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

Fernanda Strozzi[1], Chiara Rossignoli[1] and José-Manuel Zaldívar[2]

CONTENTS

1. Introduction
2. Data provision and treatment
   2.1. Data provision
      2.1.1. Electricity spot prices
      2.1.2. Disturbances and total consumption
   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
   2.3. Principal component analysis
3. Time series analysis
   3.1. Correlation matrix
   3.2. Cross correlation function
   3.3. Cross recurrence plot
4. Conclusions

[1] Cattaneo University-LIUC, C.so Matteotti, 22, 21053 Castellanza (VA), Italy
[2] European Commission, Via E. Fermi, 1, Joint Research Centre, 21028 Ispra (VA), Italy

1. Introduction

A considerable amount of literature is dedicated to the properties of electricity prices such as high volatility (Simonsen, 2005; Regnier, 2007; Strozzi *et al*., 2008; Bask and Widerberg, 2008) and long range correlation (Weron and Przybyłowicz, 2000; Simonsen, 2003; Bask et al. 2007). An empirical analysis of electricity prices in the European Energy Exchange and the Nord Pool Power Exchange (www.nordpool.com) was performed by Redl *et al.* (2008). Erzgräber *et al*. (2008) checked long range correlation by using different methods to calculate the Hurst exponent whereas volatility measurements based on Recurrent Quantification Analysis (RQA) were introduced by Strozzi *et al.* (2008).

Similarly, many studies on disturbances in electric power systems have been conducted. Carreras et al. (2004) studied 15-year time series of North American electric power transmission system blackouts and recognised that they show several characteristics of complex time series such as fat-tail distributions and self organised criticality (SOC). Moreover, they found that rescaled range analysis of the time series shows moderate long-time correlations. The fat tail distribution of these data was also confirmed by Weron and Simonsen (2006).

As far as we know, the relationship between prices and grid disturbances has never been analysed in detail although being recognised that it should exist (Zhao, 2007; Bjørndal, Jörnsten, 2006). In this work we have searched for correlations between electricity prices and disturbances using the data of the Nordic electricity market. The study of the interdependence between the physical network and the market for electricity is a preliminary step towards a deeper understanding of this market's mechanism. In particular, the knowledge of the influence of economic factors on disturbances could help in their prevention or, at least, in their management.

Nordel (www.nordel.org) is the organisation of the Transmission System Operators (TSOs) of Denmark, Finland, Iceland, Norway and Sweden. The core duties of the TSOs include: ensuring the operational security of the power system, maintaining the balance between supply and demand, adjusting the short-term and long-term adequacy of the transmission system and enhancing the efficient functioning of the electricity market. The participants (apart from Iceland, which will therefore be excluded from our analysis) benefit from a common Nordic wholesale electricity market (Nord Pool) 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. Nord Pool ASA - the Nordic Power Exchange - is the world's only multinational exchange for trading electric power. The ownership of Nord Pool is shared by the Nordic transmission system operators (TSOs). The power exchange Nord Pool Spot organises the physical trade of

electricity, the day-ahead market *Elspot* in the Nordic countries and KONTEK (the TSO area of Vattenfall Europe Transmission GmbH) in Germany. It is a part of the Nord Pool Group and provides a market place for producers, distributors, industrial companies, energy companies, trading representatives, large consumers and TSOs on which they can buy and sell physical power. (For a detailed description of the electricity price formation see www.nordpool.com.)

The degree of integration of the different national markets that constitute the Nordic electricity market has been studied by Amundsen and Bergman (2007). There are considerable differences in the methods by which electricity is generated in these countries. In Norway, nearly all electricity is generated from hydropower. Sweden and Finland each use a combination of hydro, nuclear and conventional thermal power, hydropower stations being located mainly in northern areas with thermal power prevailing in the south. Denmark relies mainly on conventional thermal power but wind power is providing an increasing part of its energy supply.

The amount of energy from the various sources changes with weather conditions and consequently the electricity prices. The disturbances associated with problems in the interconnection or in the overload of the electric grids are similarly affected by the weather, in particular by lightning. Equipment faults and repairs are also important causes of grid disturbances. In Nordel Net reports, they define a disturbance as an "*outage, forced or unintended disconnection or failed reconnection as a result of faults in the power grid*". A disturbance may consist of a single fault or, for example, an initial fault followed by secondary faults. Economic factors could also be a source of disturbances. As pointed out by Mork (2001), the emergence of a financial market for electricity can increase price volatility due to speculations on future movements. Moreover, the risk of default is increased by the reduced incentive for the producers to maintain reserve capacity and by the presence on the market of sellers with no generating capacity.

In this work we were interested in addressing the following questions:

- Are monthly spot prices correlated with disturbances?
- Can we use the evolution of one time series to anticipate the behaviour of the other?
- Can we detect windows of correlation between the two series?

To address these questions, we used data on monthly spot prices, disturbances and consumption in the Nordic region (i.e. Denmark, Finland, Norway and Sweden) from the beginning of January 2000 until the end of December 2006.

This paper is organized as follows. In Section 2 we have described the preliminary data treatment. It included the elimination of trends by applying the difference operator and subtracting the regression line. In addition, we considered the price volatility and the volatility of disturbances and of total consumption. Starting from the three initial time series of prices,

disturbances and consumption we obtained a set of twelve time series. As is well known, the correlation may change if we observe it on different data windows. For this reason we have considered the mean of the time series (or the standard deviation in the case of volatilities) on different time windows overlapped or not. In Section 3 we have studied the correlation matrices of the twelve time series. The analysis of the set of time series becomes deeper when we calculate the eigenvectors of the correlation matrices i.e. the principal components (Jolliffe, 1996) that allow one to identify how many degrees of freedom, i.e. independent variables, a possible model of the twelve time series may have. The main correlations between prices and disturbances using the mean on some time windows are underlined. Then we checked if the correlation increases by 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 at least 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 conclusions are presented.

