

Article

Complexity Analysis of Global Temperature Time Series

António M. Lopes ^{1,*}  and J. A. Tenreiro Machado ^{2,†} 

¹ UISPA–LAETA/INEGI, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200–465 Porto, Portugal

² Department of Electrical Engineering, Institute of Engineering, Polytechnic of Porto, R. Dr. António Bernardino de Almeida, 431, 4249–015 Porto, Portugal; jtm@isep.ipp.pt

* Correspondence: aml@fe.up.pt; Tel.: +351-913-499-471

† These authors contributed equally to this work.

Received: 26 March 2018; Accepted: 2 June 2018; Published: 5 June 2018



Abstract: Climate has complex dynamics due to the plethora of phenomena underlying its evolution. These characteristics pose challenges to conducting solid quantitative analysis and reaching assertive conclusions. In this paper, the global temperature time series (TTS) is viewed as a manifestation of the climate evolution, and its complexity is calculated by means of four different indices, namely the Lempel–Ziv complexity, sample entropy, signal harmonics power ratio, and fractal dimension. In the first phase, the monthly mean TTS is pre-processed by means of empirical mode decomposition, and the TTS trend is calculated. In the second phase, the complexity of the detrended signals is estimated. The four indices capture distinct features of the TTS dynamics in a 4-dim space. Hierarchical clustering is adopted for dimensional reduction and visualization in the 2-dim space. The results show that TTS complexity exhibits space-time variability, suggesting the presence of distinct climate forcing processes in both dimensions. Numerical examples with real-world data demonstrate the effectiveness of the approach.

Keywords: complexity; Lempel–Ziv complexity; sample entropy; temperature time series

1. Introduction

Understanding climatic variability is extremely important nowadays. The Earth's climate is changing and we are facing an increasing number of extreme weather events, like floods, droughts and anomalous temperatures. Since the beginning of the 20th century the Earth's surface average temperature has increased by nearly 0.8 °C [1–3], and several projections based on climate models estimate that during this century, the average temperature may increase by between 1.1 °C and 6.4 °C [1]. It has been verified that the global surface temperature has risen at an average rate of about 0.17 °C per decade since 1970 [1]. Moreover, the annual mean Arctic sea-ice extent decreased at a rate of about 3.5% to 4.1% per decade over the period from 1979 to 2012 [4]. The Earth is warming and it seems that human activity is one of the main causes for this phenomenon [5,6]. Estimates from the Intergovernmental Panel on Climate Change (IPCC) revealed a considerable increase of anthropogenic greenhouse gas emissions since the pre-industrial era [1]. In the period from 1970–2010, the annual growth rate of emissions was above 1.3% [7]. Some impacts of global warming, such as record high temperatures, melting glaciers and severe flooding are becoming increasingly common around the world [8–12]. Besides the direct effect on temperature, warming leads to the modification of wind patterns, the development of humidity, and significant changes in the rates of precipitation. These phenomena are under intensive research due to their major impacts on the social, economic and health aspects of human life [13–16].

The analysis of temperature time series (TTS) is an important topic that can help us understanding better climate dynamics. Several indicators of climate change can be used, but the time evolution of the Earth's surface temperature is probably the more straightforward measure to follow [17]. Furthermore, reliable temperature data records comprising information from meteorological stations around the world are available to support data processing [18,19].

We can find in the literature several strategies for analyzing the TTS [20]. Grieser et al. [21] studied the monthly mean temperatures of European meteorological stations. They applied statistical tools to decompose the TTS into significant components and found that the phase of the annual cycle shifts within the year, backwards and forwards in Western and Eastern Europe, respectively. The authors observed that the occurrence of extreme events increased significantly and corresponded mostly to cold peaks in winter. They also found large harmonic components with a period of 92.3 months in data from Northern and Western Europe. Additionally, this period, corresponding to about 7.7 years, was also observed in the North Atlantic Oscillation. Hughes et al. [22] pointed out that the Antarctic Peninsula region is warming faster than the rest of the world. They analyzed the minimum and maximum TTS at the Faraday/Vernadsky (Latitude 65.25 S, Longitude 64.25 W) station using a multiple regression model with non-Gaussian correlated errors. It was calculated that there was an increase in the minimum monthly temperature of 6.78 °C during the years 1951–2003. Viola et al. [23] adopted nonlinear processing tools to analyze TTS. The method of delay coordinates was applied for state space reconstruction and the delay parameters were determined by means of two methods, namely the average mutual information and the false nearest neighbors. Temperatures were extrapolated until 2028 by means of simple nonlinear prediction, which, after the state space reconstruction, uses a limited number of neighboring samples to perform predictions. Founda et al. [24] employed linear regression fitting to analyze the air TTS in Athens from 1897 until 2001 and observed a tendency towards warmer years. Ge et al. [25] studied the temperature variation in China over 20 centuries, revealing not only the cold periods from 1620–1710 and 1800–1860, but also the warming registered during the 20th century. Deser et al. [26] examined the trends of the sea surface and marine air temperatures and observed warming in all cases, with the exception of the Northwestern Atlantic Ocean. The largest warming trends were found in the middle latitudes of both the Northern and Southern hemispheres. Oñate and Pou [27] processed TTS from eleven meteorological stations on the Iberian Peninsula. Trends were identified by means of the Mann–Kendall test, and multidimensional scaling provided automatic clustering. More recently, Stephenson and Doblus-Reyes [28] used multidimensional scaling as an exploratory tool to describe ensembles of forecasts. Multidimensional scaling is a computational technique for exploring similarities between groups of data and easing the visualization of hidden patterns [29]. Lopes and Machado [5] adopted multidimensional scaling to study complex correlations between global TTS.

Earth surface temperatures exhibit slow space-time system dynamics that embed long memory effects and complex relationships. Climate data can be analyzed in a straightforward manner by calculating the means and variances for different levels of spacial and temporal aggregation, but these methods neglect long range dynamic phenomena. Furthermore, while many studies have addressed the problem of global warming, the aspect of the increasing complexity of the data has been somewhat overlooked. In fact, the complexity of the TTS is embedded into the measurements, but needs to be tackled independently of the problem of increasing temperatures. The complexity point of view may lead to a deeper insight into the predictability of future TTS evolution.

In this paper, global TTS are viewed as manifestations of a dynamic system. For each TTS, the trend is calculated by means of empirical mode decomposition (EMD), and the complexity of the detrended series is assessed by means of four indices, namely the Lempel–Ziv complexity (LZC), the sample entropy (*SampEn*), the signal harmonics power ratio (η), and the fractal dimension (*FracDim*). The four indices capture distinct features of the TTS in a multidimensional space that is projected in 2D through hierarchical clustering. The space-time variability of the complexity measures suggests that the inputs of climate vary in both dimensions. In particular, the increasing complexity of the TTS in recent years

shows that temperatures are becoming more difficult to predict. Numerical examples using real-world data for one century as well as the four complexity indices, illustrate the effectiveness of this new point of view for tackling TTS.

