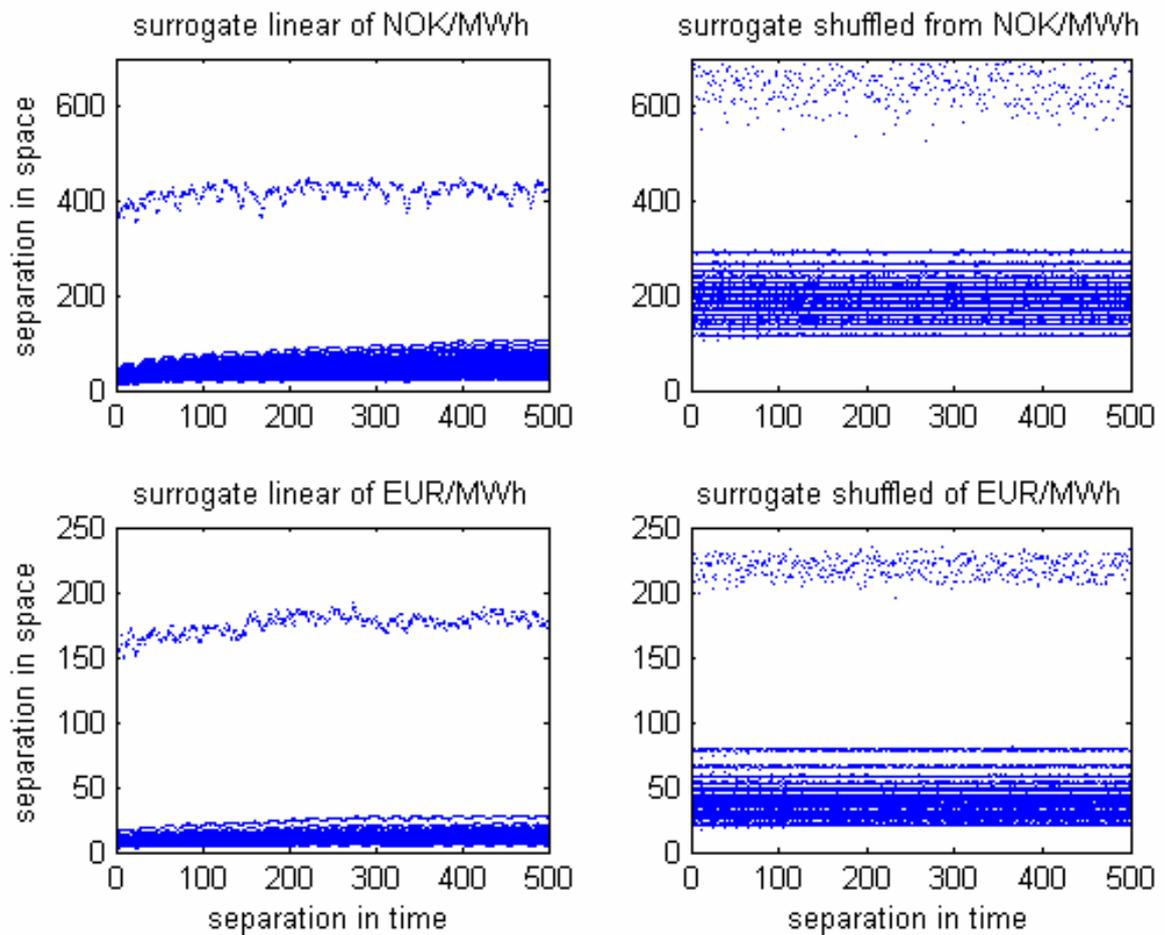


Figure 17. Space-time separation plot (stp) of the surrogate time series.

4.4. Testing for non-linearity

The former tests using surrogate data sets concerning the Hurst exponent, power spectrum and the stable distribution give an idea of the characteristics of the original series when compared with their surrogates. However, we have not tested the original time series for the existence of determinism. For this we need some parameter that is related with low dimensional determinism in the series. In order to test the null hypothesis that the series is a linear Gaussian random series with a 95% significance level, we have used the surrogate data sets for each Nord pool spot prices time series and as parameter, we have considered the error in the nonlinear one step ahead prediction (Farmer and Sidorowich, 1987). For both Nord Pool time series, the null hypothesis can be rejected since the prediction error is found to be smaller in the original time series than that of the surrogate data sets. These results are in agreement with the previous findings of the space time separation plot in which one can see that the curves saturate which means that the system is in principle not completely stochastic. However, we have also carried out another test based on time reversal symmetry statistic and in this case the null hypothesis, i.e. that a linear Gaussian random processes, cannot be rejected since the time asymmetry of the data was found to be not significantly different from that of the *surrogates*. These inconclusive results are typical of financial time series (Strozzi *et al.*, 2002).

4.5. Recurrence quantification analysis (RQA)

The actual methods developed in non-linear time series analysis assume that the data series under analysis have reach their attractors and that there are not in a transient phase, that they are autonomous and that their lengths are much longer than the characteristic time of the system in question. In the case of Nord Pool spot prices time series these assumptions are not clearly confirmed by the preliminary analysis and it may be useful to have another procedure to analyse these data.

Eckmann *et al.* (1987) introduced a new graphical tool, which they called a recurrence plot (RP). The recurrence plot is based on the computation of the distance matrix between the reconstructed points in the phase space, i.e. $\mathbf{s}_i = \{s(t), s(t-\tau), s(t-2\tau), \dots, s(t+(d_E-1)\tau)\}$,

$$d_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\| \quad (29)$$

This produces an array of distances in a $N \times N$ square matrix, \mathbf{D} , being N the number of points under study. Once this distance matrix is calculated, in the original paper of Eckmann *et al.* (1987), it was displayed by darkening the pixel located at specific (i,j) coordinates which corresponds to a distance value between i and j lower than a predetermined cutoff, i.e. a ball of radius ε centered at \mathbf{s}_i . Requiring $\varepsilon_i = \varepsilon_j$, the plot is symmetric and with a darkened main diagonal correspondent to the identity line. The darkened points individuate the recurrences of the dynamical systems and the recurrent plot provides insight into periodic structures and clustering properties that are not apparent in the original time series.

4.5.1. Selection of the threshold or cutoff value ε

A crucial parameter of a recurrence plot is the threshold ε . If ε is chosen too small, there may be almost no recurrence points and we will not be able to learn about the recurrence structure of the underlying system. On the other hand, if ε is chosen too large, almost every point is a neighbour of every other point. A too large ε includes also points into the neighbourhood which are simple consecutive points on the trajectory. Hence, we have to find a compromise for the ε value. Moreover, the influence of noise can bring us to choose a larger threshold, because noise would distort any existing structure in the RP. At higher threshold, this structure may be preserved. Several “rules of thumb” for the choice of the threshold ε are present in the literature between them (Marwan *et al.*, 2007):

a/ it should not exceed 10% of the mean or the maximum phase space diameter (Koebbe and Mayer-Kress, 1992; Zbilut and Webber, 1992)

b/ it should be such that the recurrence point density in RP is approximately 1% (Zbilut *et al.*, 2002)

c/ in order to avoid problem related to noise, ε has to be chosen such that it is five time larger than the standard deviation of the observational noise, i.e. $\varepsilon > 5\sigma$ (Thiel *et al.*, 2002)

Nevertheless, the choice of ε depends strongly on the considered system under study.

In Fig. 20 we have plotted the RP for both Nord Pool time series. We choose the 10% of the maximum phase space diameter as cutoff value. Several regime shifts are evident in both time series. A regime shift can be identified by squares structures of points separated by empty spaces (Zaldívar *et al.* 2007). However, in spite

of the differences, it is not evident how to connect the RPs with important facts in the dynamic of the underlying process. For doing this we need recurrence quantification parameters provided by RQA.

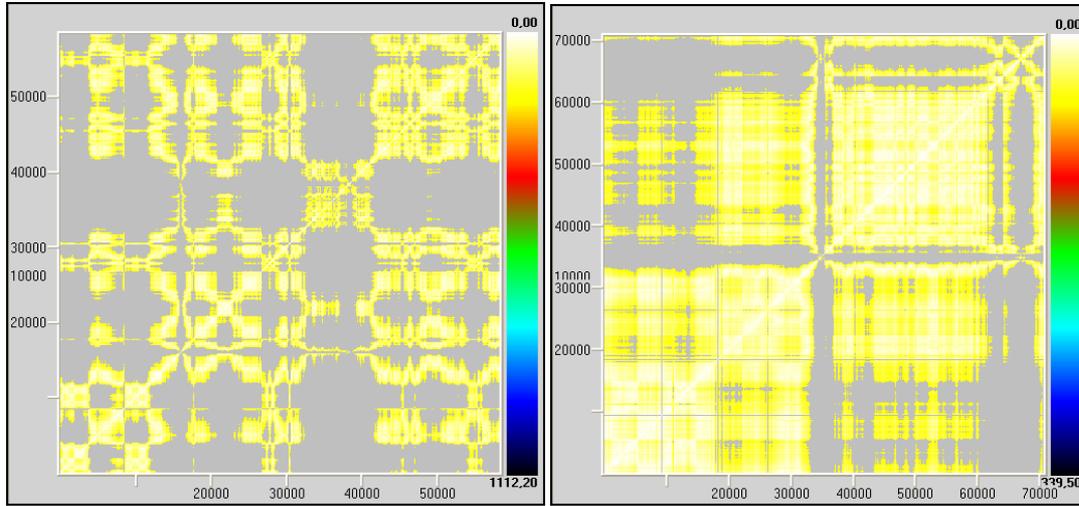
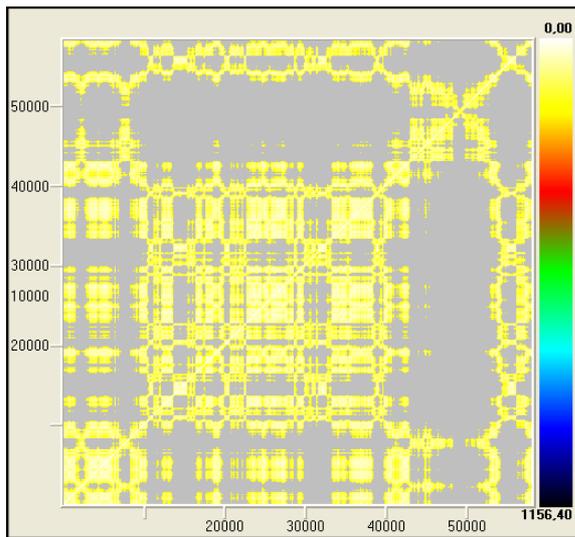
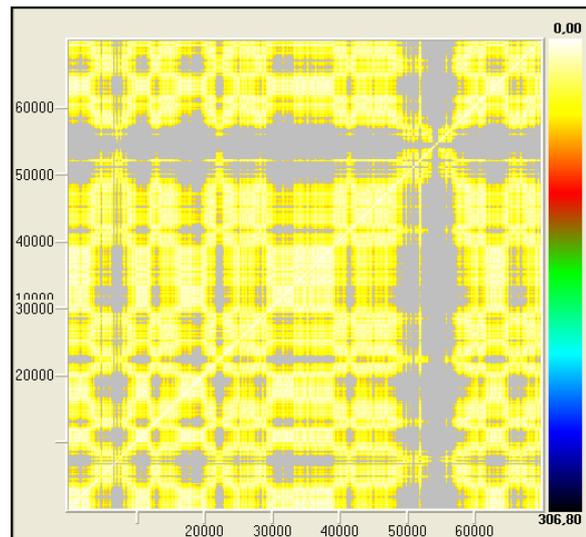


Figure 20 a). RP for NOK/MWh $\tau=15$, $d_E=10$, $\epsilon=40$ (left) and for EUR/KMh $\tau=13$, $d_E=10$, $\epsilon=10$ (right)



a/



b/

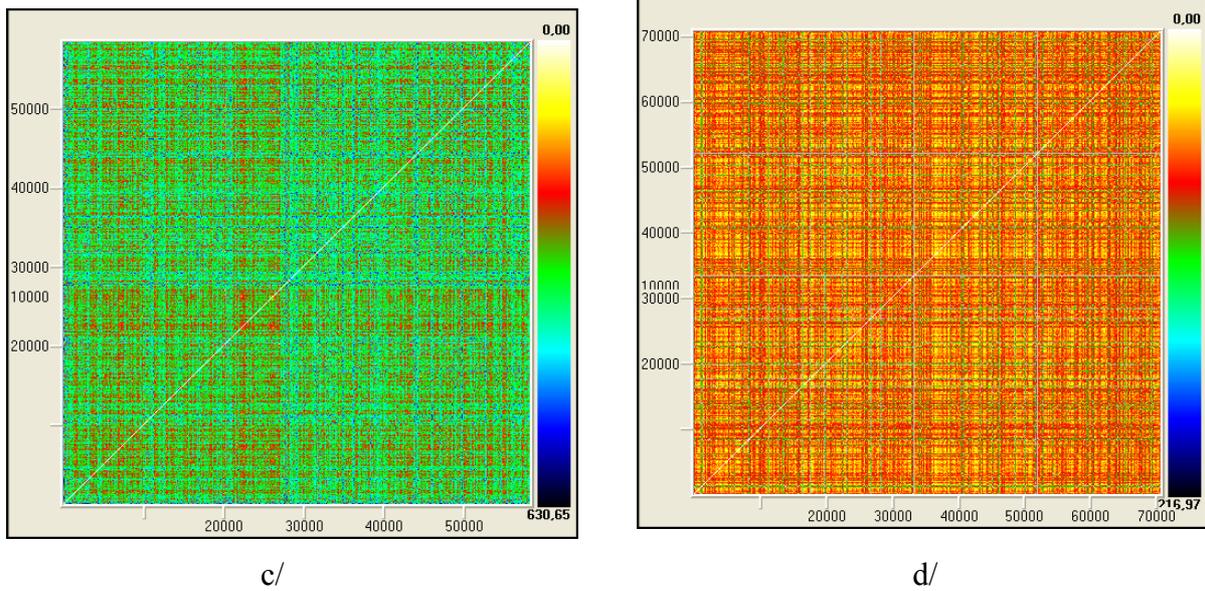


Figure 21. RPs of linear Gaussian surrogates: a/ NOK/MWh $\tau=15$, $d_E=10$, $\varepsilon=40$; b/ EUR/KMh $\tau=13$, $d_E=10$, $\varepsilon=10$. RPs of shuffled surrogates: c/ NOK/MWh $\tau=15$, $d_E=10$, $\varepsilon=40$; d/ EUR/KMh $\tau=13$, $d_E=10$, $\varepsilon=10$.

In fig 21 we have plotted the RPs for Gaussian linearly correlated and shuffled surrogate time series, respectively. Looking to fig. 20a and 20b, it can be observed that RPs of linear surrogate are qualitatively similar in the structure as those of the real time series, whereas RPs of shuffled data have no particular structures, see fig. 21c and 21d. RPs are then a tool for detecting correlations in the dynamics, but the question on how quantifying these RPs arises. This is necessary in order to distinguish, for example figures such as 20 and 21.

4.5.2. Quantification of the Recurrence Plots

Zbilut and Webber (1992) developed a methodology called Recurrence Quantification Analysis (RQA) with the aim of quantifying RP's structures. As a result, they defined several measures of complexity to quantify the small scale structures in RP. These measures are based on the recurrence point density and the diagonal and vertical line structures of the RP. A computation of these measures in small windows (sub-matrices) of the RP moving along the main diagonal yields the time dependent behaviour of these variables (Weber and Zbilut, 1994). Some studies based on RQA measures show that they are able to identify bifurcation points, especially chaos-order transitions (Trulla *et al.*, 1996). The vertical structures in the RP are related to intermittency and laminar states: those measures quantifying the vertical structures enable to detect chaos-chaos transitions (Marwan *et al.*, 2002). The measures to quantify complexity of RPs are the following:

a/ Measures based on recurrence density

%recurrence (RR) is the percentage of darkened pixels in recurrence plot:

$$RR(\varepsilon) = \frac{1}{N^2} \sum_{i,j=1}^N \mathbf{R}_{i,j}(\varepsilon) \quad (30)$$

where $\mathbf{R}_{i,j}(\varepsilon)$ is one if the state of the system at time i and the one at time j have a distance less than ε and zero otherwise.

It is a measure of the density of recurrence points in RP. Note that it corresponds to the definition of the correlation integral, Eq. (24), except that the points of the main diagonal usually are not included.

b/ Measures based on diagonal lines

Let $P(\varepsilon, l)$ be the histogram of diagonal lines of length l . If we assume we have obtained the right value of ε then we can consider $P(\varepsilon, l) = P(l)$. Processes with uncorrelated or weakly correlated behaviour cause none or very short diagonals, whereas deterministic processes cause longer diagonals. It is called *%determinism* (DET) the ratio of recurrence points that form diagonal structures (of at least length l_{min}) to all recurrence points

$$DET = \frac{\sum_{l=l_{min}}^N lP(l)}{\sum_{l=1}^N lP(l)} \quad (31)$$

%determinism (DET) is then the percentage of recurrent points forming diagonal line structures. If $l_{min} = 1$ the determinism is one. For the choice of l_{min} we have to take into account that the histogram $P(l)$ can become sparse if l_{min} is too large, and, thus, the reliability of DET decreases.

Another RQA measure considers the length L_{max} of the longest diagonal line found in the RP, or its inverse, the divergence (DIV)

$$L_{max} = \max\left\{l_i\right\}_{i=1}^{N_l}, \text{ respectively } DIV = \frac{1}{L_{max}} \quad (32)$$

where $N_l = \sum_{l \geq l_{min}} P(l)$ is the total number of diagonal lines.

These measures are related to the exponential divergence of the phase space trajectory. The faster the trajectory segments diverge, the shorter are the diagonal lines and the higher is DIV.

The measure *entropy* (ENTR) refers to the Shannon entropy of the probability $p(l) = P(l) / N_l$ to find a diagonal line of length l in RP.

$$ENTR = - \sum_{l=l_{min}}^N p(l) \ln p(l) \quad (33)$$

ENTR reflects the complexity of the RP in respect of the diagonal lines. For uncorrelated noise the value of ENTR is rather small, indicating its low complexity.

Trend is a measure of the paling recurrence points away from the central diagonal. It is a linear regression coefficient over recurrence point density of the diagonals parallel to main diagonal as a function of the time distance between these diagonals and the main diagonal. It provides information about non-stationarity in the process, especially if a drift is present in the trajectory. Trend will depend strongly on the size of the window and may yield ambiguous results for different window sizes.

c/ Measures based on vertical lines

We can find vertical lines in presence of laminar states in intermittence regimes. Let the total number of vertical lines of length v in RP is given by the histogram $P(v)$ and, analogous to the definition of the determinism, the ratio between the recurrence points forming the vertical structures and the entire set of recurrence points can be computed:

$$LAM = \frac{\sum_{v=v_{\min}}^N vP(v)}{\sum_{v=1}^N vP(v)} \quad (34)$$

This it is called *laminarity*. The computation of LAM is realised for those v that exceed a minimal length v_{\min} . LAM represents the occurrence of laminar states in the system without describing the length of these laminar phases. LAM will decrease if the RP consists of more single recurrence points than vertical structures.

The average length of vertical structures is given by

$$TT = \frac{\sum_{v=v_{\min}}^N vP(v)}{\sum_{v=v_{\min}}^N P(v)} \quad (35)$$

and is called *Trapping Time*. TT estimates the mean time that the system will abide at a specific state or how long the state will be trapped.

In contrast to the RQA measures based on diagonal lines, these measures are able to find chaos-chaos transitions. Since periodic dynamics the measures quantifying vertical structures are zero, chaos-order transition can be identified (Marwan et al., 2002).

For a recent overview of the quantifying techniques and their applications, the reader is referred to Marwan et al. (2007).

4.5.3. Analysing the complete time series

In order to check if RQA measures are able to distinguish between real data and their surrogates (linear Gaussian processes) we calculated all of them for both. The results are summarized in Tables 11-12 for NOK/MWh and EUR/MWh time series, respectively.

Table 11. RQA measures for NOK/MWk original time series ant its surrogates linear correlated.

Data set	%recur	%deter	maxline	entropy	trend	% laminar	TrapTime
Bpr	16.095	67.13	3545	8.593	-8.687	69.994	308.044
Surr001	8.150	6.129	4808	6.740	2.306	1.796	123.511
Surr002	1.926	4.521	1355	4.913	-0.142	0.000	-1
Surr003	2.807	8.026	4808	6.028	-1.616	0.000	-1
Surr004	30.218	36.309	4808	7.994	-3.360	35.521	214.805
Surr005	1.735	13.216	1844	6.117	-0.983	0.055	110
Surr006	1.007	32.018	1178	6.287	-0.752	16.980	166.134
Surr007	4.785	13.279	2674	6.895	-0.802	7.533	153.511
Surr008	14.122	17.880	4350	7.357	-4.479	9.293	154.815
Surr009	5.934	13.528	3130	7.195	-2.458	6.301	159.498
Surr010	1.193	5.900	1064	4.696	-0.677	0.347	119.500
Surr011	4.860	51.638	4808	7.918	-1.542	52.444	266.162
Surr012	31.899	52.675	4808	8.415	12.407	54.522	218.168
Surr013	4.795	9.417	4808	6.880	0.524	0.724	143.393
Surr014	5.725	9.169	4154	6.783	-2.980	1.774	144.963
Surr015	4.972	6.340	2370	6.606	-2.341	1.720	114.988
Surr016	18.050	23.404	4808	7.678	-4.183	12.845	161.470
Surr017	10.846	43.188	4614	8.799	-7.222	38.298	338.899
Surr018	4.956	8.523	4808	6.596	-2.654	3.484	141.553
Surr019	6.323	4.462	4808	6.184	-1.989	0.375	114.208

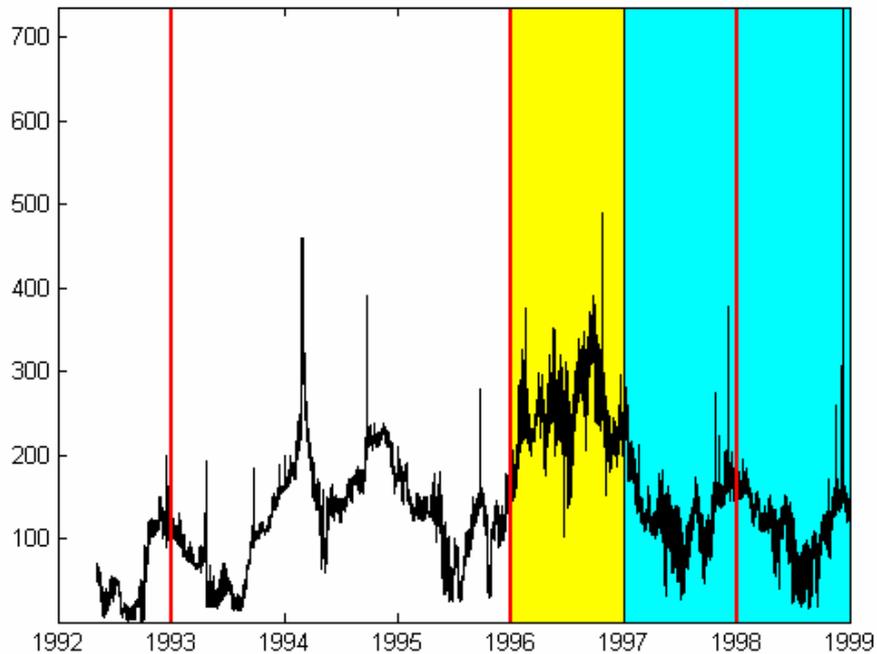
Trapping Time = -1 means that no vertical lines were found.

Table 12. RQA measures for EUR/MWk time serie ant its surrogates linear corelated.

Data set	%recur	%deter	maxline	entropy	trend	% laminar	TrapTime
Beur	7.12	35.33	2094	7.658	-4.587	33.94	263.525
Surr001	12.524	3.665	3340	6.355	-6.259	2.539	149.367
Surr002	1.643	5.894	2238	5.270	-1.100	1.872	119.367
Surr003	3.840	1.397	2150	4.533	-0.998	0.000	-1
Surr004	4.377	1.105	1324	3.970	-0.286	0.000	-1.000
Surr005	10.677	1.825	4187	5.730	-5.483	1.527	126.613
Surr006	8.658	18.813	4826	7.538	-5.638	9.854	146.364
Surr007	0.491	3.888	690	2.807	-0.346	0.000	-1.000
Surr008	23.790	11.105	4826	7.509	-7.639	9.252	162.159
Surr009	30.269	10.831	4826	7.393	-1.830	7.108	151.053
Surr010	20.536	4.700	4826	6.845	-7.466	6.416	150.611
Surr011	2.336	3.777	1888	5.094	-1.160	1.529	134.161
Surr012	3.715	1.475	3517	4.059	-1.627	0.108	117.250
Surr013	4.994	3.736	3721	5.972	-3.343	1.886	135.457
Surr014	21.649	9.020	4826	7.162	-2.900	9.664	154.810
Surr015	20.052	8.142	2669	7.247	-4.243	4.171	146.484
Surr016	6.811	5.384	3998	6.574	-4.098	0.758	125.312
Surr017	3.161	4.113	1964	5.641	-2.076	0.715	131.650
Surr018	7.809	3.369	2429	6.204	-0.473	2.766	132.437
Surr019	12.185	1.330	4826	5.429	1.503	0.088	125.600

By looking to Tables 11-12 we can observe that *%recurrence*, *maxline*, *entropy*, *trend* or *Trapping Time* parameters cannot distinguish, with a 95% of confidence, between a linear gaussian dynamic and the dynamic behind the financial time series. Of course this does not implies that are not useful for their quantification, but only that the values of the parameters are in some case higher and in other cases smaller than those of the original time series. On the contrary, using *%determinism*, *%laminarity* we obtain values which are always smaller for surrogate data in comparison with original data sets. The fact that these two parameters are able to distinguish between the original time series and the surrogate time series points toward the explanation that the original series have more diagonal and vertical lines, and therefore their state remain near or at the same place longer in time more often than for their surrogate linear Gaussian process and that they posses a different decaying of the autocorrelation function. It could be interesting to generate surrogate data using stable distributions and then compare the values of RQA parameters.

If we apply RQA to shuffled surrogates, the RQA measures do not detect any structure giving for example *%determinism*, *%recurrence* and *%laminarity* equal to zero for all cases.



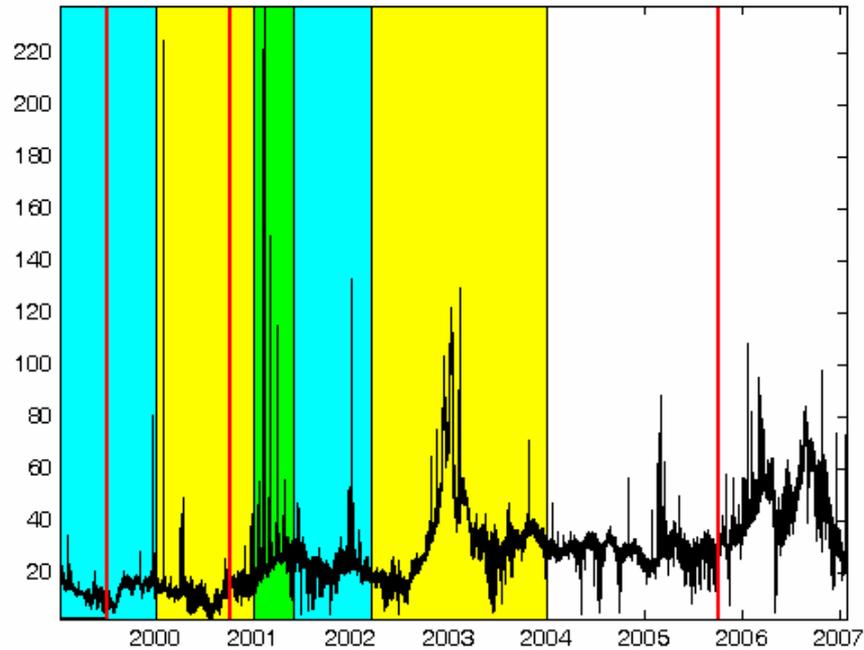


Figure 22. NOK/MWh (top) and EUR MWh (bottom) and the dates from Tables 1-3

4.5.4. RQE analysis

Let now compute RQA measures on a moving window. In this way, we obtain a time dependent profile of RQA measures. We would like to see if RQA measures are able to detect some events that are not clear from a direct inspection of the time series. For example we are interested to observe if some changes in the RQA parameters occur in correspondence of the entry of a new country in the Nord Pool (Table 1) or in correspondence with the starting of the deregulation processes (Table 2), or in correspondence with dry and wet years (Table 3). Figure 22 shows the two Nord Pool time series plotted with the dates or periods indicated in Tables 1-3, whereas in Figs. 23-24 the behaviour of RQA parameters is plotted with a moving window of one month shifted of one month for NOK/MWh and EUR/MWh, respectively.

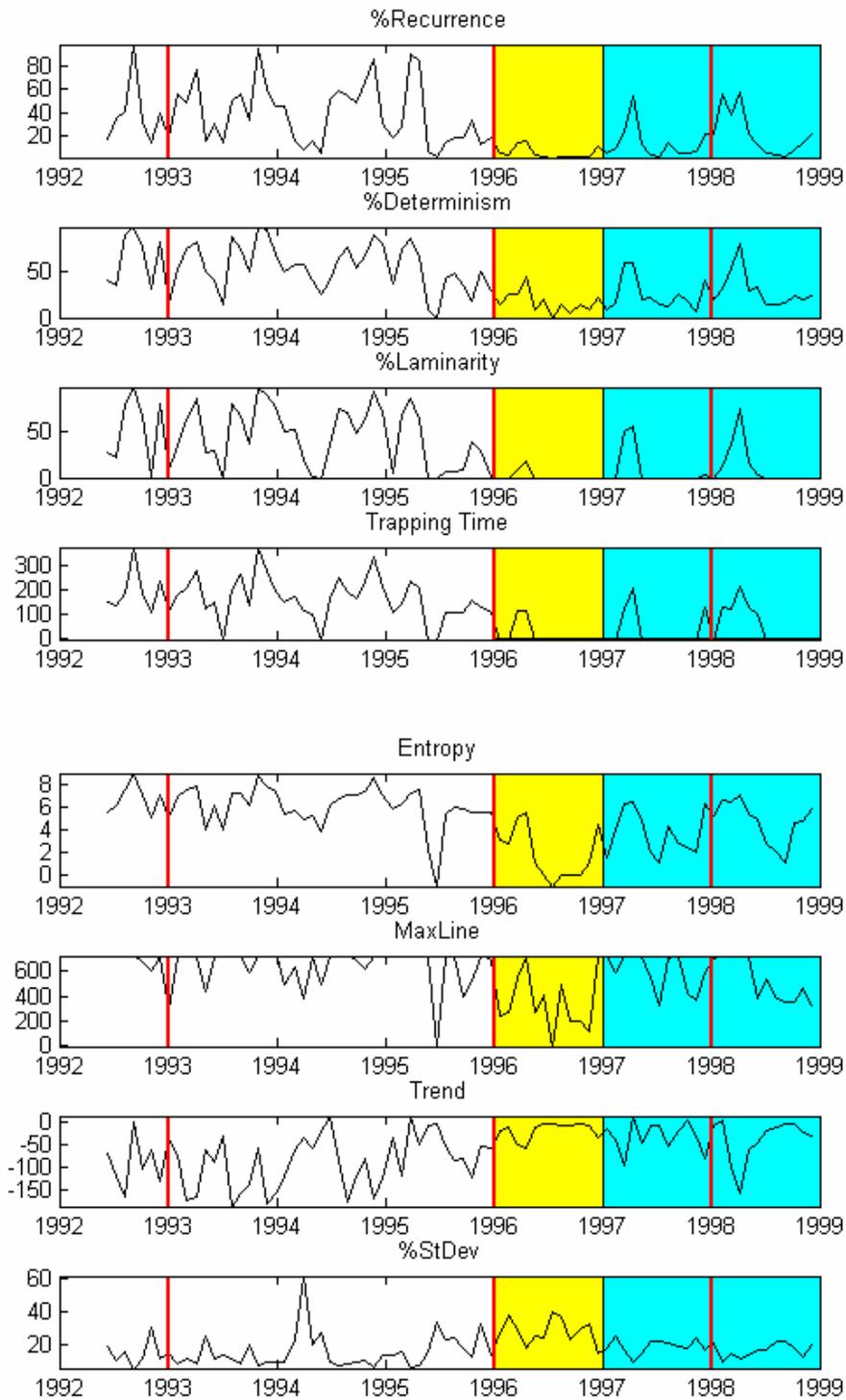


Figure 23. Nonlinear metrics of the Nord Pool spot prices time series in NOK/MWh: Values are computed from a 720 point window (one week), data are shifted 720 points. RQA parameters: $\tau=15$, $d_E=10$, distance cutoff: max. distance between points/10, line definition: 100 points (~4 days). Vertical lines correspond to the following dates: 1st January 1993, 1st January 1996, 29th December 1997 and 1st July 1999 (see historical background).