2. Data provision and treatment

2.1 Data Provision

In Table 1 we observe the evolution of the composition of Nord Pool during these years. The data sets are monthly electricity prices (*S*), monthly disturbances (*D*) and total monthly consumption (*T*) in Denmark, Finland, Norway and Sweden from January 2000 to December 2006. All the data are public. The Electricity spot prices are available on the Nord Pool (Nordic Power Exchange) web page:

www.nordpool.com/nordpool/financial/index.html.

The disturbances and total consumption are available on Nordic statistics of electricity faults in the Nordel web page:

www.nordel.org/content/Default.asp?PageID=214.

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

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

2.1.1. Electricity Spot Prices

In Figure 1 the monthly mean electricity spot prices (*S*) of Nordic Region (Nord Pool countries) and the standard deviation are plotted. The series lasts from 1st January 1999 until 26th January 2007, the prices are expressed in EUR/MWh. This price is the "system" price, also called "the unconstrained market clearing price", because it is the price that balances sales and purchases in the exchange area whilst ignoring every transmission constraints.

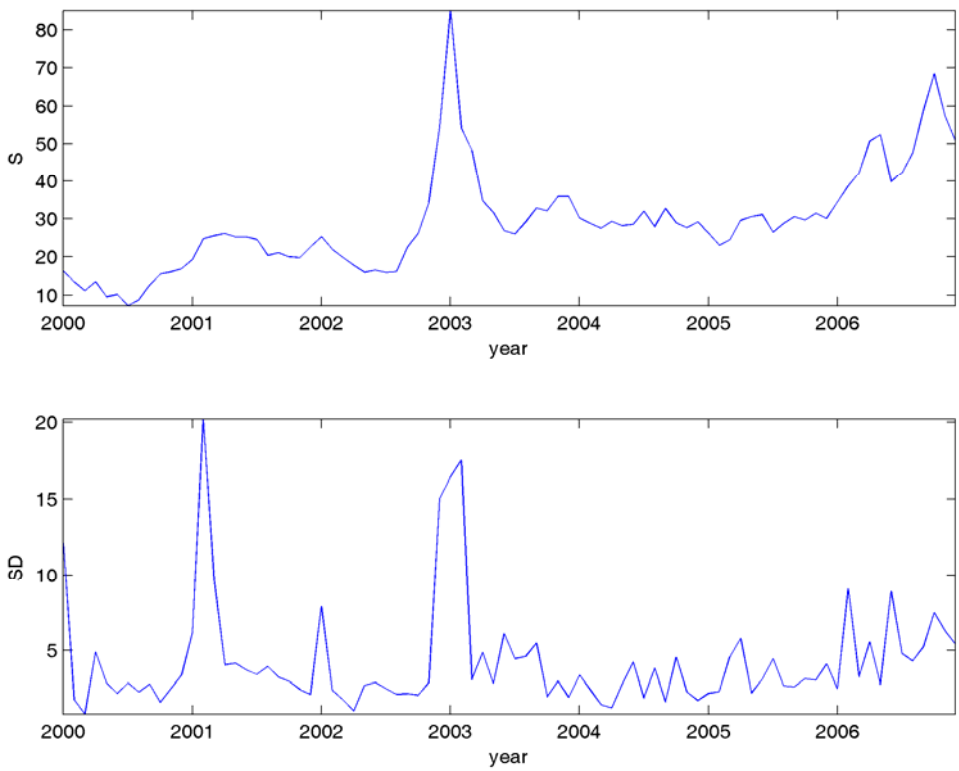


Figure 1. Monthly mean spot prices (in EUR) from the hourly time series data and standard deviation (SD).

2.1.2. Disturbances and Total Consumption

In Figure 2 the number of grid disturbances (*D*) from the beginning of 2000 until the end of 2006, as a function of time is represented. The grid considered is the 100-400kV network. In all the countries, the number of disturbances increases during the summer period. This is normally caused by lightning.

In Figure 3 the monthly Total Consumption (*T*) in the Nordic region between January 2000 until December 2006 is represented (Iceland was not considered because it is not a member of Nord Pool and it is not included in our price data).

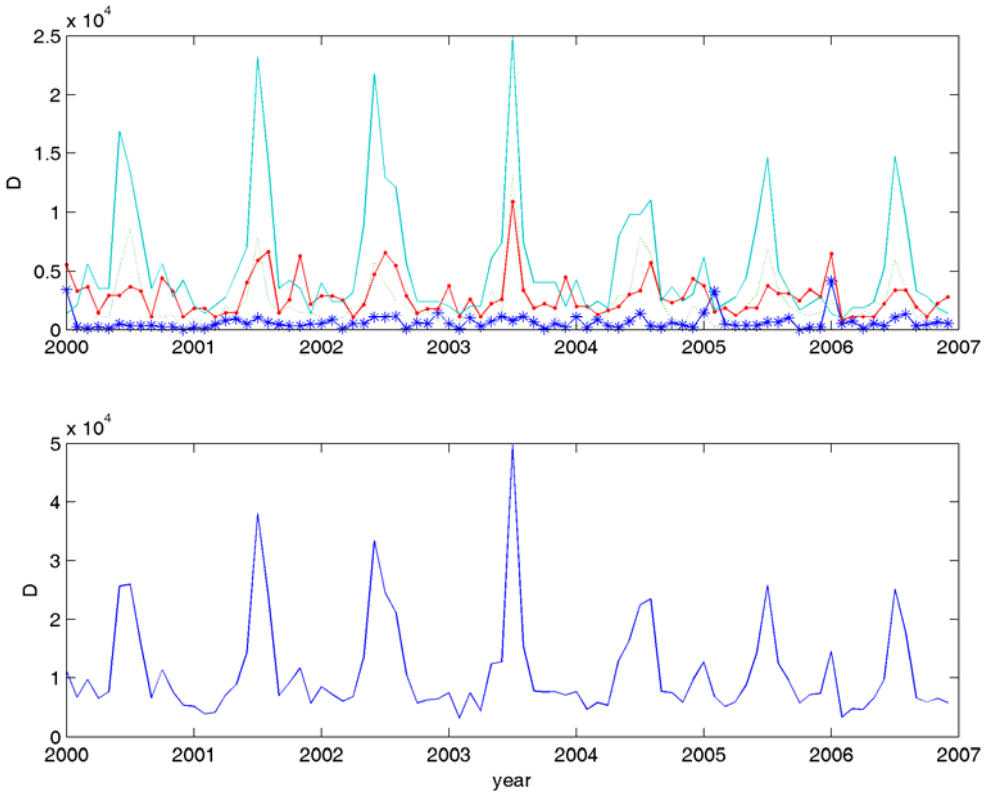


Figure 2.Top: Number of disturbances in Denmark(*), Finland(¨), Norway(.-) and Sweden(-); bottom: Total number of disturbances (D) in Denmark Finland Norway and Sweden.

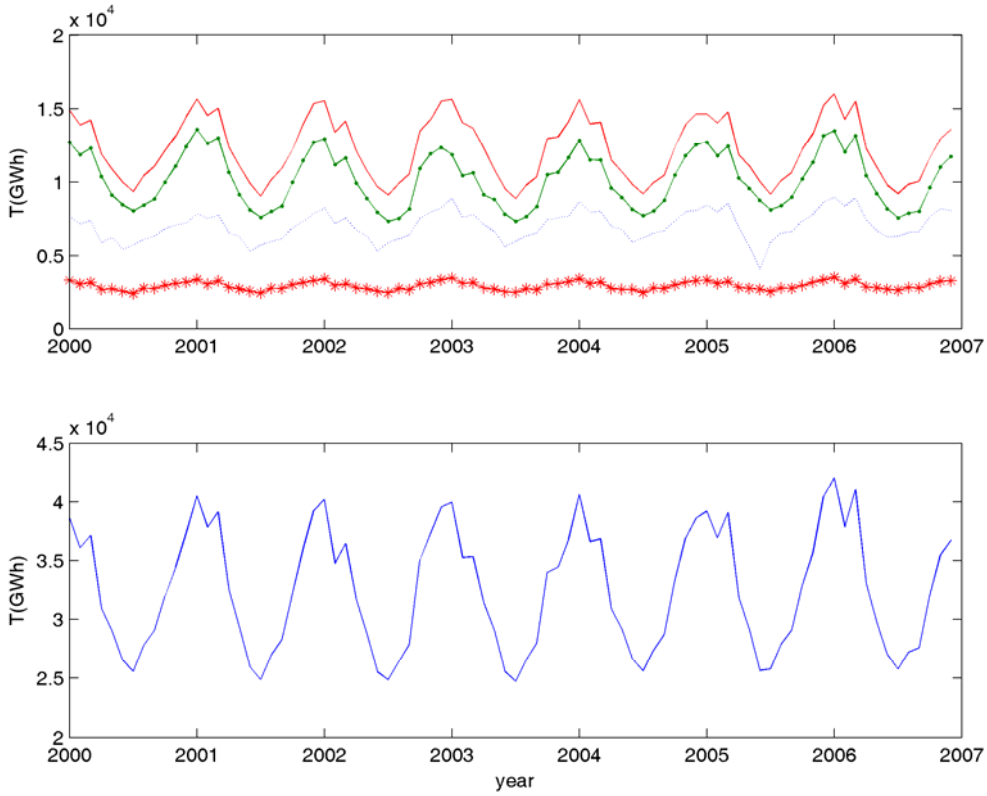


Figure 3. Top: Total Consumption of Denmark(*), Finland(:), Norway(.-) and Sweden(-); bottom: Total Consumption (T) in Nordel States except Iceland.

2.2. Data treatment

2.2. 1. Data trend and seasonality

We have treated the data (S, *D* and *T*) by subtracting the linear trend and the seasonality. The trend has been removed by subtracting the linear regression line and the seasonality by subtracting the mean value of the given time series in the corresponding month of every year. In the first column of Figure 4 the resulting time series are plotted.

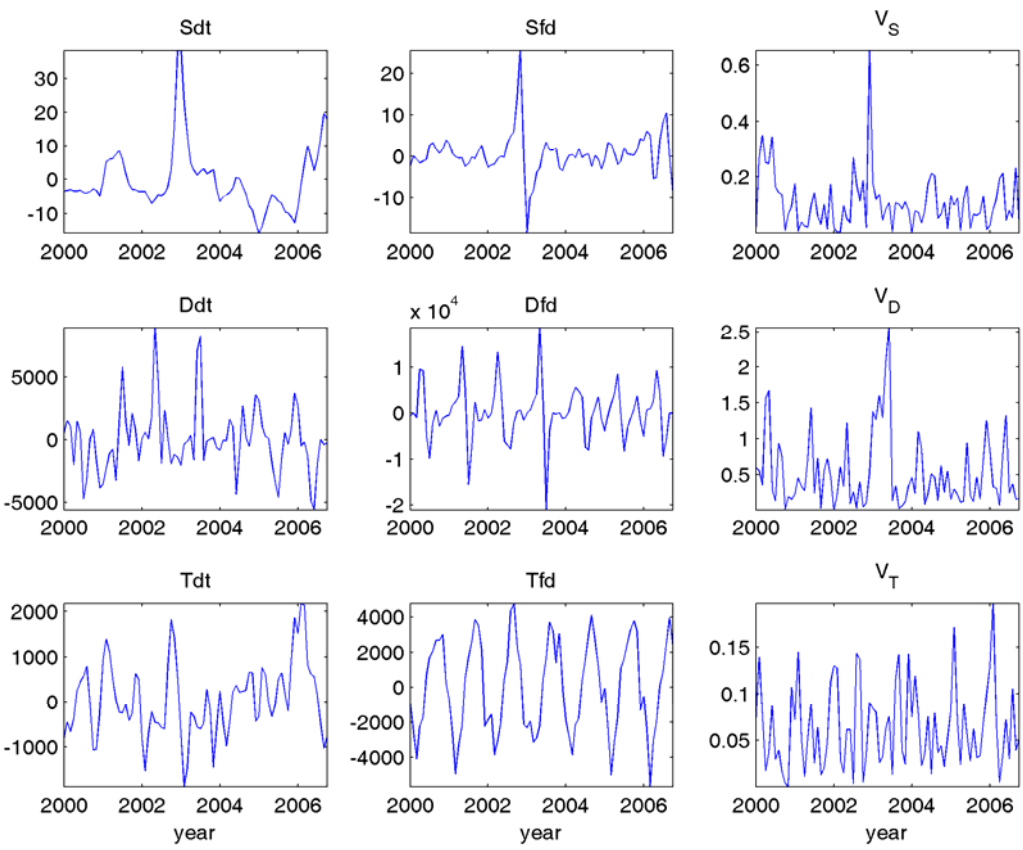


Figure 4. First Column: detrended data sets from *S*, *D* and *T*; second column: first differences; third column: volatilities (Δt=1 month), *w*=2 months, *sh*=1 month

2.2.2. Data first differences

The difference operator is normally applied to eliminate the trend. In the second column of Figure 4 we have represented the first differences of the monthly mean Spot prices, Disturbances and Total Consumption.

2.2.3. Data Volatilities

Price volatility in the Nordic electricity marked was analysed by Strozzi *et al.*, 2008. Volatility can be calculated using the sample standard deviation *SD*, defined as: $SD(X) = \frac{1}{n-1}\sum_{t=1}^{n}\left(X_t - \overline{X}\right)^{1/2}$, of the spot price returns (S), Total Consumption (T) and Disturbances (D):

$$V_X = SD\left(\frac{X(t) - X(t-\Delta t)}{X(t-\Delta t)}\right) \qquad (1)$$

where Δt is one month and the window on which we calculate the standard deviation changes from one month to twelve months. In the case of a one month window where the standard deviation cannot be calculated, we consider simply:

$$V_X = \frac{X(t) - X(t-\Delta t)}{X(t-\Delta t)} \qquad (2)$$

which are linear approximations respectively of: $ln(X(t)/X(t-\Delta t))$ i.e. the logarithm first differences. As an example, in the third column of Figure 4 the three volatilities are represented on a window (*w*) of two months translated by a shift (*sh*) of one month.

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

Table 2. Notation of the twelve time series analyzed.

Label	Definition	Label	Definition
S	Mean monthly spot prices	**Sfd**	S first differences
D	Monthly disturbancies	**Dfd**	D first differences
T	Monthly Total Consumption	**Tfd**	T first differences
Sdt	S detrended	$\mathbf{V_S}$	w=1 diff(ln(S)) w>1 Volatility of S
Ddt	D detrended	$\mathbf{V_D}$	w=1 diff(ln(D)) w>1 Volatility of D
Tdt	T detrended	$\mathbf{V_T}$	w=1 diff(ln(T)) w>1 Volatility of T

2.2.4. Time windows and shifts

We have analysed 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 have been 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) present in several of these time series. 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 methodology.

Table 3. Windows and shifts considered in this work. The unit time of windows (w) and shifts (sh) are months.

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

2.3. Principal Component Analysis

To assess the amount of new information introduced by this nine time series (*Sdt*, *Ddt*, *Tdt*, *Sfd*, *Dfd*, *Tfd*, V_S, V_D and V_T) with respect to *S*, *D* and *T*, Principal component analysis (PCA) (Jackson, 1991; Jolliffe, 2002) has been performed on the twelve time series using the windows and shifts represented in Table 3. The eigenvalues and the percentage of variance explained by each component when *w* and *sh* change have been calculated. The main results of this analysis are presented in Table 4. 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 a hyperplane of dimension three can fit the data. This is expected 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 in column six that we always need more than 3 principal components. Sometimes even 8 principal components are necessary.

Table 4. Summary of PCA results. The unit time of windows (w) and shifts (sh) are months.

w	**sh**	**# points**	**#PC to explain 50% variance**	**% variance explained**	**#PC to explain 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

3. Time Series Analysis

First we studied the linear correlation coefficients i.e. the entries of the correlation matrix; then we checked if these linear correlation coefficients could increase by shifting one series with respect to the other by measuring the Cross Correlation Function (CCF). Finally we

applied the Cross Recurrence Plot (CRP) analysis (Marwan *et al.*, 2007) which provides a tool- Line Of Synchronization (LOS) - that allows one to identify time windows in which two time series are linearly correlated. 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:

$$R(X,Y) = \frac{COV(X,Y)}{\sqrt{VAR(X)VAR(Y)}} \qquad (3)$$

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

$$COV(X,Y) = \frac{\sum_{t=1}^{n}\left(X_t - \overline{X}\right)\left(Y_t - \overline{Y}\right)}{n-1} \qquad (4)$$

where $\overline{X}$ and $\overline{Y}$ are the means of the two variables and n is the components number.

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

We calculated the matrix R of the correlation coefficients R for all the time series of Table 2 and for each window and shift indicated in Table 3. The correlation coefficients and their significance using t-test (O'Mahony, Michael,1986) have been calculated with Matlab® command *corrcoef* that transform the correlation to create a t-statistic having n-2 degrees of freedom, where n is the size of data set. The confidence bounds are based on an asymptotic normal distribution of 0.5*ln((1+R)/(1-R)). In Table 5 the entries of the correlation matrix R for which the correlation values are higher than 0.7071 (i.e. a determination coefficient $R^2 > 0.5$) with a significance level of 95% i.e. $P < 0.05$ are presented.

Table 5. Significant linear correlations coefficient *R* and t-test values *P* between data sets for the analyzed windows (*w*) and shifts (*sh*). The unit time of windows (w) and shifts (sh) are months.

Time series	**w**	**sh**	**R**	**P**
S- Sdt	1	1	0.7317	0.0000
Sfd-V_S	1	1	0.8607	0.0000
Dfd-V_D	1	1	0.8761	0.0000
Tfd-V_T	1	1	0.9896	0.0000
D-T	2	1	-0.7354	0.0000
S-Sdt	2	1	0.7195	0.0000
D-T	3	1	-0.8057	0.0000
D-T	3	3	-0.8514	0.0000
D-T	6	1	-0.9044	0.0000
Dfd-Tfd	6	1	-0.8010	0.0000
D-Dfd	6	6	-0.8503	0.0002
T-Tfd	6	6	-0.8686	0.0001
V_D*-Dfd*	6	6	0.7698	0.0021
D-T	6	6	-0.8594	0.0002
D-Tfd	6	6	0.776	0.0022
T-Dfd	6	6	0.7752	0.0019
D-T	12	1	-0.7807	0.0000
D-Tdt	12	1	0.8060	0.0000
T-Tdt	12	1	0.9904	0.0000
T-Tdt	12	12	0.9842	0.0004
T-V_D	12	12	-0.9057	0.0129
Tdt-V_D	12	12	-0.9014	0.0141
***Sfd-*V_D**	**6**	**6**	**-0.7778**	**0.0017**
***Sdt-*V_D**	**12**	**1**	**0.7567**	**0.0000**
***Sdt-*V_D**	**12**	**12**	**0.8138**	**0.0488**

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 (Vs, 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.

Considering the correlations between prices and disturbances we can conclude that it exists only for *w*=12 and *sh* = 1 or *sh* = 12, particularly between the mean Spot prices de-trended and the volatility of the disturbances.

3.2. 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 on of them in respect to the other (Orfanidis, 1996). Let us consider two time series X_t and Y_t where t=1, 2...n. The cross correlation *R* at delay *d* is defined as

$$R(X,Y,d) = \frac{\sum_t [(X_t - \overline{X})(Y_{t-d} - \overline{Y})]}{\sqrt{\sum_t (X_t - \overline{X})^2}\sqrt{\sum_t (Y_t - \overline{Y})^2}} \quad (5)$$

where $\overline{X}$ and $\overline{Y}$ are the means of the corresponding series. If Eq. 5 is computed for all delays *d* =-(n-1),..0, 1, 2,...(n-1) then it results in a cross correlation series of values of twice the length of the original series. For *d* = 0 this formula reduces to 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. The maximum values obtained are listed in Table 6 together with the correlation coefficients without delay, *R*(i,j,0), and the *P* values of the t-tests. It is possible to observe that there are significant correlations between Spot Prices and Disturbances (or their derived time series) only on windows of 6 or 12 months. In addition, the increase of correlation between price volatility, V_D, and V_S, *Sdt* and *Sfd* becomes higher than *R*=0.8. As an example, in Figure 5 we have plotted CCFs obtained using *D-S* and *D-Sfd* in which 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 as well as the static properties. Similar results are observed between *D-T*, *D-Tfd*, using *w*=2 and *sh*=1 where, in both cases, the correlation becomes higher than 0.6 (results not shown).

Table 6. Results from the cross correlation analysis.
The unit time of windows (w) and shifts (sh) are months.

Time series	w	sh	R(0)	P	delay (months)	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

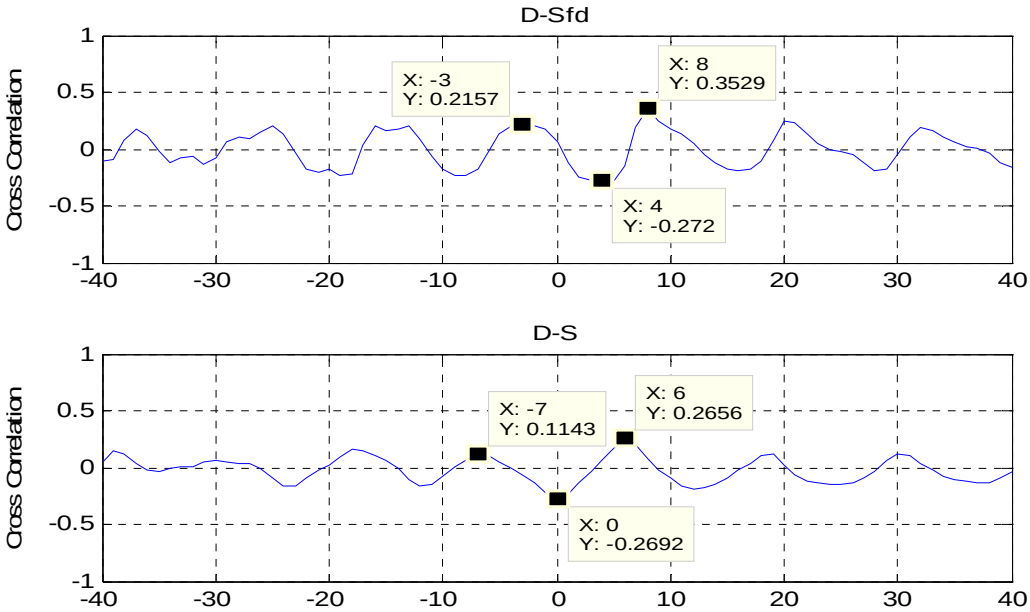


Figure 5. Cross Correlation functions for Disturbances with $w = 2$, $sh = 1$.
The unit time of windows (w) and window shift (sh) are months.

3.3. Cross Recurrence Plot (CRP)

CRP was introduced to analyse the dependencies between two different time series by comparing their cross recurrence (Marwan and Kurths, 2002). It can be considered as a generalization of the linear cross-correlation function (Marwan et al., 2007), see Figure 6.

If we have two time series, $\overrightarrow{x_i}, \overrightarrow{y_j}$, the CRP matrix is defined by:

$$\mathbf{CR}_{i,j}^{\overrightarrow{x_i}\overrightarrow{y_j}}(\varepsilon) = \Theta(\varepsilon - \left\|\overrightarrow{x_i} - \overrightarrow{y_j}\right\|) \qquad (6)$$

where *i=1, …, n; j=1, …m.*

To quantify 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 major 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 dynamic systems which cannot be detected by the usual cross-correlation. If a time dilatation or compression of one of the trajectories is applied, then a distortion in the diagonal lines appears in the CRP.

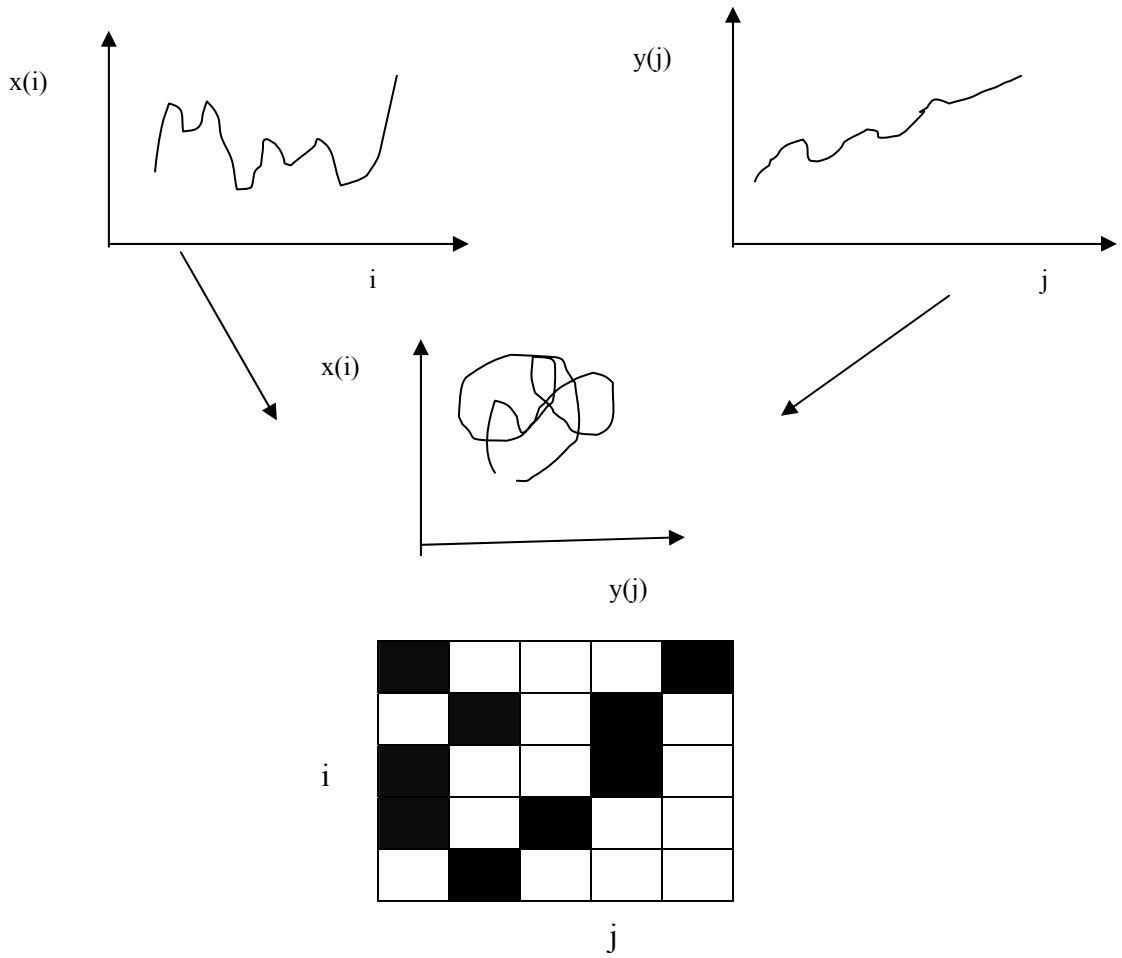


Figure 6. Cross Recurrence plot (CRP) construction.

To understand the potentiality of this technique we discuss some examples (see also Marwan and Kurths, 2002). If we plot two identical time series the CRP will contain the main diagonal line called Line of Identity (LOI). If we consider a time distortion in the second time series then the LOI will not be the main diagonal but it will have a different angle. This new line is called the line of synchronization (LOS). The LOS will be a straight line but not parallel to the main diagonal and the local slope in the CRP will correspond to the transformation of the time axes of the two considered time series. A time shift between the trajectories causes a dislocation of the LOS, therefore LOS allows one to find the rescaling function between different time series (Marwan and Kurths, 2002). For these reasons CRP correlation detection is particularly useful

when the correlation between two time series changes in time. This is because the technique is capable to detect the correlation window. Let us consider the following example:

- *x (t)=sin(πt)*, if t=[-1:0.01:1];
- *y (t)=cos(πt)*, if t=[-1:0.01:0] and *y (t)=cos(πt/2)* if t=[0:0.01:1].

The CRP and the LOS are shown in Figure 7. The linear correlation coefficient of the two complete series is *R*= 0.1687 with *P*=0.0000. Using the window suggested by LOS, i.e. x(10:109), x(1:100), R becomes: 0.9048 with *P*=0.0000. If we calculate *R* on the remaining parts: x(110:201), y(101:201-9) *R* becomes 0.2851 with *P*=0.0059.

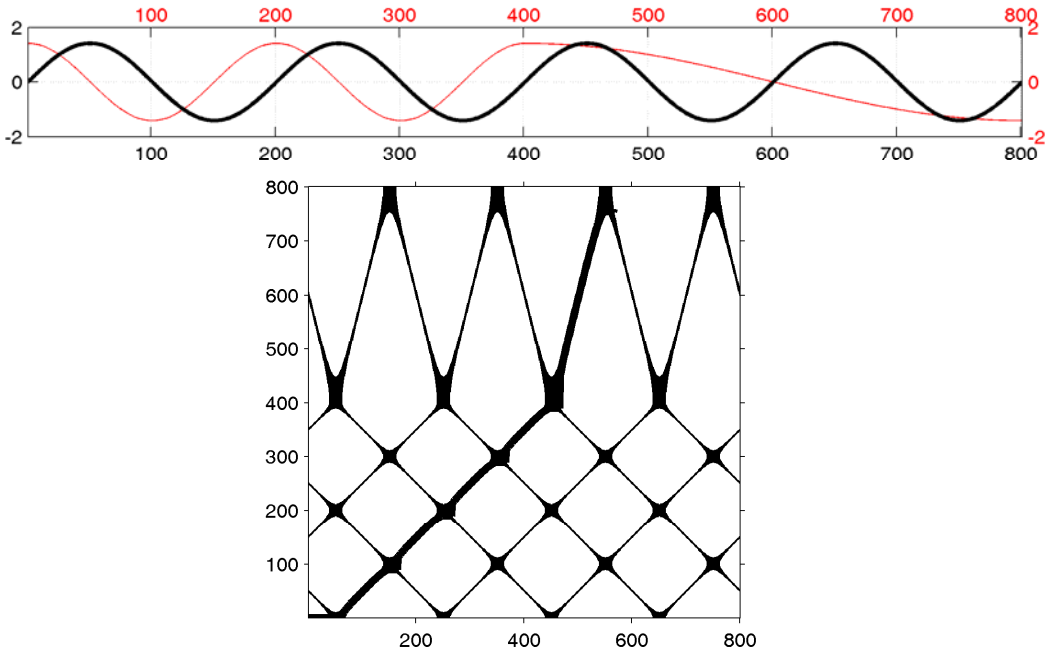


Figure 7. CRP when a change in the correlation between two time series occurs.
The two time series: sinusoidal (bold) and cosinusoidal, with a change in frequency after 400 unit of independent variable, are represented (top). CRP is represented together with LOS (bold line) (bottom)

The Line of Synchronization algorithm is presented in Marwan *et al.* (2007) and it consists of an iterative search of recurrent points in the 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 then 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 they introduced the following indicator as the LOS Quality:

$$Q = \frac{Nt}{Nt + Ng} \cdot 100 \qquad (7)$$

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

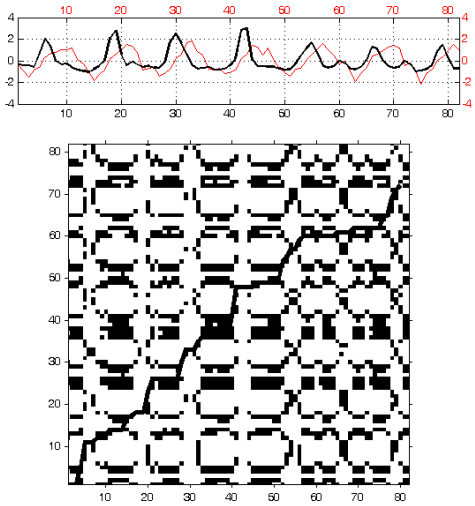
The only disadvantage of using CRP is that to obtain a good LOS quality, which means that information given by LOS shows real changes in the correlation properties, there is the need of a certain minimum number of points. In this work we have been able to obtain good LOS quality using only data with *w*= 2 and *sh* =1; in the other cases there were not enough points to perform this analysis. Therefore, even though the CCF showed correlations between V_D and *Sfd*, V_D and *Sdt*, V_D and V_S considering window of six or twelve months, these time series do not have enough points to show a reliable Line of Synchronization.

In Figure 8 we have represented CRP for the series of Disturbances (*D*) in respect of 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 correlations between the time series on some time windows, correlations that are not clear from the correlation function. To confirm the hypothesis that LOS allows 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 7.

Table 7. Correlation coefficient for different portion of the time series. *R*: correlation Coefficient of the entire time series 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*.

Q	**months considered for each time serie**	**R**	**R_{CCF}**	**Months Interval suggested by LOS**	**R_{LOS}**	**Dates correspondent to the points considered**
69.89	D(1:83); T(1:83)	-0.7354	-0.7354	D(1:30); T(1:30)	-0.8037	Feb 00-Sept 01
64.32	D(1:83); Sfd(1:83)	0.0702	-0.3529	D(1:30); Sfd(1:30)	-0.3953	Feb 00-July 02
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

Moreover, looking to Table 7, we can observe that LOS allows one to identify the time in which Spot Prices change for the start of the dry period (July 2002) and in which the prices increase due to the dependence on external energy sources.



a) Tfd-D, Q=80.40

b) Sfd-D, Q=64.32

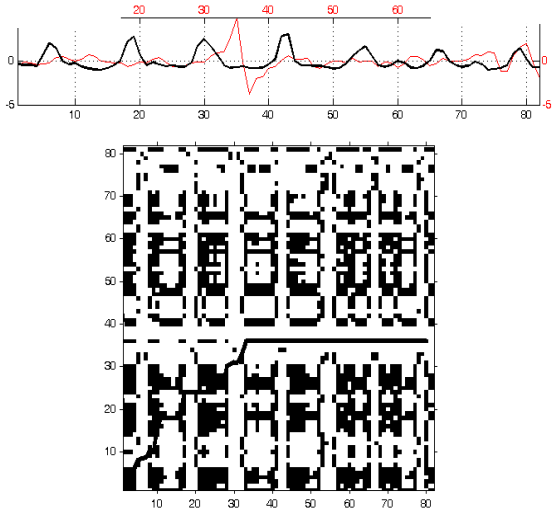


Figure 8. CRP with *w=2; sh=1*; underlying time series on the top (D in bold).
The LOS can be observed in the CRPs (bold line).

4. Conclusions

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 (Table 2). The set of 12 time series is then grouped using different time windows and translated by different shifts (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 and six months (Table 5) between the volatility of disturbances (V_D) the de-trended spot price (*Sdt)* and

with first price differences (*Sfd*). A t-test of significance using *corrcoef* Matlab® command is applied and the corresponding *p* values are listed. In Table 6 it is shown how some of these correlation values can increase shifting the corresponding time series of a delay suggested by Cross Correlation Function. Only in the case of a window of 12 translated of 12, where we have 6 data points (see Table 4), doubts exists that this correlation exists even if the correlation values and the corresponding significance tests detect it.

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, Total Consumption and Spot prices).

Disturbance volatility is not so much linearly correlated with the price itself, but this does not exclude a non linear correlation. In Figure 5 we have plotted the cross correlation functions between *D-Sfd* and *D-S* which, even if they never reache values higher than 0.4, they has a regular oscillating behaviour in respect to the delay and this can be a sign of similarity between the dynamics *D-Sfd* and *D-S* respectively.

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 number 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 7 it is shown how, using CRP, higher correlated data windows are detected using Line of Syncronization (LOS). However, we do not have enough points to apply CRP and obtain reliable information with fewer data points. For this reason it would be interesting to repeat the analysis performed in this work using daily data of disturbances and consumption (electricity spot prices are hourly).

Acknowledgements:

FS is grateful to MANMADE NEST Project (Contract No 043363) and Fondazione Cariplo for their financial support. We have used the CRP toolbox which is free downloadable from www.agnld.uni-potsdam.de/~marwan/toolbox/ .

References

Amundsen E.S., Bergman L. (2007), Integration of multiple national markets for electricity: The case of Norway and Sweden', *Energy Policy*, **35(6)**, 3383-3394.

Bask M., Liu T., Widerberg A. (2007), 'The stability of electricity prices: Estimation and inference of the Lyapunov exponents', *Physica*, **A 376**, 565-572.

Bask M., Widerberg A. (2008), 'Market structure and the stability and volatility of electricity prices', *Energy Economics* (in press).

Bjørndal M., Jörnsten K., (2007), 'Benefit from coordinating congestion management - The Nordic power market', *Energy Policy*, **35**, 1978-1991.

Carreras B.A. *et al.* (2004), 'IEEE Trans.', *Circuits and Systems*, **51(9)**, 1733-1740.

Erzgräber H. et al. (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., Kurths, **J.** (2002), 'Nonlinear analysis of bivariate data with cross recurrence plots', *Physics Letters*, **A 302**, 299–307.

Marwan N. *et al.* (2007), 'Recurrence Plots for the Analysis of Complex Systems', *Physics Reports*, **438**, 237–329.

Mork E., (2001), 'Emergence of financial markets for electricity: a European perspective', *Energy Policy*, **7-15**.

O'Mahony M. (1986), *Sensory Evaluation of Food: Statistical Methods and Procedures*, CRC Press, pp. 487. ISBN 0-824-77337-3.

Orfanidis S.J. (1996), *Optimum Signal Processing. An Introduction*, 2nd Edition, Prentice-Hall, Englewood Cliffs, NJ.

Redl C. *et al.* (2008) *Price formation in electricity forewards markets and relevance of systematic forecast errors. Energy Economics* (in press).

Regnier E. (2007), 'Oil and Energy price volatility', *Energy Economics*, **29(3)**, 405-427

Weron R., Przybyłowicz B. (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. *et al.* (2008), 'Measuring volatility in the Nordic spot electricity market using Recurrence Quantification Analysis', *Eur. Phys. J. Special Topics*, **164**, 105-115.

Weron R., Simonsen I., (2006), *Blackouts, risk and fat-tailed distributions,* arXiv:physics/0510077v1 [physics.soc-ph].

Zhao J. (2007), *Electricity market management and analysis using advanced data mining and statistical methods*, PhD thesis, wired at: http://espace.library.uq.edu.au/view/UQ:158511