Research Article

Analysis of Forest Fires by means of Pseudo Phase Plane and Multidimensional Scaling Methods

J. A. Tenreiro Machado¹ and António M. Lopes²

¹ Institute of Engineering, Polytechnic of Porto, Rua Dr. António Bernardino de Almeida 431, 4200-072 Porto, Portugal
² Institute of Mechanical Engineering, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frías, 4200-465 Porto, Portugal

Correspondence should be addressed to J. A. Tenreiro Machado; jtenreiromachado@gmail.com

Received 2 February 2014; Accepted 20 May 2014; Published 5 June 2014

Academic Editor: Riccardo Caponetto

Copyright © 2014 J. A. Tenreiro Machado and A. M. Lopes. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Forest fires dynamics is often characterized by the absence of a characteristic length-scale, long range correlations in space and time, and long memory, which are features also associated with fractional order systems. In this paper a public domain forest fires catalogue, containing information of events for Portugal, covering the period from 1980 up to 2012, is tackled. The events are modelled as time series of Dirac impulses with amplitude proportional to the burnt area. The time series are viewed as the system output and are interpreted as a manifestation of the system dynamics. In the first phase we use the pseudo phase plane (PPP) technique to describe forest fires dynamics. In the second phase we use multidimensional scaling (MDS) visualization tools. The PPP allows the representation of forest fires dynamics in two-dimensional space, by taking time series representative of the phenomena. The MDS approach generates maps where objects that are perceived to be similar to each other are placed on the map forming clusters. The results are analysed in order to extract relationships among the data and to better understand forest fires behaviour.

1. Introduction

Forest fires, being caused by natural factors, human negligence, or human intent, consume every year vast areas of vegetation. Fire compromises ecosystems, has direct impact upon economy due to the destruction of property and infrastructures, raises the carbon dioxide emissions to the atmosphere, affects the water cycle, contributes to soil erosion, and has long-term economic implications associated with the climate change. In many regions and countries, like the United States, Australia, Russia, Brazil, China, and the Mediterranean Basin, fire is a major concern nowadays, demanding efficient policies for fire prevention and suppression and recovery of the affected areas.

Climate conditions, terrain orography, and type of vegetation are important factors that condition fire propagation and the total burnt area. The efficacy of detection and suppression strategies is fundamental in order to mitigate fire impact. However, fires caused by incendiaries contribute to increasing the complexity of the phenomena. Understanding forest fires behaviour and the underlying patterns in terms of fire size and spatiotemporal distributions may help the decision makers to take preventive measures beforehand, identifying possible hazards and deciding strategies for fire prevention, detection, and suppression.

Classical statistical tools have been used to investigate forest fires. However, those methods do not capture neither all characteristics underneath forest fires dynamics nor the fire dynamics along the years [1]. Forest fires are complex phenomena that exhibit intricate correlations in terms of fire size, location, and time. Forest fires dynamics unveil long range memory, self-similarity, and absence of a characteristic length-scale [2–10], which are features also found in fractional order systems [11–18].

In this paper we look at forest fires from the perspective of dynamical systems. A public domain forest fires catalogue containing data of events that occurred in Portugal, in the period from 1980 up to 2012, is tackled. The data is analysed in an annual basis, modelling the occurrences as sequences of Dirac impulses with amplitude proportional to
the events. Therefore, we are not modelling the dynamics of each particular forest fire. Otherwise, we are describing the global fire dynamics along several decades. The time series are viewed as the output of a dynamical system and are interpreted as a manifestation of the system dynamics. In the first phase, we use the pseudo phase plane (PPP) technique. The optimal time delay for the PPP is determined by means of the autocorrelation function. The PPP portraits are compared using an appropriate metric and the results are visualized through phylogenetic trees, generated by hierarchical clustering algorithms. In the second phase, the multidimensional scaling (MDS) tools are adopted to compare and extract relationships among the data.

Having these ideas in mind, the paper is organized as follows. In Section 2 we briefly describe the forest fire catalogue used in this work. In Section 3 we address the problem by means of the PPP and visualization of trees generated by hierarchical clustering algorithms. In Section 4 we use the MDS method. The approach is applied to the data and the main results are interpreted and analysed. Finally, in Section 5, we outline the main conclusions.