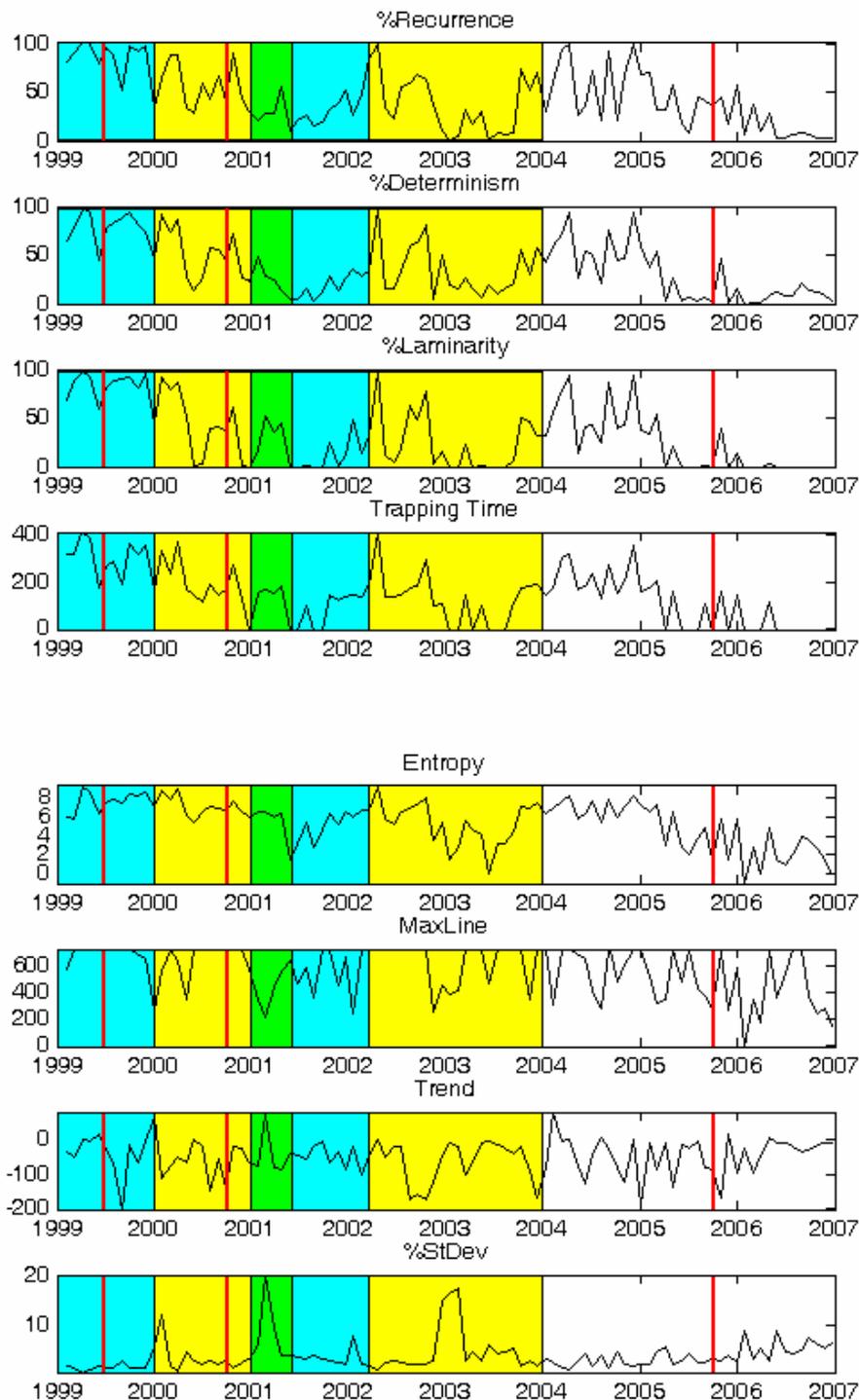


Figure 24. RQA measures of EUR/MWh: Values are computed from a 720 point window (one month), shifted of 720 points. RQA parameters: $\tau = 13$, $d_E = 10$, distance cutoff: max. distance between points/10, line definition: 100 points (~4 days). Vertical lines correspond to the following dates: 1st October 2000, 5th October 2005 (see historical background).

By looking at Figs. 23-24 we can observe a qualitative agreement between the RQA measurements: *%recurrence*, *%determinism*, *%laminarity* and *trapping time*, for both time series. Furthermore, most of the times, it is possible to observe an inflection in correspondence of the entrance of a new state in Nord Pool (red

lines in Figs. 23 and 24 and Table 1). These lines also sometimes coincide with the starting of deregulation processes in other countries (see Table 2). However, there is no clear evidence and also inflections are visible in other parts of the time series.

In addition by looking at the RQA parameters (*%recurrence*, *%determinism*, *%laminarity* and *trapping time*) we can observe that, in correspondence of dry periods (yellow periods), the parameters tend to have smaller values and/or a negative trend. This is more clear in the first time series where hydroelectric power was more important for the Nord Pool. In these dry periods, due to the high dependence from the oil, the volatility of the price increases.

It is well-known that high volatility periods are those in which it is more difficult to make forecast. Higher *%determinism* and *%laminarity* mean that the states of the system stay closer in time for longer periods forming diagonal or vertical segments in RP. Then we can assume that higher *%determinism* or *%laminarity* implies smaller volatility. To study the relationship with volatility, we have compared the profiles of these quantities with the inverse of standard deviation between 0 and 100 (see figs. 25-26). The main difference between *%determinism* and *%laminarity* is that, in the periods of high volatility, *%laminarity* reaches zero values which gives a more clear signal of volatility periods.

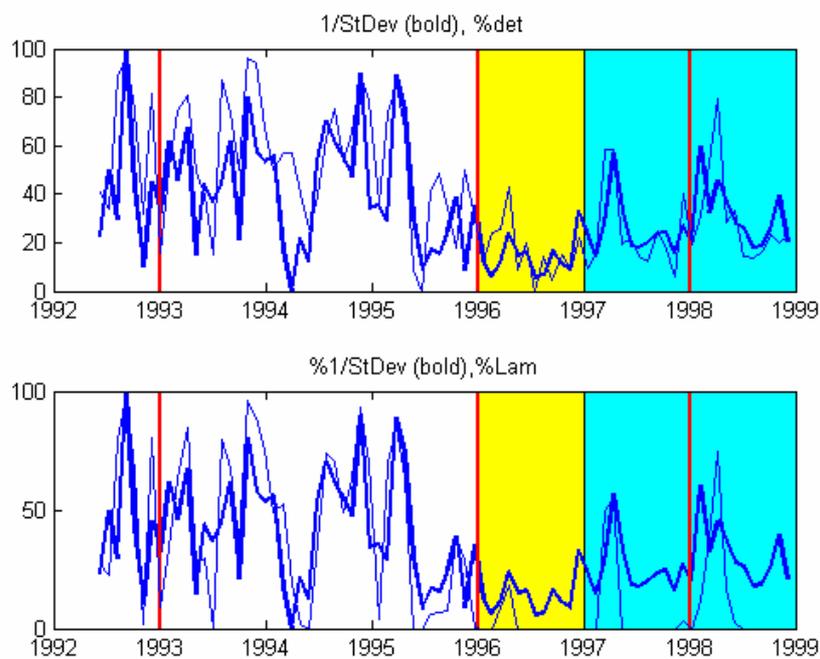


Figure 25. Inverse of standard deviation and *%determinism* (top) and *%laminarity* (bottom) for NOK/MWh

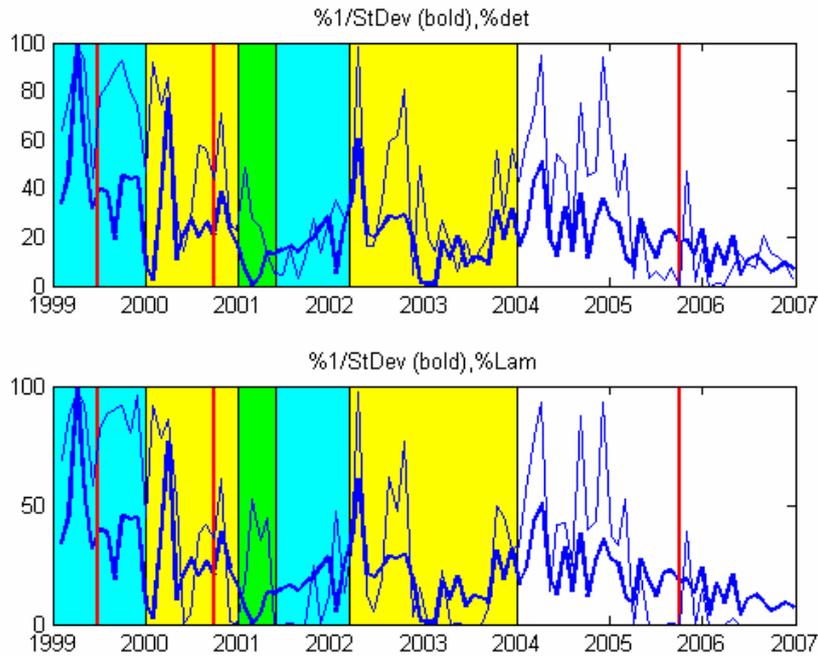


Figure 26. Inverse of standard deviation and *%determinism* (top) and *%laminarity* (bottom) for EUR/MWh

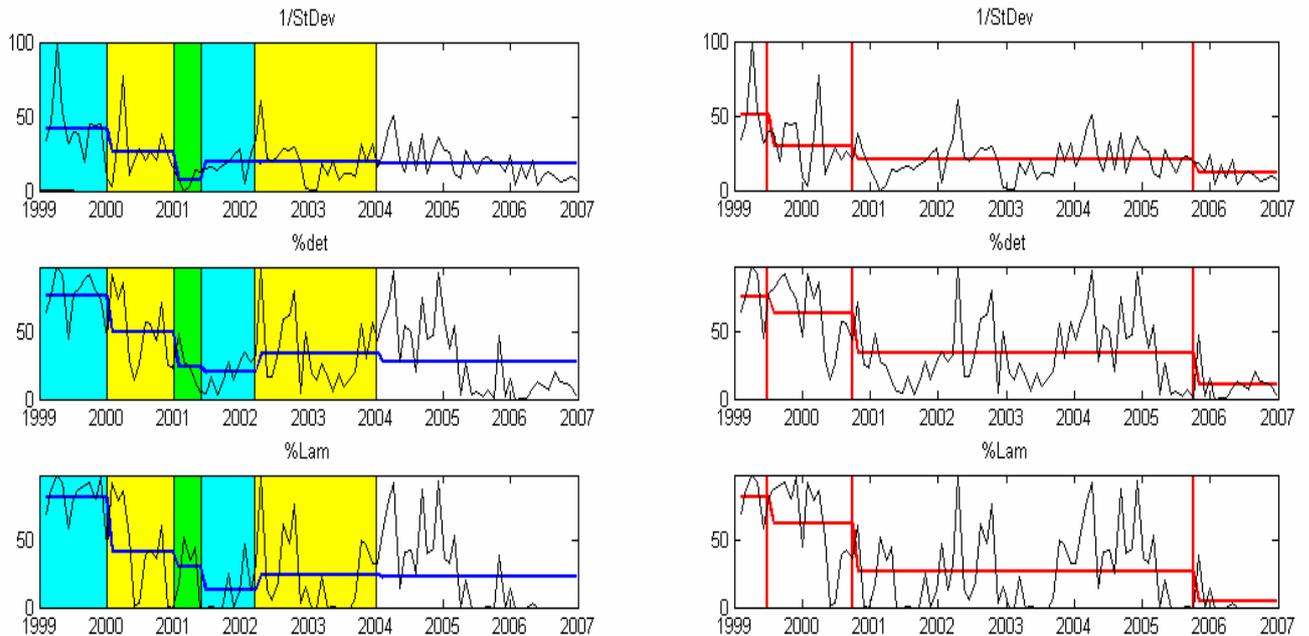


Figure 27. RQA measures of EUR/MWh: Values are computed from a 720 point window (one month), shifted of 720 points. RQA parameters: $\square=13$, $dE=10$, distance cutoff: max. distance between points/10, line definition: 100 points (~4 days). Vertical lines correspond to the following dates: 1st October 2000, 5th October 2005 (see historical background).

In order to extract more information from RQA measures, we have compared the mean values of *%determinism* and *%laminarity* with the mean values of the inverse of the standard deviation (StDev) during the periods between changes in weather conditions (for EUR fig 27, left, and fig. 28 left for NOK) and the periods between the entrance of new states in Nord Pool (fig 27 right for EUR, fig 28 right for NOK). In both

cases, it is possible to observe that using RQA measures the changes in the means are more evident (the steps higher) than using standard deviation. Then using the RQA measures it is possible to improve the detection of changes in the time series analyzed.

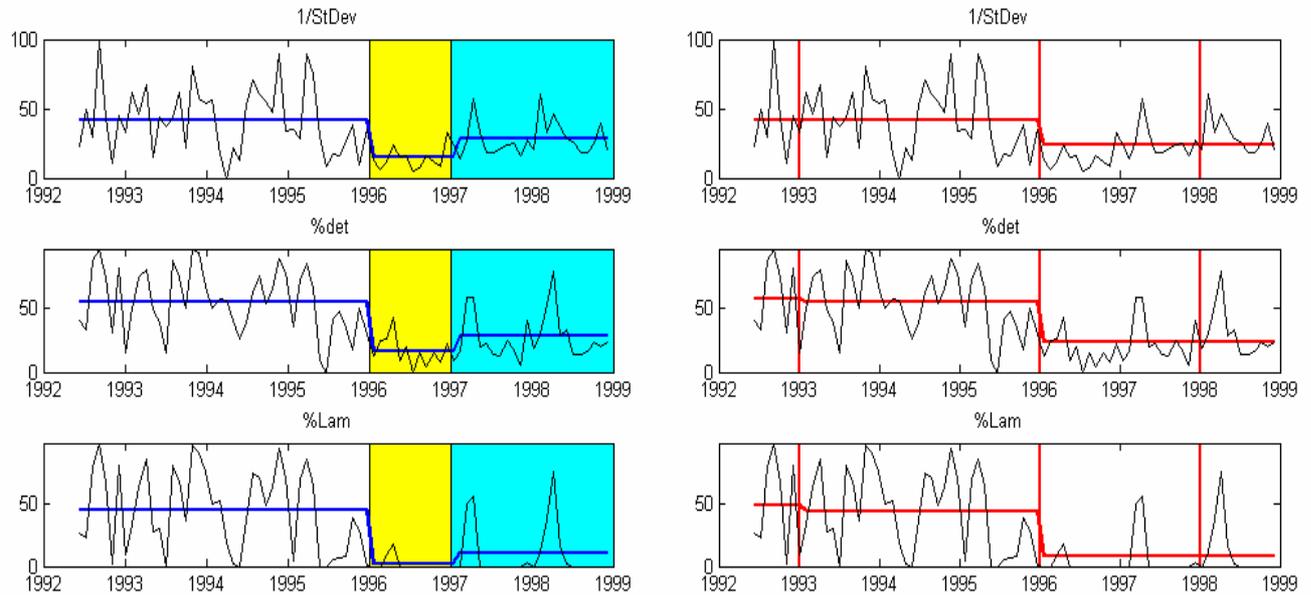


Figure 28. Nonlinear metrics of the Nord Pool spot prices time series in NOK/MWh: Values are computed from a 720 point window (one month), data are shifted 720 points. RQA parameters: $\tau=15$, $d_E=10$, distance cutoff: max. distance between points/10, line definition: 100 points (~4 days). Vertical lines correspond to the following dates: 1st January 1993, 1st January 1996, 29th December 1997 and 1st July 1999 (see historical background).

5. Conclusions

Nonlinear time series analysis has been carried out for the Nord Pool time series. Preliminary analysis confirms already published work concerning the antipersistence, $H < 0.5$, of these type of data sets. The power spectral density shows a scaling behaviour typical of financial time series. On the contrary, like in other high frequency time series such as exchange rates, the saturation in the space time separation plot shows that the time series may be considered as stationary and hence, the application of the surrogate data tests, that assumes two kinds of null hypothesis: stationary Gaussian linear process or no correlation at all, is adequate.

Stable distributions have been proposed as a model for many types of physical and economic systems because many large data sets exhibit heavy tails and skewness. It is possible to observe a clear distinction between the first period time series in NOK and the second in EUR. In the first case there is a considerable number of zeros in the first difference of the series that create some problems in fitting the parameters for a stable distribution, whereas the problem does not exist in the second case. In general terms the series seems to have long tails and be more similar to a Levy distribution than to a Gaussian one. Also in this case, linear surrogate data produce different values when fitted with stable distributions being more similar to a Gaussian ($\alpha=1.7$ instead of 1.3) and having more symmetry, with β closer to 0 than their original series that have more skew.

The application of RQA shows some critical points in the series that loosely correspond with some historical periods; however it is difficult to assign a one to one correspondence. Also in this case some RQA measures are able to distinguish between linear and shuffled surrogates time series and the original ones. We have also found a correspondence between *%determinism* and *%laminarity* with the inverse of the standard deviation, therefore, these parameters can give another method to measure volatility in time series analysis. We have compared the mean values of these three quantities calculated between the periods in which there were important changes in weather conditions or in correspondence of which there was the incorporation of new states into the Nord Pool. We have shown that *%determinism* and *%laminarity* detect these changes more clearly than standard deviation and then they provide an alternative measure of volatility.

The future developments of this work will be to find a correlation between market prices (or some related variable such as volatility) and the likelihood of blackouts. In this work, candidate parameters have been assessed.

Acknowledgements. The authors gratefully acknowledge the financial support of the MANMADE NEST Project (Contract No 043363) and Dr. H. Sivonen (NESAs) who kindly provided the data sets analysed.

References

- Abarbanel, H.D.I., *Analysis of Observed Chaotic Data*, 1996, Springer-Verlag, New-York.
- Amundsen, E.S. and Bergman, L., 2007, Integration of multiple national markets for electricity: The case of Norway and Sweden, *Energy Policy*, doi.1016/j.enpol.2006.12.014.
- Andreadis, I., 2000, Self-criticality and stochasticity of an S&P 500 index time series. *Chaos, Solitons and Fractals* **11**, 1047-1059.
- Badii, R., Broggi, G., Derighetti, B., Ravani, M., Ciliberto, S., Politi, A. and Rubio, M. A., 1988, Dimension increase in filtered signals. *Phys. Rev. Lett.* **60**, 979-982.
- Bak, P. and Chen, K., 1991, Self-organized criticality. *Scientific American* **264**, 26-33.
- Bask, M. Liu, T. and Widerberg, A., 2007. The stability of electricity prices: Estimation and inference of the Lyapunov exponents. *Physica A* **376**, 565-572.
- Breeden, J. L and N. H. Packard, 1994, A learning algorithm for optimal representation of experimental data, *Int. J. of Bifurcations and Chaos* **4**, 311- 326.
- Brock, W. A., Hsieh, D. A., LeBaron, B., 1991, *Nonlinear dynamics, chaos and instability: Statistical theory and Economic evidence*. MIT Press, Massachusetts, MA.
- Broomhead, D. S. and G. P. King, 1986, Extracting qualitative dynamics from experimental data, *Physica D* **20**, 217-236.
- Burden, R. L. and Faires, J. D. 1996. *Numerical Analysis* , 3rd Ed., PWS, Boston.
- Byström H.N. E. 2005. Extreme value theory and extremely large electricity price changes. *International Review of Economics and Finance* **14**, 41-55.
- Cannon, M. J. Percival, D. B., Caccia, D. C., Raymond, G. M. and Bassingthwaite, J. B., 1997, Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series. *Physica A* **241**, 606-626.
- Cao, L. ,1997, Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D* **110**, 43-50.
- Cao, L., 2002, Method of False Nearest Neighbors, in *Modelling and Forecasting Financial Data*, A. S. Soofi and L. Cao (Eds.), Kluwer, Boston.
- Casdagli, M., Eubank, S., Farmer, J. D., Gibson, J., 1991, State space reconstruction in the presence of noise. *Physica D* **51**, 52-98.
- Davies, M., 1994, Noise reduction schemes for chaotic time series, *Physica D* **79**, 174-192.
- Diks, C., 1999, *Nonlinear Time Series Analysis: Methods and applications*. World Scientific, Singapore.
- Eckmann, J. P., Kamphorst, S. O. and Ruelle, D., 1987, Recurrence plots of dynamical systems, *Europhys. Lett.* **4**, 973-977.
- Farmer, J.D. and Sidorowich, J. 1987. Predicting chaotic time series. *Phys. Rev. Lett.* **59**, 845-850.
- Fraser, A. and Swinney, H., 1986, Independent coordinates for strange attractors from mutual information. *Phys. Rev. A* **33**, 1134-1140.
- Friederich, R., Peinke, J. and Renner, Ch., 2000, How to quantify deterministic and random influences on the statistics of the foreign exchange market. *Phys. Rev. Lett.* **84**, 5224-5227.
- Gilmore, R., 1998, Topological analysis of chaotic dynamical systems. *Rev. Mod. Phys.* **70**, 1455-1529.
- Grassberger, P., 1983, Generalized dimension of strange attractors. *Phys. Lett A* **97**, 227-230.
- Haldrup, N., Nielsen, M. Ø., 2006. A regime switching long memory model for electricity prices. *Journal of econometrics* **135**, 349-376.

- Hegger, R., Kantz, H., Schreiber, T., 1999, Practical implementation of nonlinear time series methods: The TISEAN package. *CHAOS* **9**, 413-. The software package is publicly available at <http://www.mpipks-dresden.mpg.de/~tisean> .
- Hsieh, D. A., Chaos and nonlinear dynamics: Application to financial markets, 1991, *The Journal of Finance* **46**, 1839-1887
- Hurst, H. E., 1951, Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civ. Eng.* **116**, 770-779.
- Johnson, N. F., Jefferies, P. & Ming Hui, P. 2003. *Financial Market Complexity*, Oxford University Press.
- Kantz H., Shreiber T., 1997. *Nonlinear Time Series Analysis*, Cambridge University Press
- Kennel, M. B., R. Brown and H. D. I. Abarbanel, 1992. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Phys. Rev. A* **45**, 3403-3411.
- Koebbe, M., Mayer-Kress, G. 1992. Use of recurrence plots in the analysis of time-series data, in M. Casdagli, S. Eubank (Eds.). *Proceedings of SFI Studies in the Science of Complexity*, vol. XXI, Redwood City, Addison-Wesley, Reading, MA, pp. 361-378.
- Kostelich, E. J. and Schreiber, T., 1993, Noise reduction in chaotic time-series data: A survey of common methods. *Phys. Rev. E* **48**, 1752-1763.
- Kristiansen, T., 2006, A preliminary assessment of the market coupling arrangement on the Kontek cable, *Energy Policy* (in press)
- Kristiansen, T., 2007. Pricing of monthly contracts in the Nord Pool market. *Energy Policy* **35**, 307-316.
- Lorenz, H. W., 1993, *Nonlinear dynamical economics and chaotic motion*. Springer, New York.
- Malkiel, B., 1990, *A random walk down Wall Street*, Norton, New York.
- Mandelbrot, B. B., 1998, *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. Springer, New York.
- Mandelbrot, B. B., *The Fractal Geometry of Nature*, 1983, W. H. Freeman. New York.
- Mantegna, R. N. & Stanley, H. E., 2000. *An introduction to Econophysics*. Cambridge University Press.
- Mantegna, R. N. and Stanley, H.E., 1995, Scaling behaviour in the dynamics of an economic index. *Nature* **376**, 46-49.
- Mantegna, R. N. and Stanley, H.E., 1996, Turbulence and financial markets. *Nature* **376**, 46-49.
- Marwan, N., Romano, M. C., Thiel, M. and Kurths, J., 2007. Recurrence plots for the analysis of complex systems. *Physics Reports* **438**, 237-329.
- Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., Kurths, J. 2002. Recurrence plot based measures of complexity and its application to heart rate variability data. *Phys. Rev. E* **66**(2), 026702.
- Mees, A. I., Rapp, P. E. and Jennings L. S., 1987, Singular value decomposition and embedding dimension. *Phys. Rev. A* **36**, 340.
- Mork, E. 2001. Emergence of financial markets for electricity: a European perspective. *Energy Policy* **29**, 7-15.
- Nolan, J.P., 1997. Numerical computation of stable densities and distribution functions. *Commun. Stat.: Stochastic models* **13**, 759-774.
- Nolan, J.P., 1999. Fitting data and assessing goodness of fit with stable distributions. In *Proceedings of the Conference on Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, American University, Washington DC, June 3-5.
- Osborne, M. F.M., 1959, Brownian motion in the stock market. *Oper. Res.* **7**, 145-173.
- Packard, N., Crutchfield, J., Farmer, D. and Shaw, R., 1981, Geometry from a time series. *Phys. Rev. Lett.* **45**, 712-715.
- Papaioannou, G. and Karytinou, A., 1995, Nonlinear time series analysis of the stock exchange: The case of an emerging market. *Int. J. of Bifurcations and Chaos* **5**, 1557-1584.

- Perelló, J., Montero, M., Palatella, L., Simonsen, I. and Masoliver, J., 2007. Entropy of the Nordic electricity market: anomalous scaling, spikes, and mean-reversion. *J. Stat. Physics* (in press).
- Peters, E. E., 1996. *Chaos and Order in the Capital Markets: a New View of Cycles, Prices and Volatility*, 2nd Edition, Wiley, New York.
- Provenzale, A., Smith, L. A., Vio, R. and Murante, G., 1992, Distinguishing between low-dimensional dynamics and randomness in measured time series. *Physica D* **58**, 31-49.
- Ruelle, D., 1990, Deterministic chaos: the science and the fiction. *Proc. R. Soc. Lond. A* **427**, 241-248.
- Scheinkman, J. and LeBaron, B., 1989, Nonlinear dynamics and stock returns. *J. Business* **62**, 311-318.
- Schreiber, T. and Schmitz, A., 2000, Surrogate time series, *Physica D* **142**, 346-382.
- Schreiber, T. and Schmitz, A., 1996, Improved surrogate data for nonlinearity tests. *Phys. Rev. Lett.* **77**, 35-38.
- Schreiber, T., 1998, Interdisciplinary application of nonlinear time series methods. *Physics Reports*.
- Shlesinger, M. F., Zaslavsky, G. M., and Klafter, J., 1993, Strange kinetics. *Nature* **363**, 31-37.
- Simonsen, I., 2003. Measuring anti-correlations in the Nordic electricity spot market by wavelets. *Physica A* **322**, 597-606.
- Soofi, A. S. and Cao, L., 2002. *Modelling and forecasting financial data: Techniques of Nonlinear Dynamics*. Kluwer Academic Publishers, Norwell.
- Stark, J., Broomhead, D.S., Davies, M.E. and Huke, J. 1997, Takens embedding theorems for forced and stochastic systems. *Nonlinear Analysis* **30**, 5303-5314.
- Strozzi, F. and Zaldívar, J. M., 2002, Embedding theory: Introduction and applications to time series analysis, in *Modelling and forecasting financial data: Techniques of Nonlinear Dynamics*. A. Soofi and L. Cao (Eds.), Kluwer Academic Publishers, Boston.
- Strozzi, F., Zaldívar, J. M., Zbilut, J., P., 2007, Recurrence quantification analysis and state space divergence reconstruction for financial time series analysis. *Physica A* **376**, 487-499
- Strozzi, F., Zaldívar, J. M., & Zbilut, J. P. 2002. Application of nonlinear time series analysis techniques to high frequency currency exchange data, *Physica A* **312**, 520-538.
- Takens, F., 1981, in *Dynamical Systems and Turbulence*, Warwick 1980, vol. 898 of Lecture Notes in Mathematics, edited by A. Rand and L.S Young, Springer, Berlin, pp. 366-381.
- Takens, F., 1996, The effect of small noise on systems with chaotic dynamics. In *Stochastic and Spatial Structures of Dynamical Systems*, S. J. van Strien and S. M. Verduyn Lunel, Verhandelingen KNAW, Afd. Natuurkunde, vol. 45, pp. 3-15. North-Holland, Amsterdam.
- Theiler, J., 1991, Some comments on the correlation dimension of $1/f^\alpha$ noise. *Phys. Lett A* **155**, 480-493.
- Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Farmer, J. D., 1992, Testing for nonlinearity in time series: the method of surrogate data. *Physica D* **58**, 77-.
- Thiel, M., Romano, M.C., Kurths, J., Meucci, R., Allaria, E. and Arecchi, F.T. 2002. Influence of observational noise on the recurrence quantification analysis. *Physica D* **171**, 138-152.
- Tong, H., *Nonlinear Time Series: a Dynamical System Approach*. 1990, Oxford University Press. Oxford.
- Trulla, L.L, A. Giuliani, J.P. Zbilut, and C.L. Webber, Jr. 1996. Recurrence quantification analysis of the logistic equation with transients. *Phys. Lett. A* **223**, 255-26.
- Vehviläinen I. and Pyykkönen, T. 2005. Stochastic factor model for electricity spot price-the case of the Nordic market. *Energy Economics* **27**, 351-357.
- Webber Jr. C. L. and Zbilut, J. P., 1994. Dynamical assessment of physiological systems and states using recurrence plot strategies. *J. Appl. Physiol.* **76**, 965-973.
- Weron, R. and Przybyłowicz, B., 2000. Hurst analysis of electricity price dynamics. *Physica A* **283**, 462-468.

- Weron, R., Bierbrauer, M. and Truck, S. 2004. Modelling electricity prices: Jump diffusion and regime switching. *Physica A* **336**, 39-48.
- Whitney, H., 1936. Differentiable manifolds. *Ann. Math.* **37**, 645-680.
- Wolf A., J. B. Swift, H. R. Swinne, J. A. Vastan, 1985. Determining Lyapunov exponents from a time series, *Physica D* **16**, 285-317.
- Zaldívar, J.M., Strozzi, F., Dueri, S., Marinov, D. and Zbilut, J. P. 2007. Recurrence quantification analysis as a method for the detection of environmental thresholds. *Ecol. Model.* (in press)
- Zbilut, J. P. and Webber Jr. C. L., 1992. Embeddings and delays as derived from quantification of recurrence plots. *Phys. Lett. A* **171**, 199-203
- Zbilut, J. P., Zaldívar, J. M., Strozzi, F., 2002. Recurrence quantification based Liapunov exponents for monitoring divergence in experimental data. *Phys. Lett. A* **297**, 173-181.

ANNEX III. *Measuring volatility in the Nordic spot electricity market using Recurrence Quantification Analysis*. Eur. Phys. J. Special Topics 164 (2008), 105-115.

Measuring volatility in the Nordic spot electricity market using Recurrence Quantification Analysis

F. Strozzi^{1,a}, E. Gutiérrez², C. Noè¹, T. Rossi¹, M. Serati¹, and J. M. Zaldívar²

¹ Università Carlo Cattaneo, Castellanza, Italy

² Joint Research Centre, European Commission, Ispra, Italy

Abstract. In this work, we have applied Recurrence Quantification Analysis (RQA) to data sets taken from the Nordic spot electricity market. Our main interest was in trying to correlate their volatility with variables obtained from the quantification of recurrence plots (RP). For this reason we have based our analysis on known historical events: the evolution of the Nord Pool market and climatic factors, i.e. dry and wet years, and we have compared several dispersion measures with RQA measures in correspondence of these events. The analysis suggests that two RQA measures: DET and LAM can be used as a measure of the inverse of the volatility. The main advantage of using DET and LAM is that these measures provide also information about the underlying dynamics. This fact is shown using shuffled and linear Gaussian surrogates of the real time series.

1 Introduction

The complex behavior of financial time series, which linear stochastic models are not able to account for [1], has been attributed to the fact that financial market time series are nonlinear stochastic, chaotic or a combination of both. Even though there is no conclusive evidence of low dimension deterministic structure, in the last few years nonlinear time series analysis has expanded rapidly in the fields of economics and finance [2]. This is also due to the fact that economic and financial time series seem to provide a promising area for the development, testing and application of nonlinear techniques, and the fact that high frequency financial time series are readily available [3]. In addition to financial market time series, energy market spot prices have also been analyzed with several nonlinear techniques.

In [4],[5] the authors established, using Hurst R/S analysis, that the electricity prices are anti-persistent with a Hurst exponent lower than 0.5, $H \simeq 0.41$. Also the Lyapunov exponents, a quantity that characterizes the mean rate of separation of infinitesimally close trajectories in a dynamical system, have been estimated in a recent study [6].

Volatility in financial markets is a dispersion measure that quantifies the degree of uncertainty about the future price. It refers to the degree of unpredictable changes over time of the price and it may be measured via the standard deviation of the returns (see Section 3). Simonsen [7] has demonstrated that power market volatility has some features in common with other financial markets, such as volatility clustering [8] (i.e. large changes tend to be followed by large changes and viceversa) and fat-tailed distributions, but there are also present some peculiarities; for example, power markets exhibit volatility levels well above other financial time series probably due to the fact that electricity cannot be stored efficiently.

In this work, we have applied Recurrence Quantification Analysis (RQA) to data sets taken from the Nordic spot electricity market. The relationship between RQA measures and some

^a e-mail: fstrozzi@liuc.it

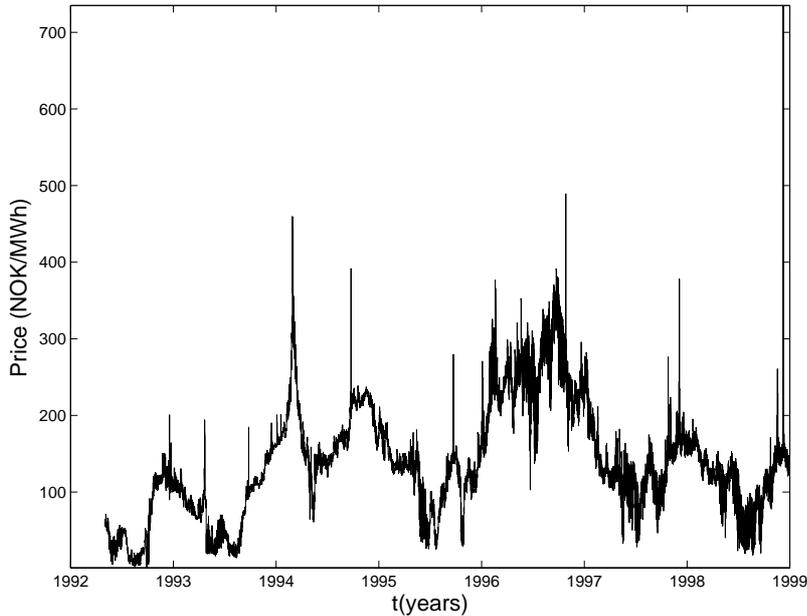


Fig. 1. Spot prices in the Nordic electricity market (Nord Pool) in NOK/MWh from May 1992 until December 1998.

dynamic features of financial time series; i.e. high frequency exchange rates, was explored in [9], [10]. Here, our main interest was in trying to correlate the volatility with variables obtained from the quantification of recurrence plots (RP). We have based our analysis on known historical events: the evolution of the Nord Pool market and climatic factors, i.e. dry and wet years. The underlying hypothesis was that the increase in the number of participants in the Nord Pool market could increase the volatility of the series and that, due to the strong dependence of hydroelectric power on climatic variability, i.e. dry-wet years, would tend also to provoke changes in the volatility of the time series. This work is a first step in the direction of exploring if there exists a correlation between the volatility in electricity prices and the frequency and intensity of blackouts. Moreover, as volatility is often used to estimate the risk associated with a financial instrument, we were interested in finding alternative measures such as the ones obtained from the application of Recurrence Quantification Analysis (RQA) [11], which allows the quantification of the Recurrence Plots (RP) [12]. The results suggest that there are two RQA measures (*DET* and *LAM* [13]) that are able to better detect salient events in comparison with other dispersion measures. In particular we analyzed the relationship between RQA measures and different dispersion measures: standard deviation of the time series, of its first differences and of the returns (i.e. financial volatility). We found a certain degree of linear correlation between these dispersion measures which is lost if we consider their linear gaussian surrogates [14]. This opens up the possibility to use these measures to assess the volatility that can take into account the non-linear dynamics that exists behind the financial data.