With these ideas in mind, the paper is organized as follows. Section 2 presents the main fundamental issues that influence global warming. Section 3 overviews the mathematical tools used for processing data. Sections 4 and 5 analyze the global TTS complexity by means of different indices. Finally, Section 6 outlines the conclusions.

2. Mathematical Fundamentals

This section introduces the main mathematical tools adopted for data processing, namely the LZC, *SampEn*, Fourier series (FS), EMD, and *FracDim*. These tools are well suited to the time series generated by many natural phenomena, such as climatic, geophysical and biological processes.

2.1. Lempel–Ziv Complexity

The LZC is a method of symbolic sequence analysis that measures the complexity of finite time series [30–32]. The LZC is based on computing the number of distinct subsequences and their recurrence rates along the series of data [33–36].

Let us consider the finite time series, $x(k)$, where $k = 1, 2, \dots, K$ represents the discrete sampling instants. In the first phase, $x(k)$ is converted into a symbolic binary sequence, $S = \{s(1), s(2), \dots, s(K)\}$, such that

$$s(k) = \begin{cases} 0, & \text{if } x(k) < T_h \\ 1, & \text{if } x(k) \geq T_h \end{cases} \quad (1)$$

where, T_h denotes a pre-defined threshold value. Usually the median of $x(k)$ is adopted for T_h , since it performs robustly against outliers [37]. In a second phase, we parse the original sequence, S , from left to right, to obtain a new (parsed) sequence, S' , formed by distinct subsequences or words. To obtain S' , different parsing methods are available [38–40]. Herein we summarize the original scheme proposed by Lempel and Ziv [30,41].

Let $S(p, q)$ denote a subsequence of S composed of the symbols between positions p and q (in our case, time instants). Thus, we have $S(p, q) = \{s(p), s(p+1), \dots, s(q)\}$, for $q \geq p$, and $S(p, q) = \{\}$ (i.e., the empty set), for $q < p$. The parsed sequence, S' , is obtained as follows:

1. Initialize:
 - (a) The parsed sequence, $S' = \{s(1)\}$;
 - (b) An auxiliary sequence, $Q = \{\}$;
 - (c) A pointer, θ , such that, $\theta(1)$ points to the symbol, $s(1)$, of S ;
 - (d) $p = 1$.
2. At the iteration p :
 - (a) Advance from $\theta(p)$ to $\theta(p+1)$ and extend the auxiliary sequence, Q , by appending the symbol $s(p+1)$;
 - (b) Check whether the current auxiliary sequence Q matches any subsequence of $S(1, p)$. If no match is found, then append, as a new word, the auxiliary sequence, Q , to the parsed sequence, S' , and reset to $Q = \{\}$;
3. Increment p ;
 - (a) If $p < K$, then go to step 2;
 - (b) If $p = K$, then append Q to S' to yield the final parsed sequence;
4. Count the number of different words, $c(K)$, in S' .

To obtain a complexity measure independent of the length of S , the count value $c(K)$ is normalized, yielding [30]

$$LZC = \frac{c(K) \log_2(K)}{K}, \quad (2)$$

which reflects the rate of occurrence of new patterns in S .

2.2. Sample Entropy

The *SampEn* is a modified version of the “approximate entropy” [42], used to measure complexity. It is defined as the symmetrical version of the logarithm of the conditional probability of two sequences that are similar for m points, remain identical for $m + 1$ points. The value of *SampEn* has little variation across the length of a time series, yielding a characteristic measure of its complexity [43]. The *SampEn* can be calculated by the five steps presented below [44–48]:

1. Generate a set of m -dimensional vectors, X_p^m ,

$$X_p^m = [x(p), x(p+1), \dots, x(p+m-1)], \quad 1 \leq p \leq K-m+1, \quad (3)$$

representing m consecutive values of $x(k)$, starting at point p ;

2. Determine the similarity, $D_{p,q}(r)$, between X_p^m and X_q^m (i.e., the vector representing m consecutive values of $x(k)$, starting at point q), by calculating

$$D_{p,q}(r) = \begin{cases} 1, & \text{if } d_{p,q} < r \\ 0, & \text{if } d_{p,q} \geq r \end{cases}, \quad (4)$$

where $d_{p,q} = \max_{n=0,1,\dots,m-1} |x(p+n) - x(q+n)|$ is a distance, and r is a tolerance value;

3. Compute the correlation sum, $B_p^m(r)$, under the constraint $p \neq q$, to exclude self-matches,

$$B_p^m(r) = \frac{\sum_{q=1, q \neq p}^{K-m} D_{p,q}(r)}{K-m-1}; \quad (5)$$

4. Find the probability, $B^m(r)$, of template matching for all vectors by

$$B^m(r) = \frac{\sum_{p=1}^{K-m} B_p^m(r)}{K-m}; \quad (6)$$

5. Calculate the *SampEn* as

$$SampEn = -\ln \frac{B^{m+1}(r)}{B^m(r)}. \quad (7)$$

The values of the parameters m and r are important in determining the *SampEn*, but no exact guidelines exist for optimizing their values. We know that the accuracy and confidence of the *SampEn* estimates increase when a small m (short template) and large r (wide tolerance) [42,46] are chosen. Moreover, when r is too small, poor conditional probability estimates are obtained, while when r is too large, detailed information is lost and *SampEn* tends to be 0. To avoid problems caused by noise r must be chosen carefully [42,46]. The generally-adopted parameter values are $m = 1$ or $m = 2$, and $r = \rho\sigma$, where σ denotes the standard deviation of the original time series, $x(k)$, and $\rho \in [0.1, 0.25]$ [46].

2.3. Fourier Analysis

The classical FS was proposed by Joseph Fourier (1768–1830) when studying problems in the area of heat conduction. The FS expresses a given signal by means of an orthonormal basis of trigonometric functions.

