Application of Recurrence Plots for Identifying Spatial Structure in Geographical Systems

A.J. Heppenstall and A.N. Ross

1School of Geography, Woodhouse Lane, University of Leeds, Leeds, LS2 9JT.

2School of the Earth and Environment, Woodhouse Lane, University of Leeds, Leeds, LS2 9JT. Tel: +44 113 343 7590; Fax: +44 113 343 6716;

E-mail: a.n.ross@leeds.ac.uk

January 2010

Number of text pages: 9 (inc references)

Number of figures: 4

Number of tables: 2
Application of Recurrence Plots for Identifying Spatial Structure in Geographical Systems

Abstract

Recurrence plots (RPs) allows both visualisation and quantification of structures hidden within data. They are particularly useful for graphically detecting hidden patterns and structural changes in data as well as examining similarities in patterns across a time series data set. However, few studies have applied RPs to detecting spatial structure in geographical systems. Of those studies, applications have been limited to uniform grids. In this paper, we demonstrate how RP’s can be used to analyse fixed spatial patterns in a non-uniformly distributed spatial system which varies with time.

Keywords: Recurrence plot; spatial structure spatio-temporal complexity
Application of Recurrence Plots for Identifying Spatial Structure in Geographical Systems

Introduction

Detecting and understanding spatial patterns within geographical systems is often a challenging problem, particularly where points are non-uniformly distributed in space. These systems frequently exhibit complicated temporal as well as spatial variations. One of the most promising techniques that has emerged within the past decade for exploring and analysing such systems is Recurrence Plots (RPs).

Recurrence plots (RPs) allows both visualisation and quantification of structures hidden within data [1]. They are particularly useful for graphically detecting hidden patterns and structural changes in data as well as examining similarities in patterns across a time series data set (where there are multiple readings at one point). RPs can be also used to study the nonstationarity of a time series as well as to indicate its degree of aperiodicity [2, 3]. Their value in further understanding complex systems has been well documented with examples ranging from climate variation [4], music [5] to heart rate variability [6].

However, application to spatial systems has to date been limited. The few published studies available [7, 8, 9, 10] have been constrained by use of uniform grids and spatial snapshots obtained at fixed times to concentrate purely on spatial patterns. Within this paper, we demonstrate how RP’s can be used to analyse fixed spatial patterns in a non-uniformly distributed spatial system which varies with time. Daily temperature data from the UK is used as an example of such a system. We demonstrate the potential of this approach for identifying spatial structures within a dynamic spatial system and offer recommendations for future work.
Recurrence plots

Consider a data set consisting of a time series of $N$ sets of measurements at $M$ different sites. The measurement at station $k$ at time $i$ is written $X_{i,k}$ and so the vector of measurements recorded at a particular time, $i$, is $X_i = (X_{i,1}, X_{i,2}, \ldots, X_{i,M})$. The RP matrix, $R_{i,j}$, is defined as [11]

$$R_{i,j} = H(\epsilon - \|X_i - X_j\|)$$

where $H$ is the Heaviside function ($H(x) = 1$ if $x > 0$ and 0 otherwise), $\| \|$ is a metric which measures how similar two vectors $X_i$ and $X_j$ are, and $\epsilon$ is a threshold value for when the metric considers the two vectors to be similar, i.e. the pattern recurs. Various metrics are used in the literature, however here we use the usual Euclidean distance so

$$\|X_i - X_j\| = \left[\sum_{k=1}^{M} (X_{i,k} - X_{j,k})^2\right]^{1/2}.$$ 

The threshold has to be prescribed for a given data set. The larger the value of the threshold, the more pairs of observations will be considered close to each other, and so the more recurrences there will be.

The RP matrix $R_{i,j}$ can be visualised as a two dimensional plot where the x and y axes correspond to the times $i$ and $j$ with values of 1 being coloured black, and 0 coloured white. Black points correspond to times which are classified as similar, i.e. the spatial pattern of measurements has recurred at the two times.

Various quantitative measures can be calculated to assess the recurrence of small-scale structures in a RP. This process is known as recurrence quantification analysis (RQA) [11]. The measures provide information on the recurrence patterns in the data and can highlight differences in recurrence structure between different data sets. The measures look at the size of both diagonal and vertical structures in
the RP. \(P(l)\) and \(P(v)\) are the number of diagonal lines of length \(l\) and the number of vertical lines of length \(v\) respectively in the RP. The calculated quantitative measures are detailed in Table 1.

<table>
<thead>
<tr>
<th>RQA measure</th>
<th>Definition</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recurrence rate</td>
<td>[ RR = \frac{1}{N^2} \sum_{i=1}^{N} R_{i,j} ]</td>
<td>The proportion of pairs of data which are recurrent (i.e. the proportion of black points in the RP)</td>
</tr>
<tr>
<td>Determinism</td>
<td>[ DET = \frac{\sum_{i=1}^{N} lP(l)}{\sum_{j=1}^{N} R_{i,j}} ]</td>
<td>Proportion of recurrence points which occur in diagonal structures which are of length (\geq l_{\text{min}}). Relates to the predictability of the system.</td>
</tr>
<tr>
<td>Laminarity</td>
<td>[ LAM = \frac{\sum_{i=1}^{N} vP(v)}{\sum_{j=1}^{N} R_{i,j}} ]</td>
<td>Proportion of recurrence points which occur in vertical structures which are of length (\geq v_{\text{min}}). Relates to the persistence of recurrent states.</td>
</tr>
<tr>
<td>Average diagonal length</td>
<td>[ L = \frac{\sum_{i=1}^{N} lP(l)}{\sum_{l=1}^{l_{\text{min}}} P(l)} ]</td>
<td>Related to the prediction length of the system.</td>
</tr>
<tr>
<td>Trapping size</td>
<td>[ TT = \frac{\sum_{v=1}^{v_{\text{min}}} vP(v)}{\sum_{v=1}^{v_{\text{min}}} P(v)} ]</td>
<td>Mean length of vertical structures. Related to the size of the area in which the system does not change.</td>
</tr>
</tbody>
</table>

Table 1: Definitions and explanations of RQA measures, based on Table 1 from [11].

For this paper, the threshold value \(\varepsilon\) was chosen based on the variations in measurements from all stations over the whole data set, quantified by calculating the standard deviation. In each case the
threshold was taken as one standard deviation of the measurements. The minimum length $l_{\text{min}}$ and $v_{\text{min}}$ for diagonal and vertical lines were both set to 5 to filter out very small recurrence structures from the calculations. For a given problem, these parameters need to be chosen depending on the data set and on the number and size of recurrent structures to be studied.

**Methodology**

The data chosen to demonstrate the power of RPs for identifying spatial structure in this paper is a record of mean daily temperatures for the UK during 2005. The data is from the National Climatic Data Centre Global Summary of the Day dataset ([http://www.ncdc.noaa.gov](http://www.ncdc.noaa.gov)). Only those stations with a full daily record for the year are used (96 out of a total of 222 stations) since the RP requires comparison of data for all possible pairs of dates during the year. The temperature on day $i$ at station $k$ is denoted by $T_{i,k}$. The time series of mean daily temperatures are dominated by the annual cycle in mean temperature (warmer in summer and cooler in winter). To focus on spatial patterns in temperature, rather than this mean signal recurrent plots were also calculated using the relative temperature, $\Delta T_{i,k}$, defined as the temperature relative to the mean temperature across all stations for that day,

$$
\Delta T_{i,k} = T_{i,k} - \frac{1}{M} \sum_{k=1}^{M} T_{i,k}.
$$

Routines for generating RPs and calculating the RQA measures were implemented in MATLAB.

**Results**

Figure 1 shows the RPs for (a) the mean daily temperature and (b) the mean daily relative temperature. Calculated values of the RQA measures are given in Table 2.
Figure 1: RPs of (a) mean daily temperature and (b) mean daily relative temperature from the 96 sites across the UK for the whole of 2005.

<table>
<thead>
<tr>
<th>RQA measure</th>
<th>Mean daily temperature</th>
<th>Mean daily relative temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>RR</td>
<td>0.49208</td>
<td>0.49253</td>
</tr>
<tr>
<td>DET</td>
<td>0.92712</td>
<td>0.70124</td>
</tr>
<tr>
<td>LAM</td>
<td>0.49028</td>
<td>0.33699</td>
</tr>
<tr>
<td>L</td>
<td>23.770</td>
<td>13.474</td>
</tr>
<tr>
<td>TT</td>
<td>22.274</td>
<td>9.3934</td>
</tr>
</tbody>
</table>

Table 2: Values of the RQA measures calculated for the RPs of temperature and relative temperature.

The RP of temperature shown in Figure 1(a) exhibits significant blocks of black along the diagonal, related to the annual cycle of temperatures. Within a season the temperature measurements tend to be similar and so are mostly recurrent. There are also off-diagonal blocks corresponding to the recurrence of temperature patterns in spring and autumn when mean temperatures are similar. There is
very little or no recurrence between days in winter and summer when temperatures tend to differ significantly. Figure 1(b) shows the RP for the relative temperature. In this case the recurrences are scattered much more uniformly through the plot. The removal of the mean temperature means that this RP focuses on spatial patterns of temperature differences. It shows that there seems to be little or no systematic variation in the patterns observed through the year.

The RQA measures given in Table 2 compare the two RPs quantitatively. Using a threshold based on standard deviation of the data ensures that the two plots have very similar recurrence rates (just under 0.5), however other RQA measures differ significantly.

The determinism for the temperature plot is 0.93. This means that the vast majority of recurrent points for the temperature RP occur in long diagonal lines (with a length $\geq 5$), reflecting the fact that on weekly to seasonal timescales temperatures tend to take similar values, with a longer time scale annual cycle dominating the differences. The temperature is deterministic in the sense that given a particular day is recurrent, there is a strong likelihood that the following days will also be recurrent, i.e. patterns of temperature tend to recur in the system. With the relative temperature RP, the determinism is slightly lower with a value 0.70. The dominant annual cycle has been accounted for by removing the daily mean temperature, hence the lower value. The fact that the determinism is still high suggests that spatial patterns of variations in temperature are still like to recur for 5 or more consecutive days (70% of cases). From a physical point of view this is perhaps not surprising as many of the reasons for variations in temperature on a given day (e.g. altitude, latitude, land use) are spatially fixed. This technique allows for these recurrent patterns to be first identified and secondly quantified.
The laminarity of the two RPs is again different (0.49 for the temperature RP and 0.34 for the relative temperature RP). The fact that approximately half of all recurrence points lie on vertical lines of length 5 or greater means that not only do temporal patterns of 5 or more days tend to recur, but often the same spatial recurrence pattern will occur on at least 5 consecutive days. Again this is not surprising for temperature where the seasonal temperatures tend to be similar. It is also well known in meteorology that a persistence forecast (i.e. the weather tomorrow will be the same as today) is quite effective to the extent that the skill of weather forecasts are evaluated by comparing them to this benchmark. Removing the annual cycle reduces the laminarity slightly, but still approximately a third of recurrent spatial patterns persist for 5 or more days across the whole of the UK.

The average diagonal length and trapping size both tell a similar story. The average diagonal length is significantly longer for the temperature RP (23.8 compared to 13.5), which might be expected given the higher determinism and again suggests that recurrent patterns will last for longer in the temperature measurements compared to the relative temperature measurements. The trapping size (related to the size of the blocks in the RP) is also larger in the temperature RP compared to the relative temperature plot (22.3 compared to 9.3). This quantifies the visual differences seen between the two plots, with the temperature RP exhibiting large seasonal blocks.

Figure 2 shows maps of the relative temperature plotted at the 96 stations across the UK for two days in 2005 (1st January and 22nd November). The 1st January is a typical day, which is recurrent with 259 days in the year (illustrated by the predominantly black column of points above day 1 in the RP in Figure 1(b). The pattern is of relatively small differences of temperature between stations (mean absolute difference of 1.1°C and a range of ±2.9 °C.) The differences are predominantly warmer in the south and cooler in the north, as would be expected. In addition to being used to identify recurrent spatial patterns in the data, RPs can be used to identify unusual or extreme spatial patterns. November
22\textsuperscript{nd} provides an example of this. It is recurrent with no other days, i.e. for day 326 (November 22\textsuperscript{nd}) the column in the RP is white, except for the day itself. This suggests that this day is unusual in terms of the spatial pattern of temperature differences, and this is confirmed by Figure 2(b). Firstly the temperature differences across the UK are much larger than for Figure 2(a), with a mean absolute difference of 2.0 °C and a range of -5.6 °C to 6.1 °C. Not only is the range unusually large, but there is also an unusual spatial distribution of temperatures. The warmest temperatures are observed around the coast (with the exception of the east coast of Scotland) but these are very localised to the coast with much lower temperatures inland. The high temperatures in the Islands off the coast of Scotland are particularly unusual. The combination of these features explains why this particular day is unique over the year.

Figure 2: Maps of the relative temperature (in °C) at the 96 stations across the UK for (a) 1\textsuperscript{st} January 2005 and (b) 22\textsuperscript{nd} November 2005.
**Discussion and Conclusions**

Current applications of RPs tend to focus on the analysis of time series. Extensions of the RP to higher dimensions to analyse spatial images are powerful in being able to detect the size and location of recurrent patterns in the data, but they are limited to regular spaced (gridded) data. These techniques are generally applied to single images to detect patterns at a fixed point in time. In many applications however observations are made at points which are non-uniformly distributed in space, for example the data set of surface temperature observations used here. Where there are fixed spatial patterns in the data which recur at difference times then RPs, and their associated quantitative measures, can be used to detect these patterns, even when the data is non-uniformly distributed. In addition to detecting patterns, the technique provides information about the temporal distribution of these patterns, and hence on the predictability of the system. Finally, RPs can highlight unusual or extreme spatial patterns in the data which might be difficult to identify using other techniques. The example of surface temperatures for the UK used in this paper has demonstrated these uses of RPs, and hence highlighted the potential for the use of RPs to analyse geographical systems.
References


10 Proulx, R.; Cote, P.; Parrott, L. Multivariate recurrence plots for visualising and quantifying the dynamics of spatially extended ecosystems. Ecological Complexity 2009 6 37 –47.