2 DATA PROVISION AND HISTORICAL BACKGROUND

We have analyzed hourly data from the Nord Pool system spot prices. The series is divided into two parts. In the first part, which lasts from 4th May 1992 until 31st December 1998 and comprises 58,392 data points, Fig. 1, the prices are indicated in Norwegian Krone (NOK)/MWh, whereas in the second time series, which lasts from 1st January 1999 until 26th January 2007

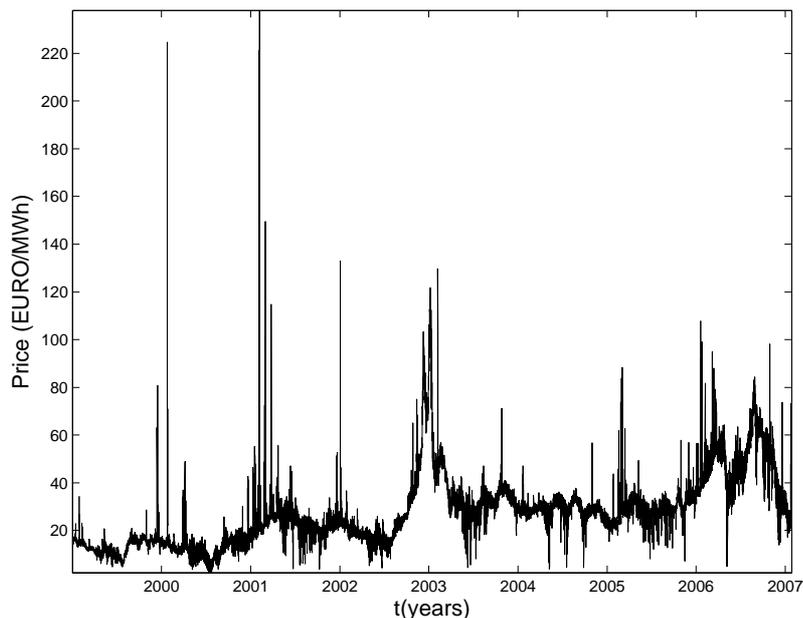


Fig. 2. Spot prices in the Nordic electricity market (Nord Pool) in EUR/MWh from January 1997 until January 2007.

Table 1. Nord Pool participating countries and dates of entry.

Countries	Date of entry
Norway	1/1/93
Norway and Sweden	1/1/96
Norway, Sweden and Finland	29/12/97
Norway, Sweden, Finland and W. Denmark	1/7/99
Norway, Sweden, Finland, W. & E. Denmark	1/10/00
KONTEK zone (Germany)	5/10/05

and comprises 70,752 data points (see Fig. 2), the prices are expressed in EUR/MWh.

The Nordic electricity market, known as Nord Pool was created in 1993 and it is owned by the two national grid companies, Statnett SF in Norway (50%) and Affrverket Svenska Kraftnt in Sweden (50%). The Nord Pool was established as a consequence of the decision in 1991 by the Norwegian Parliament to deregulate the market for power trading. Between 1992 and 1995 only Norway contributed to the market, in 1996 a joint Norwegian-Swedish power exchange was started-up and the power exchange was renamed Nord Pool ASA. Finland started a power exchange market of its own in 1996; it joined Nord Pool in 1997, and on the 15th of June 1998, Finland became an independent price area on the Nord Pool Exchange. The western part of Denmark (Jutland and Funen) has been part of the Nordic electric power market since 1st July 1999, whereas the eastern part of Denmark entered after 1st October 2000. On 5th October 2005 the German area KONTEK was added in the Nord Pool exchange market. Table 1 summarizes the historical evolution of the Nord Pool.

The spot market operated by Nord Pool is an exchange market where participants trade power contracts for physical delivery the next day, and is thus referred to as a day-ahead market. When no grid congestion exists there will be a single identical price across all the area. However, when there is insufficient transmission capacity in a sector of the grid, grid congestion

will arise and the market system will establish different "price areas". Sometimes the prices are of the entire Nordic region, sometimes more than one price area exists [15],[16]. In this work we will only consider the "system price".

The variation of the prices in the Nord pool system is well correlated with the variations in precipitation in Norway and Sweden because of its strong dependence on hydropower generation. Usually the definition of dry and wet refers to the deviation from normal in TWh (1012 Wh). When this value is negative the correspondent period is considered dry and viceversa wet when it is positive. Meteorological data analysis reported in [17] found that the 1996 was a "dry" year, while 1997-2000 was a series of "wet" years, the 2000 was not very "wet" and the first part of 2001 was quite "dry" but the autumn was very rainy and 2001 started well with a water reservoir above the normal level. During the autumn and winter season of 2002-2003 there was a sharp decline of precipitation. This was a rare event with a frequency of only one in every 100-200 years. This event resulted in the spot prices increasing in 2003. By looking into Figs. 1 and 2, we can observe that weather conditions have effects on the electricity prices. However, they are not able to explain all the features of the time series. Moreover spot prices can increase tenfold during a single hour. These spikes, which are normally quite short lived, tend to be more severe during high price periods [5].

3 DATA ANALYSIS AND RESULTS

3.1 Volatility measures

Several measures of volatility has been used in literature [5],[18],[19], between them we consider:

$$V_1 = SD(s_t) \quad (1)$$

$$V_2 = SD(s_t - s_{t-1}) \quad (2)$$

$$V_3 = SD((s_t - s_{t-1})/s_{t-1}) \quad (3)$$

where s_t and SD refer to the time series values and the standard deviation, respectively. To calculate standard deviation we used the formula:

$$SD(s_t) = \frac{1}{n-1} \cdot \sqrt{\sum_{i=1}^n (s_i - \bar{s})^2} \quad (4)$$

$\bar{s} = \frac{1}{n} \cdot \sum_{i=1}^n (s_i)$ and n is the number of points considered. In V_3 the argument of SD is an approximation of $\ln(s_t/s_{t-1})$ which is often used to measure financial volatility. In order to compare these quantities with RQA measures we invert and normalize them between 0 and 100 as follows:

$$IV_i = 1/V_i, i = 1..3 \quad (5)$$

$$nIV_i = \frac{IV_i - \min(IV_i)}{\max(IV_i) - \min(IV_i)} \cdot 100, i = 1..3 \quad (6)$$

Since there is a considerable amount of noise in financial time series, we assume that an increase of the dispersion measures should be correlated with a decrease of RQA measures that account for the predictability of the underlying dynamical system.

3.2 Finding the time delay and embedding dimension

The theory of embedding is a mathematical method that allows a temporal time series of measurements to be represented in a state space "similar" -in a topological sense- to that

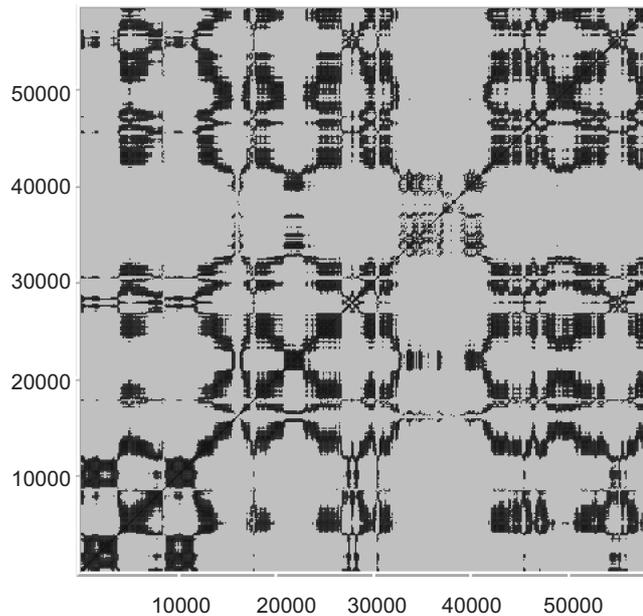


Fig. 3. Recurrence Plot of NOK/MWh. $\Delta t = 15$, $d_E = 10$, distance cutoff: 40, axis units: hour.

of the underlying dynamical system we are interested in analyzing. State space reconstruction techniques were introduced in [20],[21]. In nonlinear time series analysis delay coordinates are usually used to reconstruct a representation of the original state space that generated the dynamics. The state at a time t of a measured variable $s(t)$ is given by $S(t) = s(t), s(t - \Delta t), s(t - 2\Delta t), \dots, s(t - (d_E - 1)\Delta t)$, whereas Δt is the time delay between data when reconstructing the state space, and d_E is the embedding dimension or the dimension of the space required to unfold the dynamics. Determining the time delay and the embedding dimension is the first step in nonlinear time series modeling and prediction. The time delay, for the Nord Pool time series, has been obtained using the first minimum of the AMI (Average Mutual Information function, [22]) with values of 15 and 13 hours, respectively. The embedding dimension has been computed using the E1&E2 method [23]. Both series give the same value, $d_E = 10$. These high values are in agreement with similar analysis carried out by Strozzi *et al.* [9] for high frequency foreign exchange time series.

3.3 Quantification of the Recurrence Plots

Eckmann *et al.* [12] introduced a new graphical tool, which they called recurrence plot (RP). The recurrence plot is based on the computation of the distance matrix between points in the reconstructed state space:

$$d_{ij} = \|S_i - S_j\|. \quad (7)$$

This produces an array of distances in a $n \times n$ square matrix, \mathbf{D} , n being the number of points under study. If this distance is lower than a predetermined cutoff, r , the pixel located at specific (i, j) coordinates is darkened. These points highlight the recurrences of the dynamical systems and the recurrence plot provides insight into periodic structures and clustering properties that are not apparent in the original time series. Figures 3 and 4, show the RPs for Nord Pool time series (generated with VRA, <http://www.myjavaserver.com/~nonlinear/vra>).

In order to extend the original concept and make it more quantitative, Zbilut and Webber [24] developed a methodology called Recurrence Quantification Analysis (RQA). Several variables to quantify RPs have been defined (see for example <http://homepages.luc.edu/~cwebber>, <http://tocsy.agnld.uni-postdam.de>), of which:

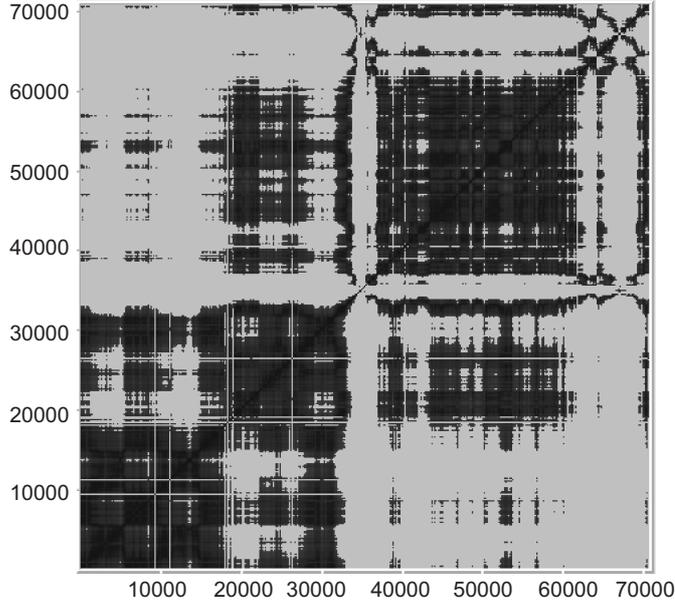


Fig. 4. Recurrence Plot of EUR/MWh. $\Delta t = 13$, $d_E = 10$, distance cutoff: 10, axis units: hour.

- RR (%recurrence): the percentage of darkened pixels in recurrence plot).
- DET (%determinism): the percentage of recurrent points forming diagonal line structures, which can be defined following [13] as:

$$DET = \frac{\sum_{l=l_{min}}^n (lP(l))}{\sum_{l=1}^n (lP(l))} \quad (8)$$

where $P(l) = P(r, l)$ is the histogram of diagonal lines of length l , l_{min} is the minimum number of points considered to have a diagonal segment and r is the distance cutoff used to have Recurrence Plot.

- L_{max} : the longest diagonal line found in the RP, which is related with the inverse of the maximum Lyapunov exponent which measures the stability of the system in the state space.
- ENTR: the Shannon entropy that quantifies the structures in RP.
- Trend: the measure of the paling recurrence points away from the central diagonal.
- LAM (%laminarity): the percentage of points forming vertical lines, which can be defined following [13] as:

$$LAM = \frac{\sum_{\nu=\nu_{min}}^n (\nu P(\nu))}{\sum_{\nu=1}^n (\nu P(\nu))} \quad (9)$$

where $P(\nu) = P(r, \nu)$ is the histogram of vertical lines of length ν , ν_{min} is the minimum number of points considered to have a vertical segment and r is the distance cutoff used to have Recurrence Plot.

- TT (the trapping time), which estimates the mean time that the system will stay at a specific state.

In order to understand if the information obtained from the RQA measures are related to statistical or dynamical properties, we performed the following test. We recalculated RQA parameters on randomly shuffled data sets of, respectively, EUR/MWh and NOK/MWh. If we use the same parameters of Figs. 3 and 4, and, in particular, for the same radius, r , the RR becomes zero instead of 7.12 (see Table 3) and the RP is empty. If we increase the radius, we have recurrent points but without any structure Fig. 5. Although the statistical distribution of the data does not change if we only shuffle them, the RQA parameters do change; however, the

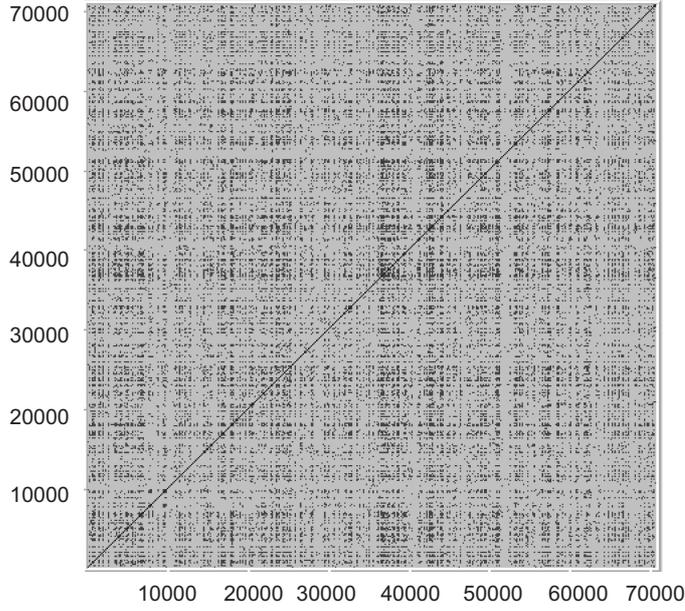


Fig. 5. RP of EUR/MWh shuffled data. $\Delta t = 13$, $d_E = 10$, distance cutoff: 25, axis units: hour.

Table 2. RQA measures for NOK/MWh original time series and its surrogates.

Data set	RR	DET	L_{max}	ENTR	Trend	LAM	TT
NOK	16.10	67.13	3545	8.59	-8.69	69.99	308
<i>Surr</i> ₀₁	8.15	6.13	4808	6.74	2.31	1.80	124
<i>Surr</i> ₀₂	1.93	4.52	1355	4.91	-0.14	0.0	-
<i>Surr</i> ₀₃	2.81	8.03	4808	6.03	-1.62	0.0	-
<i>Surr</i> ₀₄	30.22	36.31	4808	7.99	-3.36	35.52	215
<i>Surr</i> ₀₅	1.74	13.22	1844	6.12	-0.98	0.06	110
<i>Surr</i> ₀₆	1.01	32.02	1178	6.29	-0.75	16.98	166
<i>Surr</i> ₀₇	4.79	13.28	2674	6.90	-0.80	7.53	154
<i>Surr</i> ₀₈	14.12	17.88	4350	7.36	-4.48	9.29	155
<i>Surr</i> ₀₉	5.93	13.53	3130	7.20	-2.46	6.30	159
<i>Surr</i> ₁₀	1.19	5.90	1064	4.70	-0.68	0.35	120
<i>Surr</i> ₁₁	4.86	51.64	4808	7.92	-1.54	52.44	266
<i>Surr</i> ₁₂	31.90	52.68	4808	8.42	12.41	54.52	218
<i>Surr</i> ₁₃	4.80	9.42	4808	6.88	0.52	0.72	144
<i>Surr</i> ₁₄	5.73	9.17	4154	6.78	-2.98	1.77	145
<i>Surr</i> ₁₅	4.97	6.34	2370	6.61	-2.34	1.72	115
<i>Surr</i> ₁₆	18.05	23.40	4808	7.68	-4.18	12.85	161
<i>Surr</i> ₁₇	10.85	43.19	4614	8.80	-7.22	38.30	339
<i>Surr</i> ₁₈	4.96	8.52	4808	6.60	-2.65	3.48	142
<i>Surr</i> ₁₉	6.32	4.46	4808	6.18	-1.99	0.38	114

RQA measures are not able to identify the statistical distribution of the data. To check if RQA measures were appropriate to analyze the spot prices dynamics, we have created surrogate time series of the real data set generated by a Gaussian linear random process with the same FFT [14]; then we have computed their RQA parameters. The results are summarized in Table 2 and Table 3 for NOK/MWh and EUR/MWh time series, respectively.

We can observe that the parameters RR, L_{max} , ENTR, Trend and TT could not distinguish (with a 95% confidence) between a linear Gaussian dynamics and the dynamics behind

Table 3. RQA measures for EUR/MWh original time series and its surrogates.

Data set	RR	DET	L_{max}	ENTR	Trend	LAM	TT
EUR	7.12	35.33	2094	7.66	-4.59	33.94	264
<i>Surr</i> ₀₁	12.52	3.67	3340	6.36	-6.26	2.54	149
<i>Surr</i> ₀₂	1.64	5.89	2238	5.27	-1.10	1.87	119
<i>Surr</i> ₀₃	3.84	1.40	2150	4.53	-1.00	0.0	-
<i>Surr</i> ₀₄	4.38	1.11	1324	3.97	-0.29	0.0	-
<i>Surr</i> ₀₅	10.68	1.83	4187	5.73	-5.48	1.53	127
<i>Surr</i> ₀₆	8.66	18.81	4826	7.54	-5.64	9.85	146
<i>Surr</i> ₀₇	0.49	3.89	690	2.81	-0.35	0.0	-
<i>Surr</i> ₀₈	23.79	11.11	4826	7.51	-7.64	9.25	162
<i>Surr</i> ₀₉	30.27	10.83	4826	7.39	-1.83	7.11	151
<i>Surr</i> ₁₀	20.54	4.70	4826	6.85	-7.47	6.42	151
<i>Surr</i> ₁₁	2.34	3.78	1888	5.09	-1.16	1.53	134
<i>Surr</i> ₁₂	3.72	1.48	3517	4.06	-1.63	0.11	117
<i>Surr</i> ₁₃	4.99	3.74	3721	6.88	0.52	0.72	144
<i>Surr</i> ₁₄	21.65	9.02	4826	7.16	-2.90	9.66	155
<i>Surr</i> ₁₅	20.05	8.14	2669	7.25	-4.24	4.17	146
<i>Surr</i> ₁₆	6.81	5.38	3998	6.57	-4.10	0.76	125
<i>Surr</i> ₁₇	3.16	4.11	1964	5.64	-2.08	0.72	132
<i>Surr</i> ₁₈	7.81	3.37	2429	6.20	-0.47	2.77	132
<i>Surr</i> ₁₉	12.19	1.33	4826	5.43	1.50	0.09	126

the Nord pool time series. Of course, this does not imply that those parameters are not useful for their quantification, but only that the values of the parameters in the surrogate time series were indistinguishable from those of the original time series. On the contrary, DET and LAM always produced values which were higher in the original data set when compared with surrogate data.

The fact that DET and LAM were able to distinguish between the original and the surrogate time series can be explained by assuming that there is more structure in the original series, and therefore the distance in state space remained closer for longer times when compared with their surrogate linear Gaussian process. It is possible to assume that during high volatility periods the sensitivity increases, and consequently the forecast becomes more difficult even for short time horizon. The measures related with the percentage of determinism (DET) of the time series will tend to decrease. Higher DET and LAM mean that the states of the system stay closer in time for longer periods, forming diagonal or vertical segments in the RPs. Thus, we may assume that higher DET values imply smaller volatility.

We have computed these RQA measures inside a moving window. For this analysis we used a one month moving window shifted by one month (720 points) for both data sets (NOK/MWh and EUR/MWh).

We were interested in observing if some changes in the RQA parameters make sense in correspondence of the entry of a new country in Nord Pool (Table 1) or in correspondence with dry and wet years. To study the relationship between DET and LAM with the dispersion measures considered (Eq. 6), we have compared their profiles, as can be seen in Figs. 6 and 7, for the meteorological conditions, where the bold lines are seasonal -3 months- averages of the plotted quantities. It is possible to observe, for some years, a seasonal behavior with a decrease during the middle of the year corresponding with spring and summer, with minimum values for 1996 (dry year) for NOK/MWh series in all measures except for nIV_3 , Fig. 6. Similarly, for EUR/MWh series, there is a qualitative agreement being 2001, 2003, 2005 and 2006 the years for which lower values are found.

To analyze the effect of the entrance of a new state in the Nord Pool on DET, LAM and on the three dispersion measures, we took the mean values of the measures between the two successive changes of the composition of the Nord Pool. In this case looking at Figs. 8 and 9 it seems that, as the number of countries in the Nord Pool system increases, there is a decrease of

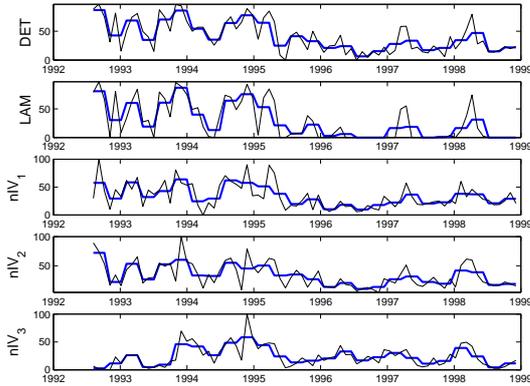


Fig. 6. RQA and dispersion measures for NOK/MWh series. RQA parameters computed from a 720 point window (one month), shifted by one month. RQA parameters: $\Delta t = 15$, $d_E = 10$, distance cutoff: 40, line definition: 100 points (≈ 4 days). In bold seasonal averaged values (three months).

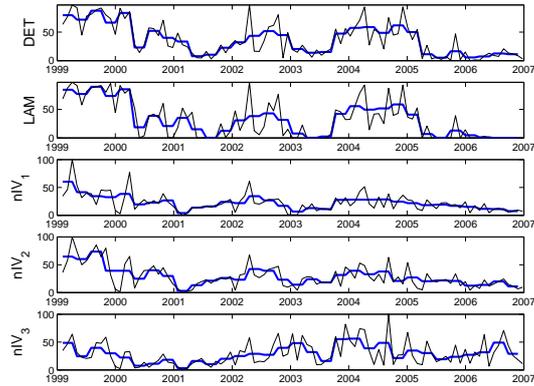


Fig. 7. RQA and dispersion measures of EUR/MWh time series. RQA measures are computed from a 720 point window (one month), shifted by one month. RQA parameters: $\Delta t = 13$, $d_E = 10$, distance cutoff: 4, line definition: 100 points (≈ 4 days). In bold seasonal averaged values (three months).

DET and LAM as well as of the inverse of the other dispersion measures, with the exception of nIV_3 . Moreover it is possible to observe that by using the RQA measures, the changes in the mean values are more evident (the steps are higher) than using the other dispersion measures. However, it is evident that these effects could not have been discovered *a priori* without the historical knowledge. In any case, this behavior suggest that volatility is influenced by external events and that high values of volatility correspond to low values of DET and LAM.

The determination coefficient R^2 , which measures the degree of linear correlation between DET and LAM and the dispersion measures given by Eq. 6, is presented in Table 4. There is a high linear relationship between DET and LAM, with a determination coefficient R^2 of 0.88 and 0.89 for NOK and EUR, respectively. The linear relationship between these values and the inverse of the dispersion measures is lower. However the R^2 values between nIV_1 (the normalized inverse of the standard deviation) and DET are 0.47 for both series. For the case of LAM we obtain 0.45 and 0.58 for EUR/MWh and NOK/MWh, respectively. The determination coefficient decreases to 0.25 for nIV_2 and is practically zero for nIV_3 . Also, as seen in Figs. 6-9, this measure of the volatility is less correlated to the others since it is defined in relative terms (see Eq. 3).

The same treatment, as applied to the linear gaussian surrogates of Nord Pool time series, produces a decrease of the linear correlation, see Table 5. There is a decrease in R^2 : 0.56 and

Table 4. Determination coefficient R^2 between dispersion measures DET and LAM.

	EUR/MWh	NOK/MWh
LAM -DET	0.89	0.88
nIV_1 - DET	0.47	0.47
nIV_2 - DET	0.25	0.25
nIV_3 - DET	0.02	0.10
nIV_1 - LAM	0.45	0.58
nIV_2 - LAM	0.26	0.30
nIV_3 - LAM	0.02	0.12

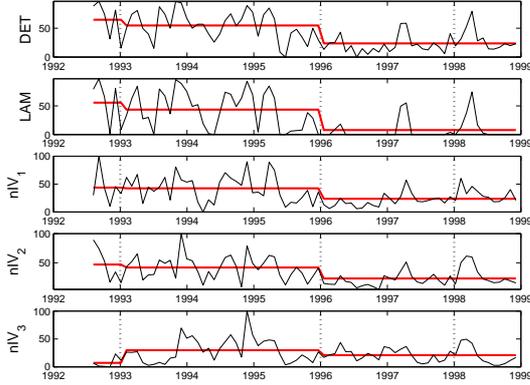


Fig. 8. RQA and dispersion measures of time series in NOK/MWh. RQA measures are computed from a 720 point window (one month), data are shifted by 720 points. RQA parameters: $\Delta t = 15$, $d_E = 10$, distance cutoff: 40, line definition: 100 points (≈ 4 days). Bold lines correspond to the mean values between two vertical lines according to the dates given in Table 1.

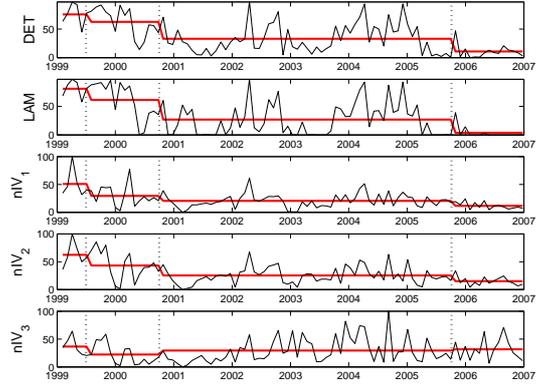


Fig. 9. RQA measures of EUR/MWh. Values are computed from a 720 point window (one month), shifted by 720 points. RQA parameters: $\Delta t = 13$, $d_E = 10$, distance cutoff: 4, line definition: 100 points (≈ 4 days). Bold lines correspond to the mean values between two vertical lines according to the dates given in Table 1.

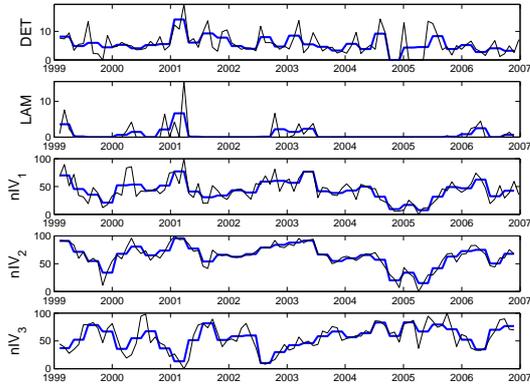


Fig. 10. RQA and dispersion measures of a EUR/MWh surrogate time series. Same parameters as Fig. 7.

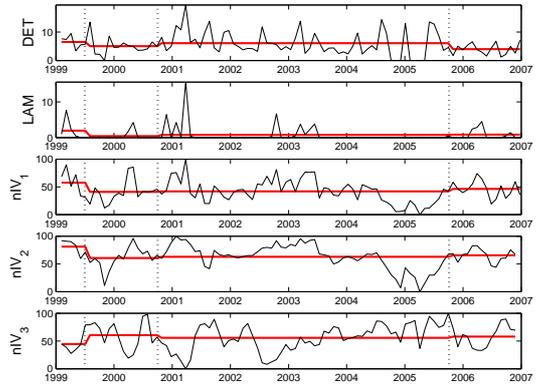


Fig. 11. RQA and dispersion measures of a EUR/MWh surrogate time series. Same parameters as Fig. 9.

0.67 compared with 0.89 and 0.88 between DET and LAM. The same occurs for nIV_1 , whereas the correlations remain practically the same for nIV_2 and nIV_3 .

It seems that DET and LAM quantify the volatility related to the underlying nonlinear dynamics in these series, whereas this effect decreases when we use surrogates, in particular the linear correlation between DET, LAM and nIV_1 .

In addition, when we compare the results (Figs. 10 and 11) with the historical data (Figs. 7 and 9) we can observe that DET, LAM and nIV_1 show no seasonal pattern, and the continuous decrease present in Fig. 9 has disappeared.

Table 5. Determination coefficient R^2 between dispersion measures DET and LAM. Mean values on 19 surrogate data sets.

	EUR/MWh	NOK/MWh
LAM -DET	0.56	0.67
nIV_1 - DET	0.18	0.16
nIV_2 - DET	0.21	0.27
nIV_3 - DET	0.09	0.08
nIV_1 - LAM	0.25	0.29
nIV_2 - LAM	0.21	0.32
nIV_3 - LAM	0.08	0.09

4 CONCLUSIONS

Nonlinear time series analysis and, in particular, Recurrence Quantification Analysis has been carried out for the Nord Pool time series, the goal being to analyze and characterize them using RQA measures. In order to assess the results, we have used historical information in our analysis. Particularly, we were interested in studying the reasons for the high volatility found in these series [7] and in finding RQA measures that could provide additional results to complement standard dispersion measures (see Eqs. 1- 3).

As a first step, we have compared the RQA measures of the original time series with two types of surrogate series: shuffled and linear Gaussian with the same FFT. We have observed that RQA measures do not characterize the probability distribution of the data, because the shuffled and the real data have the same mean and variance, but different values of RQA measures. In addition, we have found that two RQA measures: DET and LAM are able to distinguish between real and linear Gaussian surrogate with 95% of confidence. For this reason and because of the hypothesis that high volatility can imply small DET and LAM, we have compared them with the inverse of the normalized dispersion measures given by Eq. 6 on a one month moving window translated of one month. We have found that these measures are correlated with the inverse of dispersion measures that are used to evaluate the volatility of financial time series, see Table 4. We have found a qualitative agreement (see Fig. 6 and Fig. 7) from the point of view of high and low values corresponding to wet and dry periods and a general decrease of the measures with the entrance of new countries in the Nord Pool (Fig. 8 and Fig. 9). The linear correlation between these measures decreases for the linear gaussian surrogates (see Table 5) as well as the agreement with historical events. We have observed that DET and LAM provide an alternative measure of dispersion of a financial time series that take into account the underlying dynamics. To see if the RQA measures have some advantages in comparison with the other dispersion measures (Eq. 6), we have observed that DET and LAM show more pronounced jumps between the periods analyzed. This behavior is lost when we apply the same treatment to surrogate data sets.

In future work we will investigate the correlation between market prices (or some related variable such as volatility, DET, LAM) and the likelihood of blackouts.

The authors gratefully acknowledge the financial support of the European Commission DG RTD funded MANMADE NEST Project (Contract No 043363) and of Fondazione Cariplo. The authors would like to acknowledge Dr. H. Sivonen (NESA) who kindly provided the data sets analyzed.