Let us assume that $x(k) = x(k + T)$, $k, T \in \mathbb{N}$, is a periodic function with period T , $T = \frac{1}{f}$, and $\omega = 2\pi f$, where f and ω denote the frequency and angular frequency, respectively. The signal $x(k)$ can be expressed by means of FS expansion [49,50]:

$$x(k) = \sum_{n=0}^{T-1} a_n e^{j2\pi f n k}, \quad (8)$$

where $j = \sqrt{-1}$, and the Fourier coefficients, a_n , are given by

$$a_n = \frac{1}{T} \sum_{k=0}^{T-1} x(k) e^{-j2\pi f n k}. \quad (9)$$

According to Parseval's theorem, the average power of $x(k)$ is given by

$$\mathcal{P}_x = \frac{1}{T} \sum_{k=0}^{T-1} |x(k)|^2 = \sum_{n=0}^{T-1} |a_n|^2. \quad (10)$$

2.4. Empirical Mode Decomposition

The EMD is a powerful signal processing tool that is able to deal with non-stationary and non-linear time series [51–53]. The basic idea is that the original signal $x(k)$ is decomposed into a finite set of simpler signals called intrinsic mode functions (IMF). The algorithm is as follows [51]:

1. Identify all extrema of $x(k)$;
2. Interpolate between minima and maxima, and find the lower and upper envelopes, $e_{min}(k)$ and $e_{max}(k)$, respectively;
3. Calculate the mean using $m(k) = \frac{1}{2}[e_{min}(k) + e_{max}(k)]$;
4. Extract the details using $d(k) = x(k) - m(k)$;
5. Iterate on the residual, $m(k)$.

For real-world data, this procedure iterates steps 1–4 until the detail signal, $d(k)$, is an intrinsic mode function (IMF), meaning that it has a zero-mean according to some criterion. It should be noted that the complete process guarantees that the number of extrema decreases between two successive residuals and that the decomposition is achieved with a finite number of modes.

The original signal can be represented by

$$x(k) = m_L(k) + \sum_{l=1}^L d_l(k), \quad (11)$$

where $d_l(k)$ represent the modes and $m_L(k)$ denotes the residual.

2.5. Fractal Dimension

A fractal is a geometrical object that displays identical patterns at different scales. The *FracDim* is a measure of how much the fractal fills the space as it is magnified from larger to smaller scales [54–56]. The box-counting method is often adopted for estimating the *FracDim*, due to its straightforward numerical implementation. For a fractal represented by a binary image, with non-zero pixels belonging to the fractal, and zero pixels forming the image background, the box counting algorithm is as follows [56]:

1. Pad the image with background pixels so that its dimensions are at a power of 2;

2. Cover the fractal object with a grid of squares with size ϵ (in the first iteration, there is just one square of equal size to the size of the image);
3. Count the number of boxes (i.e., squares), $N_\epsilon \in \mathbb{N}$, needed to cover the object;
4. If $\epsilon > 1$, then make $\epsilon \leftarrow \frac{\epsilon}{2}$ and repeat step 2.
5. Estimate the *FracDim* as the slope of the log-log plot, N_ϵ versus ϵ , calculated by means of the least squares method.

The estimate of *FracDim* obtained by the algorithm is a good approximation to

$$\text{FracDim} = - \lim_{\epsilon \rightarrow 0^+} \frac{\log N_\epsilon}{\log \epsilon}. \quad (12)$$

3. Dataset

The University of Delaware “Terrestrial Air Temperature: 1900–2014 Gridded Monthly Time Series (V 4.01)” was used. The dataset is publicly available at the website http://climate.geog.udel.edu/~climate/html_pages/download.html#T2014, and was retrieved on 1 March 2018. The archive consists of monthly mean air TTS for the period from January 1900 up to December 2014. The spatial coverage corresponds to land areas, with one TTS per cell, on a $0.5^\circ \times 0.5^\circ$ latitude \times longitude grid of resolution. The dataset was compiled by means of interpolation techniques on data gathered from several sources, including the Global Historical Climatology Network Monthly Version 3 (GHCN3) dataset [18]; the Daily Global Historical Climatology Network (GHCN-Daily) archive [57]; the Atmospheric Environment Service/Environment Canada archive; records from the State Hydrometeorological Institute of St. Petersburg, Russia; data for Greenland provided by the GC-Net [58]; records from the Automatic Weather Station Project (Charles R. Stearns, University of Wisconsin—Madison); the Global Synoptic Climatology Network archive (Dataset 9290c, of the National Climatic Data Center); and observations contained within the Global Surface Summary of Day (GSOD) [59].

Herein, we illustrate the usability of the dataset, while highlighting some aspects of the global warming revealed by the temperature anomalies [13]. Given a reference period, $P_{12} = [P_1, P_2]$, from year P_1 up to year P_2 , the temperature anomaly for the period $P_{34} = [P_3, P_4]$ is

$$A_{12}^{34} = \frac{1}{P_4 - P_3} \sum_{n=P_3}^{P_4} y(n) - \frac{1}{P_2 - P_1} \sum_{n=P_1}^{P_2} y(n), \quad (13)$$

where $y(n)$ represents the average temperature of year n .

Temperature anomalies represent deviations from a reference value, or long-term average [13]. Positive (negative) anomalies mean that the observed temperatures are higher (lower) than the reference value. Figure 1 depicts the annual land temperature anomalies for the years $P_{34} = [1970, 1974]$, $P_{34} = [1990, 1994]$, $P_{34} = [2000, 2004]$ and $P_{34} = [2010, 2014]$, relative to the period $P_{12} = [1951, 1980]$. The results show the Earth’s warming clearly, with emphasis on certain regions, namely in the Northern Hemisphere.

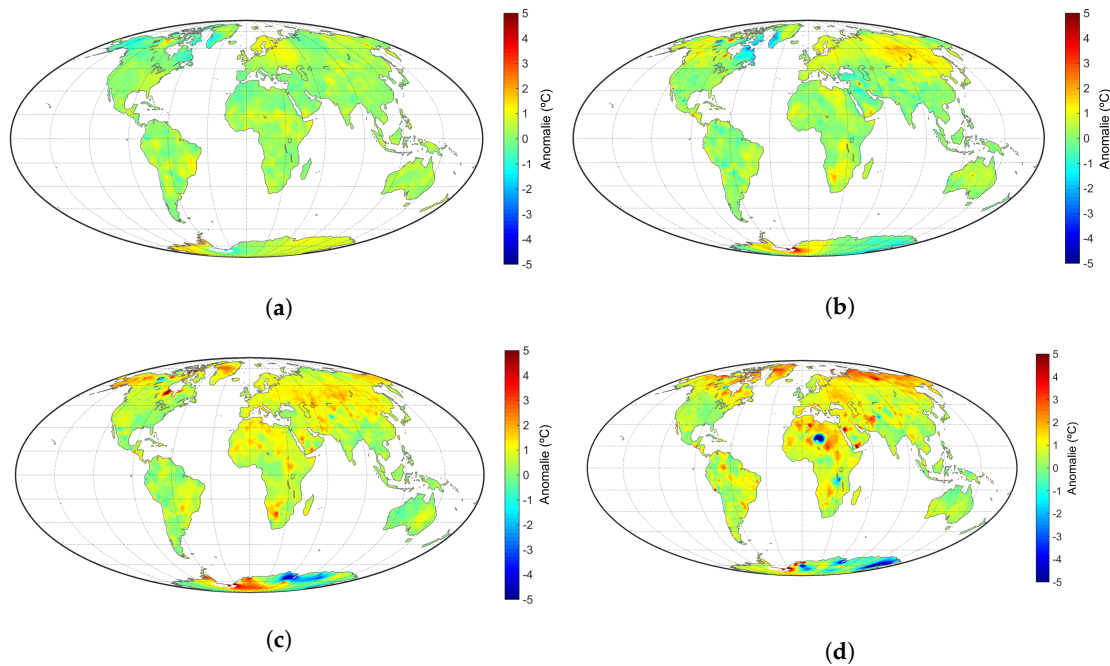


Figure 1. Worldwide land temperature anomalies relative to $P_{12} = [1951, 1980]$ for the 5-year periods: (a) $P_{34} = [1970, 1974]$; (b) $P_{34} = [1990, 1994]$; (c) $P_{34} = [2000, 2004]$; (d) $P_{34} = [2010, 2014]$.

4. On the Complexity of TTS

The TTS, $x(k), k = 1, \dots, K$, exhibits a trend. Here, k represents time with a 1-month resolution, and $k = 1$ and $k = K = 1380$ correspond to the initial and final instants—January 1900 and December 2014, respectively. The EMD proved to be a good method for determining the trends of data from nonstationary and nonlinear processes, without the need to know the trends' functional forms [53]. Within this approach, the TTS trend was estimated by the residuals of the EMD.

Figure 2 depicts the original data, $x(k)$, the IMF modes, $d_l(k), l = 1, \dots, 7$, and the residuals, $m_L(k)$, obtained for a TTS of a station with latitude/longitude coordinates $(71.25^\circ, 89.75^\circ)$, following the procedure reported in [60]. Figure 3 illustrates the original data, $x(k)$, and the trend, $\bar{x}(k) = m_L(k)$.

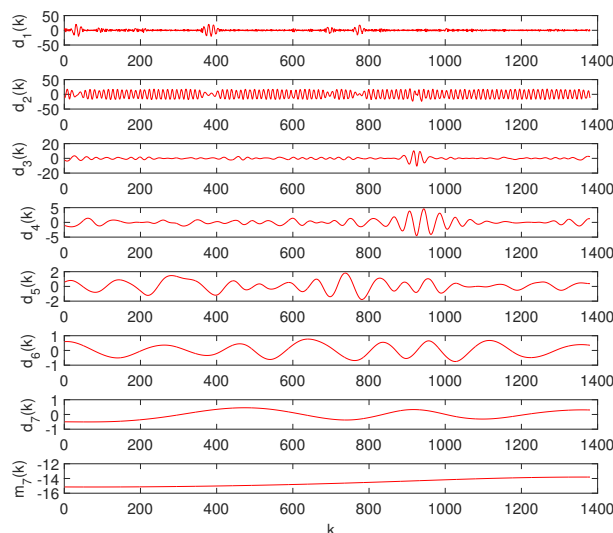


Figure 2. The intrinsic mode functions (IMF), $d_l(k), l = 1, \dots, 7$, and the residuals, $m_7(k)$, obtained for a temperature time series (TTS) of a station with latitude/longitude coordinates $(71.25^\circ, 89.75^\circ)$.

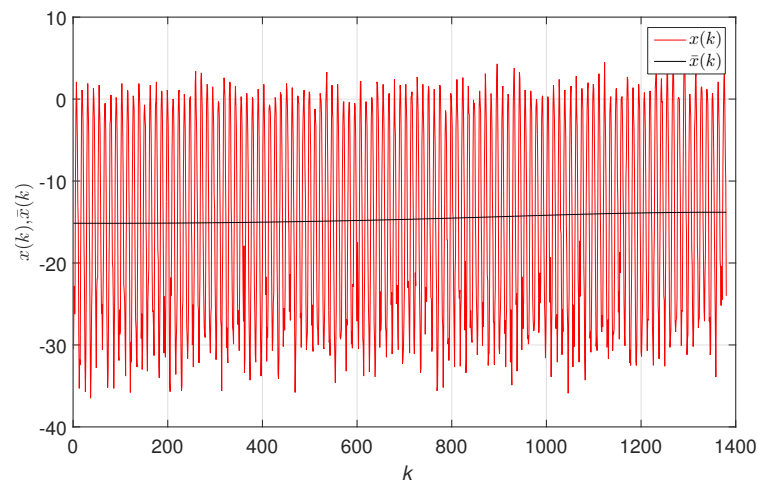


Figure 3. Monthly mean TTS, $x(k)$, and trend approximation, $\bar{x}(k)$, for the TTS at the latitude/longitude coordinates $(71.25^\circ, 89.75^\circ)$.

4.1. Lempel–Ziv Complexity of the TTS

Before determining the LZC, we first detrended the original TTS by calculating $\tilde{x}(k) = x(k) - \bar{x}(k)$, where $\bar{x}(k)$ was determined by EMD. Figure 4 exemplifies the detrending pre-processing of the TTS with latitude/longitude coordinates $(71.25^\circ, 89.75^\circ)$.

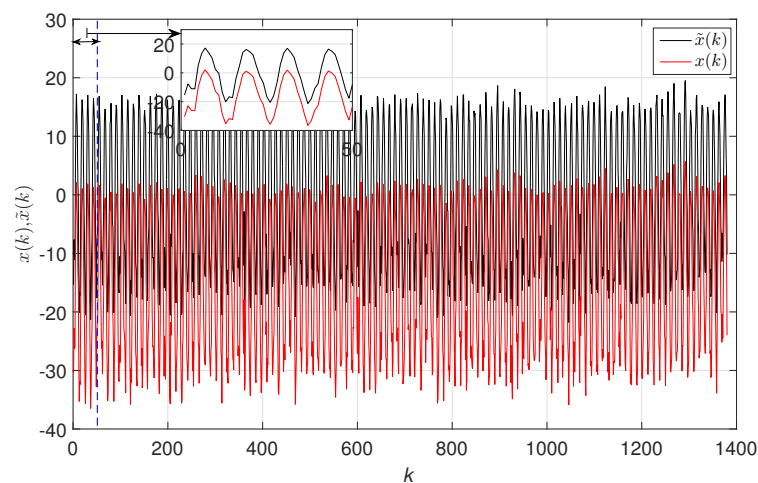


Figure 4. The detrending pre-processing of the TTS with latitude/longitude coordinates $(71.25^\circ, 89.75^\circ)$ by means of EMD.

The LZC algorithm shown in Section 2.1 was then applied to each $\tilde{x}(k)$ on the regular grid (i, j) covering land areas, totaling $L = 85,690$ TTS. Figure 5 represents the LZC_{ij} on a contour map. Using the LZC index, we verified that the monthly mean TTS values in the Northern Hemisphere and in certain regions of South America and Australia are more complex than those in other regions of the Earth. This spatial variability of complexity suggests that climate dynamics are driven by distinct regional forcing processes.

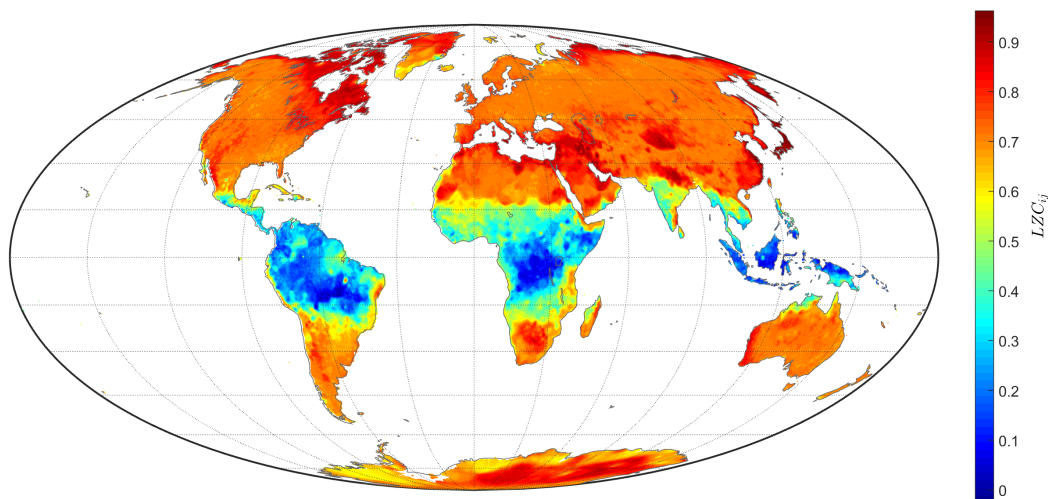


Figure 5. Contour map depicting LZC_{ij} for the total $L = 85,690$ TTS of the grid (i, j) on land.

4.2. Sample Entropy of the TTS

To calculate the *SampEn*, we applied the algorithm from Section 2.2 to the detrended TTS, $\tilde{x}(k)$, with $m = 2$ and $r = 0.5$, as discussed in Section 2.2. Figure 6 represents the $SampEn_{ij}$ for each TTS of the grid (i, j) on land.

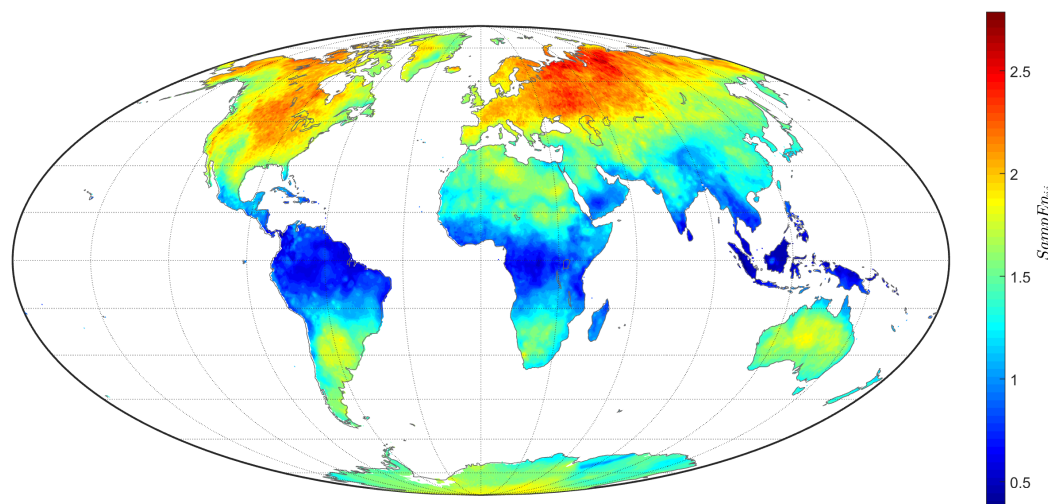


Figure 6. Contour map depicting $SampEn_{ij}$ for the total $L = 85,690$ TTS of the grid (i, j) on land.

To compare the results obtained with LZC_{ij} and $SampEn_{ij}$ we calculated the two-dimensional correlation, R , as

$$R = \frac{\sum_i \sum_j (LZC_{ij} - \langle LZC \rangle) \cdot (SampEn_{ij} - \langle SampEn \rangle)}{\sqrt{\sum_i \sum_j (LZC_{ij} - \langle LZC \rangle)^2} \cdot \sqrt{\sum_i \sum_j (SampEn_{ij} - \langle LZC \rangle)^2}}, \tag{14}$$

where $\langle \cdot \rangle$ denotes the mean of all values in a matrix. The correlation value obtained, $R = 0.97$, quantified the similarities observed between Figures 5 and 6, showing that both indices yield similar spatial patterns. The results thus suggest the existence of distinct regional climate forcing processes.

4.3. Harmonic Content of the TTS

In this sub-section, we propose an alternative index for measuring the complexity of TTS. The index, η , is given by the harmonics power ratio of the signals:

$$\eta = \left(1 - \frac{\mathcal{P}_h}{\mathcal{P}_t}\right) \times 100\%, \tag{15}$$

where

$$\mathcal{P}_h = |a_0|^2 + |a_1|^2, \tag{16a}$$

$$\mathcal{P}_t = \frac{1}{T} \sum_{k=0}^{T-1} |\tilde{x}(k)|^2, \tag{16b}$$

when $T = 12$ months.

Following Parseval’s theorem (10), \mathcal{P}_h and \mathcal{P}_t denote the signal power captured by the DC and first harmonic components, and the total power of $\tilde{x}(k)$, respectively. Therefore, η measures the signal power contained on the higher harmonics of $\tilde{x}(k)$.

Figure 7 represents the η_{ij} for the TTS of the grid (i, j) on land. We verified that tropical and equatorial regions are characterized by larger values of η .

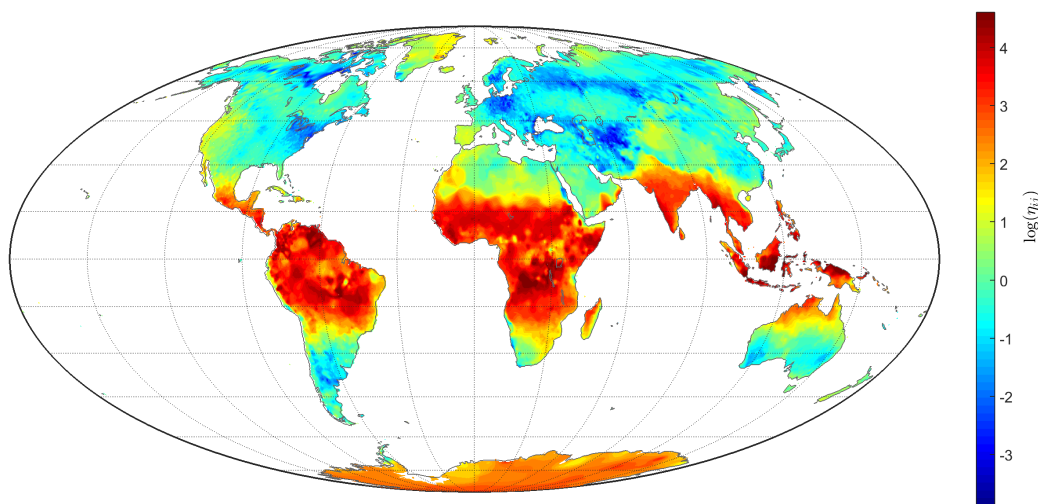


Figure 7. Contour map depicting η_{ij} for the total $L = 85,690$ TTS of the grid (i, j) on land.

4.4. Fractal Dimension of the TTS

Here we calculate the *FracDim* for the phase portrait of the TTS. A n -dimensional dynamical system can be represented by a set of first-order differential equations governing n state variables. By knowing the state variables at time $k = k_0$ and the system inputs for $k \geq k_0$, the system behavior for $k \geq k_0$ can be determined. The state space consists of the set of all possible states, each one corresponding to a unique point. As k evolves, sequences of points describing trajectories in the state space are obtained. The set of trajectories is known as the state space portrait. The phase portrait (PP) corresponds to the two-dimensional state space representation [61,62].

We consider the detrended TTS, $\tilde{x}(k)$, and its numerical time derivative, $\dot{\tilde{x}}(k) = \tilde{x}(k + 1) - \tilde{x}(k)$, to be the state variables. Figure 8 represents the PP $(\tilde{x}, \dot{\tilde{x}})$ of the detrended TTS with latitude/longitude coordinates $(71.25^\circ, 89.75^\circ)$.

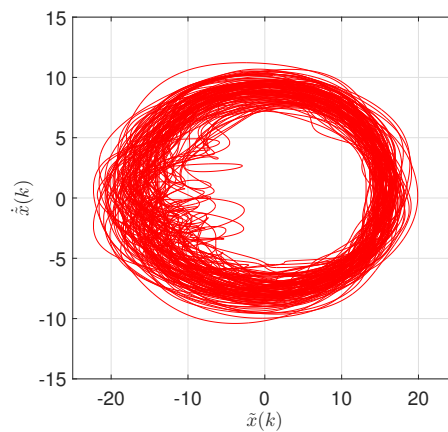


Figure 8. The phase portrait (PP) (\hat{x}, \hat{x}) of the detrended TTS with latitude/longitude coordinates $(71.25^\circ, 89.75^\circ)$.

Figure 9 represents the $FracDim_{ij}$ for all TTS of the grid (i, j) on land.

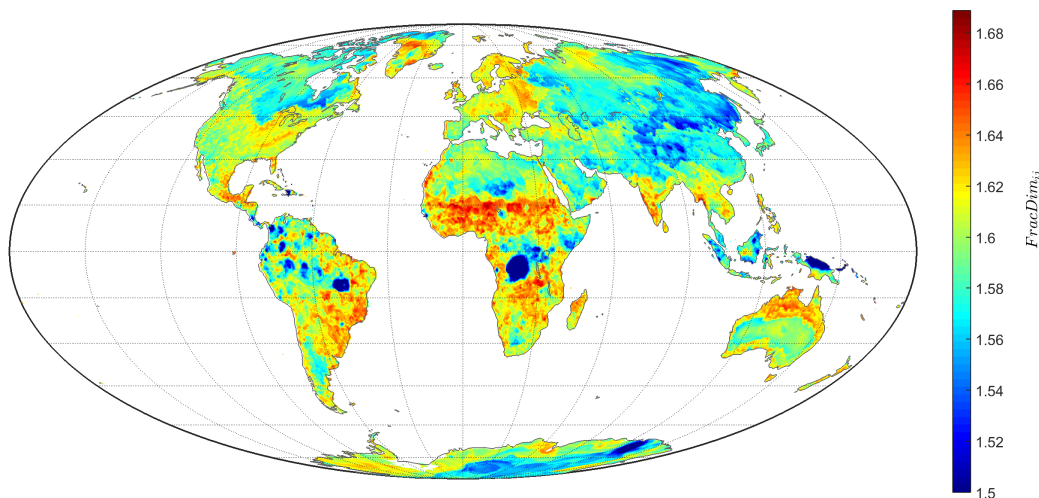


Figure 9. Contour map depicting $FracDim_{ij}$ for the total $L = 85,690$ TTS of the grid (i, j) on land.

5. Temporal Dynamics of Global Warming

In this section we describe the study of the temporal complexity of global warming. In the first phase, TTS measured on land (within a total of $L = 85,690$) was divided into $W = 43$ time intervals, $w_t, t = 1, \dots, W$, by means of a 30-year length sliding window, with a 28-year overlap. For each time interval, w_t , we calculated the L -dimensional vectors, LZC_{w_t} , $SampEn_{w_t}$, η_{w_t} and $FracDim_{w_t}$, representing the complexity of the detrended TTS at the time intervals w_t , captured in the perspective of each of the four indices. Experiments demonstrated that the window adopted established a good compromise between time discrimination and statistical significance of the data within each time interval. In a second phase, for each indice and w_t we determined the relative frequency histograms. Figure 10 depicts the histograms for the four indices versus w_t on contour maps, where the years correspond to the center of the time windows. The LZC_{w_t} and η_{w_t} maps exhibited two modes that increased with w_t almost monotonically. Regarding the LZC_{w_t} index, higher modes were mainly due to the TTS in the regions of Antarctica, South Africa, East Australia, North Africa, the Middle East and some setentrional areas of America and Asia. On the other hand, lower modes were obtained from TTS located in the tropical regions of Earth. In the η_{w_t} , the higher and lower modes were due to TTS from equatorial and tropical regions, and to the TTS from North Australia, North Africa, East Asia and

Northwest America. The $SampEn_{w_t}$ and $FracDim_{w_t}$ indices had single mode. For the $SampEn_{w_t}$, most TTS were from North Africa, South Asia, South America, Australia and Antarctica. For the $FracDim_{w_t}$, the main contribution was due to TTS from tropical regions.