2. Forest Fires Dataset

Data of forest fires collected at the Portuguese Institute of Nature and Forest Conservation (INCF), available online at http://www.icnf.pt/portal/florestas/dfc/ini/estatisticas, is used [19]. The INCF dataset contains events since 1980 and up to 2012. Ignitions might have different sources, as natural causes, human negligence, or human intentionality, among others. The data was retrieved in December, 2013. Each data record contains information about the events date, time (with one minute resolution), geographic location, and size (in terms of burnt area). We discard small size events, as those are prone to measurement errors. Moreover, some small events may be missing because probably they were not reported. For that purpose we adopt a cutoff threshold value of 10 hectares for the burnt area. Experiments showed this value as a good trade-off between catalogue completeness and results accuracy.

The evolution of the burnt area and the cumulative number of events are depicted in Figures 1 and 2, respectively. In Figure 3 we depict the Lorenz curve relating to the cumulative burnt area and the cumulative number of events. The Gini coefficient, given by the double of the Gini area, measures the inequality among values of burnt area, being equal to 0.5968.

The time series representative of the occurrences is shown in Figure 4, where we can note the yearly periodicity of the events, with the peaks of burnt area occurring in summer. During the period covered by the catalogue, stronger fire activity has been verified around the middle of the decade 2000–2009.

Using the Fourier transform (FT), the forest fires data is analyzed in the frequency-domain. For each annual time series (33 in total) the amplitude spectra are computed and approximated by a power law (PL) function. The PL parameters are interpreted as the signature of the system dynamics. For example, Figure 5 depicts the amplitude spectra for year...
Mathematical Problems in Engineering

3. Analysis of Forest Fires by means of PPP

The PPP is a particular case of the pseudo phase space (PPS), which is justified by Takens’ embedding theorem [20]. The PPS allows the representation of system dynamics in a higher dimensional space, by taking a smaller sample of signals representing measurements of the system time history [21–23]. The PPS is useful in analysing signals with nonlinear behaviour and systems where complete information about all system states is unavailable. When compared to the classical phase space technique, the PPS reconstruction has the advantage of being more robust to signal noise.

In practical terms, we construct a \( n \)-dimensional phase space, \( U(t) \):

\[
U(t) = [s(t), s(t+\tau), s(t+2\tau), \ldots, s(t+(n-1)\tau)],
\]

(1)

where \( n \in \mathbb{N} \) and \( \tau \in \mathbb{R}^+ \) represent the time delay and embedding dimension, respectively. The matrix \( U(t) \) is usually plotted in a \( n \)-dimensional diagram. For \( n = 2 \) a two-dimensional time delay space is obtained and the PPS reduces to the PPP. In this case we have \( U(t) = [s(t), s(t+\tau)] \) related to the model given by the state vectors \([s(t), \dot{s}(t)]\). The choice of the time delay \( \tau \) is critical and must be accomplished adopting some criterion.

In this section we analyse forest fires in an annual basis, representing the events of the \( i \)-th year \((i = 1980, \ldots, 2012)\) by

\[
x_i(t) = \sum_{k=1}^{r} S_k \delta(t-t_k)
\]

(2)

leading to 33 one-year length time series.

This means that the events are modelled as Dirac impulses, where \( S_k \) represents fire size, \( t_k \) is the instant of occurrence, parameter \( t \) represents time, and \( T \) is the total time length, in minutes, corresponding to year \( i \).

The signals \( x_i(t) \) are then normalized according to the following equation:

\[
\bar{x}_i(t) = \frac{x_i(t) - \mu}{\sigma},
\]

(3)

where \( \mu \) and \( \sigma \) represent the global mean and standard deviation values, that is, the values calculated for the whole set of events registered during the time period 1980–2012, with minimum magnitude equal to 10 ha.

To implement the PPP, we firstly integrate the signals \( \bar{x}_i(t) \), (to be denoted \( \bar{X}_i(t) = \int_{t}^{t+T} \bar{x}_i(u) du, 0 \leq t \leq T \), where \( t = 0 \) corresponds to the first minute of each year, \( i \)) that represent the normalized time series of the occurrences in every year \((i = 1980, \ldots, 2012)\). The correlation function, \( r_{\tau}[\bar{X}_i(t), \bar{X}_i(t-\tau)] \), is then used to correlate \( \bar{X}_i(t) \) with its time delayed version \( \bar{X}_i(t-\tau) \):

\[
r_{\tau}(\tau) = \frac{\sum_{t=1}^{T} \bar{X}_i(t) \cdot \bar{X}_i(t-\tau)}{\sqrt{\sum_{t=1}^{T} \bar{X}_i(t)^2 \cdot \sum_{t=1}^{T} \bar{X}_i(t-\tau)^2}},
\]

(4)

\( i = 1980, \ldots, 2012 \).

For each case, values within the interval \( \tau \in [1440, 288000] \) minute (i.e., \( \tau \in [1, 200] \) days) are tested and the optimal time delay, \( \tau_m \), is computed, corresponding to the time at which the correlation function has its first point of inflection.

Figure 6 depicts, for example, the signals \( \bar{X}_{2000}(t) \) and \( \bar{X}_{2000}(t-\tau) \), representative of the events that occurred in year 2000. The normalized time series, \( \bar{X}_{2000}(t) \), reveals a “noisy” nature, while the corresponding integral, \( \bar{X}_{2000}(t) \), is much smoother, showing more clearly possible correlations between multiple points. The larger discontinuities observed in the amplitude correspond to instants of sudden increase in fire activity.

In Figure 7, the correlation function for year \( i = 2000 \), \( r_{\tau}[\bar{X}_{2000}(t), \bar{X}_{2000}(t-\tau)] \) versus the time delay, \( \tau \) is presented. In this case, the optimal time delay yields \( \tau_{m_{2000}} = 106 \) days. The optimal time delays calculated for the 33 one-year length time series are summarized in Table 1.

Figure 8 gives a global perspective of the PPP portraits, \( \bar{X}_i(t-\tau_m) \) versus \( \bar{X}_i(t) \), for the 33 time series. Figure 9, serving
as an example, details the results obtained for the initial and final time series, that is, years 1980 and 2012, respectively. Both figures reveal complex patterns that resemble those found in chaotic systems, demonstrating the rich dynamics of forest fires.

To compare the 33 PPP patterns we calculate a $33 \times 33$ similarity matrix $E = [e_{ij}]$, based on the 2-dimensional correlation, $d_{ij}$ ($i, j = 1980, \ldots, 2012$), between the PPP curves, defined by

$$
\begin{align*}
    d_{ij} &= \left| \left( \sum_{t=1}^{T} [\overline{X}_i(t) \cdot \overline{X}_j(t) + \overline{X}_i(t - \tau_{m_i}) \cdot \overline{X}_j(t - \tau_{m_j})] \right) \\
    &\quad \times \left| \left( \sum_{t=1}^{T} [\overline{X}_i^2(t) + \overline{X}_j^2(t - \tau_{m_j})] \right)^2 \right|^{1/2} \\
    &\quad \cdot \left| \left( \sum_{t=1}^{T} [\overline{X}_i^2(t) + \overline{X}_j^2(t - \tau_{m_j})] \right)^2 \right|^{-1/2} \right|,
\end{align*}
$$

(5)

$$
e_{ij} = \frac{d_{ij}}{\max \{d_{ij}\}}.$$

Table 1: Optimal time delays for the 33 time series during the period 1980–2012.

<table>
<thead>
<tr>
<th>Year</th>
<th>$\tau_{m_i}$ (days)</th>
<th>$r_{ij} (\tau_{m_i})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>99</td>
<td>0.7926</td>
</tr>
<tr>
<td>1981</td>
<td>99</td>
<td>0.8577</td>
</tr>
<tr>
<td>1982</td>
<td>97</td>
<td>0.8503</td>
</tr>
<tr>
<td>1983</td>
<td>146</td>
<td>-0.1871</td>
</tr>
<tr>
<td>1984</td>
<td>129</td>
<td>0.6520</td>
</tr>
<tr>
<td>1985</td>
<td>112</td>
<td>0.7487</td>
</tr>
<tr>
<td>1986</td>
<td>48</td>
<td>0.8941</td>
</tr>
<tr>
<td>1987</td>
<td>155</td>
<td>0.5918</td>
</tr>
<tr>
<td>1988</td>
<td>106</td>
<td>0.5567</td>
</tr>
<tr>
<td>1989</td>
<td>163</td>
<td>0.3985</td>
</tr>
<tr>
<td>1990</td>
<td>34</td>
<td>0.6606</td>
</tr>
<tr>
<td>1991</td>
<td>163</td>
<td>0.3458</td>
</tr>
<tr>
<td>1992</td>
<td>122</td>
<td>0.7998</td>
</tr>
<tr>
<td>1993</td>
<td>114</td>
<td>0.7617</td>
</tr>
<tr>
<td>1994</td>
<td>101</td>
<td>0.7726</td>
</tr>
<tr>
<td>1995</td>
<td>116</td>
<td>0.9506</td>
</tr>
<tr>
<td>1996</td>
<td>117</td>
<td>0.7622</td>
</tr>
<tr>
<td>1997</td>
<td>98</td>
<td>0.9423</td>
</tr>
<tr>
<td>1998</td>
<td>172</td>
<td>0.9097</td>
</tr>
<tr>
<td>1999</td>
<td>100</td>
<td>0.6130</td>
</tr>
<tr>
<td>2000</td>
<td>106</td>
<td>-0.0433</td>
</tr>
<tr>
<td>2001</td>
<td>173</td>
<td>0.7722</td>
</tr>
<tr>
<td>2002</td>
<td>158</td>
<td>-0.2180</td>
</tr>
<tr>
<td>2003</td>
<td>165</td>
<td>-0.8656</td>
</tr>
<tr>
<td>2004</td>
<td>168</td>
<td>-0.7299</td>
</tr>
<tr>
<td>2005</td>
<td>165</td>
<td>-0.9852</td>
</tr>
<tr>
<td>2006</td>
<td>62</td>
<td>0.1518</td>
</tr>
<tr>
<td>2007</td>
<td>40</td>
<td>0.9445</td>
</tr>
<tr>
<td>2008</td>
<td>111</td>
<td>0.9682</td>
</tr>
<tr>
<td>2009</td>
<td>111</td>
<td>0.8566</td>
</tr>
<tr>
<td>2010</td>
<td>144</td>
<td>-0.8174</td>
</tr>
<tr>
<td>2011</td>
<td>55</td>
<td>0.9683</td>
</tr>
<tr>
<td>2012</td>
<td>200</td>
<td>-0.8595</td>
</tr>
</tbody>
</table>

Figure 10 depicts $E$ as a contour map. To facilitate the comparison, the cases $i = j$ (i.e., those with maximum correlation value) are removed from the graph, due to their higher values.


As an alternative method to visualize and to compare results, a hierarchical clustering algorithm is adopted [24–26]. A phylogenetic tree and circular phylogram are generated, using the successive (agglomerative) clustering and
average-linkage method (Figure 11). The software PHYLIP was used for generating both graphs (http://evolution.genetics.washington.edu/phylip.html).

Figure II unveils groups of objects (years) in such a way that objects in the same group (cluster) are more similar to each other than to those in other groups. For example, we can easily identify clusters composed by years \( \mathcal{A} = \{1998, 2002, 2006, 2012\} \), \( \mathcal{B} = \{1999, 2000, 2007\} \), \( \mathcal{C} = \{2004, 2010\} \), and \( \mathcal{D} = \{1991, 2003, 2005\} \). Years in the same cluster have identical time-amplitude fire pattern. Both representations of Figures 10 and 11 can be used to visualise the clusters of forest fires, on an annual basis. Figure II leads to a result that is easier to interpret, as it identifies groups of objects that are similar, while Figure 10 just maps similarities between pairs of objects.

4. MDS Analysis and Visualization

In this section we adopt the MDS tool to visualize the relationships between forest fires events. An appropriate metric is proposed and the generated MDS graphs are analysed.

The MDS is a statistical technique for visualizing data. The MDS approach generates maps where objects that are perceived to be similar to each other are placed on the map forming clusters. The maps are indeterminate with respect to translation, rotation, and reflection and the axes have no special meaning. The algorithm requires the definition of a similarity measure (or, inversely, of a distance) and the construction of a \( s \times s \) symmetric matrix of similarities (or distances) between each pair of \( s \) objects. MDS reproduces the observed similarities by assigning a point to each object in a \( m \)-dimensional space. For \( m = 2 \) or \( m = 3 \) dimensions the points may be displayed on a “map” [27–33].

We adopt the \( 33 \times 33 \) similarity matrix \( \mathbf{E} = [e_{ij}] \), defined by (5). The MDS map for \( m = 3 \) is depicted in Figure 12. A shorter (larger) distance between two points on the map means that the corresponding objects are more
The MDS map of Figure 12 exposes the clusters that were previously identified by the hierarchical clustering (Figure 11). Comparing the MDS maps and the visualization trees, we conclude that both allow easy interpretation of the results and that there is no multiannual pattern. The MDS maps have the advantage of being more intuitive, mainly when dealing with a large number of objects.
5. Conclusion
This paper analysed forest fires data, adopting tools normally used in dynamical systems analysis. The data consisted in a public domain forest fires catalogue, containing information for Portugal and covering 33 years during the period 1980–2012. The events were modelled as time series of Dirac impulses with amplitude proportional to the burnt area. The data was analysed in an annual basis using the PPP and MDS tools. The PPP was used to model forest fires dynamics. Based on an appropriate correlation index, the MDS was adopted to compare annual patterns. Those tools allow different perspectives over forest fires that may be used to better understand such a complexity phenomenon.

Conflict of Interests
The authors declare that there is no conflict of interests regarding the publication of this paper.

References