References

1. R. N. Mantegna, H. E Stanley, *An introduction to Econophysics* (Cambridge University Press, Cambridge, 2000)
2. A. S. Soofi, L. Cao (Eds.), *Modelling and forecasting financial data: Techniques of Nonlinear Dynamics* (Kluwer Academic Publishers, Norwell, 2002)

3. M. M. Dacorogna, R. Genay, U. Mller, R. B. Olsen, O. V. Pictet, *An Introduction to High-Frequency Finance* (Academic Press, San Diego, 2001)
4. R. Weron, B. Przybylowicz, *Physica A* **283**, 462 (2000)
5. I. Simonsen, *Physica A* **322**, 597 (2003)
6. M.Bask, T. Liu, A. Widerberg, *Physica A* **376**, 565 (2007)
7. I. Simonsen, *Physica A* **355**, 10 (2005)
8. A.N. Shiryaev, *Essential of Stochastic Finance*, World Scientific Publishing, (1999).
9. F. Strozzi, J. M. Zaldívar, J. P. Zbilut, *Physica A* **312**, 520 (2002)
10. F. Strozzi, J. M. Zaldívar, and J. P. Zbilut, *Physica A* **376**, 487 (2007)
11. C. L. Webber, J. P. Zbilut, *J. Appl. Physiol.* **76**, 965 (1994)
12. J. P. Eckmann, J. O. Kamphorst, D. Ruelle, *Europhys. Lett.* **4**, 973 (1987)
13. N. Marwan, M.C. Romano, M. Thiel, J. Kurths, *Physics Reports* **438**, 237 (2007)
14. H. Kantz, T. Shreiber, *Nonlinear Time Series Analysis* (Cambridge University Press, Cambridge, 1997)
15. T. Kristiansen, *Energy Policy* **35**, 307 (2007)
16. N. Haldrup, M. Nielsen, *Journal of econometrics* **135**, 349 (2006)
17. R. Weron, M. Bierbrauer, S. Truck, *Physica A* **336**, 39 (2004)
18. S.D.H. Hsu, B.M. Murray, *Physica A* **380**, 366 (2007)
19. A. Figueiredo, I. Gleria, R. Matsushita, S. Da Silva, *Physica A* **346**, 484 (2005)
20. N. Packard, J. Crutchfield, D. Farmer, R. Shaw, *Phys. Rev. Lett.* **45**, 712 (1981)
21. F. Takens, *Dynamical Systems and Turbulence*, Lecture Notes in Mathematics vol. 898, A. Rand and L.S Young (eds.), (Springer, Berlin, 1981), pp. 366-381
22. A. Fraser and H. Swinney, *Phys. Rev. A* **33**, 1134 (1986)
23. L. Cao, *Physica D* **110**, 43 (1997)
24. J.P. Zbilut, C.L. Webber, *Phys. Lett. A* **171**, 199 (1992)

ANNEX III. *Time series analysis and long range correlations of Nordic spot electricity market data.* Physica A 387 (2008), 6567-6574



Time series analysis and long range correlations of Nordic spot electricity market data

Hartmut Erzgräber^{a,*}, Fernanda Strozzi^b, José-Manuel Zaldívar^c, Hugo Touchette^a, Eugénio Gutiérrez^c, David K. Arrowsmith^a

^a School of Mathematical Sciences, Queen Mary / University of London, Mile End Road, London E1 4NS, UK

^b Quantitative Methods Institute, LIUC-Università Cattaneo, Corso Matteotti 22, 21053 Castellanza (VA), Italy

^c European Commission, Joint Research Centre, Via E. Fermi 1, 21020 Ispra (VA), Italy

ARTICLE INFO

Article history:

Received 13 February 2008

Received in revised form 23 June 2008

Available online 7 August 2008

PACS:

05.45.Tp

05.40.Jc

Keywords:

Hurst exponent

Nonlinear time series analysis

Long range correlations

ABSTRACT

The electricity system price of the Nord Pool spot market is analysed. Different time scale analysis tools are assessed with focus on the Hurst exponent and long range correlations. Daily and weekly periodicities of the spot market are identified. Even though space time separation plots suggest more stationary behaviour than other financial time series, we find large fluctuations of the spot price market which suggest time-dependent scaling parameters.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The complex behaviour of financial time series has been the object of a considerable amount of studies [1,2]. It has been demonstrated that linear stochastic models are not able to capture properly this complexity and therefore it has been attributed to the fact that financial markets are nonlinear stochastic, chaotic or a combination of both. Specifically, in the last decades there has been a considerable amount of discussion about the characterisation of financial time series using the theory of Brownian motion [3,4], fractional Brownian motion [5], nonlinearity [6], chaos and fractals [7–9], scaling behaviour [10,11], and self organised criticality [12,13]. Most of the tests developed in the area of economic theory provide evidence of nonlinear dynamics, which may be deterministic or not deterministic. There is no convincing evidence of deterministic low-dimensionality in price series [14,15], and the claims of low-dimensional chaos have never been well-justified [16,17,11]. Nevertheless, in the last few years nonlinear time series analysis has expanded rapidly in the fields of Economics and Finance. This is also due to the fact that economic and financial time series seem to provide a promising area for the development, testing and application of nonlinear techniques [18] and the fact that high frequency financial time series are readily available.

Among these time series, energy spot prices have been analysed with several nonlinear techniques. Weron and Przybyłowicz [19] studied the electricity prices using Hurst R/S analysis and showed anti-persistent behaviour with a Hurst exponent lower than 1/2. Using another technique, the average wavelet coefficient method [20], a Hurst exponent with value $H = 0.41$ has been obtained which is in agreement with other energy spot price time series. The question of modelling

* Corresponding address: School of Engineering Computing and Mathematics, University of Exeter, North Park Road, Exeter EX4 4QF, UK.
E-mail address: h.erzgraber@exeter.ac.uk (H. Erzgräber).

electricity spot prices has also been addressed by several researchers. Because of the high volatility in Nord Pool electricity prices, Byström [21] applied extreme value theory to investigate the tails of the price change distribution and then used the peaks-over-threshold method to analyse the data that exceed the threshold. Perelló et al. [22] proposed a GARCH model for the spot price. Weron et al. [23] fitted a jump diffusion and regime switching model to Nord Pool spot prices. Vehviläinen and Pyykkönen [24] developed a stochastic factor based approach to mid-term modelling of spot prices taking into account climate data, hydro-balance, base load supply and the underlying mechanisms in spot price generation. The model was able to provide simulated values for the fundamental data, demand and supply information, and pricing strategies.

Here we are mainly concerned with quantifying long range correlations in energy spot price market data in terms of Hurst exponents. Such a concept has been widely used for the analysis of economic time series at the level of different quantities. For instance, Simonsen [25] analyses the volatility of the Elspot electricity market. Volatility clustering is observed, and relations between electricity markets and traditional financial markets are described. Main differences to traditional financial markets are a general high level of volatility and a possible dependence of the volatility on the price itself. Reference [26] contains a brief discussion on the application of standard financial tools to electricity markets. In particular, it appears that the price for electricity is more volatile compared to other commodities because electricity cannot be stored in an efficient way. The Spanish electricity market is analysed in Ref. [27], using multifractal detrended fluctuation analysis. The Hurst exponent is estimated to $H = 0.16 \pm 0.01$. Ref. [28] analyses different energy prices, using the detrended moving average technique. In particular, crude oil, natural gas, heating oil, unleaded gasoline, and propane gas are considered. Focus is on the decay process of shocks in the return process.

We investigate in detail correlation properties of the Nord Pool electricity market. Some basic features about the time series are reviewed in Section 2. In addition, we address the question whether such data can be described as a stationary process, at least on the time scales covered by the data set. To keep the presentation self-contained, we recall in Section 3 some basic facts about the Hurst exponent and algorithms for the estimation of such a quantity. We then compare in Section 4 results obtained by these different algorithms and evaluate in more detail fluctuation properties related with such exponents. We check, in particular, if surrogate time series with the same power spectrum but originated by a linear Gaussian process may have the same Hurst exponent. Some comments on large fluctuations are contained in the conclusion, Section 5.

2. Data set and time series analysis

The Nordic electricity market, known as Nord Pool (<http://www.nordpool.no>) was created in 1993 and is owned by the two national grid companies, Statnett SF in Norway (50%) and Affärverket Svenska Kraftnät in Sweden (50%). The market was established as a consequence of the decision in 1991 by the Norwegian parliament to deregulate the market for power trading. Therefore, between 1992 and 1995 only Norway contributed to the market, in 1996 a joint Norwegian-Swedish power exchange was started-up and the power exchange was renamed Nord Pool ASA. Finland started a power exchange market of its own, EL-EX, in 1996 and joined Nord Pool in 1997. Beginning of 15th June 1998, Finland became an independent price area on the Nord Pool Exchange. The western part of Denmark (Jutland and Funen) has been part of the Nordic electric power market since 1 July 1999, whereas the eastern part of Denmark entered after 1st October 2000. On 5th October 2005 also the German area KONTEK was added in the Nord Pool exchange market.

The spot market operated by Nord Pool is an exchange market where participants trade power contracts for physical delivery the next day. Thus, it is referred to as a day-ahead market. The spot market is based on an auction with bids for purchase and sale of power contracts of one hour duration covering the 24 h of the following day. At the deadline for the collection of all buy and sell orders the information is gathered into aggregate supply and demand curves for each power-delivery hour. From these supply and demand curves the equilibrium spot prices – referred to as the system prices – are calculated.

We have analysed hourly data from the Nord Pool system spot prices. The series is divided into two parts. The first part, from 4th May 1992 to 31st December 1998, comprises 58,392 data points. The prices are indicated in Norwegian Krone (NOK)/MWh, whereas the second part of time series, from 1st January 1999 to 26th January 2007, comprises 70,752 data points with prices being expressed in EUR/MWh. We have considered the time series $s(t)$ as well as the corresponding returns over the time horizon Δ , defined as

$$r_{\Delta}(t) = \ln(s(t)/s(t - \Delta)). \quad (1)$$

Fig. 1 shows the hourly returns for the two parts of the time series considered. For both parts we have also computed the distribution function, using the program STABLE for univariate data [29]. The result resembles in each case a stable distribution $S(\alpha, \beta, \gamma, \delta)$ where the fit yields parameter values $\alpha = 1.116$, $\beta = 0.127$, $\gamma = 0.242$, $\delta = -0.05$ for the first part of the time series and $\alpha = 1.315$, $\beta = 0.173$, $\gamma = 0.272$, $\delta = -0.07$ for the second part. Thus the observed distribution resembles a Cauchy distribution ($\alpha = 1$, $\beta = 0$) and differs considerably from a Gaussian ($\alpha = 2$).

Stationarity, that means broadly speaking that long time averages like mean values, variances, distribution functions, or correlation functions do not depend on the initial time, is a property normally required for a statistical analysis. There are prominent examples in physics, like disordered systems or glasses, which show failure of stationary behaviour because of intrinsic properties of the dynamics (cf. e.g. Ref. [30] for a simple dynamical model of ageing). Furthermore, economic time series normally depend on external factors and may suffer from pronounced non-stationary behaviour [31]. It is also worth

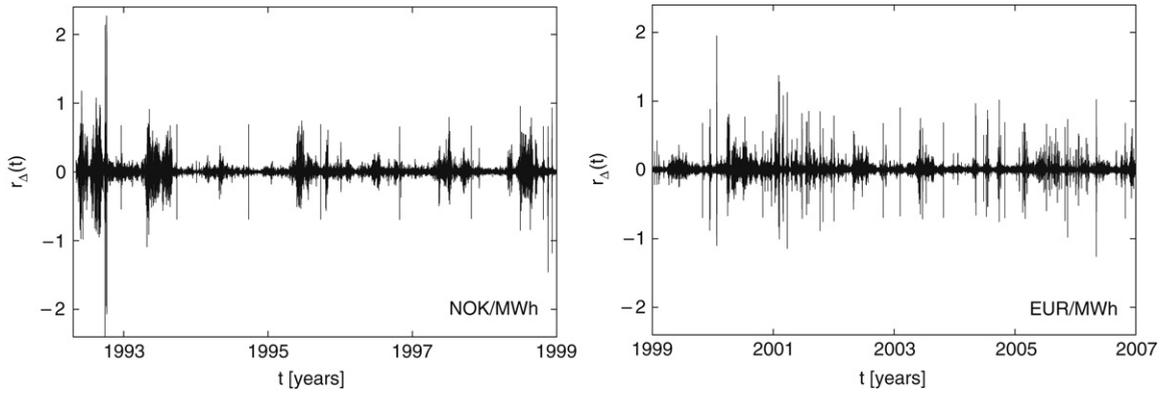


Fig. 1. Left: Hourly logarithmic return, Eq. (1), for the spot prices in the Nordic electricity market (Nord Pool) from May 1992 until December 1998. Right: Hourly logarithmic return for the spot prices from January 1999 until January 2007.

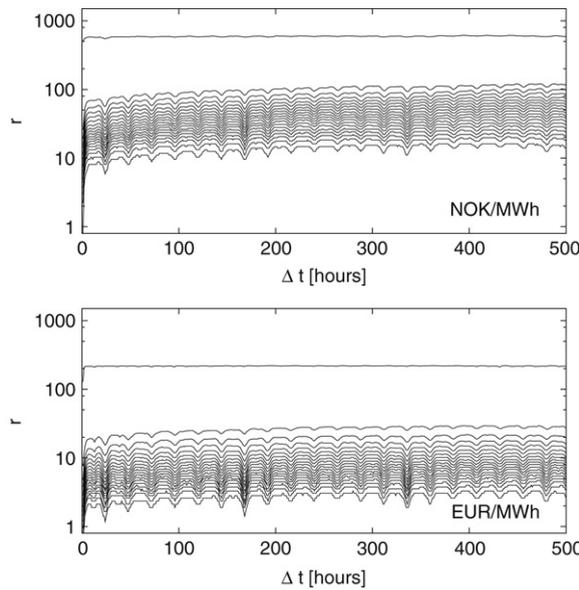


Fig. 2. Top: Space-time separation plot of the Nord Pool spot prices (NOK/MWh). Bottom: Space-time separation plot of the Nord Pool spot prices (EUR/MWh).

stressing that the non-stationary components, such as the trend, may sometimes be of more interest than the stationary residual. While it is almost impossible to test for stationary behaviour in a rigorous way, we can still check whether on time scales of interest our data set behaves essentially like a stationary process. We here report on a relatively simple stationarity test, called space time separation plot [32]. For this purpose one evaluates the probability

$$P(r, \Delta t) = \text{Prob}(\|x(t + \Delta t) - x(t)\| < r) \tag{2}$$

that phase space points, separated in time by an interval Δt , have distance less than r . If the process is stationary and if Δt exceeds the correlation time then such a quantity becomes independent of the time lag Δt and coincides with the correlation integral which is frequently used for estimating fractal properties of chaotic attractors. Since phase space coordinates are normally not accessible, one employs standard delay embedding techniques to estimate the required probability function. We have used the program *stp* of the Tisean software package [33] which returns level lines of $P(r, \Delta t)$ for $P = 0.05, 0.1, 0.15, \dots$ Horizontal level lines in such a contour plot indicate the required independence on Δt and are thus a signature of a stationary time series. Fig. 2 shows the results of the test of the time series under consideration. In those graphics, the separation time Δt is represented in the horizontal axis whereas the the separation in space, r , is represented in the vertical axis. As can be observed from Fig. 2 the contour plot obtained from the Nord Pool time series consists essentially of horizontal lines apart from a weak 24 h periodicity. Thus $P(r, \Delta t)$ is essentially independent of Δt . While financial time series are normally not stationary, the results from space time separation plots demonstrate that electricity market data, here the corresponding returns, are more stationary than all financial series analysed so far, i.e. foreign market exchange data sets [31]. Furthermore, it should be noted that the changes in the market, e.g. the appearance of new participants

(cf. beginning of this section) does not show up as a strong violation of stationary behaviour, at least on the considered time scales. The results shown in Fig. 2 support the conjecture that the data set we are dealing with is more stationary than other financial time series. Further evidence comes from the observation that the contour plots do not change qualitatively when being based on shuffled surrogate data, even though the weak 24 h periodicity disappears, as expected for proper shuffled surrogates.

3. Long range correlations

An algebraic decay of the autocorrelation function $C(\tau) = \langle r_\Delta(t)r_\Delta(t + \tau) \rangle \sim |\tau|^{-\beta}$ on large time scales or the corresponding power law behaviour of the power spectrum $S(\omega) \sim |\omega|^{\beta-1}$ in the low frequency domain may be characterised in terms of the Hurst exponent $H = 1 - \beta/2$. A power law scaling of the correlation function on small time scales can be related with the fractal dimension of the corresponding stochastic process, and both quantities, the fractal dimension of the process and the Hurst exponent are in general independent quantities [34].

A tool for studying long-term memory and fractality of a time series is the rescaled range or R/S analysis first introduced by Hurst [35] in hydrology. Mandelbrot [36] argued that R/S analysis is a more powerful tool in detecting long range dependence when compared to more conventional methods like autocorrelation analysis, variance ratios, and spectral analysis. The range R of a time series with a finite sampling rate is defined by

$$R(\tau) = \max_t(X(t, \tau)) - \min_t(X(t, \tau)), \quad (3)$$

where $X(t, \tau)$ denotes the sum of the deviation of the time series $s(t)$ from its mean value $\langle s \rangle_\tau$ over some time interval τ

$$X(t, \tau) = \sum_{\ell=t}^{t+\tau} (s(\ell) - \langle s \rangle_\tau(t)). \quad (4)$$

Moreover, $S(\tau)$ denotes the standard deviation of the time series over the time window τ . Computed for different sizes of the time window, the rescaled range $R(\tau)/S(\tau)$ shows a power law scaling

$$R(\tau)/S(\tau) \sim \tau^H \quad (5)$$

with exponent H . The Hurst exponent is equal to $1/2$ for Brownian motion, while $H < 1/2$ or $H > 1/2$ indicate anti-correlated and positively correlated increments, respectively.

Improved methods to estimate the Hurst exponent have been proposed to take care of non-stationary components of the time series. The detrended moving average (DMA) uses the scaling behaviour of the standard deviation

$$\sigma_{DMA}(\tau) = \sqrt{\frac{1}{N-\tau} \sum_{\ell=\tau}^N (s(\ell) - \langle s \rangle_\tau(\ell))^2} \quad (6)$$

about a moving average $\langle s \rangle_\tau(t)$ of a time series $s(t)$ of length N for different sizes τ of the moving average window (cf. Refs. [37,38]). The power law scaling of this standard deviation with the window size, $\sigma_{DMA}(\tau) \sim \tau^H$, yields the Hurst exponent. The generalised multifractal detrended fluctuation analysis (MF-DFA) pursues a similar idea. Here, the standard deviation is computed with regards to a low-order polynomial fit of the time series. One divides the time series $s(t)$ into n non-overlapping windows of equal size τ . For each window a polynomial fit to the time series is computed. The standard deviation

$$\sigma_{DFA}(\tau) = \sqrt{\frac{1}{N} \sum_{\ell=1}^N (s(\ell) - s_\tau^{poly}(\ell))^2} \quad (7)$$

quantifies the variation of the time series $s(t)$ about the polynomial fit $s_\tau^{poly}(t)$ where the order m of the polynomials determines the order of the MF-DFA. Different orders differ in their ability to eliminate trends in the time series; see e.g. Refs. [39,40,42]. Again, the scaling of the standard deviation with the window size yields the Hurst exponent, $\sigma_{DFA}(\tau) \sim \tau^H$. The approach has been generalised by introducing a spectrum of Hurst exponents to take multifractal properties of the time series into account as well.

4. Results

We now report on the calculation of the Hurst exponent for the Nordic Pool Spot data using some of the methods just described. First, we have used the standard scaled windowed variance method [41] to estimate the Hurst exponent by linear regression of $\ln(R/S)$ versus $\ln(\tau)$. Fig. 3 shows the data evaluated separately for the two parts of the time series, currencies in NOK and in EUR. We obtain a quite pronounced scaling range for the first part of the time series with Hurst exponent $H_{NOK} = 0.44$. For the second part the scaling behaviour is slightly less convincing and corrupted by rather large fluctuations.

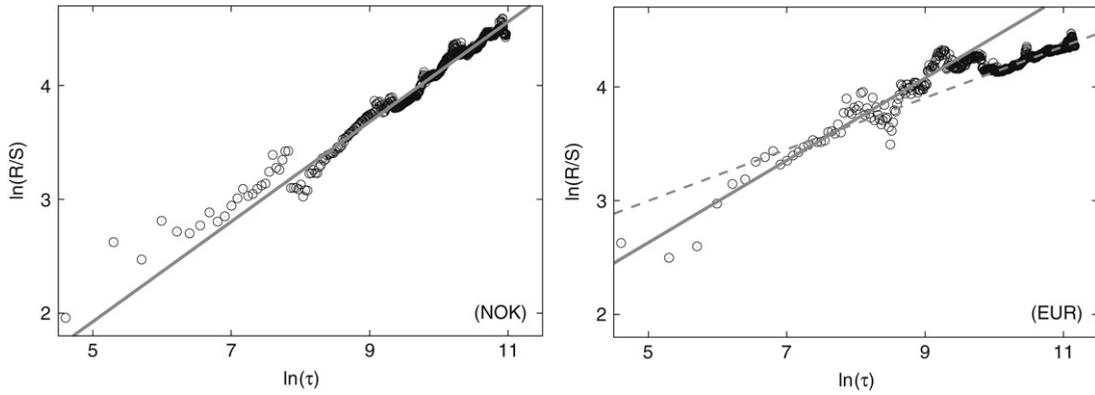


Fig. 3. Rescaled range as a function of the window size τ for spot prices (NOK: left) from May 1992 to December 1998 and (EUR: right) January 1999 till January 2007 on a double logarithmic scale (arbitrary units). Straight lines indicate a fit to the data with slope $H_{NOK} = 0.44$ and $H_{EUR} = 0.36$, respectively. The broken line indicates a data fit over a different, smaller interval with slope $H_{EUR} = 0.23$.

The actual value for the slope depends on the range for the data fit. One can produce values between $H_{EUR} = 0.36$ and $H_{EUR} = 0.23$. Nevertheless, as it can be seen both parts of the time series show anti-persistence, $H < 1/2$. This has already been found by several researchers [19,20,22], amongst others. We generated two types of surrogate time series. The first type is a Gaussian linear stochastic process with the same mean, variance and power spectrum of the original data, the second is obtained by a random shuffling of the original time series. The Hurst exponent of these two types of surrogates together with the standard deviation measured over 20 realisations are $H_{NOK}^{(G)} = 0.36 \pm 0.07$, $H_{NOK}^{(S)} = 0.51 \pm 0.07$, $H_{EUR}^{(G)} = 0.33 \pm 0.06$, and $H_{EUR}^{(S)} = 0.53 \pm 0.06$. The Hurst exponents of the original time series differ slightly from the linear surrogate but this does not mean that the value of H helps us to distinguish between the original time series and their surrogates because of the quite large errors that come with the numerical values. In fact the exponents seem to be in the range 0.4 ± 0.1 although the Hurst exponent for the second part of the time series is quite low. For the shuffled surrogate time series we obtain, as expected, Hurst exponents close to $1/2$.

The difference between the two Hurst exponents H_{NOK} and H_{EUR} is a clear signature of non stationary behaviour most likely to be caused by the appearance of new market participants. It is thus not surprising that these values differ. On the other hand one should also keep in mind that the actual error bars for the exponents are likely to be substantial. For the shuffled time series one would expect a Hurst exponent of $H = 1/2$ and the obtained values through surrogate data suggest an error of the order of 0.1. Plain statistical confidence intervals, as usual, may grossly underestimate the error bars. Furthermore, the Gaussian surrogates which preserve the power spectrum yield results for the exponents which are quite similar for both parts of the time series. Thus, there could be a feature among the second part of the time series which is not captured by the autocorrelation function of the data. But in view of the aforementioned error estimates, the deviations could be a signature of the inaccuracy of the numerical values. The precise numerical value of the Hurst exponent might only be of limited significance for the quantitative description of dynamical behaviour in real systems with finite length of time series. However, it may allow to distinguish qualitatively between persistent and anti-persistent dynamical behaviour. Indeed, our analysis consistently predicts a anti-persistence, i.e. $H < 1/2$, for the Nord Pool spot market.

As we have seen from the discussion so far, estimators which describe the decay of correlations in a real world process, such as the Nordic electricity spot market can vary quite substantially. This may limit the accuracy and the interpretation of those results. One way to resolve this dilemma is to characterise certain trends of the Hurst exponent as some control parameters are changed, rather than estimating a single value. Since for the electricity spot market there are no controllable parameters, an 'educated' resampling of the given time series is a sensible way to identify trends in the correlation decay. In particular, we looked at the system price at a certain fixed hour of each day from 1 am to 12 pm. Such a resampling results in 24 different time series of smaller size. Fig. 4(a) shows the power spectrum for the system price at 1 am of each day where no distinct peaks can be identified. On the other hand, Fig. 4(b) shows the power spectrum for 8 am where distinct peaks can be seen which correspond to the weekly periodicity of the system price. Indeed, this suggests that the system price during night hours is not affected by the 7-day interval of our industrial society, whereas there are strong correlations during daily working hours. This behaviour is summarised in Fig. 4(c) where all 24 power spectra are shown in a three-dimensional representation.

The Hurst exponents estimated from these power spectra are shown in Fig. 5 where different methods for the computation of the exponents have been compared. While the different methods yield quite distinct numerical values all methods essentially produce dips at around 9 am and 6 pm indicating that at these times the correlations in the system price are strongly dominated by the 7-day interval imposed on the market. Hurst exponents estimated from the asymptotic decay of the correlation function (diamonds) are practically constant although we expect such a method to be the least reliable one. The R/S-method (triangles) and the MF-DFA-method (circles) give practically constant results for the Hurst exponents as well, although with a different value. These features might be attributed to the intrinsic averaging of the respective methods.

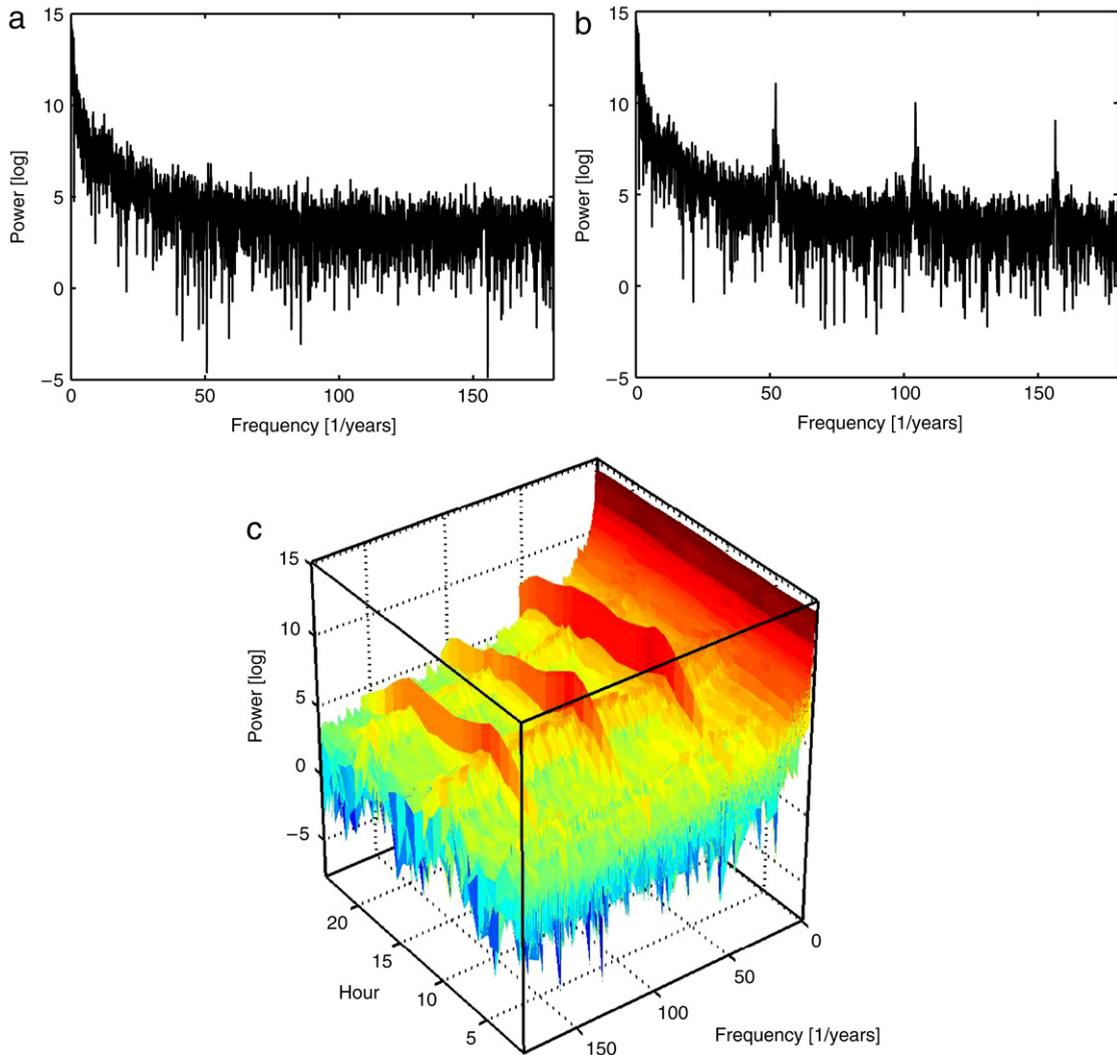


Fig. 4. Weekly periodicities in the system price along different times of the day. Panel (a): Spectrum of the system price at 1 am of each day, panel (b): spectrum at 8 am of each day, panel (c): three-dimensional representation of the power spectrum for each hour of the day.

Hurst exponents evaluated from the power spectrum (crosses) and those obtained by the DMA-method (squares) display clearly the daytime dependence. It should be noted that all methods give different results for the estimated Hurst exponent when applied to a real-world time series, whereas they give identical results when applied to an ideal self-affine process. Nevertheless, all methods show a dependence on daytime, although at different scales.

5. Conclusion

A perfectly self-affine process can be characterised completely by a single Hurst exponent. However, such a mathematical property is rarely shared by a real world time series. It is therefore sensible to apply more sophisticated data analysis tools, one of which is the generalised multifractal detrended fluctuation analysis, as introduced by Ref. [39]. Our analysis of the electricity system price of the Nordic spot market has shown considerable variations of the Hurst exponent, although the results are consistent with a mainly anti-persistent time series as shown by the traditional R/S-method applied to the original time series and to Gaussian linear surrogates with the same mean, variance, and power spectrum. Anti-persistence is preserved while a shuffled time series yields Hurst exponents close to $1/2$.

To illustrate the large fluctuation properties of the Nord Pool data more clearly we may compute a time-dependent Hurst exponent as well. To this end we resample the complete time series in overlapping time windows of different length (1000 h, 5000 h, and 10,000 h) and use the power spectrum to estimate the Hurst exponent of the respective time window. On the one hand this allows us to estimate the fluctuations of the Hurst exponent on different time scales. On the other hand this gives an estimation for the accuracy of the Hurst exponent when only a finite number of data points is available. Fig. 6

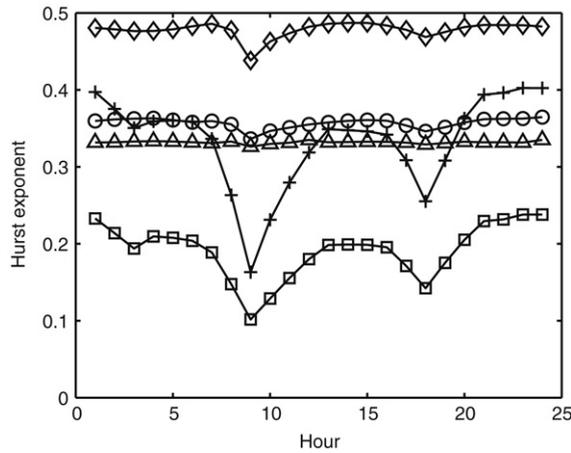


Fig. 5. Evolution of the Hurst exponent during the day estimated from the different spectra obtained from the time series and from the autocorrelation function. Crosses (+): power law scaling of the power spectrum, diamonds (◇): evaluation of the decay of the correlation function, triangles (Δ): R/S method, squares (□): DMA, and circles (○): MF-DFA with first-order interpolation of the time series.

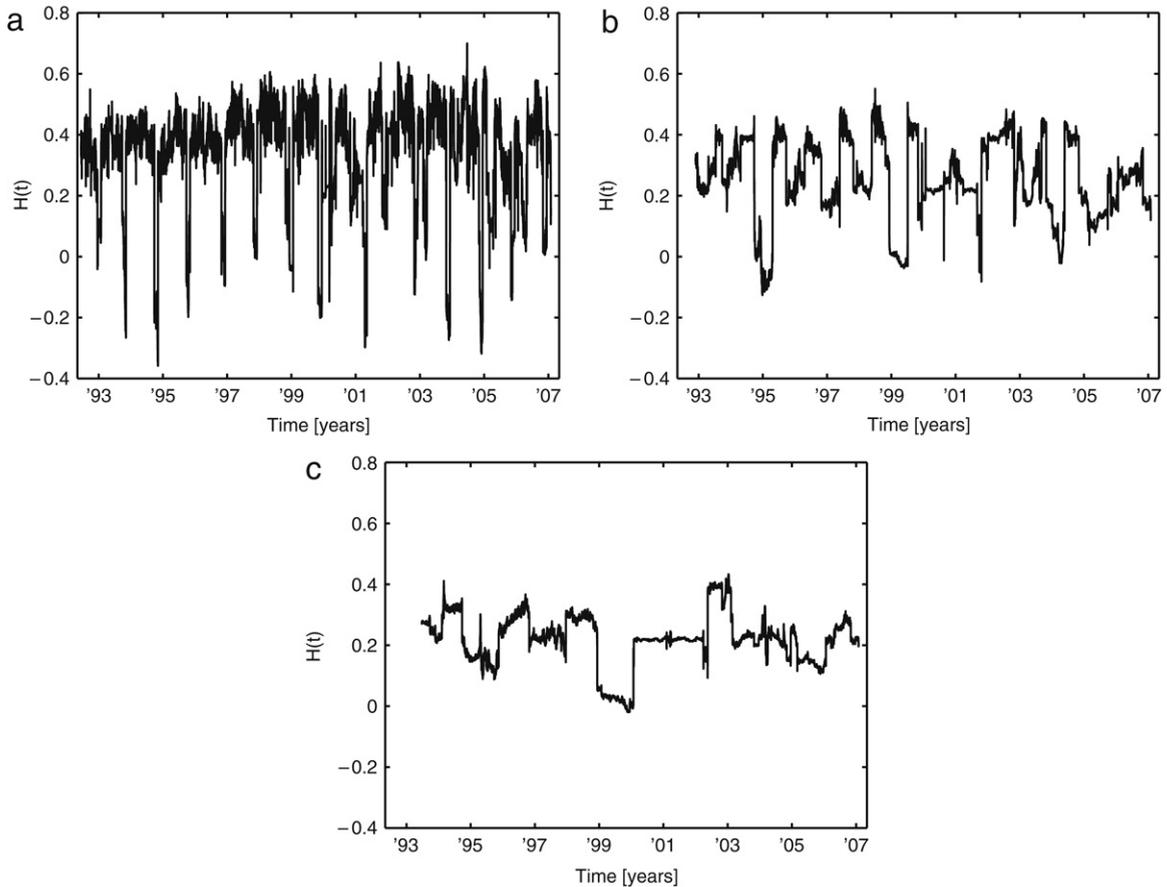


Fig. 6. Time fluctuations of the Hurst exponent for different sizes of a moving window obtained from the decay in the power spectrum. Window sizes are (a) 1000 h, (b) 5000 h, and (c) 10,000 h.

shows the results for the system price. Large fluctuations of the Hurst exponent can be seen when computed from windows of size 1000 h. Fluctuations become smaller as the length of the time window increases; however, even for time windows of 10,000 h fluctuations are still substantial. This indicates that the actual value of H may vary strongly, depending on the time for which it is estimated. Only time windows of more than 100,000 h give a practically constant Hurst exponent similar to

the value obtained in Section 3. In other words, for the estimation of the Hurst exponent or any other quantitative measure of the correlation decay the finite length of the time series may have important consequence on the outcome.

Some of the fluctuations of the Hurst exponents encountered in our analysis reflect rather obvious changes in the market. For instance, the daily variations described in Section 4 are a feature which could be detected as well by a straightforward correlation analysis. Thus, such variations are a signature of an incomplete suppression of non-stationary trends of the time series rather than a true modulation of the anti-persistence. However, such variations indicate that the error bars for the Hurst analysis are quite substantial and may amount up to 20% of the numerical value. In this context it is quite remarkable that the rescaled range method yields almost constant values so that such a straightforward approach could be conjectured to produce more accurate values than some of the more sophisticated techniques. However, one should bear in mind that a single exponent is not likely to capture the whole complexity of a real world dynamical process.

Acknowledgements

The authors gratefully acknowledge financial support through the European Commissions NEST programme by the project MANMADE (Contract No 043363). The authors are grateful to NESAs, in particular to Dr. H. Sivonen, for providing the data sets, to Fondazione Cariplo, and to W. Just for useful discussions.

References

- [1] R.N. Mantegna, H.E. Stanley, *An Introduction to Econophysics*, Cambridge Univ. Press, 2000.
- [2] N.F. Johnson, P. Jefferies, P. Ming Hui, *Financial Market Complexity*, Oxford Univ. Press, 2003.
- [3] M.F.M. Osborne, Brownian motion in the stock market, *Oper. Res.* 7 (1959) 145.
- [4] B. Malkiel, *A Random Walk Down Wall Street*, Norton, New York, 1990.
- [5] B.B. Mandelbrot, *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*, Springer, New York, 1998.
- [6] W.A. Brock, D.A. Hsieh, B. LeBaron, *Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Economic Evidence*, MIT Press, Massachusetts, MA, 1991.
- [7] D.A. Hsieh, Chaos and nonlinear dynamics: Application to financial markets, *J. Finance* 46 (1991) 1839.
- [8] H.W. Lorenz, *Nonlinear Dynamical Economics and Chaotic Motion*, Springer, New York, 1993.
- [9] E.E. Peters, *Chaos and Order in the Capital Markets: A New View of Cycles, Prices and Volatility*, 2nd edition, Wiley, New York, 1996.
- [10] R.N. Mantegna, H.E. Stanley, Scaling behaviour in the dynamics of an economic index, *Nature* 376 (1995) 46.
- [11] R.N. Mantegna, H.E. Stanley, Turbulence and financial markets, *Nature* 383 (1996) 587.
- [12] P. Bak, K. Chen, Self-organized criticality, *Sci. Amer.* 264 (1991) 26.
- [13] M.F. Shlesinger, G.M. Zaslavsky, J. Klafter, Strange kinetics, *Nature* 363 (1993) 31.
- [14] J. Scheinkman, B. LeBaron, Nonlinear dynamics and stock returns, *J. Business* 62 (1989) 311.
- [15] G. Papaioannou, A. Karytinos, Nonlinear time series analysis of the stock exchange: The case of an emerging market, *Internat. J. Bifur. Chaos* 5 (1995) 1557.
- [16] I. Andreadis, Self-criticality and stochasticity of an S&P 500 index time series, *Chaos, Solitons Fractals* 11 (2000) 1047.
- [17] R. Friederich, J. Peinke, C. Renner, How to quantify deterministic and random influences on the statistics of the foreign exchange market, *Phys. Rev. Lett.* 84 (2000) 5224.
- [18] A.S. Soofi, L. Cao, *Modelling and Forecasting Financial Data: Techniques of Nonlinear Dynamics*, Kluwer Academic Publishers, Norwell, 2002.
- [19] R. Weron, B. Przybyłowicz, Hurst analysis of electricity price dynamics, *Physica A* 283 (2000) 462.
- [20] I. Simonsen, Measuring anti-correlations in the Nordic electricity spot market by wavelets, *Physica A* 322 (2003) 597.
- [21] H.N.E. Byström, Extreme value theory and extremely large electricity price changes, *Int. Rev. Econ. Finance* 14 (2005) 41.
- [22] J. Perelló, M. Montero, L. Palatella, I. Simonsen, J. Masoliver, Entropy of the Nordic electricity market: Anomalous scaling, spikes, and mean-reversion, *J. Stat. Mech.* 2006 (2006) P11011.
- [23] R. Weron, M. Bierbrauer, S. Truck, Modelling electricity prices: Jump diffusion and regime switching, *Physica A* 336 (2004) 39.
- [24] I. Vehviläinen, T. Pyykkönen, Stochastic factor model for electricity spot price—the case of the Nordic market, *Energy Econom.* 27 (2005) 351.
- [25] I. Simonsen, Volatility of power markets, *Physica A* 355 (2005) 10.
- [26] R. Weron, Energy price risk management, *Physica A* 285 (2000) 127.
- [27] P. Norouzzadeh, W. Dullaert, B. Rahmani, Anti-correlation and multifractal features of Spain electricity spot market, *Physica A* 380 (2007) 333.
- [28] A. Serletis, A.A. Rosenberg, The Hurst exponent in energy futures prices, *Physica A* 380 (2007) 325.
- [29] J.P. Nolan, Numerical computation of stable densities and distribution functions, *Comm. Stat. Stoch. Mod.* 13 (1997) 759.
- [30] G. Radons, Anomalous transport in disordered dynamical systems, *Physica D* 187 (2004) 3.
- [31] F. Strozzi, J.M. Zaldívar, J.P. Zbilut, Application of nonlinear time series analysis techniques to high frequency currency exchange data, *Physica A* 312 (2002) 520.
- [32] A. Provenzale, L.A. Smith, R. Vio, G. Murante, Distinguishing between low-dimensional dynamics and randomness in measured time series, *Physica D* 58 (1992) 31.
- [33] R. Hegger, H. Kantz, T. Schreiber, Practical implementation of nonlinear time series methods: The TISEAN package, *Chaos* 9 (1999) 413.
- [34] T. Gneiting, M. Schlather, Stochastic models that separate fractal dimension and the Hurst effect, *SIAM Rev.* 46 (2004) 269.
- [35] H.E. Hurst, Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civ. Eng.* 116 (1951) 770.
- [36] B.B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman, New York, 1983.
- [37] E. Alessio, A. Carbone, G. Castellì, V. Frappietro, Second-order moving average and scaling of stochastic time series, *Eur. Phys. J. B* 27 (2002) 197.
- [38] A. Carbone, G. Castellì, H. Stanley, Time-dependent Hurst exponent in financial time series, *Physica A* 344 (2004) 267.
- [39] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, A.L. Goldberger, Mosaic organization of dna nucleotides, *Phys. Rev. E* 49 (1994) 1685.
- [40] A. Bunde, S. Havlin, J.W. Kantelhardt, T. Penzel, J.-H. Peter, K. Voigt, Correlated and uncorrelated regions in heart-rate fluctuations during sleep, *Phys. Rev. Lett.* 85 (2000) 3736.
- [41] M.J. Cannon, D.B. Percival, D.C. Caccia, G.M. Raymond, J.B. Basingthwaight, Evaluating scaled windowed variance methods for estimating the Hurst coefficient of time series, *Physica A* 241 (1997) 606.
- [42] J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, H.E. Stanley, Multifractal detrended fluctuation analysis of nonstationary time series, *Physica A* 316 (2000) 87.

ANNEX IV. *Correlation analysis between electricity spot prices and faults in the electricity grid in the Nordic region.* LIUC paper xxx (2009)

Correlation analysis between electricity spot prices and faults in the electricity grid of the Nordic region.

Fernanda Strozzi and José Manuel Zaldívar Comenges*

Cattaneo University-LIUC, * European Commission IHCP-JRC Ispra (VA) Italy

Analisi della correlazione tra il prezzo dell'energia ed i disturbi della rete elettrica della regione nordica.

Sommario

La regione nordica considerata in questo lavoro comprende Danimarca, Finlandia, Norvegia, Svezia. La finestra temporale dei dati va da gennaio 2000 a dicembre 2006. In questo lavoro viene analizzata la correlazione tra il prezzo dell'energia, i disturbi nella rete elettrica ed i consumi totali. Sono stati considerati i prezzi dell'energia ottenuti da medie mensili di dati ad alta frequenza e confrontati con il numero mensile di disturbi e di consumi totali nello stesso periodo di tempo. Il trattamento preliminare dei dati include l'eliminazione dei trend lineari e delle componenti cicliche. Dalle serie dei prezzi è stata ricavata la loro volatilità, e, similmente, la volatilità dei disturbi e dei consumi totali. Le domande a cui si vuole rispondere sono le seguenti. Il prezzo dell'energia è correlato al numero dei disturbi in modo da poter usare una serie temporale per anticipare l'andamento dell'altra e prevenire eventi avversi? Nel caso in cui la correlazione non sia evidente su tutto l'intervallo di tempo considerato, si possono identificare traslazioni o finestre temporali in cui la correlazione aumenti ed in modo che gli estremi di tali finestre corrispondano ad eventi documentati?

L'analisi dei dati viene fatta dapprima misurando la correlazione lineare e la relativa significatività applicando il t-test, successivamente calcolando la funzione di correlazione per vedere se, traslando una serie rispetto all'altra, la correlazione possa aumentare. Infine è stata applicata la Cross Recurrence Analysis (CRP) per evidenziare possibili finestre di correlazione lineare. La conclusione principale del lavoro è che esiste una correlazione tra la volatilità dei prezzi dell'energia e quella dei disturbi ed in particolare l'aumento della prima è seguita da un aumento della seconda. L'analisi CRP fornisce risultati molto interessanti individuando finestre di correlazione corrispondente ad eventi rilevanti noti, tuttavia questi risultati non sono sempre significativi dal punto di vista statistico e per confermarli servirebbe una quantità di dati maggiore di quelli disponibili per questo lavoro.

Correlation analysis between faults in the electricity grid and spot prices in the Nordic region.

Summary

In this work we have analyzed possible correlations between electricity prices and disturbances using the data of the Nordic electricity market. We have used the monthly spot prices, disturbances and consumption from the beginning of January 2000 until the end of December 2006 in the Nordic region, i.e. Denmark, Finland, Norway and Sweden. The preliminary treatment of the data include the elimination of the trends applying the difference operator and subtracting the regression line. In addition, we have considered the price volatility and similarly the volatility of disturbances and of total consumption. The questions we were interested in addressing were the following: Are the monthly spot prices correlated with disturbances? Can we increase the correlation by shifting the time series and can we use the evolution of one time series to anticipate the behaviour of the other and/or to prevent adverse events? Can we detect windows of correlation and find a correspondence of the starting and ending point with some know events? The main conclusion of this work is that a correlation between disturbances and prices exists. Using the Cross Correlation function we have found a strong correlation between the volatility of disturbances and detrended prices, but only on windows of six or twelve months. To analyse the information on a shorter period we have applied Cross Recurrence Plot (CRP) analysis and we have shown that the advent of external events are able to change the correlation properties of the time series, in this case the volatility of disturbances and of prices. However, CRP analysis would need more data points that the available at the moment. To improve the results it would be necessary to repeat the analysis using at least daily data of disturbances and consumption.

Correlation analysis between faults in the electricity grid and spot prices in the Nordic region

Fernanda Strozzi and José Manuel Zaldívar Comenges*

Cattaneo University-LIUC, * European Commission IHCP-JRC Ispra (VA) Italy

CONTENTS

1. INTRODUCTION
2. DATA PROVISION AND TREATMENT
 - 2.1 Data Provision
 - 2.1.1. Disturbances and Total Consumption
 - 2.1.2. Electricity Spot Prices
 - 2.2 Data treatment
 - 2.2.1. Data trend and seasonality
 - 2.2.2. Data first differences
 - 2.2.3. Data volatilities
 - 2.2.4. Time windows and shifts
3. TIME SERIES ANALYSIS
 - 3.1. Correlation matrix
 - 3.2. Principal component analysis
 - 3.3. Cross Correlation Function
 - 3.4. Cross Recurrence Plot
4. CONCLUSIONS

REFERENCES

APPENDIX 1: Correlation matrices and t-test

APPENDIX 2: Principal Component Analysis (PCA)

1. Introduction

The electricity market deregulation has caused stress in the market and in the electricity grids. In fact, the competition in the electricity market, together with its volatility, stressed the electricity grids with the variation of the flow in the physical network. The analysis of possible correlations between prices and disturbances is the goal of Task 5.3 of the MANMADE project and the subject of this report.

The correlations, once detected, can help in the prevention of the disturbances acting on the electricity price or, at least, in the management of the contingency. The forecasting of the disturbances will be the goal of another Task of the project.

There are a considerable number of studies in open literature on the properties of the electricity prices such as the high volatility (Simonsen, 2005, Strozzi *et al*, 2008), their long range correlation (Weron and Przybyłowicz, 2000; Simonsen, 2003; Bask *et al*. 2007). Erzgräber *et al*. 2008 checked the long range correlation using different methods to calculate Hurst exponent. Volatility measurement based on Recurrent Quantification Analysis were introduced by Strozzi *et al*. 2008, but, as far as we know, the relationship between the prices and grid disturbances has never been analysed in detail. Nevertheless it has been previously recognized that there should exist a relationship (Zhao, 2007).

In this work we have analysed possible correlations between electricity prices and disturbances using the data of the Nordic electricity market, which are publicly available. We have used the monthly spot prices, disturbances and consumption from the beginning of January 2000 until the end of December 2006 in the Nordic region, i.e. Denmark, Finland, Norway and Sweden. The preliminary treatment of the data include the elimination of the trends applying the difference operator and subtracting the regression line. In addition, we have considered the price volatility and similarly the volatility of disturbances and of total consumption. Starting from the initial three time series of prices, disturbances and consumption, we have obtained in this way a set of twelve time series. The questions we were interested in addressing were the following:

- Are the monthly spot prices, or one of its related time series, correlated with disturbances, or one of its related time series, on some time windows and with some shift?
- Do the twelve time series contain new information with respect to the three original ones?
- Can we increase the correlation by shifting the time series and can we use the evolution of one time series to anticipate the behaviour of the other and/or to prevent adverse events?

- Since some external events could change existing correlations or create new ones, can we detect windows of correlation and find a correspondence of the starting and ending point with some historical know events?

To answer the above mentioned questions we have proposed the following methodology. In Section 2 we have described the preliminary data treatment to generate the twelve time series. As it is well know, the correlation may change if we observe it on different data windows. For this reason, we have grouped the twelve time series considering their mean (or the standard deviation in the case of the volatilities) on different time windows overlapped or not. In Section 3 we have studied the correlation matrices of the twelve time series. The main correlations between prices and disturbances in correspondence of some time windows are underlined. The analysis of the correlation matrices becomes deeper when we calculate their eigenvectors i.e. the principal components (Jolliffe, 1996) that allow identifying how many degrees of freedom, i.e. independent variables, may have a possible model of the twelve time series. After, we have checked if we can increase the correlation, shifting one time series in respect to the other, i.e. we have calculated the Cross Correlation Function (CCF), the correlation in respect to a shift (Orfanidis, 1996). As Marwan *et al.* (2007) pointed out, the concept of CCF can be extended using Cross Recurrence Plot (CRP), which is a tool that, by measuring the recurrence of two time series can calculate the Line Of Synchronization (LOS), and detect if, even a portion of the two time series, is linearly correlated with a portion of the other and which translation is necessary. Finally, in Section 4 the main results and the conclusions are presented.

2. Data provision and treatment

2.1 Data Provision

The data sets considered are related with the Electricity grid and market in the Nordic region (see Fig. 1) to detect possible correlations between disturbances and electricity prices. The data sets are monthly disturbances, the monthly total consumption and the monthly Electricity price in Denmark, Finland, Norway and Sweden from January 2000 and December 2006. All the data are public. The disturbances and Total Consumption are available on Nordic statistics of electricity faults in the Nordel web page: <http://www.nordel.org/content/Default.asp?PageID=214>. The Electricity spot prices are available on the Nord Pool (Nordic Power Exchange) web page: <http://www.nordpool.com/nordpool/financial/index.html>.



Figure 1. Transmission grid in Nordic countries (<http://www.nordel.org>).

Nordel is the collaboration organisation of the Transmission System Operators (TSOs) of Denmark, Finland, Iceland, Norway and Sweden. The core duty of the TSOs includes (<http://www.nordel.org>):

- Ensuring the operational security of the power system.
- Maintaining the instantaneous balance between supply and demand.
- Ensuring and maintaining the short-term and long-term adequacy of the transmission system.
- Enhancing the efficient functioning of the electricity market.

Nordel's objectives are (<http://www.nordel.org>):

- Development of an adequate and robust transmission system aiming at few large price areas.
- Seamless cooperation in the management of the daily system operations to maintain the security of supply and to use the resources efficiently across the borders.
- Efficient functioning of the North-West European electricity market with the aim to create larger and more liquid markets and to improve transparency of the TSO operations
- establishment of a benchmark for European transparency of the TSO information.

The market participants can benefit from a common Nordic wholesale electricity market consisting of a day-ahead market, intra-day market and regulating power market. In these markets power can be traded 24 hours a day throughout the year.

Nordel's co-operation in market facilitation aims to create and prepare for an efficient Nordic wholesale electricity market by balance settlement, congestion management, market coupling and monitoring

The main risk factors in the Nordic power and energy balances are (<http://www.nordel.org>):

- temperature
- availability of the Nordic power plants
- precipitation
- transmission capacities

The ownership of Nord Pool, the Nordic Power Exchange, is shared by the Nordic transmission system operators (TSOs) and Nord Pool ASA. Nord Pool ASA - The Nordic Power Exchange - is the world's only multinational exchange for trading electric power. The Nord Pool Group is headquartered in Oslo, Norway with offices in Sweden, Finland, Denmark, the Netherlands and Germany. The vision of a truly pan-Nordic power exchange was realised when Eastern Denmark was fully integrated into the Nordic market 1st October 2000, and all the Nordic nations operate in a joint market. Western Denmark was integrated into the Nordic Power in 1st

July 1999. Sweden and Norway became a single power exchange area in 1996. Finland joined the Nordic Power Exchange area in 1998.

Electric production differs considerably among the Nordic countries. In Norway, nearly all electricity is generated from hydropower. Sweden and Finland use a combination of hydropower, nuclear power, and conventional thermal power. Hydropower stations are located mainly in northern areas, whereas thermal power prevails in the south. Denmark relies mainly on conventional thermal power, but wind power is providing an increasing part of the demand for energy.

The power exchange Nord Pool Spot, organizes the physical trade of electricity, the day-ahead market *Elspot* in the Nordic countries and KONTEK in Germany (the TSO area of Vattenfall Europe Transmission GmbH). Nord Pool Spot is a part of the Nord Pool Group and is owned 20% by Nord Pool ASA and the Nordic Transmission System Operators: Statnett SF, Svenska Kraftnät, Fingrid Oyj and Energinet.dk own 20 % each.

Nord Pool Spot provides a market place to producers, distributors, industrial companies, energy companies, trading representatives, large consumers and TSOs on which they can buy or sell physical power.

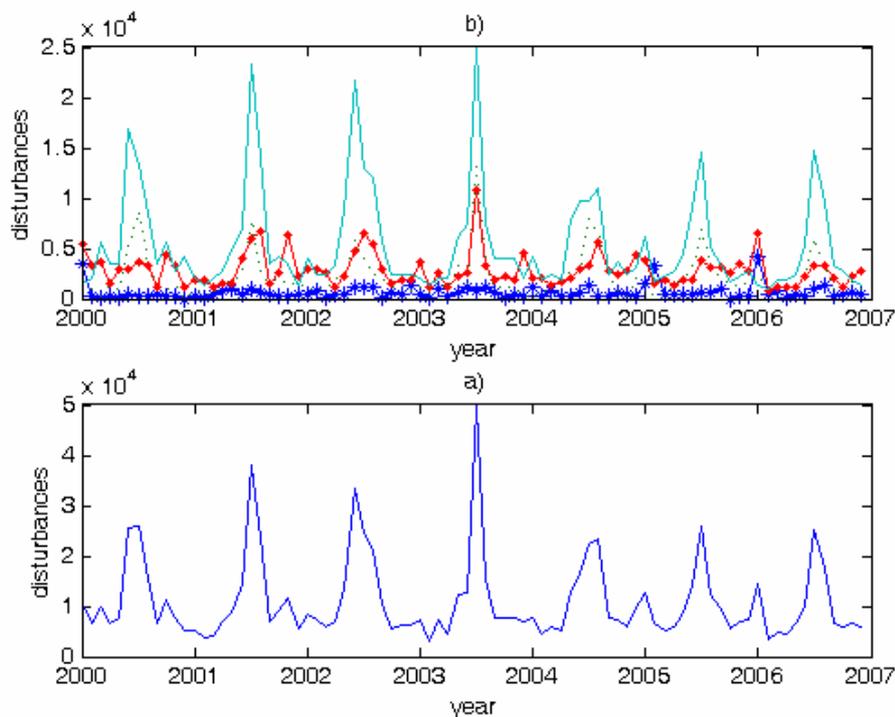


Figure 2. a) Number of disturbances in Denmark(*), Finland(·), Norway(-.-) and Sweden(-). b) Total number of disturbances in Denmark Finland Norway and Sweden.

2.1.1. Disturbances and Total Consumption

A disturbance may consist of a single fault but it can also contain many faults, consisting for example of an initial fault followed by some secondary faults. A disturbance is defined in

Nordel net reports (<http://www.nordel.org>) by an “*outage, forced or unintended disconnection or failed reconnection as a result of faults in the power grid*”.

In Figure 2 the number of grid disturbances from the beginning of 2000 until the end of 2006, according to months are represented. The grid considered is the 100-400kV network (Fig. 1).

For all the countries the number of disturbances is usually greatest during summer period. This is caused by lightning. Apart from lightening, the other causes of grid disturbances are other natural phenomena, operation, maintenance and faults in technical equipment.

In Figure 3 the monthly Total Consumption (TC) in Nordic region except Iceland between January 2000 until December 2006 is represented.

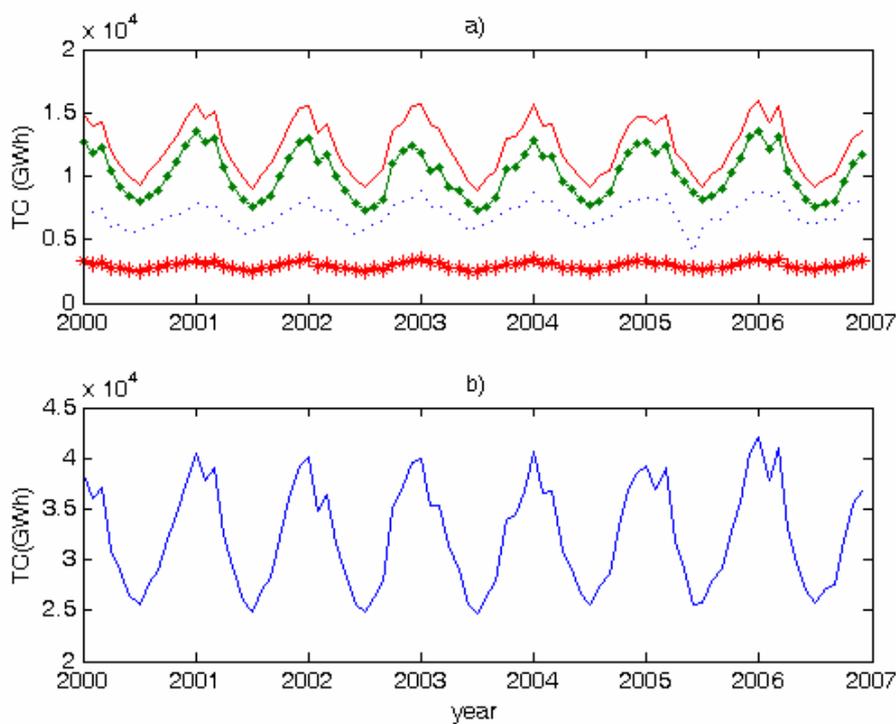


Figure 3. a) Total Consumption of Denmark(*), Finland(:), Norway(-) and Sweden(-). b) Total Consumption in Nordel States except Iceland.

We did not consider the disturbances and Total consumption of Iceland because it is not a member of Nord Pool and we were interested in analyzing the correlation between prices and disturbances.

2.1.2. Electricity Spot Prices

In Figure 4 the Electricity spot prices of Nordic Region (Nord Pool countries) considered are plotted. They are hourly data from the Nord Pool system spot prices. The series lasts from 1st January 1999 until 26th January 2007 and comprises 70,752 data points (Figure 4), the prices are expressed in EUR/MWh.

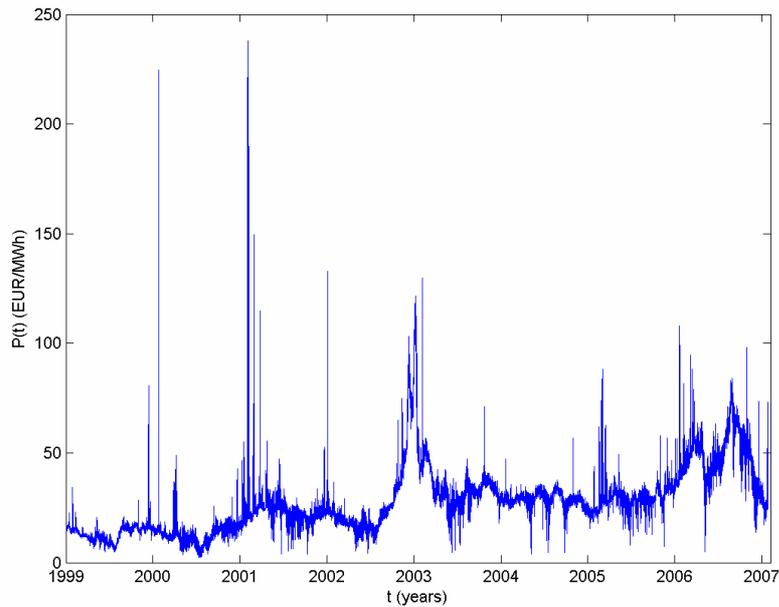


Figure 4. Spot prices in the Nordic electricity market (Nord Pool) from January 1997 to January 2007.

In Figure 4 the system price is represented. The system price is also denoted "the unconstrained market clearing price", because it is the price that balances sale and purchase in the exchange area while not considering any transmission constraints.

In Table 1 it is possible to observe the evolution of the composition of Nord Pool during these years.

Table 1. Nord Pool participating countries and dates of entry.

Countries	Date of entry of new country (dd/mm/yy)
Norway	1/1/93
Norway and Sweden	1/1/96
Norway, Sweden and Finland	29/12/97
Norway, Sweden, Finland and western Denmark	1/7/99
Norway, Sweden, Finland, western and eastern Denmark	1/10/00
KONTEK (Germany)	5/10/05

As already mentioned, to have a comparable data set, we have excluded Iceland from the disturbance and total consumption data. In addition, to be consistent with the data frequency of the other time series – disturbances and total consumption-, we will consider only monthly mean spot prices, see Figure 5.

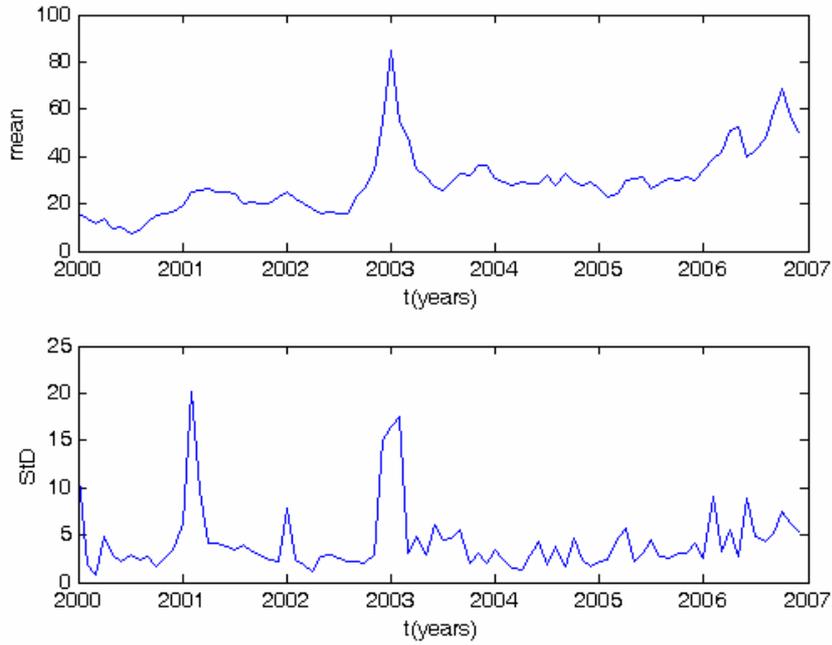


Figure 5. Monthly mean spot prices obtained from Fig. 4.

2.2. Data treatment

2.2. 1. Data trend and seasonality

We have treated the data (monthly mean Spot Price, Total Consumption and Disturbances) subtracting the linear trend and the seasonality. The trend is calculated using the linear regression line and the seasonality is removed subtracting the mean value of the given time series on the correspondent month of every year. In Figure 6 the trends and the seasonality subtracted are represented and, in Figure 7, the resulting time series are plotted.

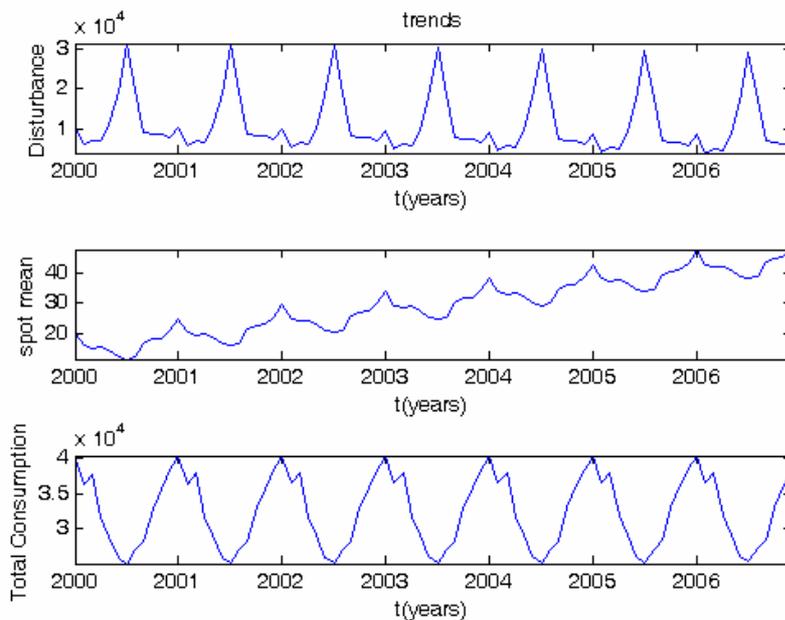


Figure 6. Subtracted trends and seasonality.

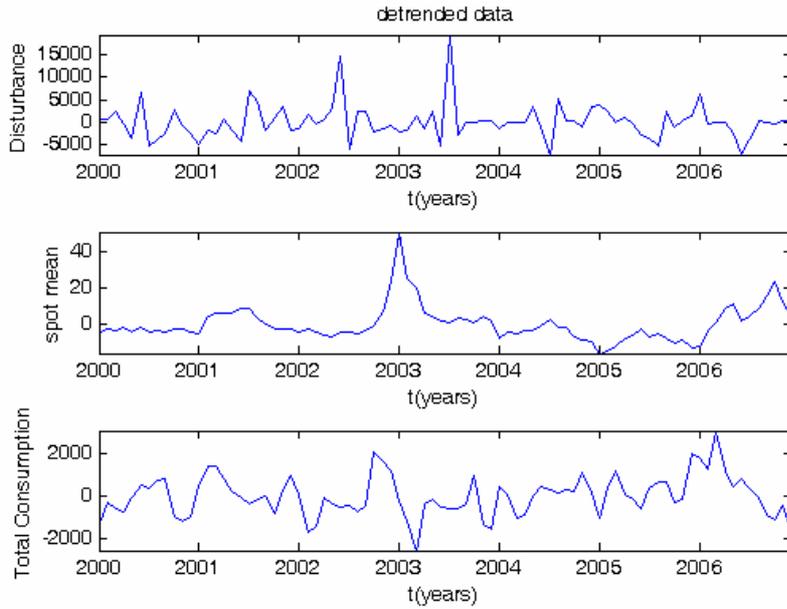


Figure 7. Detrended data sets.

2.2.2. Data first differences

The difference operator is often applied to eliminate the trend. In Figure 8 we have represented the first differences of the monthly mean Spot prices, Total Consumption and Disturbances.

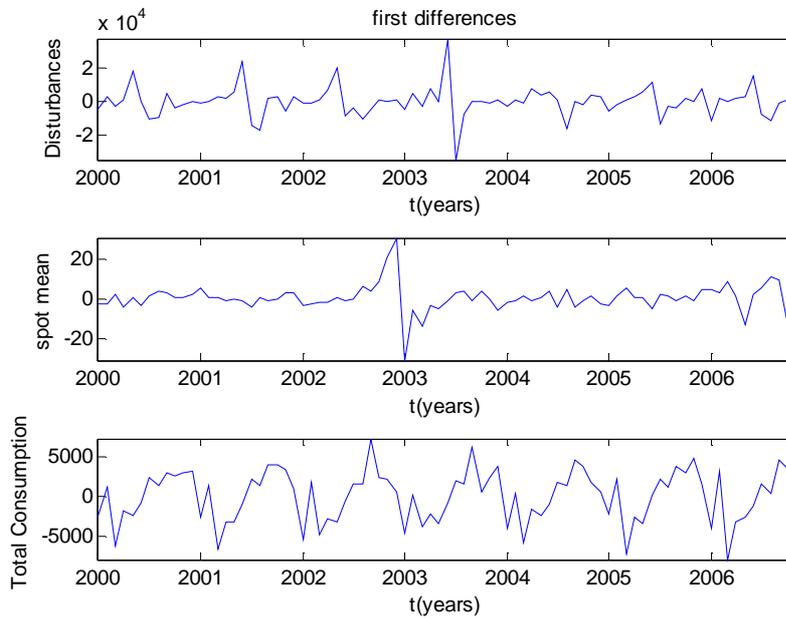


Figure 8. First differences time series.

2.2.3. Data Volatilities

Prices volatility in the Nordic electricity market, using the same data, it was analysed by Strozzi *et al.*, 2008. Volatility can be calculated using standard deviation SD, defined as

$SD(X) = \frac{1}{n-1} \sum_{t=1}^n (X_t - \bar{X})^{1/2}$, of the spot price returns (S), Total Consumption (T) and

Disturbances (D):

$$V_S = SD\left(\frac{S(t) - S(t - \Delta t)}{S(t - \Delta t)}\right) \quad (1)$$

$$V_D = SD\left(\frac{D(t) - D(t - \Delta t)}{D(t - \Delta t)}\right) \quad (2)$$

$$V_T = SD\left(\frac{T(t) - T(t - \Delta t)}{T(t - \Delta t)}\right) \quad (3)$$

Where Δt is 1 month and the window on which we will calculate the Standard Deviation will change from 1 month to 12 months. In the case of 1 month window we cannot calculate standard deviation because we have only one point then in this case we will consider simply:

$$V_S = \frac{S(t) - S(t - \Delta t)}{S(t - \Delta t)} \quad (4)$$

$$V_D = \frac{D(t) - D(t - \Delta t)}{D(t - \Delta t)} \quad (5)$$

$$V_T = \frac{T(t) - T(t - \Delta t)}{T(t - \Delta t)} \quad (6)$$

which are linear approximations respectively of: $\ln(S(t)/S(t-\Delta t))$, $\ln(D(t)/D(t-\Delta t))$ and $\ln(T(t)/T(t-\Delta t))$ i.e. the logarithm first differences.

In Figure 9 the three volatilities are represented on a window (w) of 2 months translated by a shift (sh) of 1 month.

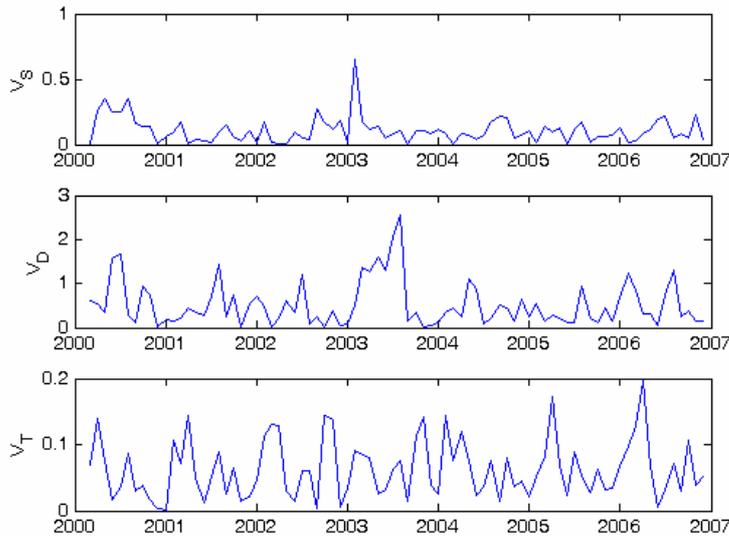


Figure 9. Volatilities of the data sets considered ($\Delta t=1$ month), $w=2$ months, $sh=1$ month.

The twelve time series considered in the rest of this work will be labelled in according with Table 2.

Table 2. Notation of the twelve time series analyzed. *Diff*= first differences

<i>Label</i>	<i>Definition</i>	<i>Label</i>	<i>Definition</i>	<i>Label</i>	<i>Definition</i>	<i>Label</i>	<i>Definition</i>
S	Mean monthly spot prices	Sdt	S detrended	Sfd	S first differences	V_S	w=1 diff(ln(S)) w>1 Volatility of S
D	Monthly disturbances	Ddt	D detrended	Dfd	D first differences	V_D	w=1 diff(ln(D)) w>1 Volatility of D
T	Monthly Total Consumption	Tdt	T detrended	Tfd	T first differences	V_T	w=1 diff(ln(T)) w>1 Volatility of T

2.2.4. Time windows and shifts

We will analyse possible correlations between mean Spot prices, Disturbances and Total consumptions considering real data, de-trended data, first differences and volatilities of the three time series. These correlations will be checked for different time windows and for different time shifts as it is presented in Table 3. The reason for this choice is based on the natural periodicity inside one year (seasonality = 3 months, semester periodicity=6 months and the annual = 12 months). The only exception is the window of two months. This choice is explained by the need to have the maximum number of points in order to apply Cross Recurrence Plot, in fact 2 is the minimum window to calculate standard deviation. (SD)

Table 3. Windows and shifts considered in this work.

w	1	3	6	12	2	3	6	12
sh	1	3	6	12	1	1	1	1

3. Time Series Analysis

First, we have studied the linear correlation coefficients i.e. the entries of the correlation matrix together with its eigenvectors (PCA). Then we have checked if these linear correlation coefficients could increase by shifting one series with respect to the other measuring the Cross Correlation Function (CCF). Finally, we have applied the Cross Recurrence Plot (CRP) analysis, which provides a tool: Line Of Synchronization (LOS) that allows to identify time windows in which two time series are linearly correlated and it represents an extension of the linear Cross Correlation Function.

3.1. Correlation matrix

The correlation coefficient matrix represents the normalized measure of the strength of linear relationship between variables. The correlation coefficient R of two variables X and Y is given by:

$$R(X,Y) = \frac{COV(X,Y)}{\sqrt{VAR(X)VAR(Y)}} \quad (7)$$

where the $COV(X,Y)$ is the covariance matrix, i.e.

$$COV(X,Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n-1} \quad (8)$$

where \bar{X} and \bar{Y} are the mean of the two variables and n is the components number.

The correlation coefficients range from -1 to 1, where values close to 1 suggest that there is a positive linear relationship between the data columns, values close to -1 suggest that one column of data has a negative linear relationship to another column of data (anticorrelation). Values close or equal to 0 suggest there is no linear relationship between the data columns.

We have applied the MATLAB[®] function *corrcoef* that produces the matrix of correlation coefficients for all the time series of Table 2 and for each window and shift indicated in Table 3. In Table 4 the entries of correlation matrix R for $w=1$ and $sh=1$ are presented, the rest, obtained with the other values of w and sh , are in the Appendix 1. To measure the significance of each correlation we have applied the t-test. The resulting P matrix for $w=1$ and $sh=1$ is presented in Table 5, and the rest can be seen in Appendix 1. In every correlation matrix R we have considered the correlation values $R(i,j)$ higher than 0.7071 (i.e. a determination coefficient $R^2 > 0.5$) with a significance level of 95% i.e. $P(i,j) < 0.05$. Such values are highlighted in yellow in Tables 4 and 5. Each $P(i,j)$ value gives the probability of getting a correlation as large as the observed value by random chance, when the true correlation is zero.

Table 4. Correlation matrix R. $w=1$ month, $sh=1$ month. Yellow if $|R(i,j)|>0.7071$ ($R(i,j)^2>0.5$)

	S	D	T	S_{dt}	D_{dt}	T_{dt}	S_{fd}	D_{fd}	T_{fd}	V_S	V_D	V_T
S	1	-0.2439	0.2014	0.7317	-0.1072	-0.0308	0.2651	-0.0307	0.0322	0.2568	-0.0484	0.0423
D		1	-0.6270	-0.0617	0.4828	-0.0885	-0.0491	0.5390	-0.0626	-0.1447	0.6106	-0.1054
T			1	-0.0227	-0.0311	0.2162	0.1139	-0.0941	0.3110	0.1349	-0.2259	0.3114
S_{dt}				1	-0.1259	-0.1649	0.2908	0.0018	-0.0833	0.2960	0.0267	-0.0747
D_{dt}					1	-0.1534	-0.0086	0.5122	-0.0238	-0.0396	0.5334	-0.0271
T_{dt}						1	0.2436	-0.0297	0.1539	0.2736	-0.0567	0.1535
S_{fd}							1	-0.0849	0.1476	0.8607	-0.0510	0.1495
D_{fd}								1	-0.1962	-0.1692	0.8761	-0.2584
T_{fd}									1	0.1485	-0.1922	0.9896
V_S										1	-0.1373	0.1633
V_D											1	-0.2568
V_T												1

Table5. t-test matrix P. $w=1$ $sh=1$.

	<i>S</i>	<i>D</i>	<i>T</i>	<i>S_dt</i>	<i>D_dt</i>	<i>T_dt</i>	<i>S_fd</i>	<i>D_fd</i>	<i>T_fd</i>	<i>V_S</i>	<i>V_D</i>	<i>V_T</i>
<i>S</i>	1	0.0263	0.0679	<u>0.0000</u>	0.3347	0.7820	0.0154	0.7826	0.7728	0.0191	0.6642	0.7039
<i>D</i>		1	0.0000	0.5798	0.0000	0.4265	0.6595	0.0000	0.5740	0.1917	0.0000	0.3429
<i>T</i>			1	0.8386	0.7802	0.0497	0.3052	0.3977	0.0042	0.2239	0.0400	0.0042
<i>S_dt</i>				1	0.2567	0.1363	0.0077	0.9868	0.4538	0.0066	0.8104	0.5019
<i>D_dt</i>					1	0.1662	0.9385	0.0000	0.8307	0.7223	0.0000	0.8077
<i>T_dt</i>						1	0.0265	0.7899	0.1649	0.0123	0.6106	0.1658
<i>S_fd</i>							1	0.4451	0.1831	<u>0.0000</u>	0.6473	0.1773
<i>D_fd</i>								1	0.0754	0.1262	<u>0.0000</u>	0.0183
<i>T_fd</i>									1	0.1802	0.1817	<u>0.0000</u>
<i>V_S</i>										1	0.2157	0.1402
<i>V_D</i>											1	0.0191
<i>V_T</i>												1

The significant correlations for all the time windows, w , and all the shifts, sh , are presented in Tables 6 and 7. The correlation between different time series are highlighted in blue.

Table 6. Significant linear correlations coefficient $R(i,j)$ between data sets for different when w equal to sh .

w=1, sh=1	w=3 (seasonal); sh=3	w=6; sh=6	w=12; sh=12
S,Sdt (0.7317)	D,T (-0.8154)	Dfd,D(-0.8503)	Tdt,T(0.9842)
Sfd,Vs(0.8607)		Tfd,T(-0.8686)	V_D,T(-0.9057)
Dfd,V _D (0.8761)		V _D ,Dfd(0.7698)	V_D,Sdt(0.8138)
Tfd,V _T (0.9896)		D,T(-0.8594)	V_D,Tdt(-0.9014)
		D, Tfd(0.776)	
		T,Dfd(0.7752)	

Table 7. Significant linear correlations coefficient $R(i,j)$ between data sets for different w and $sh=1$.

w=2; sh=1	w=3 (seasonal); sh=1	w=6; sh=1	w=12; sh=1
T,D (-0.7354)	T,D (-0.8057)	T,D(-0.9044)	T,D(-0.7807)
S, Sdt(0.7195)		Tfd,Dfd(-0.8010)	D,Tdt(-0.7586)
			D,Ddt(0.8060)
			T,Tdt(0.9904)
			V_D-Sdt (0.7567)

The main findings of this analysis are:

- $w=1, sh=1$. There are only expected correlations between Spot prices (S) and their first differences (Sdt), the first differences of spot prices (Sfd), Disturbances (Dfd) and Total Consumptions (Tfd) and their logarithms (V_s, V_D, V_T).
- $w=2$ and $w=3$. A strong correlation (higher than 0.7) appears between Total Consumption (T) and Disturbances (D).
- $w=6$. A strong correlation between T and D is still preserved both for $sh=1$ and $sh=6$. Moreover, for $sh=6$ a correlation between their first differences appears, but for $sh=1$ the relation is between D and T respectively and the first differences of T and D .
- $w=12$. the relationships between T and D are confirmed even for $sh=12$ and for $sh=1$. For both sh values volatility of disturbances starts to be correlated with the Spot prices and Total Consumption de-trended.

Since we are interested mostly in the correlations between price and disturbances we can conclude that it exists only for $w=12$ and $sh=12$ or $sh=1$, particularly between the volatility of disturbances and the mean Spot prices de-trended.

3.2. Principal Component Analysis (PCA)

Principal component analysis (PCA) is a technique used to reduce multidimensional data sets (Jackson, 1991, Jolliffe, 2002). It is a way to identify patterns (linear) in data and then to compress them by reducing the number of dimensions without much loss of information.

The main steps in the case of m time series X^1, X^2, \dots, X^m of length n are the followings. The mean is subtracted to each time series in order to have data sets with mean 0:

$$\begin{bmatrix} X_1^1 - \overline{X^1} & X_1^2 - \overline{X^2} & \dots & X_1^m - \overline{X^m} \\ \dots & \dots & \dots & \dots \\ X_n^1 - \overline{X^1} & X_n^2 - \overline{X^2} & \dots & X_n^m - \overline{X^m} \end{bmatrix} \quad (9)$$

where $X^i = (X_1^i, \dots, X_n^i)$ and $\overline{X^i}$ is the mean value of X^i . Then the covariance matrix it is calculated as follows:

$$COV = (c_{ij}), i = 1, \dots, m, j = 1, \dots, m \quad (10)$$

where:

$$c_{ij} = \frac{\sum_{t=1}^n (X_t^i - \overline{X^i})(X_t^j - \overline{X^j})}{n-1} \quad (11)$$

The eigenvalues and eigenvectors of COV are calculated too. With the eigenvectors of the covariance matrix it is possible to extract lines that characterize the data. The eigenvector with the highest eigenvalue is the *principal component* of the data set. The columns of the eigenvector matrix and eigenvalue matrix are sorted in order of *decreasing* eigenvalues. A subset of the eigenvectors is selected as basis vectors: the more significant and the others are cancelled. Usually those eigenvalues which sum is 90% of the sum of all eigenvalues are considered. The original data matrix without means is represented in the new basis.

The first principal component is that linear combination of the original variables which accounts for the maximum amount of variance in a single line. It is the line of best fit through the data,

and the residual variance about this line is then a minimum for the data set. The second principal component is that line which is orthogonal to the first principal component and accounts for the maximum amount of the remaining variance in the data. The first two components therefore represent the plane of best fit through the data. All remaining principal components are defined similarly, such that the lowest order components normally account for very little variance and can usually be ignored. The eigenvalues obtained from Principal Components Analysis are equal to the variance explained by each of the principal components, in decreasing order of importance. The eigenvectors are weightings with *loadings* that, when applied to the original data, obtain principal component scores for the observations. A large positive or negative value indicates a variable that is correlated, either in a positive or a negative way, with the component. The function *princomp* of MATLAB[®] is applied to calculate loadings, the eigenvalues and then the percentage of variance explained by each component when *w* and *sh* vary in accordance with Table 3.

Table 8. Summary of PCA results.

<i>w</i>	<i>sh</i>	# points	#PC to explain at least 50% variance	% variance explained	#PC to explain at least 90% variance	% variance explained
1	1	83	3	63.39	6	91.03
2	1	82	3	53.08	8	91.08
3	1	81	3	56.35	8	92.67
6	1	78	3	62.62	7	92.85
12	1	72	2	54.80	6	93.26
3	3	27	3	56.55	7	91.69
6	6	13	2	61.44	5	94.38
12	12	6	2	65.63	4	96.75

The main results of this analysis are presented in Table 8 and in Appendix 2. In the first two columns there are the values of *w* and *sh* and, in the third, the number of points of each time series considered in calculating PCA. In the fourth column the number of principal components able to explain at least the 50% of variance is listed. It seems that an hyper plane of dimension three can fit the data. This is not so strange since we built the twelve time series starting from three of them (S, D, T), but if we are interested in explaining at least 90% of variance we can see that we need always more than 3 principal components. Sometimes even 8 principal components are necessary i.e. the original time series and their first difference, for example, do not contain still all the independent information.

3.3. Cross Correlation Function (CCF)

Cross correlation is a generalization of the correlation coefficient and a standard method of estimating the degree to which two series are correlated when we shift them one in respect to the others (Orfanidis, 1996). Let we consider two time series X_i and Y_i where $i=1,1,2\dots n$. The cross correlation R at delay d is defined as

$$R(X, Y, d) = \frac{\sum_i [(X_i - \bar{X})(Y_{i-d} - \bar{Y})]}{\sqrt{\sum_i (X_i - \bar{X})^2} \sqrt{\sum_i (Y_i - \bar{Y})^2}} \quad (12)$$

Where \bar{X} and \bar{Y} are the means of the corresponding series. If Eq. 12 is computed for all delays $d = -(n-1), \dots, 0, 1, 2, \dots, (n-1)$ then it results in a cross correlation series of twice the length of the original series. For $d=0$ it becomes the linear correlation coefficient $R(X, Y)$.

We have calculated the cross correlation function for every window, w , and every shift, sh , of Table 3 using the MATLAB[®] function *xcov*. In Figure 10 we have plotted only CCFs obtained using disturbances, spot prices or their modification and for which the maximum of correlation function reach a value of at least 0.5.

In Figure 10 it is possible to observe that there are significant correlations between Spot Prices and Disturbances (or their modifications) only on windows of 6 or 12 months. The maximum values obtained are listed in Table 9 together with the correlation coefficients without delay, $R(0)$, and the p values of the t-tests.

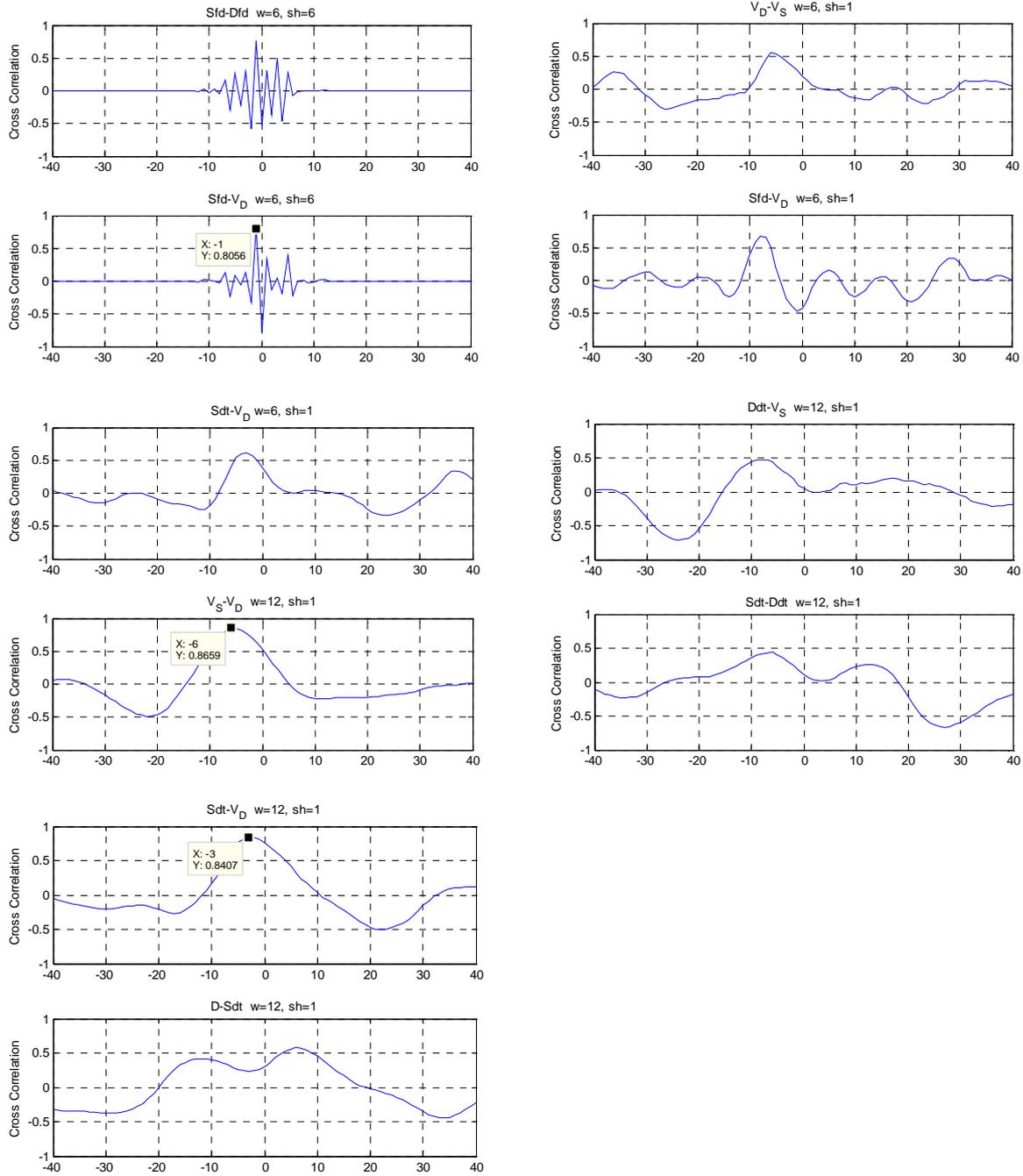


Figure 10. CCFs as a function of the delay.

Table 9. Results from the cross correlation analysis.

<i>Time series</i>	<i>w</i>	<i>sh</i>	<i>R</i> (0)	<i>p</i>	<i>delay</i> (months)	<i>R</i> (<i>delay</i>)	<i>p</i>
$V_S V_D$	12	1	0.5183	0.0000	-6	0.8906	0.0000
$V_S V_D$	6	1	0.1855	0.1040	-6	0.5959	0.0000
$Sdt V_D$	12	1	0.7567	0.0000	-3	0.8536	0.0000
$Sfd V_D$	6	1	-0.4273	0.0001	-8	0.7430	0.0000
$Sfd V_D$	6	6	-0.7778	0.0017	-1	0.8725	0.0002

V_D is correlated with price volatility, price first difference and price de-trended but only considering windows of six or twelve months. It is not so correlated with the price itself. In Figure 11 we have plotted the correlation function, using $w=2$ and $sh=1$, between D-T, D-Tfd, and, in both cases, the correlation becomes higher than 0.6. Moreover, one can observe the regularity of the damping of the correlation function, which is even more important than the correlation value itself, because it detects a similarity in the dynamic and not only in the static properties. For these reasons, we have also plotted the correlation function between D-Sfd and D-S which, that, even if it never reaches R-values higher than 0.4, it has a regular oscillating behaviour in respect to the delay.

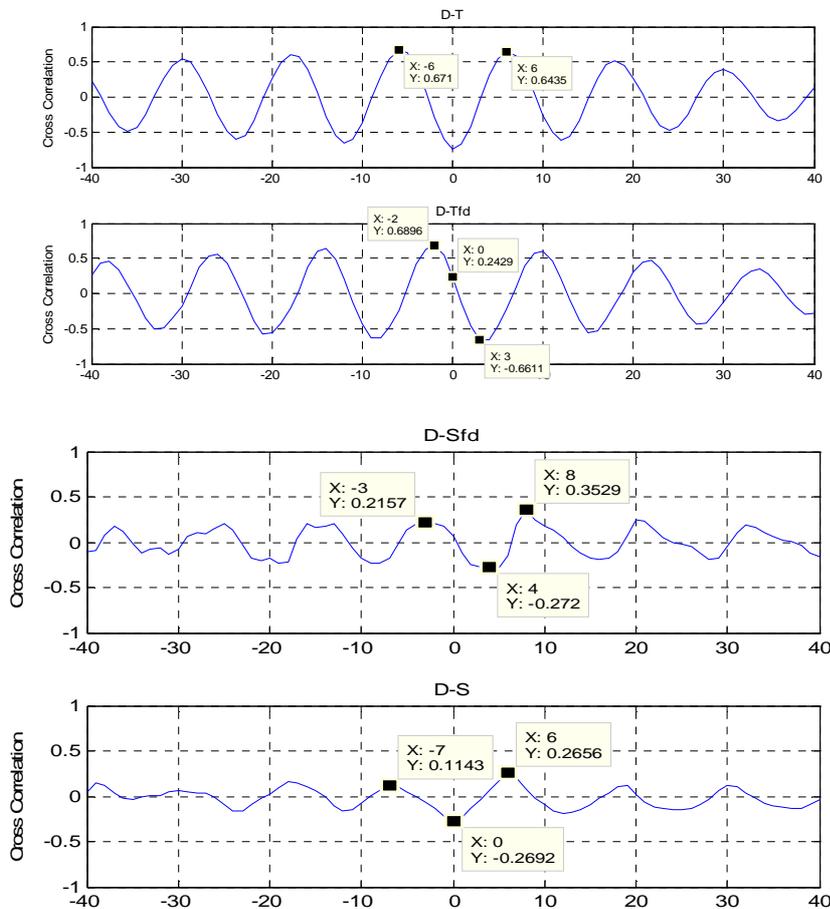


Figure 11. Cross Correlation functions for Disturbances with $w = 2$, $sh = 1$.

3.4. Cross Recurrence Plot (CRP)

CRP is a bivariate extension of RP and was introduced to analyse the dependencies between two different time series by comparing their joint recurrence (Marwan and Kurths, 2002). It can be considered as a generalization of the linear cross-correlation function (Marwan *et al.* 2007).

If we have two dynamical systems, see Fig. 12, which trajectories are respectively \vec{x}_i, \vec{y}_j where $i=1, \dots, n, j=1, \dots, m$, the CRP matrix is defined by:

$$\mathbf{CR}_{i,j}^{\overline{x_i}, \overline{y_j}}(\varepsilon) = \Theta(\varepsilon - \|\overline{x_i} - \overline{y_j}\|) \quad (13)$$

where $i=1, \dots, n; j=1, \dots, m$.

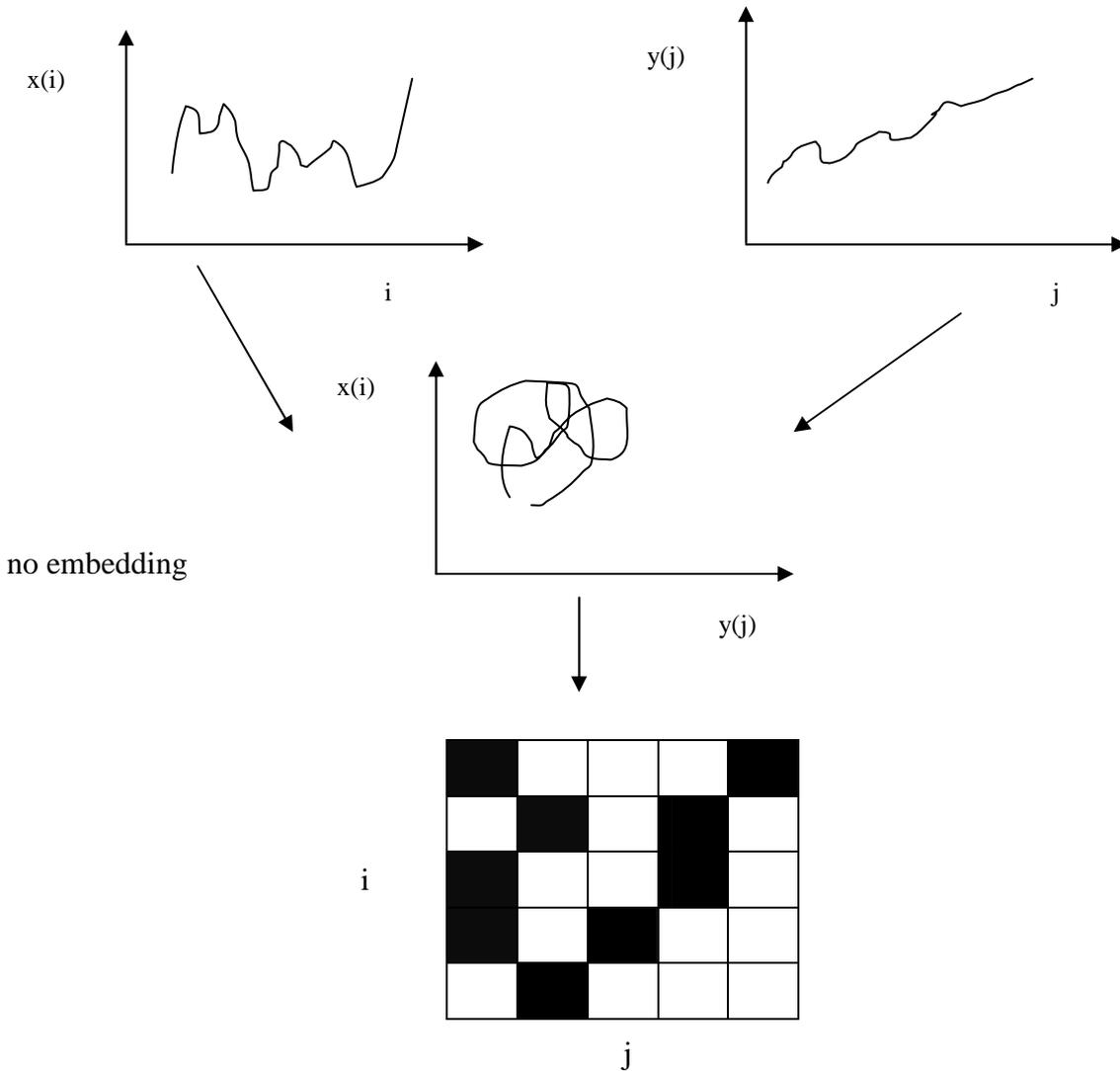


Figure 12. Cross Recurrence plot (CRP) construction.

To quantify the CRP, different measures were introduced based on the percentage of the number of recurrent points forming diagonal, vertical or orthogonal lines. The lines which are diagonally oriented are of mayor interest in fact they represent segments of both trajectories, which run parallel for some time. The frequency and length of these lines are related to a similarity between the two dynamical systems which cannot be detected by the common cross-correlation. If a time dilatation or compression of one of the trajectories is applied then a distortion of the diagonal lines appear in the CRP. In the following analysis we have applied CRP toolbox which is free downloadable from <http://www.agnld.uni-potsdam.de/~marwan/toolbox/>.

In order to better understand the potentiality of this representation let we consider some examples. In figure 13 a) we have applied CRP to two identical time series: $\sin(\pi t)$ then the

CRP contains the main diagonal line of identity (underlined in red). If we consider a time distortion in the second trajectory in such a way that it becomes $\sin(\pi + \sin(1.5t))$ then the LOI will be distorted and the new line is called line of synchronization (LOS), see Fig 13 b. If we stretch or compress the second trajectory and it becomes $\sin(3\pi)$, the LOS will be a straight line but not parallel to the main diagonal, see Figure 13 c). The local slope in CRP corresponds to the transformation of the time axes of the two considered trajectories. A time shift between the trajectories causes a dislocation of the LOS. Hence LOS allows finding the rescaling function between different time series.

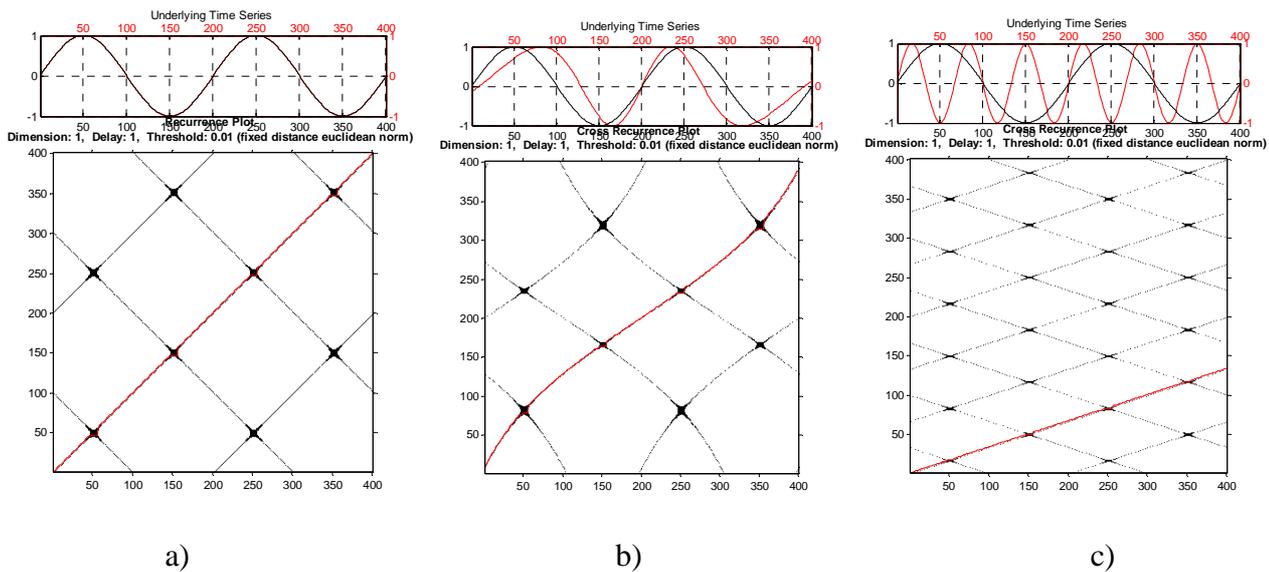


Figure 12. a) $\text{CRP}(\sin(\pi t), \sin(\pi t))$; b) $\text{CRP}(\sin(\pi t), \sin(\pi t + \sin(1.5t)))$; c) $\text{CRP}(\sin(\pi t), \sin(3\pi t))$.

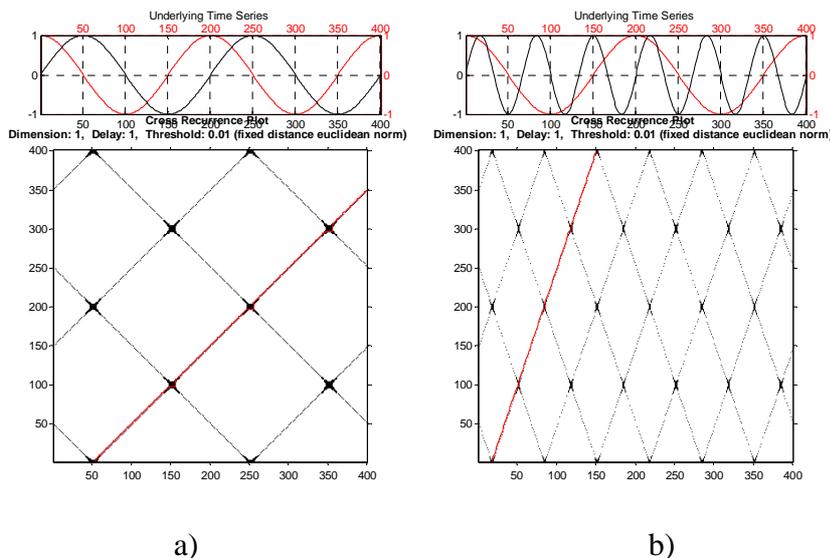


Figure 14: a) $\text{CRP}(\sin(\pi t), \cos(\pi t))$, b) $\text{CRP}(\sin(3\pi t), \cos(\pi t))$.

In Fig 14 a) the LOS has slope 1 and it detects a linear correlation between the two time series, if we shift them by 50 units (which corresponds to $\pi/2$ in the unit considered) then we can synchronise the system. In Figure 14 b) the series detect two dynamics shifted and with different

speeds (the first: $\sin(3\pi t)$ is faster), we can read this information on the LOS slope (higher than one) and on its translation.

- *Detection of changes in the correlation using LOS: An example*

CRP correlation detection can be even more useful when the correlation between two time series change in time. This is because the technique is able of detecting the window of correlation. In fact, if we consider two time series in which we introduce a break in the correlation:

- first time series: $y_1(t)=\sin(\pi \cdot t)$, if $t=[-1:0.01:1]$;
- second time series: $y_2(t)=\cos(\pi \cdot t)$, if $t=[-1:0.01:0]$ and $y_2(t)=\cos(\pi \cdot t/2)$ if $t=[0:0.01:1]$;

The CRP and the LOS are shown in Figure 15.

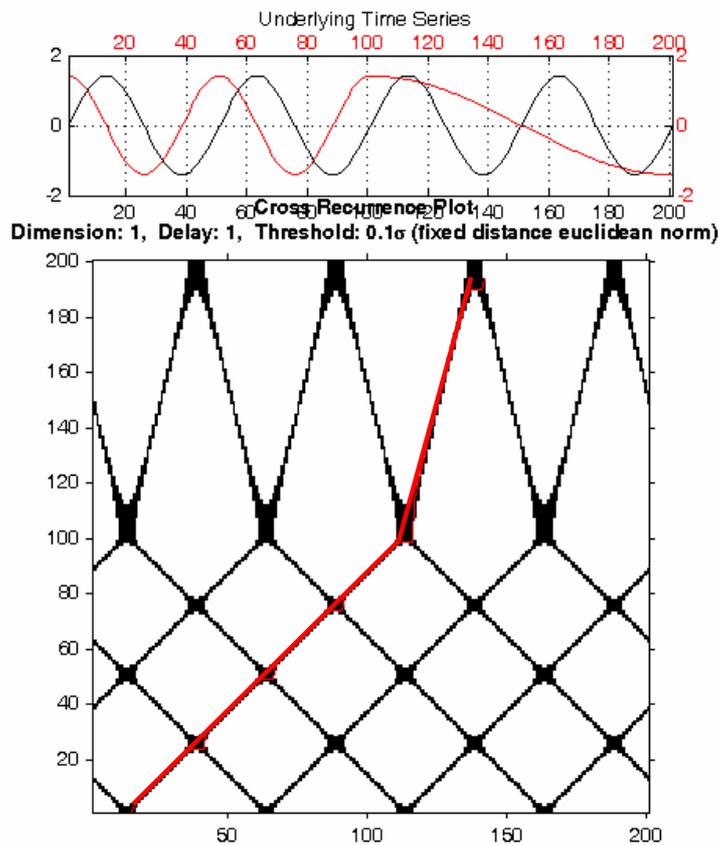


Figure 15. CRP when a change in correlation occurs.

The linear correlation coefficient of the two complete series is $R = 0.1687$ with a t-test with $P=0.0000$. Using the window suggested by LOS, i.e. $y_1(10:109)$, $y_2(1:100)$, R becomes: 0.9048 with $P=0.0000$. If we calculate R on the remaining parts: $y_1(110:201)$, $y_2(101:201-9)$ R becomes 0.2851 with $P=0.0059$.

A disadvantage of using CRP is that in order to obtain a good LOS quality, which means that information given by LOS show real changes in the correlation properties, there is the need of a

certain minimum amount of points. In this work we have been able to obtain good LOS quality using only data with $w=2$ and $sh=1$; in the other cases there were not enough points to perform this analysis.

- Line Of Synchronization:: Algorithm and Quality

The Line of Synchronization algorithm is presented in Marwan et al. (2007) and it consists on an iterative search of recurrent points in CRP starting from the first point next to the axes origin and then looking in a predefined window. If this window does not contain other recurrence points, it is increased. If there are subsequent recurrence points in y-direction (x-direction) the window size is iteratively increased in y-direction (x-direction) until a predefined size $dx*dy$ or until no recurrence points are found.

Moreover Marwan et al. (2007) introduced the following indicator as the LOS Quality:

$$Q = \frac{Nt}{Nt + Ng} * 100 \tag{14}$$

where Nt is the number of target points and Ng the number of gap points. The larger is Q the better is LOS.

- LOS calculation of real time series.

In this section, using Cross Correlation Function, we have observed correlations between V_D and Sfd , V_D and Sdt , V_D and Sfd . These time series do not have enough points to show a reliable Line of Synchronization. For this reason we have to consider only small windows of data. The smallest window that we can use is $w=2$ in order to calculate standard deviation and the minimum shift is one. In Figures 16-17 the time series considered in these analyses are plotted as a function of time unit.

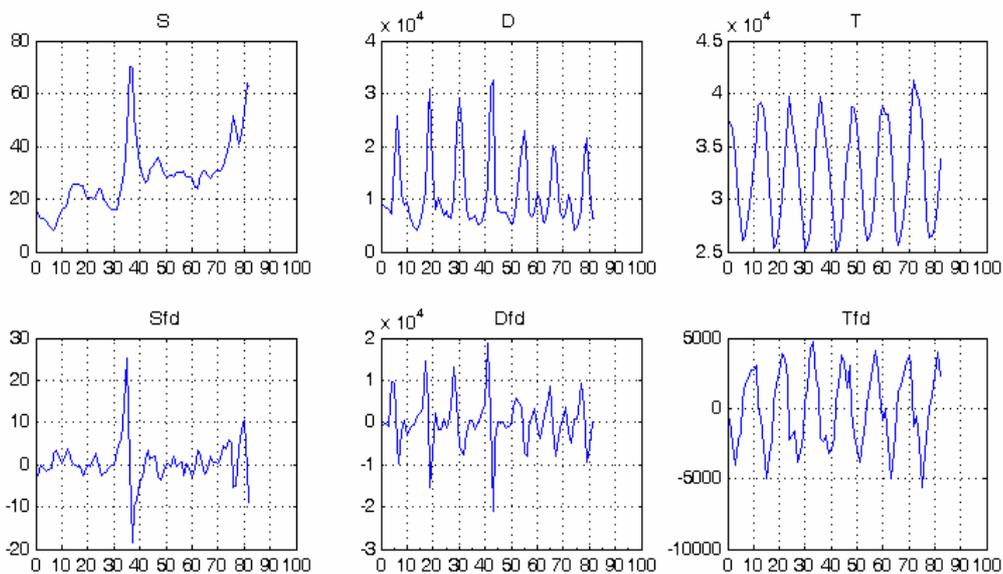


Figure 16. Time series for $w = 2, sh=1$

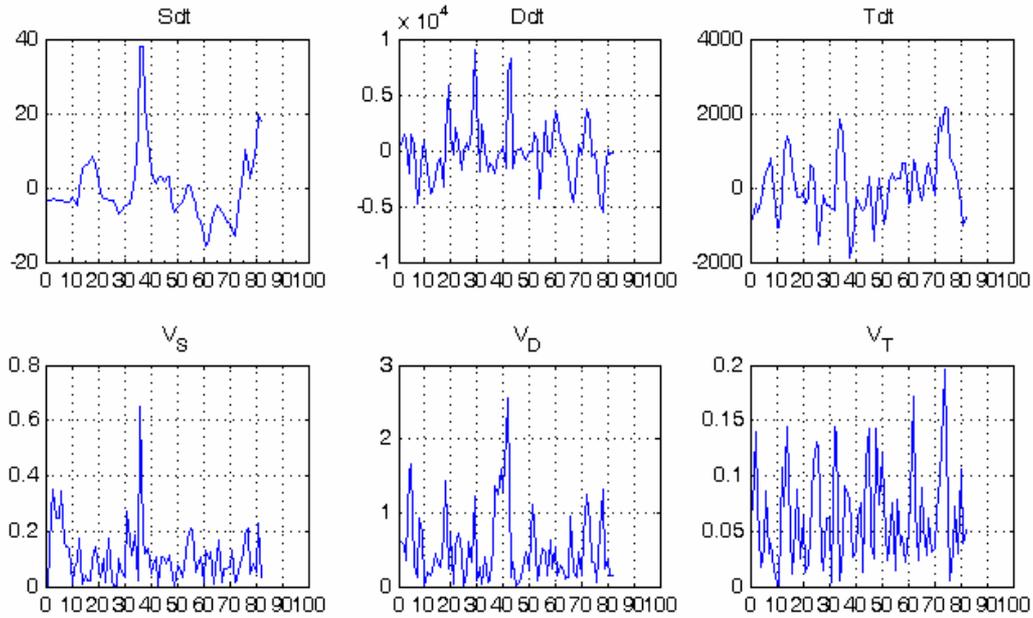
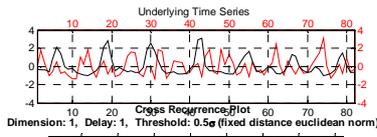


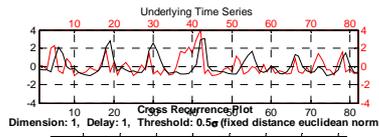
Figure 17. time series for $w = 2, sh = 1$.

It is interesting to see that when the price increase between 30 and 40 time units, due to the dry period, the volatility of disturbance increases too with a certain delay. Looking only to Disturbances and Spot prices such relationship is less evident.

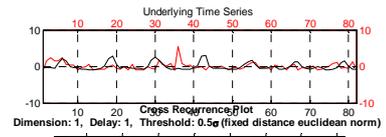
In Figure 18 we have represented CRP for the series of Disturbances (D) in respect of the other time series of Table 2 with ($w=2, sh=1$) together with their LOS, in order to see if it is possible to extract information about possible correlations between the time series on some time windows that are not clear from the correlation function. In Figure 19 we have plotted the CRP considering the volatility of disturbances (V_D) instead of the disturbances themselves.



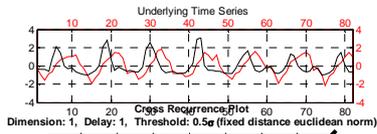
a) $V_T-D, Q=74.91$



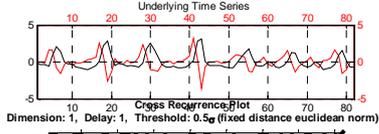
b) $V_D-D, Q=82.82$



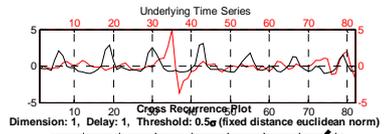
c) $V_S-D, Q=72.43$



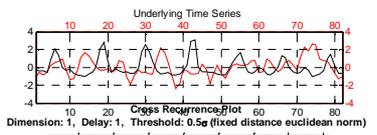
d) $Tfd-D, Q=80.40$



e) $Dfd-D, Q=81.69$



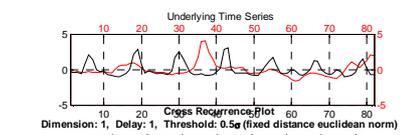
f) $Sfd-D, Q=64.32$



g) $Tdt-D, Q=76.23$



h) $Ddt-D, Q=74.76$



i) $Sdt-D, Q=78.08$

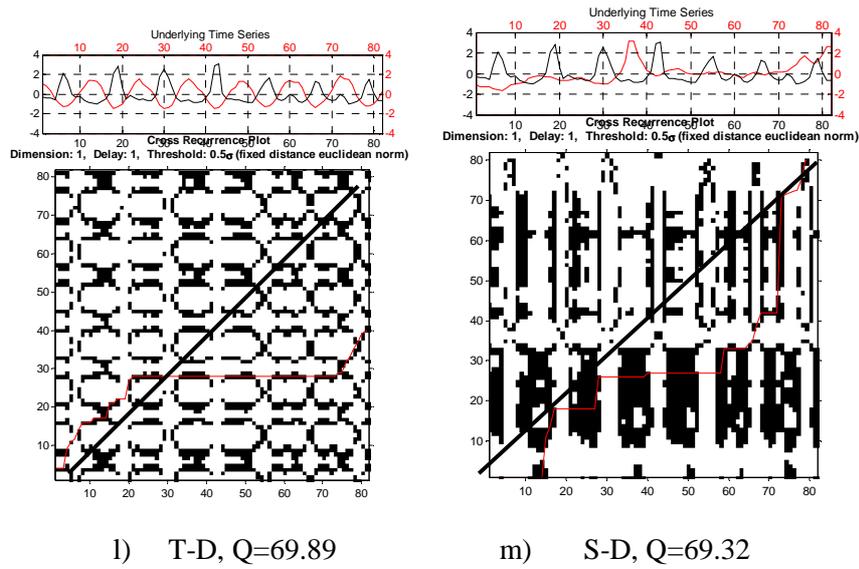


Figure 18. CRP with $w=2$; $sh=1$;

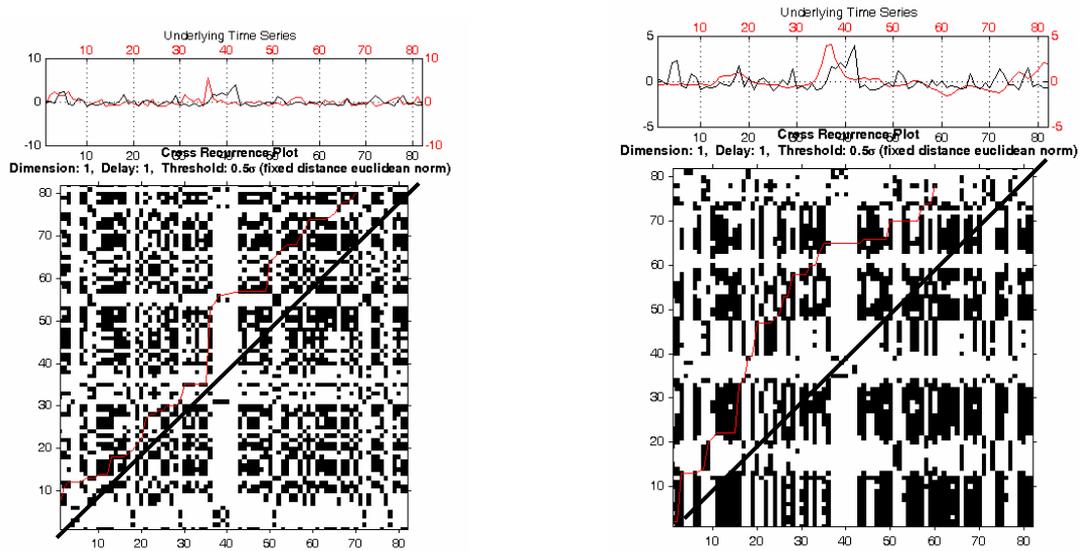


Figure 19. CRP of V_S-V_D (left) and CRP of $Sdt-V_D$ (right).

Looking to Figures 18 and 19 it is possible to observe that many cross recurrent plots have a white horizontal band between 30 and 40 points as indicated in Table 10.

Table 10: white band in CRP.

Serie	Band points	Real date
Sfd_D	30-40	July 02-April 03
VD-D	36-42	January 03- July 03
VS-D	30-40	July 02- April 03
Tdt-D	33-40	October 02-April 03
S-D	33-40	October 02-April 03

The periods of Table 10 correspond to the ones in which the price increases due to a dry period and then a high dependence from external sources of energy appeared and probably a change in the correlation properties. The presence of this wide white bands causes, in general, a change in the properties of LOS because it is not possible to find near recurrence points and the algorithm double the step search. This is evident for example in Figure 18 for the first differences of prices (*Sfd*) in respect of disturbances (*D*). It seems that the correlation properties i.e. the LOS parallelism to the main diagonal of CRP, change after crossing the band. The same happens for the correlation between some variables such as *S*, *Tdt*, *VS*, *Sfd* with *D*. To confirm the hypothesis that LOS allows in detecting windows of higher linear correlation, we have compared the correlation of the entire time series with the one obtained using only the portion of the data in which the LOS is parallel to the main diagonal (R_{LOS}) and with the one suggested by the correlation function (R_{CCF}) i.e. obtained translating the entire time series. All the results are shown in Table 11.

Table 11. Correlation coefficient for different portion of the time series. *ns*: not significative. *R*: correlation Coefficient of the entire time series and without shift. R_{CCF} : max correlation obtained using Cross Correlation Function. R_{LOS} : Correlation coefficient of the portion of the time series suggested by LOS.

Figure	Q	Serie and total points considered	R	R_{CCF}	Interval suggested by LOS	R_{LOS}	Date correspondent to the points considered
m)	69.32	D(1:83); S(1:83);	-0.2692	-0.2692	D(18:28); S(18:28);	0.3979 (ns)	July 01- May 02
l)	69.89	D(1:83); T(1:83)	-0.7354	-0.7354	D(1:30); T(1:30)	-0.8037	Feb 00- Sept 01
f)	64.32	D(1:83); Sfd(1:83)	0.0702	-0.3529	D(1:30); Sfd(1:30)	-0.3953	Feb 00- July 02
e)	81.69	D(1:83); Dfd(1:83)	-0.4119	-0.6809	D(1:19); Dfd(2:20)	-0.7021	Feb 00- June 02 March 00 -July 01
d)	80.40	D(1:83); Tfd(1:83)	0.2429	0.6896	D(1:60); Tfd(3:62)	0.7455	Feb 00-July 06 May 00-Dec 06
	76.4632	VD(1:83) Sdt(1:83)	0.1545	0.4418	VD(1:35); Sdt(7:41)	-0.2248 (ns)	Feb 00-Dec 02

When, looking to CRP of fig 18-19 a portion of LOS parallel to the main diagonal (listed in Table 11) detected a window of higher linear correlation (see column 7) in comparison with the

one calculated using all the points (column 4) or the one obtained shifting the two time series as the Cross Correlation Function suggested (see Fig 11 and column 5). Moreover, looking to Table 11, we can observe that LOS allows identifying the time in which Spot Prices changes for the starting of the dry period (July 2002) and in which the prices increase due to the dependence from external sources.

Moreover, as we can see from Figure 17 a spike in V_S is followed by a spike in V_D and we can observe that the two dynamics can be correlated even after the dry period but with a delay. Looking at Figure 19, after the white band, we see that the LOS is still parallel to the main diagonal but with a shift. We have measured the correlation coefficients on the intervals indicated by LOS and we have obtained again an improvement (results not shown) with respect to considering all the time series, eventually shifted, but, perhaps due to the small amount of points, the correlation values are not significative.

4. Conclusions

In this work we have analysed possible correlations between electricity prices and disturbances in the Nordic Region (Denmark, Finland, Norway and Sweden) from January 2000 until December 2006. By a preliminary treatment of the three original time series we have obtained other nine time series: three without trends, three first differences and three volatilities (Eqs. 1-3). The set of 12 time series is then grouped using different time windows and translated by different shifts (see Table 3).

First we have analysed the time series obtained using the linear correlation coefficient R . We have found a strong linear correlations, i.e. R higher than 0.7, for windows of twelve months (see Table 6) between the volatility of disturbances and the de-trended spot price.

Applying the Principal Component Analysis to the 12 time series, we have observed that more than 3 PCs are necessary to explain at least the 90% of variance, therefore the treated time series contain independent information in comparison with the first three ones (Disturbances, Spot prices and Total Consumption).

In Figure 10 we have plotted some Cross Correlation Functions and we have seen that the linear correlation between the time series can be increased shifting them. The maximum correlation values obtained are listed in Table 9. We have found that disturbance volatility is correlated with price volatility, price first difference and de-trended price, but only considering window of six or twelve months. Disturbance volatility is not so much correlated with the price itself. In Figure 11 we have plotted the correlation functions between $D-Sfd$ and $D-S$ which, even if it

never reaches values higher than 0.4, it has a regular oscillating behaviour in respect to the delay and this can be a sign of similarity between the two dynamics.

Finally we have applied Cross Recurrence Plot analysis, which gives an extension of the Cross Correlation Function and it helps to detect portion of the time series that are linear correlated. The only problem in performing this analysis is the amount of points necessary. For this reason we have analysed only the case of $w=2$ (minimum to calculate standard deviation) and $sh=1$. In Table 11 it is shown how, using CRP, higher correlated data windows are detected. A relationship between the volatility of Spot prices and the volatility of disturbances appears in Fig 17, in which the spike in the price in correspondence of a dry period is followed, with a certain delay, by a spike in the disturbances. This correlation is confirmed by CRP representation given in Figure 19, where it seems that the two time series are synchronised before the dry period but after a delay appears between them. The only disadvantage of CRP is that we can apply it and extracting reliable information only if we have a minimum amount of data. It would be interesting to repeat the analysis performed in this work using daily data of disturbances and consumption.

References

- Bask, M., Liu, T. and Widerberg, A. 2007. The stability of electricity prices: Estimation and inference of the Lyapunov exponents, *Physica A* 376, 565-572.
- Erzgräber, H., Strozzi, F., Zaldívar, J.M., Touchette, H., Gutiérrez, E. and Arrowsmith, D.K. 2008. Time series analysis and long range correlations of Nordic spot electricity market data. *Physica A* 387, 6567-6574.
- Jackson, J. E., 1991. *A User's Guide to Principal Components*, John Wiley and Sons.
- Jolliffe, I. T. 2002. *Principal Component Analysis*, 2nd edition, Springer.
- Marwan, N. and Kurths, J. 2002. Nonlinear analysis of bivariate data with cross recurrence plots, *Physics Letters A* 302, 299–307.
- Marwan, N., Romano, M. C., Thiel, M. and Kurths J. 2007. *Recurrence Plots for the Analysis of Complex Systems*. *Physics Reports* 438, 237–329.
- Orfanidis, S.J., 1996. *Optimum Signal Processing. An Introduction*. 2nd Edition, Prentice-Hall, Englewood Cliffs, NJ.
- R. Weron and B. Przybyłowicz, 2000. Hurst analysis of electricity price dynamics, *Physica A* 283, 462-468.
- Simonsen, I. 2003. Measuring anti-correlations in the Nordic electricity spot market by wavelets, *Physica A* 322, 597-606.
- Simonsen, I. 2005. Volatility of power markets, *Physica A* 355, 10-20.
- Strozzi, F., Gutiérrez, E., Noè, C., Rossi, T., Serati, M. and Zaldívar, J.M. 2008. Measuring volatility in the Nordic spot electricity market using Recurrence Quantification Analysis. *Eur. Phys. J. Special Topics* 164, 105-115.
- Zhao Junhua, 2007. *Electricity market management and analysis using advanced data mining and statistical methods*. PhD thesis. <http://espace.library.uq.edu.au/view/UQ:158511>

Appendix 1 Correlation matrices and t-test

- Window=3 (seasonal), shift=3;

R =

1.0000	-0.2593	0.1333	0.6728	-0.2086	0.0817	-0.1413	0.0169	-0.1247	-0.1156	-0.0025	0.0752
-0.2593	1.0000	-0.8154	-0.0508	0.4328	-0.1900	0.0650	-0.5204	0.4788	0.0342	0.0082	-0.0455
0.1333	-0.8154	1.0000	-0.0724	-0.0414	0.2034	0.0331	0.0773	-0.3500	0.0031	-0.1754	0.3753
0.6728	-0.0508	-0.0724	1.0000	-0.2381	-0.1965	-0.2660	0.0180	-0.1039	0.0357	0.2702	-0.1705
-0.2086	0.4328	-0.0414	-0.2381	1.0000	-0.2613	-0.0555	-0.2481	0.0516	-0.0845	0.0407	0.1951
0.0817	-0.1900	0.2034	-0.1965	-0.2613	1.0000	0.5799	0.0528	-0.0040	0.0304	-0.0543	0.0647
-0.1413	0.0650	0.0331	-0.2660	-0.0555	0.5799	1.0000	-0.1939	0.3568	-0.0085	-0.4753	0.0245
0.0169	-0.5204	0.0773	0.0180	-0.2481	0.0528	-0.1939	1.0000	-0.5531	-0.1126	0.5404	-0.3603
-0.1247	0.4788	-0.3500	-0.1039	0.0516	-0.0040	0.3568	-0.5531	1.0000	0.0373	-0.4220	-0.2978
-0.1156	0.0342	0.0031	0.0357	-0.0845	0.0304	-0.0085	-0.1126	0.0373	1.0000	0.0490	0.1020
-0.0025	0.0082	-0.1754	0.2702	0.0407	-0.0543	-0.4753	0.5404	-0.4220	0.0490	1.0000	-0.2325
0.0752	-0.0455	0.3753	-0.1705	0.1951	0.0647	0.0245	-0.3603	-0.2978	0.1020	-0.2325	1.0000

P =

1.0000	0.1915	0.5073	0.0001	0.2964	0.6854	0.4820	0.9332	0.5355	0.5658	0.9902	0.7095
0.1915	1.0000	0.0000	0.8014	0.0241	0.3424	0.7473	0.0054	0.0115	0.8654	0.9676	0.8218
0.5073	0.0000	1.0000	0.7197	0.8374	0.3090	0.8700	0.7015	0.0735	0.9879	0.3816	0.0538
0.0001	0.8014	0.7197	1.0000	0.2318	0.3260	0.1800	0.9291	0.6059	0.8597	0.1729	0.3951
0.2964	0.0241	0.8374	0.2318	1.0000	0.1880	0.7833	0.2121	0.7982	0.6753	0.8401	0.3294
0.6854	0.3424	0.3090	0.3260	0.1880	1.0000	0.0015	0.7936	0.9840	0.8804	0.7880	0.7484
0.4820	0.7473	0.8700	0.1800	0.7833	0.0015	1.0000	0.3325	0.0677	0.9665	0.0122	0.9033
0.9332	0.0054	0.7015	0.9291	0.2121	0.7936	0.3325	1.0000	0.0028	0.5760	0.0036	0.0649
0.5355	0.0115	0.0735	0.6059	0.7982	0.9840	0.0677	0.0028	1.0000	0.8534	0.0283	0.1313
0.5658	0.8654	0.9879	0.8597	0.6753	0.8804	0.9665	0.5760	0.8534	1.0000	0.8083	0.6128
0.9902	0.9676	0.3816	0.1729	0.8401	0.7880	0.0122	0.0036	0.0283	0.8083	1.0000	0.2432
0.7095	0.8218	0.0538	0.3951	0.3294	0.7484	0.9033	0.0649	0.1313	0.6128	0.2432	1.0000

- Window = 6; shift = 6;

R =

1.0000	-0.4047	0.3100	0.5402	-0.1778	0.2332	-0.2855	0.2637	-0.1955	-0.0971	0.3099	0.3250
-0.4047	1.0000	-0.8594	-0.1892	0.6352	-0.3381	0.2550	-0.8503	0.7665	-0.0158	-0.5359	-0.3012
0.3100	-0.8594	1.0000	0.0219	-0.2717	0.4335	-0.1567	0.7752	-0.8686	0.0018	0.4301	0.5935
0.5402	-0.1892	0.0219	1.0000	-0.2351	-0.1793	-0.3495	0.3018	-0.1575	0.2447	0.5391	0.0171
-0.1778	0.6352	-0.2717	-0.2351	1.0000	-0.4916	-0.1291	-0.3401	0.0532	-0.2625	-0.1439	0.1929
0.2332	-0.3381	0.4335	-0.1793	-0.4916	1.0000	0.5172	-0.0417	0.0633	0.2126	-0.3194	0.3222
-0.2855	0.2550	-0.1567	-0.3495	-0.1291	0.5172	1.0000	-0.5204	0.4549	0.4273	-0.7778	0.2901
0.2637	-0.8503	0.7752	0.3018	-0.3401	-0.0417	-0.5204	1.0000	-0.8924	-0.0864	0.7698	0.2625
-0.1955	0.7665	-0.8686	-0.1575	0.0532	0.0633	0.4549	-0.8924	1.0000	0.1011	-0.6590	-0.4628
-0.0971	-0.0158	0.0018	0.2447	-0.2625	0.2126	0.4273	-0.0864	0.1011	1.0000	-0.0014	0.3053
0.3099	-0.5359	0.4301	0.5391	-0.1439	-0.3194	-0.7778	0.7698	-0.6590	-0.0014	1.0000	-0.0562
0.3250	-0.3012	0.5935	0.0171	0.1929	0.3222	0.2901	0.2625	-0.4628	0.3053	-0.0562	1.0000

P =

1.0000	0.1702	0.3026	0.0567	0.5611	0.4433	0.3444	0.3839	0.5221	0.7524	0.3028	0.2786
0.1702	1.0000	0.0002	0.5358	0.0197	0.2585	0.4004	0.0002	0.0022	0.9593	0.0591	0.3173
0.3026	0.0002	1.0000	0.9434	0.3693	0.1389	0.6091	0.0019	0.0001	0.9954	0.1424	0.0325
0.0567	0.5358	0.9434	1.0000	0.4394	0.5579	0.2417	0.3162	0.6074	0.4203	0.0573	0.9557
0.5611	0.0197	0.3693	0.4394	1.0000	0.0879	0.6741	0.2555	0.8631	0.3863	0.6391	0.5278
0.4433	0.2585	0.1389	0.5579	0.0879	1.0000	0.0703	0.8923	0.8371	0.4856	0.2874	0.2831
0.3444	0.4004	0.6091	0.2417	0.6741	0.0703	1.0000	0.0683	0.1184	0.1453	0.0017	0.3363
0.3839	0.0002	0.0019	0.3162	0.2555	0.8923	0.0683	1.0000	0.0000	0.7791	0.0021	0.3863
0.5221	0.0022	0.0001	0.6074	0.8631	0.8371	0.1184	0.0000	1.0000	0.7424	0.0143	0.1112
0.7524	0.9593	0.9954	0.4203	0.3863	0.4856	0.1453	0.7791	0.7424	1.0000	0.9965	0.3104
0.3028	0.0591	0.1424	0.0573	0.6391	0.2874	0.0017	0.0021	0.0143	0.9965	1.0000	0.8553
0.2786	0.3173	0.0325	0.9557	0.5278	0.2831	0.3363	0.3863	0.1112	0.3104	0.8553	1.0000

- Window =12; shift =12

R=

1.0000	-0.1224	-0.1710	0.4665	0.5750	-0.3117	-0.5668	0.5599	-0.1739	-0.4327	0.3599	0.4001
-0.1224	1.0000	-0.5911	0.6509	0.6041	-0.4968	0.2140	-0.4577	-0.3629	0.7903	0.4244	0.4273
-0.1710	-0.5911	1.0000	-0.6735	-0.3835	0.9842	0.4892	0.6237	-0.0832	-0.5198	-0.9057	0.0590
0.4665	0.6509	-0.6735	1.0000	0.5149	-0.6494	-0.5304	-0.0414	-0.3383	0.1474	0.8138	0.2562
0.5750	0.6041	-0.3835	0.5149	1.0000	-0.4414	0.0954	0.0754	-0.3796	0.4136	0.2583	0.7939
-0.3117	-0.4968	0.9842	-0.6494	-0.4414	1.0000	0.5488	1.0000	-0.1192	-0.1006	0.4843	-0.9014
-0.5668	0.2140	0.4892	-0.5304	0.0954	0.5488	1.0000	-0.1192	-0.1006	0.4843	-0.7498	0.3983
0.5599	-0.4577	0.6237	-0.0414	0.0754	0.5369	-0.1192	1.0000	-0.5067	-0.7924	-0.3952	0.1674
-0.1739	-0.3629	-0.0832	-0.3383	-0.3796	-0.1049	-0.1006	-0.5067	1.0000	0.0679	0.1192	-0.1203
-0.4327	0.7903	-0.5198	0.1474	0.4136	-0.4368	0.4843	-0.7924	0.0679	1.0000	0.1885	0.3123
0.3599	0.4244	-0.9057	0.8138	0.2583	-0.9014	-0.7498	-0.3952	0.1192	0.1885	1.0000	-0.0917
0.4001	0.4273	0.0590	0.2562	0.7939	0.0205	0.3983	0.1674	-0.1203	0.3123	-0.0917	1.0000

P=

1.0000	0.8172	0.7460	0.3510	0.2326	0.5476	0.2408	0.2479	0.7418	0.3915	0.4835	0.4319
0.8172	1.0000	0.2166	0.1615	0.2041	0.3161	0.6839	0.3614	0.4796	0.0614	0.4017	0.3981
0.7460	0.2166	1.0000	0.1425	0.4530	0.0004	0.3247	0.1857	0.8754	0.2905	0.0129	0.9115
0.3510	0.1615	0.1425	1.0000	0.2959	0.1628	0.2791	0.9380	0.5119	0.7805	0.0488	0.6241
0.2326	0.2041	0.4530	0.2959	1.0000	0.3809	0.8574	0.8872	0.4579	0.4150	0.6212	0.0594
0.5476	0.3161	0.0004	0.1628	0.3809	1.0000	0.2594	0.2721	0.8432	0.3865	0.0141	0.9693
0.2408	0.6839	0.3247	0.2791	0.8574	0.2594	1.0000	0.8220	0.8497	0.3304	0.0861	0.4342
0.2479	0.3614	0.1857	0.9380	0.8872	0.2721	0.8220	1.0000	0.3050	0.0602	0.4381	0.7512
0.7418	0.4796	0.8754	0.5119	0.4579	0.8432	0.8497	0.3050	1.0000	0.8984	0.8221	0.8204
0.3915	0.0614	0.2905	0.7805	0.4150	0.3865	0.3304	0.0602	0.8984	1.0000	0.7206	0.5468
0.4835	0.4017	0.0129	0.0488	0.6212	0.0141	0.0861	0.4381	0.8221	0.7206	1.0000	0.8628
0.4319	0.3981	0.9115	0.6241	0.0594	0.9693	0.4342	0.7512	0.8204	0.5468	0.8628	1.0000