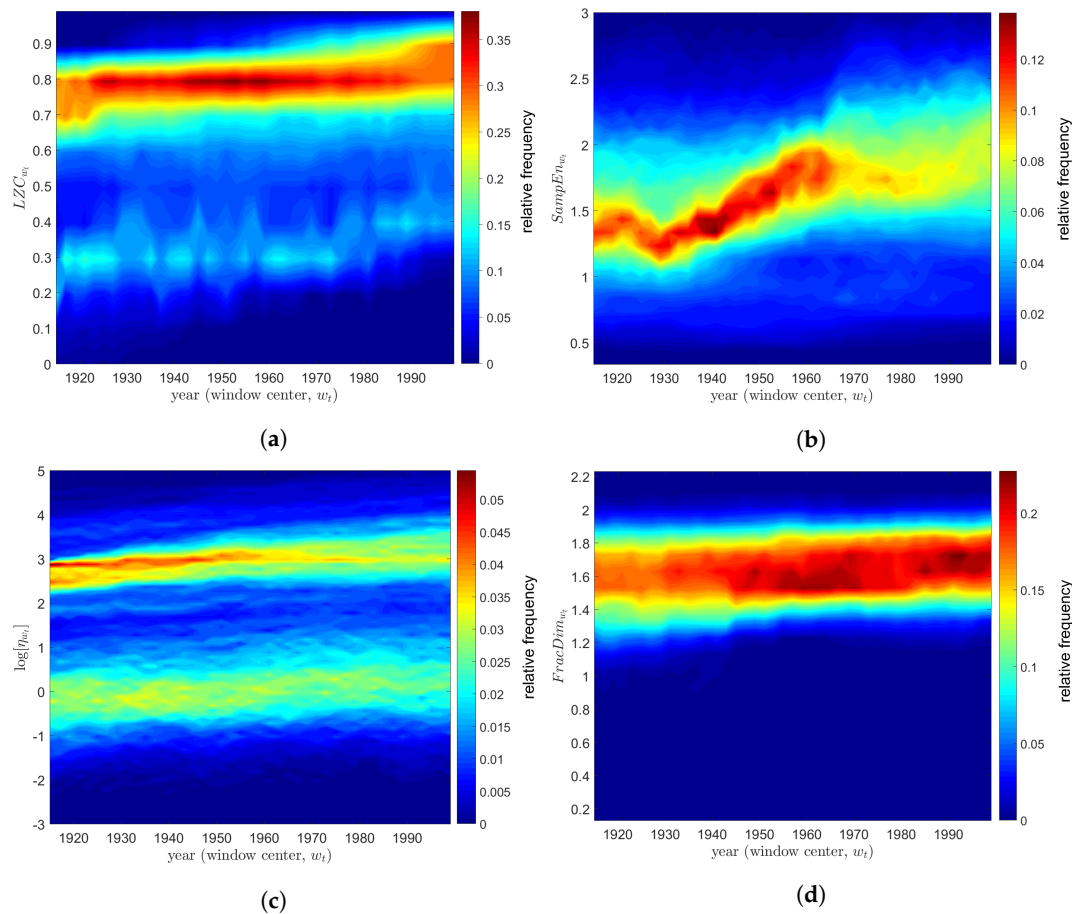


Figure 10. Contour maps of the histograms of the complexity indices versus w_t , $t = 1, \dots, W$: (a) LZC_{w_t} ; (b) $SampEn_{w_t}$; (c) $\log[\eta_{w_t}]$; (d) $FracDim_{w_t}$. Each year corresponds to the center of a time interval.

In the follow-up, to unveil temporal patterns, we adopted a hierarchical clustering (HC) analysis and visualization trees. The main objective of HC is to group together objects that are similar to one another in some sense [63,64]. The input of HC is a $W \times W$ symmetric matrix $[d_{tu}]$, $t, u = 1, \dots, W$, where d_{tu} represents the Canberra distance between the time periods, (w_t, w_u) [29,65]:

$$d_{tu} = \sum_{l=1}^L \left(\frac{|LZC_{w_t}(l) - LZC_{w_u}(l)|}{|LZC_{w_t}(l)| + |LZC_{w_u}(l)|} + \frac{|SampEn_{w_t}(l) - SampEn_{w_u}(l)|}{|SampEn_{w_t}(l)| + |SampEn_{w_u}(l)|} + \frac{|\eta_{w_t}(l) - \eta_{w_u}(l)|}{|\eta_{w_t}(l)| + |\eta_{w_u}(l)|} + \frac{|FracDim_{w_t}(l) - FracDim_{w_u}(l)|}{|FracDim_{w_t}(l)| + |FracDim_{w_u}(l)|} \right). \tag{17}$$

In the HC algorithm, the successive (agglomerative) clustering and average-linkage methods were adopted. Figure 11 depicts the corresponding graph generated by the software, PHYLIP (<http://evolution.genetics.washington.edu/phylip.html>). We identified three clusters of TTS complexity patterns, closely coincident to time periods characterized by different temperature trends [66], namely the warming period, $\mathcal{A} = [1915, 1947]$, the cooling trend in the middle of the 20th century, $\mathcal{B} = [1947, 1977]$, and the new warming period, $\mathcal{C} = [1977, 1999]$. Moreover, in \mathcal{A} , the HC unveiled two groups, namely $\mathcal{A}_1 = [1915, 1927]$ and $\mathcal{A}_2 = [1927, 1947]$. In the clusters \mathcal{A}_2 , \mathcal{B} and \mathcal{C} ,

we noted the emergence of some smaller clusters. In \mathcal{A}_2 , we found subclusters composed of the years {1929, 1931, 1933, 1935} and {1937, 1939, 1941, 1945, 1947}. In \mathcal{B} , we verified the appearance of the groups {1949, 1951, 1953, 1955, 1957, 1959} and {1961, 1963, 1965, 1967, 1969, 1971, 1973, 1975, 1977}. Finally, in \mathcal{C} , there were {1979, 1981, 1983, 1985, 1987} and {1991, 1993, 1995, 1997, 1999}. These subclusters include between 4 and 6 time windows, suggesting some kind of periodicity. However, to the authors' best knowledge, there have been no studies addressing these fluctuations with the standard perspective of the rise in temperature. Therefore, we question whether there is some hidden correlation between the increasing temperatures and their complexity that has somehow been overlooked in previous approaches. This open question will require additional studies with more detailed data analyses in time and space.

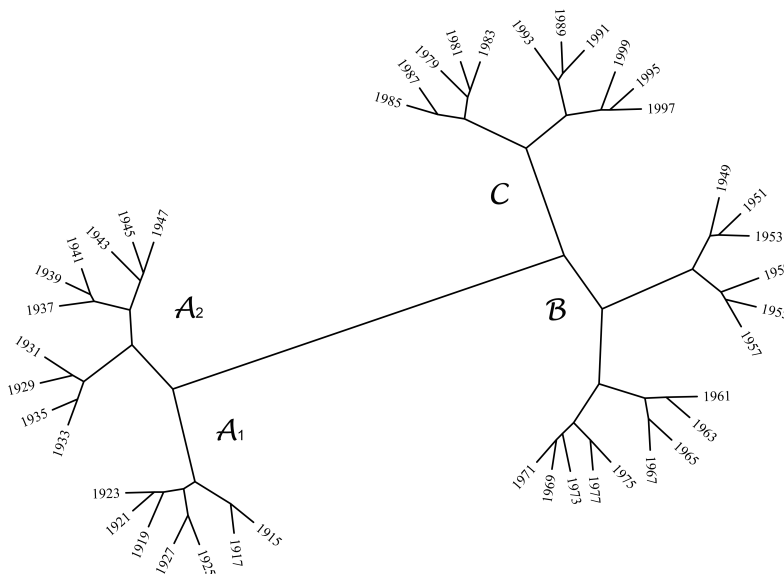


Figure 11. The hierarchical tree generated by the hierarchical clustering (HC) algorithm based on (17).

In summary, we verified that the complexity of the TTS is increasing, that a solid assessment is accomplished when a combination of several distinct measures is used, and that complexity is an important issue to consider when analyzing global warming.

6. Conclusions

The TTS embeds rich information contributed by a multitude of distinct factors at different scales. The complexity of the TTS is related to climate forcing factors. In this paper, information from worldwide land meteorological stations was processed and its complexity was calculated in space and time by means of different indices. The increasing complexity of the TTS in recent years demonstrates that there has been a change in the dominating factors that drive the climate. On one hand, we have natural forces, such as the orbital effect, solar radiation, volcanic eruptions, changes in land cover, greenhouse gas, and aerosols. On the other hand, we find anthropogenic factors, including the rise in greenhouse gas and aerosols caused by human activities, as well as sulfate air pollutants, reactive nitrogen, dust, urban heat islands, ozone change, and land-surfaces changes also due to human activities. Four complexity measures were adopted. However, no single measure captured all effects and a multi-measurement perspective led to a superior assessment of the system state. The resulting multidimensional set of measurements was synthesized by means of the HC visualization technique. The obtained clusters revealed clear patterns in the increasing complexity embedded in the TTS. More detailed quantification of these patterns by means of spatial statistics, namely spatial Fourier power spectrum, will be addressed in further research.

Author Contributions: J.A.T.M. and A.M.L. conceived, designed and performed the experiments, analyzed the data and wrote the paper.

Acknowledgments: The authors acknowledge the University of Delaware for the data used in this paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Solomon, S.; Qin, D.; Manning, M.; Chen, Z.; Marquis, M.; Averyt, K.B.; Tignor, M.; Miller, H.L. *Contribution of Working Group I to the Fourth Assessment Report Of the Intergovernmental Panel on Climate Change*; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2007.
- Root, T.L.; Price, J.T.; Hall, K.R.; Schneider, S.H.; Rosenzweig, C.; Pounds, J.A. Fingerprints of global warming on wild animals and plants. *Nature* **2003**, *421*, 57–60. [[CrossRef](#)] [[PubMed](#)]
- Huang, J.; Yu, H.; Dai, A.; Wei, Y.; Kang, L. Drylands face potential threat under 2 °C global warming target. *Nat. Clim. Chang.* **2017**, *7*, 417–422. [[CrossRef](#)]
- Desch, S.J.; Smith, N.; Groppi, C.; Vargas, P.; Jackson, R.; Kalyaan, A.; Nguyen, P.; Probst, L.; Rubin, M.E.; Singleton, H.; et al. Arctic ice management. *Earth's Future* **2017**, *5*, 107–127. [[CrossRef](#)]
- Lopes, A.M.; Machado, J.A.T. Analysis of temperature time-series: Embedding dynamics into the MDS method. *Commun. Nonlinear Sci. Numer. Simul.* **2014**, *19*, 851–871. [[CrossRef](#)]
- Alexiadis, A. Global warming and human activity: A model for studying the potential instability of the carbon dioxide/temperature feedback mechanism. *Ecol. Model.* **2007**, *203*, 243–256. [[CrossRef](#)]
- Working Group I. *The Scientific Basis. Climate Change*; Third Assessment Report of the Intergovernmental Panel on Climate Change; IPCC: Geneva, Switzerland, 2001.
- Dai, A. Drought under global warming: A review. *Wiley Interdiscip. Rev. Clim. Chang.* **2011**, *2*, 45–65. [[CrossRef](#)]
- You, Q.; Kang, S.; Pepin, N.; Flügel, W.A.; Sanchez-Lorenzo, A.; Yan, Y.; Zhang, Y. Climate warming and associated changes in atmospheric circulation in the eastern and central Tibetan Plateau from a homogenized dataset. *Glob. Planet. Chang.* **2010**, *72*, 11–24. [[CrossRef](#)]
- Jevrejeva, S.; Moore, J.C.; Grinsted, A. Sea level projections to AD2500 with a new generation of climate change scenarios. *Glob. Planet. Chang.* **2012**, *80*, 14–20. [[CrossRef](#)]
- Giannakopoulos, C.; le Sager, P.; Bindl, M.; Moriondo, M.; Kostopoulou, E.; Goodess, C. Climatic changes and associated impacts in the Mediterranean resulting from a 2 °C global warming. *Glob. Planet. Chang.* **2009**, *68*, 209–224. [[CrossRef](#)]
- Zhu, Y.; Wang, H.; Zhou, W.; Ma, J. Recent changes in the summer precipitation pattern in East China and the background circulation. *Clim. Dyn.* **2011**, *36*, 1463–1473. [[CrossRef](#)]
- Hansen, J.; Ruedy, R.; Sato, M.; Lo, K. Global surface temperature change. *Rev. Geophys.* **2010**, *48*, RG4004. [[CrossRef](#)]
- Brohan, P.; Kennedy, J.J.; Harris, I.; Tett, S.F.; Jones, P.D. Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res. Atmos.* **2006**, *111*. [[CrossRef](#)]
- Menne, M.J.; Williams, C.N., Jr. Homogenization of temperature series via pairwise comparisons. *J. Clim.* **2009**, *22*, 1700–1717. [[CrossRef](#)]
- Wu, Z.; Huang, N.E.; Wallace, J.M.; Smoliak, B.V.; Chen, X. On the time-varying trend in global-mean surface temperature. *Clim. Dyn.* **2011**, *37*, 759–773. [[CrossRef](#)]
- Rohde, R.; Muller, R.; Jacobsen, R.; Perlmutter, S.; Rosenfeld, A.; Wurtele, J.; Curry, J.; Wickhams, C.; Mosher, S. Berkeley Earth temperature averaging process. *Geoinf. Geostat. Overv.* **2013**, *1*, 1–13. [[CrossRef](#)]
- Lawrimore, J.H.; