Recurrence Quantification Analysis for time series

Strozzi F.*, Pozzi R.*

*Scuola di Ingegneria Gestionale, Corso Matteotti 22, 21053 - Castellanza – Italy (fstrozzi@liuc.it, rpozzi@liuc.it)

Abstract: Recurrence is a fundamental property of dynamical systems and studying it helps in deepen our understanding of the systems and predict their evolution. Recurrence has been studied in many real systems, such as financial exchange rates, damage detection and neuroscience data, but, as far as we know have never been applied with the aim of detecting demand patterns needed to develop a reliable forecast model. The analysis is performed using the Recurrence Plot (RP) together with the tool called Recurrence Quantification Analysis (RQA). RP is based on the computation of a distance matrix between reconstructed points in phase space. RQA provides a set of measures for RP quantification. In this work we analyse the RPs corresponding to several simulated demand time series and the relationship between Determinism (DET, a RQA measure) and parameters traditionally used to assess demand patterns. The visualization of the RP and the calculation of DET of the simulated demand time series provide positive results when used to detect demand patterns.

Keywords: Recurrence Quantification Analysis; Recurrence Plot; Trend; Seasonality

1. Introduction

Forecasting future value of time series is a topic of primary importance for all applied science (Tratar et al., 2016). Accordingly, the time series literature is ample and interest in time series forecasting has been stable over time (De Gooijer and Hyndman, 2006). In the field of operations management, forecasting is a key tool to manage inventory planning, capacity management and the design of the customer service levels (Bayraktar et al., 2008). In fact, demand forecasts represent the gross requirements of each finished product in each time bucket of the horizon, input of the materials requirement planning procedure (Rossi et al., 2016) and, in case of EOQ-based inventory management systems, estimating demand and characterize it is essential to apply the correct methodology to manage the case of single machine-multi item systems (Rossi et al., 2016). In the supply chain management field, the impact of forecasting methods on the bullwhip effect has been extensively studied and method selection to help managers yielding the reduction of inventory costs of a supply chain has been discussed (Zhang, 2004, Strozzi et al., 2008).

In literature forecasting has been performed qualitatively, quantitatively or combination of both. The qualitative methods are suitable when there is not enough data to apply quantitative methods, which involve either the development of mathematical models concerning causal variables rather than observation of historical data at regular intervals (Arunraj and Ahrens, 2015). Forecasting techniques in inventory management and sales estimating deal mainly with exponential smoothing methods, which reliably take into account the trend and seasonal patterns of the time series by means of simple formulae (Tratar et al., 2015).

Model performance changes on conditions, such as the nature of the data (De Gooijer and Hyndman, 2006) and a model that suit the forecast of a class of items offered by a company may obtain poor performance when forecasting other items offered by the same company. The issue of selecting the appropriate model based on demand pattern has been ignored in practice (Gardner and McKenzie, 1988; Gardner and Díaz-Saiz, 2002). Usually in companies items classification based demand patterns is missing and seasonality is attributed to demand series with no information from statistical testing as well as it is often assumed multiplicative (Gardner and Díaz-Saiz, 2002). In literature, classical references have narrowed the issue of model selection to visual inspection of the data plot (Gardner and McKenzie, 1988), but, recently, it is in need of attention (Tratar et al., 2015).

Patterns are strategically significant in forecasting and their identification and analysis is needed before developing a forecasting model (Chen and Ebrahimpour, 1991). As visual inspection of the data plot can lead to poor rather than incorrect explanations (Gardner and McKenzie, 1988), ratio-to-moving average procedure (Makridakis et al., 1998) and the study of the autocorrelation of the time series (Wei, 1990; Darbellay and Slama, 2000) are mostly applied to detect time series pattern.

According to Marwan et al. (2007), the investigation of recurrence reveals distinctive properties of the system and might help to predict its future behaviour. Recurrence Plot (RP) is a tool to visualize the recurrent behaviour of a dynamical system. One of the main advantages of recurrence plots is that it provides useful information even for short and non-stationary data, where other methods fail. Moreover, using Recurrence Quantification Analysis (RQA) it is possible to detect in advance changes
Recurrence Quantification Analysis has applied to analyse data of very different origin. Zbilut et al. (2002) have applied RQA for diagnosing changes in nonstationary cardiac signals; the work by Strozi et al. (2002) has dealt with high frequency currency exchange data and Zaldivar et al. (2008) characterized regime shift in environmental time series; more recently, Begun et al. (2016) have used RQA to classify different neurological disorders; Meng et al. 2016, have studied the pressure signals; Acharaya et al. (2017) have applied RQA to the identification of EEG signals; the use by Xiong et al. (2017) concerns the traffic flow data. Applications to experimental data have expanded the utilisation of RP to the investigation of similarity and transitions in data series (Marwan et al. 2007).

Given the paramount importance of demand forecasting in operations and supply chain management and the significance of pattern identification with this purpose, we propose the use of RP and RQA, already successfully applied to many research areas, to provide a new approach for demand time series pattern identification. In particular, we have first applied the RP and the RQA to a set of simulated data representing possible real demand time series (i.e. weekly and monthly data affected by trend, seasonality and noise) and the results have been compared with the identification of seasonality performed with the classical study of the autocorrelation of the time series. Then, an empirical evaluation using the database of the largest forecasting competition to date, the M3-Competition (Makridakis and Hibon, 2000), is undertaken in order to obtain insights into the performances of the proposed application of RP and the RQA for demand time series pattern identification.

The paper is organized as follows, in Section 2 details on RQA and RP are given; in Section 3 both the simulated and real data from M3-Competition are described and the application of RQA and RP to the corresponding time series is outlined; in Section 4 the obtained results are discussed and conclusions are provided in Section 5.

2. From time series to Recurrence Quantification Analysis (RQA)

2.2 Space state reconstruction

In this section, it will be explained, first, how to transform a time series in RP and, second, the measures for its quantification, namely the RQA.

To build a RP from a time series the dynamics of the series should be reconstructed in a new space: the embedded space. 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. If the equations of the system are not known, and it is not possible (or too expensive) to measure all the variables, the real state space of the system is not accessible to us. However, the Takens theorem (Takens, 1981) states that, by measuring few variables, we can reconstruct a one-to-one correspondence between a reconstructed state space and the original. The reconstructed space can be obtained using delayed versions of the time series a certain number of times. The number of the displaced time series with a time delay $\tau$ is called the embedding dimension $d_e$. In the embedding space, many invariants of the dynamics are preserved (Marwan et al. 2007).

To estimate time delay $\tau$ and the embedding dimension $d_e$, different approaches have been proposed (Cao, 1997). The most common techniques used to calculate the time delay are the first zero of the autocorrelation function or the first minimum of the mutual information function (Fraser and Swinney, 1986). The first minimum of the mutual information function is the lag to allow the enough independence between the different translations of the time series but not so independent as to have no connection at all. Usually the time delay given by the mutual Information function is lower than the one given by the first zero of the autocorrelation function. In this work the False Nearest Neighbour algorithm (Kennel, et al., 1992) is used to calculate $d_e$ which is based on the increasing the dimension until the percentage of neighbours of the points in the reconstructed space stops to decrease.

2.3 Recurrence Plot and Recurrence Quantification Analysis

The concept of recurrence in conservative systems was introduced by Poincaré (1890). One hundred years later, in 1987, Eckmann et al. introduced the RP as tool to visualize the recurrence. Recurrence Plot is a square matrix of zero and one to visualize the recurrence of a time series with a $\varepsilon$-threshold. Strozzi et al. (2002) has dealt with high frequency currency exchange data

$$R_{i,j} = \begin{cases} 1 & \text{if } x_i \approx x_j \\ 0 & \text{if } x_i \not\approx x_j \end{cases}$$

were $x_i \approx x_j$ means distance up to a distance $\varepsilon$, i.e. the matrix compares the state of a system at time $i$ and $j$. The RP is the visualization of recurrence matrix, which gives hints about the characteristic time evolution of these trajectories (Marwan et al., 2007).

The threshold $\varepsilon$ is another important parameter to build the RP. If it is chosen too small no recurrence points will be found and the RP will be empty, if too large nearly all the points will be recurrent and the RP will be completely full. Several “rules of thumb” have been developed in the literature, as a general rules it should not exceed 10% of the mean or maximum phase space diameter (Marwan et al. 2007; and Zbilut and Webber, 1992).

Figure 1 depicts the synthesis of the above-described steps. A series of data values $s(t)$ collected from a starting
time ‘t’ with frequency ‘f’ constitutes the original time series, plotted in the bi-dimensional space. The representation of same time series in the embedded space, gives a new dynamical series, which is a curve lying in a \( d_x \)-dimensional time-delayed space. The axes of the RP correspond to the time which is given by pursuing a state on the trajectory and the recurrence of a state, e.g. \( x(t) \), in the \( d_x \)-dimensional space is marked with points in the corresponding RP.

According to Eckmann et al. (1987) and Marwan (2003) the appearance of RPs gives hints about the characteristic time evolution of these trajectories. Regarding the topology of the RP four characteristics of the structure can be identified: homogeneity, periodicity, drifts, white areas (Eckmann et al., 1987). A homogeneous RP identifies a uniformly distributed system. Periodic recurrent structures in RP identify an oscillating system. Fading (or drifts) to the upper left and lower right corners of the RP characterise a system with slowly varying parameters. White bands or areas characterize states of the system far from the normal. Focus on textures, such as single dots, diagonal, horizontal and vertical lines, reveals small scale structures (Marwan, 2003). Single isolated points represent states that do not persist over time, or fluctuations. Diagonal lines represent the visit of the same region of the state in the embedded space at different times, the longest the duration of intervals with similar local behaviour of the trajectory segment the longest the length of the diagonal lines when a diagonal line is parallel to the main diagonal line of the RP the evolution of the state system is defined similar at different times; when a diagonal line is orthogonal to the main diagonal line of the RP the evolution of the system state is similar with inverse time. Vertical or horizontal lines and clusters identify states that have slow or no change over time.

In order to quantify objectively the small scale structure of RP and extract information on the underlying time series several measures have been proposed (Zbilut and Webber, 1992; Webber and Zbilut, 1994; and Marwan et al., 2002), and they are collected under the name of ‘Recurrence Quantification Analysis’. These measures are based on the recurrence point density and they quantify the number and length of the diagonal and vertical lines in RP. In this work the measure so called Determinism (DET), i.e. the ratio of recurrence points forming diagonal structures (at least of length \( l_{\min} \)) to all recurrent points, is considered:

\[
DE\text{T} = \frac{\sum_{i=1}^{N} l_{P}(i)}{\sum_{i=1}^{N} l_{P}(i)}
\]

Where \( l_{P}(i) \) is the probability of finding lines of length bigger than \( l_{\min} \). The presence of diagonal line means that the trajectory of the system in the embedded space runs near to itself at different time, i.e. it repeats its behaviour.

### 3. RQA applied to simulated data

In this section the RP and DET are calculated for a set of simulated data sets (see Table 1 and Figure 2) using Matlab crp toolbox (http://tocsy.pik-potsdam.de). The time series were generated using different sampling frequencies: \( 2\pi/600, 2\pi/52, 2\pi/12 \) to simulate a sampling rate with a very high, weekly or monthly frequency respectively over three years.

<table>
<thead>
<tr>
<th>Function</th>
<th>Reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \sin(x) )</td>
<td>seasonality</td>
</tr>
<tr>
<td>( \frac{1}{10} x + \sin(x) )</td>
<td>trend + seasonality</td>
</tr>
<tr>
<td>( \frac{1}{10} x \cdot \sin(x) )</td>
<td>trend \cdot seasonality</td>
</tr>
<tr>
<td>( \text{rand} + 2 )</td>
<td>random demand</td>
</tr>
<tr>
<td>( x + \text{rand} )</td>
<td>linear trend with additive noise</td>
</tr>
<tr>
<td>( x \cdot \text{rand} )</td>
<td>linear trend with multiplicative noise</td>
</tr>
</tbody>
</table>

![Graphs](image-url)
3.1 RPs of simulated demand time series

Figure 3 depicts the RPs of the simulated time series with sampling frequency equal to $2\pi/600$ for three years. The RPs related to the time series obtained through other sampling frequencies are here omitted as similar to the ones depicted by Figure 3.

![Figure 3: RPs of simulated data of Table 1 with a step of $2\pi/600$.](image)

According to the patterns and textures described by Eckmann et al. (1987) and Marwan (2003), the characteristics of the depicted RPs are here derived. A periodic recurrent structure is visible in the RP corresponding to the seasonal function a) and such recurrent behaviour is highlighted by diagonal lines parallel to the main diagonal line of the RP; as the diagonal lines are the longest possible, this would mean that even the DET will assume its maximum value. The recurrent structure visible in the RP corresponding to the seasonal function with additive trend b) shows a modified seasonal factor and visible clusters identify the additive trend component of a demand time series. The fading to the upper left corner of the RP corresponding to the seasonal function with multiplicative trend c) shows the more important contribution of the trend component of the demand time series, while the parallel lines still unveil seasonality. The homogeneous RP corresponding to the random function (d) identifies a uniformly distributed demand time series, and single dots characterize demand values that do not persist over time. Drifts to the upper left and lower right corners of diagonal lines combined with large white areas characterising the RP corresponding to the trend with additive noise function e) describe a system with slowly varying parameters, i.e. trend, and noise. The multiplicative role of noise in function f) is captured by changes in the texture of the corresponding RP, as lines are orthogonal. Lines are interrupted by noise, nevertheless the low resolution of plot does not allow the visual inspection of this interruption.

3.2 RQA of simulated demand time series

In order to quantify objectively the small scale structure of RP and extract quantitative information on the underlying time series, DET measure is calculated for all the considered time series and, along with this value, the classical autocorrelation value ($r(k)$) calculated according to Sianesi (2014) and the corresponding lag ‘$k$’ are given. Note that $r(k)$ is always calculated on the three years time window. Different time windows (one, two, three years) of the generated series are also considered to check if the length of the data series may impact on the comparison of DET and classical autocorrelation (see Tables 2–4).

Table 2 outlines the DET values calculated according to the three time windows, $r(k)$ and ‘$k$’ of the high frequency sampling (step of $2\pi/600$) simulated demand time series. The calculated values of DET present slight differences over the three different time windows and coherence between the obtained values and $r(k)$ is found. DET value is equal or nearly 1 in the case of function a), b), c) and (e) simulating seasonality, seasonality with additive trend, seasonality with multiplicative trend and trend respectively. The high values of determinism mean that for a certain time, the trajectory of the dynamical system repeats or goes near to former (periodicity) or actual (trend) visited states, corresponding to the many diagonal lines in the RPs of time series a) b) and c) (Figure 2). The value of ‘$k$’ equal to 600 means that the seasonality calculated by classical autocorrelation is annual, while 1 means that the autocorrelation finds a trend component. The lowest value is for d) time series, i.e. the random one, and an intermediate value for f) time series in which the noise is multiplicative and partially destroys the trend component. With reference to d), the difference between the values of DET and $r(k)$ might be in the fact that in the calculation of recurrence and DET the states of the systems near to themselves are considered.

<table>
<thead>
<tr>
<th></th>
<th>Three years</th>
<th>Two years</th>
<th>One years</th>
<th>$r(k)$</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>1,0000</td>
<td>1,0000</td>
<td>1,0000</td>
<td>1,0000</td>
<td>600</td>
</tr>
<tr>
<td>b)</td>
<td>1,0000</td>
<td>0,9999</td>
<td>1,0000</td>
<td>1,0000</td>
<td>600</td>
</tr>
<tr>
<td>c)</td>
<td>1,0000</td>
<td>1,0000</td>
<td>1,0000</td>
<td>1,0000</td>
<td>600</td>
</tr>
<tr>
<td>d)</td>
<td>0,3881</td>
<td>0,3980</td>
<td>0,4157</td>
<td>0,0594</td>
<td>8</td>
</tr>
<tr>
<td>e)</td>
<td>0,9943</td>
<td>0,9911</td>
<td>0,783</td>
<td>0,9975</td>
<td>1</td>
</tr>
<tr>
<td>f)</td>
<td>0,7206</td>
<td>0,7147</td>
<td>0,7396</td>
<td>0,4487</td>
<td>8</td>
</tr>
</tbody>
</table>
Table 3 outlines the DET values calculated according to the three time windows, \( r(k) \) and 'k' of the weekly frequency sampling (step of \( 2\pi/52 \)) simulated demand time series. Again, the calculated values of DET present slight differences over the three different time windows and coherence between the obtained values and \( r(k) \) is found. With reference to the simulated demand time series c), seasonal and multiplicative trend, when considering weekly data the highest value of \( r(k) \) is obtained for \( k = 1 \), i.e. trend factor is predominant, nevertheless still value near 1 for \( r(52) \) is found.

\[
\begin{array}{cccccc}
\text{Three years} & \text{Two years} & \text{One year} & \text{r(k)} & k \\
\hline
a) & 1,0000 & 1,0000 & 1,0000 & 1,0000 & 52 \\
b) & 1,0000 & 0,9904 & 1,0000 & 1,0000 & 52 \\
c) & 0,9962 & 0,9957 & 1,0000 & 0,9925 & 1 \\
d) & 0 & 0,0995 & 0 & 0,1834 & 25 \\
e) & 0,9639 & 0,9873 & 0,8268 & 0,9979 & 2 \\
f) & 0,3711 & 0,3793 & 0,3818 & 0,4776 & 1 \\
\end{array}
\]

Table 4 presents the DET values calculated according to the three time windows, \( r(k) \) and 'k' of the monthly frequency sampling (step of \( 2\pi/12 \)) simulated demand time series. The calculated values of DET over the two time windows of one and two years present slight differences and coherence between the obtained values and \( r(k) \) is confirmed.

\[
\begin{array}{cccccc}
\text{Three years} & \text{Two years} & \text{One year} & \text{r(k)} & k \\
\hline
a) & 1,0000 & 1,0000 & 1,0000 & 1 & 12 \\
b) & 1,0000 & 0,9904 & 1,0000 & 1 & 12 \\
c) & 0,9962 & 0,9957 & 1,0000 & 0,9925 & 12 \\
d) & 0 & 0,0995 & 0 & 0,1834 & 8 \\
e) & 0,9639 & 0,9873 & 0,8268 & 0,9979 & 1 \\
f) & 0,3711 & 0,3793 & 0,3818 & 0,4776 & 7 \\
\end{array}
\]

3.4. RQA of real data

Makridakis and Hibon (2000) launched a competition, called M3-Competition, in which many researchers had to test different forecasting techniques to see if more sophisticated methods perform better than simple ones. They collected a huge set of real data i.e. 3003 time series of different origin (micro, industry, macro, finance, demographic, etc.) with different sampling frequency (yearly, quarterly, monthly, etc.). In this paper we verified the results of the former section on some of these data. We choose four time series of industrial kind and with a monthly sampling frequency. We selected the series to test RQA results on different patterns: no seasonality and no trend, seasonality with additive and multiplicative trend, seasonality without trend, trend (see Figure 4).

Plotting their recurrences (Figure 5) the concentration of RP points along the main diagonal unveils a similar evolution of the states over time, i.e. the presence of an additive trend, while many lines parallel to the main diagonal are a sign of seasonality. Fading to the upper left and lower right corners reveals a change in the amplitude of the cycles and then can be interpreted as a sign of trend and drift.

![Figure 4](image-url)

Figure 4. Monthly industrial time series from M3-Competition dataset. a) no seasonality and no trend, b) seasonality with additive and multiplicative trend, c) seasonality, d) trend.

![Figure 5](image-url)

Figure 5. Recurrence Plot of time series in Figure 4. ‘nz’ is the number of dark points.

Figure 5b) and 5d) show a band of points around the main diagonal and then the presence of additive trend. In 5b) the multiplicative trend is showed by the fading in the upper left and lower right corners. The same pattern was found in simulated data, respectively in Figure 3c) and 3e)
i.e. seasonality with multiplicative trend (in this case the additive trend is missing) and trend. The ability to detect seasonality in real data using RP is proved by the similar texture of RP in figure 5 c) (real data with seasonality) and RP of Figure 3a) in which the seasonality was simulated. The absence of clear seasonality or trend of time series in Figure 4a) is confirmed by the dispersion of its recurrence points (see Figure 5a).

In Table 5 the DET and the r(k) of the whole real time series were calculated and the results are in accordance as for simulated data. The difference between DET and r(k) for time series b) and c) and seasonality can be due to the presence of noise and the not perfect shape of the cycle that in RP causes a slight dispersion of points.

<table>
<thead>
<tr>
<th>DET</th>
<th>r(k)</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>0,4365</td>
<td>0,48</td>
</tr>
<tr>
<td>b)</td>
<td>0,7087</td>
<td>0,98</td>
</tr>
<tr>
<td>c)</td>
<td>0,7730</td>
<td>0,94</td>
</tr>
<tr>
<td>d)</td>
<td>0,9445</td>
<td>0,99</td>
</tr>
</tbody>
</table>

### 4. Discussion of the results

The visualization of the RP and the calculation of DET of the simulated demand time series have provided positive results when used to detect demand patterns. While DET provides the quantitative value representing the pattern of recurrence, this measure alone is insufficient to discriminate whether the system repeats or goes near to former (seasonality) or actual (trend) visited states. To distinguish seasonality and trend the combination of quantitative and qualitative information (from the RP pattern and texture) has been used.

The values provided by DET are coherent with the obtained values from the classical r(k) calculus, proving the proposed tool to be a good alternative to perform pattern identification. Moreover, in most cases the DET values calculated with less than 2 years of data (needed by the autocorrelation (Sianesi, 2014) and ratio-to-moving average (Dekker et al., 2004)) detect annual seasonality by means of recurrence.

One limitation of RP and RQA is the necessity of enough points (r ∙ d^l) to build the embedded space, depending on the embedding parameters.

### 5. Conclusions

The present work proposes a new approach to perform pattern identification in demand time series based on recurrence, a fundamental property of dynamical systems. In particular, we propose the combined use of RP, a tool to visualize the recurrent behaviour of a dynamical system, and one measure of the RQA, known in literature as a possible way to detect in advance changes in the states of a dynamical system. In particular, we focus on the DET measure, which deals with the quantification of the diagonal lines, representing the similar evolution of the state in the embedded space at different times, of the RP.

The proposed tool and measure proves to successfully detect seasonal and trend patterns characterizing both simulated and four time series from the industrial monthly subset of the M3-Competition. Moreover, when enough point to build the embedded space are available, succeeds in requiring the availability of a shorter time window than classical methods to detect seasonality need.

As a future research step, we propose the application of other RQA measures to both simulated and real demand time series to detect other properties.

### References