- Window =2; shift =1

R =

1.0000	-0.2692	0.1775	0.7195	-0.1748	0.0543	-0.1541	0.0242	-0.0508	0.1353	0.0195	0.0696
-0.2692	1.0000	-0.7354	-0.0689	0.3791	-0.1185	0.0702	-0.4119	0.2429	0.0730	0.1785	-0.1720
0.1775	-0.7354	1.0000	-0.0331	-0.0376	0.2229	0.0249	0.0811	-0.2756	-0.0683	-0.1530	0.3387
0.7195	-0.0689	-0.0331	1.0000	-0.2024	-0.1431	-0.2229	0.0175	-0.0808	0.3428	0.1545	-0.0496
-0.1748	0.3791	-0.0376	-0.2024	1.0000	-0.2039	-0.0204	-0.2593	0.0282	-0.1574	0.1379	0.0332
0.0543	-0.1185	0.2229	-0.1431	-0.2039	1.0000	0.5048	0.0152	-0.0871	0.0372	-0.1166	0.1482
-0.1541	0.0702	0.0249	-0.2229	-0.0204	0.5048	1.0000	-0.2136	0.2398	-0.0919	-0.2885	0.0416
0.0242	-0.4119	0.0811	0.0175	-0.2593	0.0152	-0.2136	1.0000	-0.4489	-0.0309	0.3995	-0.0508
-0.0508	0.2429	-0.2756	-0.0808	0.0282	-0.0871	0.2398	-0.4489	1.0000	-0.0466	-0.2556	-0.1539
0.1353	0.0730	-0.0683	0.3428	-0.1574	0.0372	-0.0919	-0.0309	-0.0466	1.0000	0.0602	0.0289
0.0195	0.1785	-0.1530	0.1545	0.1379	-0.1166	-0.2885	0.3995	-0.2556	0.0602	1.0000	-0.0059
0.0696	-0.1720	0.3387	-0.0496	0.0332	0.1482	0.0416	-0.0508	-0.1539	0.0289	-0.0059	1.0000

P =

1.0000	0.0145	0.1107	0.0000	0.1162	0.6283	0.1669	0.8293	0.6507	0.2254	0.8622	0.5343
0.0145	1.0000	0.0000	0.5385	0.0004	0.2890	0.5306	0.0001	0.0279	0.5147	0.1086	0.1223
0.1107	0.0000	1.0000	0.7676	0.7373	0.0441	0.8244	0.4686	0.0122	0.5419	0.1700	0.0019
0.0000	0.5385	0.7676	1.0000	0.0682	0.1997	0.0441	0.8763	0.4704	0.0016	0.1659	0.6579
0.1162	0.0004	0.7373	0.0682	1.0000	0.0662	0.8560	0.0187	0.8017	0.1578	0.2165	0.7672
0.6283	0.2890	0.0441	0.1997	0.0662	1.0000	0.0000	0.8925	0.4367	0.7400	0.2969	0.1841
0.1669	0.5306	0.8244	0.0441	0.8560	0.0000	1.0000	0.0540	0.0300	0.4113	0.0086	0.7106
0.8293	0.0001	0.4686	0.8763	0.0187	0.8925	0.0540	1.0000	0.0000	0.7826	0.0002	0.6506
0.6507	0.0279	0.0122	0.4704	0.8017	0.4367	0.0300	0.0000	1.0000	0.6778	0.0205	0.1675
0.2254	0.5147	0.5419	0.0016	0.1578	0.7400	0.4113	0.7826	0.6778	1.0000	0.5910	0.7964
0.8622	0.1086	0.1700	0.1659	0.2165	0.2969	0.0086	0.0002	0.0205	0.5910	1.0000	0.9578
0.5343	0.1223	0.0019	0.6579	0.7672	0.1841	0.7106	0.6506	0.1675	0.7964	0.9578	1.0000

- Window =3; shift=1

R =

1.0000	-0.2857	0.1794	0.6949	-0.2155	0.0994	-0.0842	0.0184	-0.0598	0.1994	-0.0372	0.0954
-0.2857	1.0000	-0.8057	-0.0689	0.3557	-0.1463	0.0446	-0.3540	0.2741	0.0571	0.3571	-0.0829
0.1794	-0.8057	1.0000	-0.0396	-0.0401	0.2288	0.0258	0.1020	-0.2768	0.0117	-0.2939	0.3271
0.6949	-0.0689	-0.0396	1.0000	-0.2502	-0.1339	-0.1734	0.0278	-0.0903	0.4318	0.1680	-0.0499
-0.2155	0.3557	-0.0401	-0.2502	1.0000	-0.2247	-0.0721	-0.1879	0.0363	-0.1691	0.2137	0.0921
0.0994	-0.1463	0.2288	-0.1339	-0.2247	1.0000	0.5424	-0.0155	-0.0321	0.0930	-0.1768	0.2017
-0.0842	0.0446	0.0258	-0.1734	-0.0721	0.5424	1.0000	-0.2489	0.3217	0.0930	-0.3052	0.1432
0.0184	-0.3540	0.1020	0.0278	-0.1879	-0.0155	-0.2489	1.0000	-0.5756	-0.1397	0.3508	-0.1934
-0.0598	0.2741	-0.2768	-0.0903	0.0363	-0.0321	0.3217	-0.5756	1.0000	-0.0007	-0.2676	-0.1487
0.1994	0.0571	0.0117	0.4318	-0.1691	0.0930	0.0930	-0.1397	-0.0007	1.0000	0.0377	0.0626
-0.0372	0.3571	-0.2939	0.1680	0.2137	-0.1768	-0.3052	0.3508	-0.2676	0.0377	1.0000	-0.1280
0.0954	-0.0829	0.3271	-0.0499	0.0921	0.2017	0.1432	-0.1934	-0.1487	0.0626	-0.1280	1.0000

P =

1.0000	0.0097	0.1091	0.0000	0.0534	0.3771	0.4548	0.8702	0.5959	0.0744	0.7416	0.3970
0.0097	1.0000	0.0000	0.5413	0.0011	0.1925	0.6924	0.0012	0.0133	0.6129	0.0011	0.4618
0.1091	0.0000	1.0000	0.7254	0.7220	0.0399	0.8194	0.3651	0.0124	0.9175	0.0078	0.0029
0.0000	0.5413	0.7254	1.0000	0.0243	0.2333	0.1216	0.8054	0.4227	0.0001	0.1338	0.6580
0.0534	0.0011	0.7220	0.0243	1.0000	0.0437	0.5226	0.0931	0.7477	0.1311	0.0554	0.4135
0.3771	0.1925	0.0399	0.2333	0.0437	1.0000	0.0000	0.8911	0.7762	0.4088	0.1144	0.0710
0.4548	0.6924	0.8194	0.1216	0.5226	0.0000	1.0000	0.0250	0.0034	0.4091	0.0056	0.2023
0.8702	0.0012	0.3651	0.8054	0.0931	0.8911	0.0250	1.0000	0.0000	0.2135	0.0013	0.0836
0.5959	0.0133	0.0124	0.4227	0.7477	0.7762	0.0034	0.0000	1.0000	0.9951	0.0157	0.1854
0.0744	0.6129	0.9175	0.0001	0.1311	0.4088	0.4091	0.2135	0.9951	1.0000	0.7383	0.5785
0.7416	0.0011	0.0078	0.1338	0.0554	0.1144	0.0056	0.0013	0.0157	0.7383	1.0000	0.2548
0.3970	0.4618	0.0029	0.6580	0.4135	0.0710	0.2023	0.0836	0.1854	0.5785	0.2548	1.0000

-Window =6; shift=1

R =

1.0000	-0.3463	0.1923	0.6105	-0.2231	0.2116	-0.0266	0.1143	-0.1268	0.2354	0.1002	0.1234
-0.3463	1.0000	-0.9044	-0.0615	0.4189	-0.2827	0.0008	-0.2796	0.3076	-0.0554	0.4163	-0.0634
0.1923	-0.9044	1.0000	-0.0495	-0.0647	0.2813	0.0437	0.1298	-0.2560	0.0236	-0.4282	0.1258
0.6105	-0.0615	-0.0495	1.0000	-0.2575	-0.1414	-0.1156	0.1331	-0.1223	0.5683	0.3841	-0.0205
-0.2231	0.4189	-0.0647	-0.2575	1.0000	-0.3015	-0.1635	-0.1700	0.0544	-0.2151	0.2066	0.1666
0.2116	-0.2827	0.2813	-0.1414	-0.3015	1.0000	0.5240	-0.0853	0.0025	0.0472	-0.2940	0.2343
-0.0266	0.0008	0.0437	-0.1156	-0.1635	0.5240	1.0000	-0.4951	0.3740	0.2386	-0.4273	0.2910
0.1143	-0.2796	0.1298	0.1331	-0.1700	-0.0853	-0.4951	1.0000	-0.8010	-0.1195	0.3950	0.0715
-0.1268	0.3076	-0.2560	-0.1223	0.0544	0.0025	0.3740	-0.8010	1.0000	-0.0179	-0.2907	-0.1661
0.2354	-0.0554	0.0236	0.5683	-0.2151	0.0472	0.2386	-0.1195	-0.0179	1.0000	0.1855	0.0059
0.1002	0.4163	-0.4282	0.3841	0.2066	-0.2940	-0.4273	0.3950	-0.2907	0.1855	1.0000	-0.0415
0.1234	-0.0634	0.1258	-0.0205	0.1666	0.2343	0.2910	0.0715	-0.1661	0.0059	-0.0415	1.0000

P =

1.0000	0.0019	0.0917	0.0000	0.0496	0.0629	0.8168	0.3190	0.2685	0.0381	0.3827	0.2817
0.0019	1.0000	0.0000	0.5926	0.0001	0.0121	0.9942	0.0132	0.0062	0.6299	0.0002	0.5814
0.0917	0.0000	1.0000	0.6669	0.5737	0.0126	0.7041	0.2575	0.0237	0.8377	0.0001	0.2725
0.0000	0.5926	0.6669	1.0000	0.0229	0.2167	0.3135	0.2453	0.2861	0.0000	0.0005	0.8583
0.0496	0.0001	0.5737	0.0229	1.0000	0.0073	0.1525	0.1368	0.6360	0.0586	0.0696	0.1448
0.0629	0.0121	0.0126	0.2167	0.0073	1.0000	0.0000	0.4577	0.9826	0.6815	0.0090	0.0389
0.8168	0.9942	0.7041	0.3135	0.1525	0.0000	1.0000	0.0000	0.0007	0.0354	0.0001	0.0098
0.3190	0.0132	0.2575	0.2453	0.1368	0.4577	0.0000	1.0000	0.0000	0.2972	0.0003	0.5339
0.2685	0.0062	0.0237	0.2861	0.6360	0.9826	0.0007	0.0000	1.0000	0.8765	0.0098	0.1462
0.0381	0.6299	0.8377	0.0000	0.0586	0.6815	0.0354	0.2972	0.8765	1.0000	0.1040	0.9589
0.3827	0.0002	0.0001	0.0005	0.0696	0.0090	0.0001	0.0003	0.0098	0.1040	1.0000	0.7183
0.2817	0.5814	0.2725	0.8583	0.1448	0.0389	0.0098	0.5339	0.1462	0.9589	0.7183	1.0000

- Window =12; shift=1

R =

1.0000	-0.3159	0.2947	0.4122	0.1272	0.1993	0.0066	-0.0013	-0.2961	0.1884	0.3200	0.4188
-0.3159	1.0000	-0.7807	0.2975	0.8060	-0.7586	-0.2624	-0.2240	0.0036	0.2320	0.2840	0.1669
0.2947	-0.7807	1.0000	-0.3883	-0.5084	0.9904	0.5101	0.0054	-0.1337	-0.2487	-0.4667	0.1228
0.4122	0.2975	-0.3883	1.0000	0.1107	-0.3651	-0.0788	0.0812	-0.3881	0.6700	0.7567	0.2342
0.1272	0.8060	-0.5084	0.1107	1.0000	-0.5645	-0.2670	-0.3143	0.0003	0.0491	0.1604	0.3787
0.1993	-0.7586	0.9904	-0.3651	-0.5645	1.0000	0.5467	0.0170	-0.1454	-0.2165	-0.4670	0.0910
0.0066	-0.2624	0.5101	-0.0788	-0.2670	0.5467	1.0000	-0.0879	-0.0586	0.3551	-0.2520	0.3595
-0.0013	-0.2240	0.0054	0.0812	-0.3143	0.0170	-0.0879	1.0000	-0.1707	-0.0558	0.1683	-0.0590
-0.2961	0.0036	-0.1337	-0.3881	0.0003	-0.1454	-0.0586	-0.1707	1.0000	-0.1223	-0.1303	-0.0504
0.1884	0.2320	-0.2487	0.6700	0.0491	-0.2165	0.3551	-0.0558	-0.1223	1.0000	0.5183	0.1933
0.3200	0.2840	-0.4667	0.7567	0.1604	-0.4670	-0.2520	0.1683	-0.1303	0.5183	1.0000	0.0837
0.4188	0.1669	0.1228	0.2342	0.3787	0.0910	0.3595	-0.0590	-0.0504	0.1933	0.0837	1.0000

P =

1.0000	0.0069	0.0120	0.0003	0.2872	0.0932	0.9563	0.9914	0.0115	0.1130	0.0061	0.0003
0.0069	1.0000	0.0000	0.0112	0.0000	0.0000	0.0260	0.0586	0.9759	0.0499	0.0156	0.1610
0.0120	0.0000	1.0000	0.0007	0.0000	0.0000	0.0000	0.9639	0.2628	0.0351	0.0000	0.3040
0.0003	0.0112	0.0007	1.0000	0.3547	0.0016	0.5104	0.4980	0.0008	0.0000	0.0000	0.0477
0.2872	0.0000	0.0000	0.3547	1.0000	0.0000	0.0234	0.0072	0.9981	0.6821	0.1784	0.0010
0.0932	0.0000	0.0000	0.0016	0.0000	1.0000	0.0000	0.8872	0.2230	0.0678	0.0000	0.4469
0.9563	0.0260	0.0000	0.5104	0.0234	0.0000	1.0000	0.4630	0.6251	0.0022	0.0328	0.0019
0.9914	0.0586	0.9639	0.4980	0.0072	0.8872	0.4630	1.0000	0.1518	0.6414	0.1576	0.6223
0.0115	0.9759	0.2628	0.0008	0.9981	0.2230	0.6251	0.1518	1.0000	0.3060	0.2752	0.6742
0.1130	0.0499	0.0351	0.0000	0.6821	0.0678	0.0022	0.6414	0.3060	1.0000	0.0000	0.1037
0.0061	0.0156	0.0000	0.0000	0.1784	0.0000	0.0328	0.1576	0.2752	0.0000	1.0000	0.4844
0.0003	0.1610	0.3040	0.0477	0.0010	0.4469	0.0019	0.6223	0.6742	0.1037	0.4844	1.0000

Appendix 2. Principal Component Analysis (PCA)

Traditionally, principal component analysis is performed on the symmetric Covariance matrix or on the symmetric Correlation matrix. These matrices can be calculated from the data matrix. The covariance matrix contains scaled sums of squares and cross products. A correlation matrix is like a covariance matrix but in which the variables, i.e. the columns, have been standardized. We will have to standardize the data if the variances of variables, or if the units of measurement of the variables differs considerably.

In this appendix we report the loadings i.e. eigenvectors matrix ordered in decreasing order in respect to the correspondent eigenvalues and the variance explained for different time window w and different shifts sh .

The columns of loadings are the principal components, the rows are the coordinates of the eigenvectors of covariance matrix in the coordinate system: $\{S, D, T, Sdt, Ddt, Tdt, Sfd, Dfd, Tfd, Vs, Vd, Vt\}$

- Loadings for windows =1; shift =1;

0.1801	-0.4014	-0.2349	-0.4308	0.0655	-0.2196	-0.2880	-0.6285	0.1153	-0.1032	0.1062	0.0023
-0.3934	-0.1531	0.2379	0.0802	-0.3920	-0.1733	-0.1521	0.1371	0.6040	-0.2497	0.3226	0.0436
0.2872	0.0083	0.1344	-0.2054	0.6711	0.2206	0.0899	0.2632	0.4220	-0.0910	0.3051	0.0378
0.1018	-0.4453	-0.3041	-0.3593	-0.1610	-0.1850	-0.0466	0.6879	-0.0987	0.1073	-0.1129	-0.0156
-0.2924	-0.1836	0.3101	-0.1076	0.1262	0.5012	-0.6616	0.0590	-0.1768	0.1009	-0.1439	-0.0518
0.1527	-0.0640	0.2032	0.4514	0.3460	-0.6499	-0.3977	0.1075	-0.0781	0.0744	-0.0624	-0.0133
0.2213	-0.4827	0.0824	0.3576	-0.0895	0.2244	0.1905	-0.1379	0.3781	0.5055	-0.2666	0.0098
-0.3963	-0.2527	0.1831	-0.1117	0.3116	-0.1412	0.3549	-0.0538	0.0358	-0.3832	-0.5822	0.0093
0.2760	0.0306	0.5399	-0.2630	-0.1996	-0.1090	0.1178	-0.0161	-0.0529	0.0168	-0.0135	-0.7023
0.2610	-0.4511	0.0570	0.3734	-0.0783	0.2393	0.0953	0.0227	-0.3546	-0.5614	0.2681	-0.0044
-0.4144	-0.2761	0.1856	-0.0793	0.1824	-0.1403	0.3181	-0.0775	-0.3328	0.4165	0.5145	0.0476
0.2985	0.0421	0.5203	-0.2574	-0.2139	-0.0787	0.0375	-0.0069	-0.1150	0.0149	-0.0860	0.7056

Eigenvalue	Cumulative sum of % variances explained
3.4124	28.4363
2.2086	46.8410
1.9863	63.3939
1.3426	74.5818
1.1692	84.3254
0.8048	91.0324
0.4502	94.7836
0.2249	96.6580
0.1709	98.0824
0.1215	99.0951
0.1021	99.9458
0.0065	100.0000

- Loadings for Windows =2; shift =1

0.3465 -0.1773 0.4449 -0.1565 0.0341 -0.4237 0.0363 -0.0616 -0.0458 -0.3012 -0.5810 0.0996
 -0.4940 -0.2600 0.0490 0.0864 0.3277 -0.0519 0.0439 -0.2625 -0.0862 0.0290 -0.0016 0.6995
 0.3910 0.3575 -0.0642 -0.3521 0.0001 0.0211 -0.2700 0.3029 -0.1582 0.2942 0.0609 0.5575
 0.2760 -0.3741 0.4499 -0.0458 0.0779 -0.2021 -0.0253 -0.0678 0.2151 0.2225 0.6563 -0.0266
 -0.2800 -0.0709 -0.2187 -0.4982 0.3065 -0.2672 -0.4257 0.2080 0.1844 -0.3854 0.1423 -0.1669
 0.0957 0.4118 0.1222 0.4031 0.3952 -0.2266 -0.1099 -0.0580 -0.5121 -0.2744 0.2686 -0.1056
 -0.1753 0.4458 0.1907 0.2928 0.2071 -0.2698 -0.0478 0.1500 0.6308 0.2799 -0.1782 0.0003
 0.3484 -0.1230 -0.4032 0.3847 -0.1046 -0.0798 0.1087 0.1566 0.3560 -0.4822 0.1553 0.3403
 -0.3465 0.0695 0.3578 -0.0484 -0.2796 -0.0330 0.3793 0.6466 -0.1350 -0.2192 0.1638 0.1022
 0.1022 -0.2047 0.3050 0.1897 0.3573 0.6713 -0.3116 0.3065 0.0833 -0.1362 -0.1534 -0.0162
 0.0874 -0.3879 -0.3302 0.1506 0.3631 -0.2649 0.2036 0.4600 -0.2198 0.3980 -0.1594 -0.1518
 0.1705 0.2179 -0.0370 -0.3694 0.4937 0.2295 0.6583 -0.1173 0.1576 -0.1125 0.0629 -0.0258

eigenvalue	Cumulative sum of % variances explained
2.5456	21.2135
2.1069	38.7706
1.7169	53.0778
1.2568	63.5516
1.1540	73.1679
0.8827	80.5238
0.6936	86.3037
0.5734	91.0819
0.4304	94.6689
0.3533	97.6127
0.1798	99.1111
0.1067	100.0000

- Loadings for Window =3; shift =1

0.3381 -0.1095 0.4278 -0.1373 0.0907 0.5344 -0.0421 -0.1302 -0.0179 0.2691 -0.5273 0.0922
 -0.5355 -0.0342 0.2023 -0.0583 -0.2785 0.0521 0.1807 -0.2099 -0.0240 -0.0115 0.0051 0.7141
 0.4711 0.1396 -0.2427 -0.2784 0.1131 -0.1118 -0.3390 0.2001 -0.1859 -0.2360 0.0220 0.5926
 0.2152 -0.2627 0.5555 -0.0806 0.0298 0.1348 -0.0382 -0.0045 0.1963 -0.3152 0.6412 -0.0080
 -0.2994 -0.0256 -0.1895 -0.5566 -0.0658 0.1897 -0.5833 -0.2368 0.1624 0.2121 0.1633 -0.1712
 0.2124 0.3548 -0.0060 0.2331 -0.5166 0.2463 -0.0919 -0.2448 -0.5142 0.1360 0.2968 -0.0884
 0.0003 0.4941 0.0785 0.2177 -0.3281 0.1448 -0.2289 0.1653 0.6106 -0.2884 -0.1947 0.0048
 0.2231 -0.3974 -0.2900 0.3355 -0.2070 0.0616 -0.0557 0.1419 0.3696 0.5352 0.2031 0.2480
 -0.2777 0.3496 0.2647 0.1042 0.3504 0.1249 -0.1499 0.5723 -0.1517 0.3957 0.2146 0.0754
 0.1123 -0.0004 0.4578 -0.0741 -0.2741 -0.7230 -0.2576 -0.0249 -0.0095 0.2878 -0.1507 -0.0424
 -0.1710 -0.4490 0.0099 -0.0447 -0.4362 0.1450 -0.1722 0.5709 -0.2873 -0.2479 -0.1928 -0.1377
 0.1649 0.2144 -0.0517 -0.5927 -0.3011 0.0134 0.5714 0.2852 0.1421 0.1961 0.1021 -0.0718

eigenvalue	Cumulative sum of % variances explained
2.5907	21.5893
2.2755	40.5516
1.8957	56.3489
1.2492	66.7591
1.2104	76.8458
0.7672	83.2390
0.6604	88.7422
0.4249	92.2834
0.4061	95.6678
0.2772	97.9778
0.1721	99.4123
0.0705	100.0000

- Loadings for Window =6; shift=1

0.3277 0.1341 0.3307 -0.0481 0.2342 -0.6310 0.0833 -0.2754 -0.0558 0.4499 -0.0007 0.1471
 -0.5359 0.0558 0.1779 -0.2073 -0.1693 -0.0620 0.0541 -0.2252 -0.1248 -0.1727 -0.0579 0.7073
 0.4560 -0.1166 -0.2488 0.0875 0.4084 0.1916 0.1906 0.2006 0.1789 -0.1388 0.0718 0.6103
 0.1733 0.2900 0.5171 0.0577 0.1746 -0.0649 -0.1514 -0.0470 0.1935 -0.7075 -0.0620 -0.1156
 -0.3104 0.0394 -0.1944 -0.3352 0.6159 0.0394 0.4066 -0.2353 0.0646 -0.0676 -0.2629 -0.2694
 0.2864 -0.2990 0.0527 -0.2796 -0.3883 -0.2060 0.5920 0.1461 -0.2682 -0.2980 -0.1142 -0.0938
 0.0703 -0.4605 0.2435 -0.2749 -0.1762 0.1809 -0.0138 -0.2948 0.6914 0.1357 0.0250 -0.0142
 0.2191 0.4454 -0.2606 -0.1157 -0.3065 0.0320 -0.0701 -0.0270 0.2316 0.0850 -0.7087 0.0923
 -0.2651 -0.3886 0.2471 0.2352 0.1725 -0.2316 -0.0768 0.5475 0.1127 0.0880 -0.5014 0.0381
 0.1429 0.0581 0.5190 -0.0181 0.0810 0.6517 0.1287 -0.0102 -0.3656 0.2851 -0.2079 0.0193
 -0.1712 0.4688 0.1756 -0.2096 -0.0777 0.0110 0.3442 0.5346 0.3453 0.1995 0.3265 -0.0098
 0.1270 -0.0755 -0.0255 -0.7514 0.1478 -0.0285 -0.5187 0.2846 -0.1962 -0.0084 0.0363 0.0179

eigenvalue	Cumulative sum of % variances explained
2.8242	23.5353
2.6929	45.9762
1.9972	62.6199
1.2969	73.4270
1.0064	81.8141
0.7694	88.2260
0.5553	92.8538
0.2996	95.3502
0.2354	97.3117
0.1868	98.8682
0.1291	99.9437
0.0068	100.0000

- Loadings for Window =12; shift =1

0.0081 -0.4542 0.0647 0.4841 -0.3653 -0.0642 0.4118 -0.1318 -0.1607 -0.1820 0.1214 -0.3959
 -0.4214 0.1498 0.2523 -0.0646 0.3019 0.0134 -0.0898 0.2128 0.2891 -0.2487 0.3693 -0.5529
 0.4547 -0.1867 0.1095 0.1071 -0.0208 -0.0615 -0.0257 0.2920 0.3268 -0.0423 0.6377 0.3599
 -0.3021 -0.4294 -0.1693 -0.0653 -0.0145 -0.1267 -0.3365 -0.2904 0.2388 -0.5696 -0.0909 0.2964
 -0.3343 0.0826 0.4432 0.3128 0.1027 0.0386 0.3783 0.3182 0.0984 -0.1195 -0.3077 0.4606
 0.4561 -0.1807 0.0857 0.0222 0.0405 -0.0728 -0.1465 0.3044 0.4184 -0.0899 -0.5825 -0.3287
 0.2274 -0.2659 0.3254 -0.5047 0.2173 0.1252 0.1314 0.1377 -0.5341 -0.3641 0.0000 0.0000
 0.0405 -0.0902 -0.4918 0.0932 0.3083 0.7302 0.2728 0.0588 0.1365 -0.1221 0.0000 -0.0000
 0.0119 0.3158 0.1362 -0.2826 -0.7378 0.3873 0.0248 0.0224 0.1838 -0.2690 0.0000 0.0000
 -0.2068 -0.3848 0.0221 -0.5295 -0.0536 -0.0900 0.4412 -0.0844 0.3776 0.4159 -0.0000 0.0000
 -0.3251 -0.2978 -0.2714 -0.0231 -0.2698 0.0405 -0.2784 0.6957 -0.2484 0.1839 -0.0000 0.0000
 -0.0511 -0.3118 0.4972 0.1522 0.0156 0.5104 -0.4234 -0.2447 0.0057 0.3603 0.0000 0.0000

Eigenvalue	Cumulative sum of % variances explained
4.0386	33.6551
2.5374	54.8002
1.6825	68.8214
1.2192	78.9810
0.8815	86.3269
0.8325	93.2644
0.2978	95.7462
0.2513	97.8404
0.1587	99.1627
0.1005	100.0000
0.0000	100.0000
0.0000	100.0000

- Loadings for Window=3; shift =3;

0.2487	0.0063	0.4608	0.4155	-0.1256	0.2470	-0.1001	0.1035	0.5573	0.1219	-0.3376	0.1271
-0.4657	0.3377	-0.0039	0.0373	0.0714	0.3174	0.1556	0.0193	-0.0351	-0.1015	0.0472	0.7235
0.2876	-0.4533	-0.1766	0.2005	-0.0634	-0.1980	-0.3958	-0.2648	-0.2272	0.0098	-0.0354	0.5634
0.2268	0.2661	0.4640	0.3451	0.0724	0.1046	-0.1703	0.1264	-0.5301	-0.0866	0.4328	-0.0793
-0.2306	0.1288	-0.4498	0.2007	-0.2050	0.2983	-0.6526	0.1766	0.1544	-0.0929	0.1494	-0.2115
0.0230	-0.4081	0.2136	-0.3033	0.1437	0.5512	-0.0448	-0.3758	0.1707	-0.2987	0.3232	-0.0691
-0.2530	-0.4112	0.2275	-0.2573	-0.0249	0.2520	-0.1638	0.5044	-0.3694	0.2472	-0.3256	-0.0048
0.4273	0.1081	-0.0896	-0.4415	-0.0789	0.0184	-0.0397	0.3966	0.2401	0.3276	0.4447	0.2731
-0.4489	0.0123	0.3290	-0.0518	-0.0506	-0.2689	-0.2639	-0.3556	0.1349	0.5725	0.2637	-0.0313
-0.0354	-0.0188	-0.0188	0.0479	0.9153	-0.1676	-0.2393	0.1888	0.1865	-0.0345	-0.0249	0.0189
0.2895	0.3946	-0.1455	-0.2021	0.1914	0.3800	-0.1518	-0.3946	-0.2324	0.3837	-0.3651	-0.0643
-0.0171	-0.2990	-0.3219	0.4769	0.1529	0.2943	0.4169	0.0294	-0.0397	0.4742	0.2473	-0.0839

eigenvalue	Cumulative sum of % variances explained
2.8608	23.8396
2.2642	42.7078
1.6610	56.5497
1.5987	69.8726
1.0800	78.8724
0.9322	86.6409
0.6053	91.6850
0.4041	95.0526
0.3170	97.6940
0.1977	99.3418
0.0744	99.9616
0.0046	100.0000

- Loadings for Window =6; shift =6

0.2264	-0.0327	0.2786	0.3265	0.6512	-0.0560	-0.4200	0.1218	-0.1774	-0.2870	0.0871	0.1537
-0.4134	0.1728	-0.0313	0.2391	-0.0091	-0.1514	0.2930	0.1477	0.2199	-0.1466	0.3045	0.6692
0.3893	-0.2370	-0.2552	-0.0364	0.0185	-0.1551	0.2178	-0.1684	-0.2429	0.0007	-0.5325	0.5320
0.1873	0.1630	0.5438	0.3797	0.0082	0.3184	0.5804	-0.1759	-0.0160	0.0344	-0.1245	-0.1104
-0.1667	0.2587	-0.4840	0.4700	0.0705	-0.2453	0.1678	0.0893	-0.1451	-0.3043	-0.2637	-0.4070
0.0479	-0.5336	0.0887	-0.1629	0.2856	-0.4910	0.4423	0.0424	0.1981	-0.0838	0.2112	-0.2601
-0.2363	-0.4891	-0.0053	0.0562	-0.1253	0.4101	0.1183	0.5506	-0.4343	-0.1098	-0.0079	0.0000
0.4309	0.0602	-0.0789	-0.1146	-0.1565	0.2193	0.0135	0.3830	0.5545	-0.5034	-0.1046	0.0000
-0.4071	-0.0310	0.3103	-0.0410	0.1603	-0.1250	-0.1268	0.2587	0.3361	0.2080	-0.6744	-0.0000
-0.0117	-0.2903	0.3254	0.3377	-0.6198	-0.3462	-0.2875	-0.1779	-0.0086	-0.2755	-0.0370	-0.0000
0.3554	0.3031	0.1385	0.0336	-0.1959	-0.4110	0.0465	0.5876	-0.2072	0.3928	0.0973	0.0000
0.1654	-0.3388	-0.3015	0.5569	0.0217	0.1631	-0.1203	0.0259	0.3854	0.5063	0.0989	-0.0000

Eigenvalue	Cumulative sum of % variances explained
4.7794	39.8285
2.5936	61.4415
1.6013	74.7859
1.2937	85.5670
1.0577	94.3811
0.3081	96.9487
0.2058	98.6635
0.0808	99.3370
0.0672	99.8965
0.0118	99.9949
0.0006	100.0000
0.0000	100.0000

- Loadings for Window =12; shift =12;

-0.1410 0.4800 -0.1966 -0.3271 -0.1978 -0.0332 -0.0019 -0.3448 -0.1609 0.1956 0.0587 -0.6143
 -0.3313 -0.2580 -0.2364 0.2799 0.2674 0.0064 -0.5004 0.0050 0.2922 -0.2630 0.1641 -0.4313
 0.4232 0.0318 -0.1780 -0.0185 0.2323 0.8334 -0.1036 0.0246 -0.1261 0.0549 0.0437 -0.0841
 -0.3716 0.1997 -0.0981 0.2317 0.5407 0.0870 0.4250 -0.0158 0.0313 0.1288 -0.5088 -0.0377
 -0.2716 0.0358 -0.4490 -0.2084 -0.3464 0.2782 0.1390 -0.1241 0.5716 -0.0527 -0.0002 0.3441
 0.4171 -0.0333 -0.1628 0.0833 0.3694 -0.3055 0.0309 -0.4702 0.3371 0.3931 0.2446 0.1077
 0.1830 -0.4426 -0.3149 -0.0252 -0.0540 -0.0902 0.6548 0.0944 -0.0825 -0.2538 0.1697 -0.3494
 0.2134 0.4305 -0.2891 0.1143 0.0006 -0.2002 -0.0168 0.7366 0.2122 0.1628 0.1103 -0.0770
 0.0294 -0.1328 0.3710 -0.7080 0.3149 0.0082 0.0179 0.2094 0.3952 -0.0121 -0.0878 -0.1874
 -0.2403 -0.4841 -0.0928 -0.0349 -0.1517 0.0722 -0.0783 0.2017 -0.1214 0.7760 -0.0392 -0.0678
 -0.3966 0.1555 0.2201 0.0086 0.2168 0.1652 0.2431 0.0691 -0.0927 0.0713 0.7736 0.1401
 -0.1028 -0.0480 -0.5085 -0.4404 0.3385 -0.2125 -0.2083 0.0542 -0.4474 -0.1317 0.0159 0.3328

Eigenvalue	Cumulative sum of % variances explained
4.9765	41.4707
2.8991	65.6301
2.6572	87.7734
1.0775	96.7527
0.3897	100.0000
0	100.0000
0	100.0000
0	100.0000
0	100.0000
0	100.0000
0	100.0000
0	100.0000
0	100.0000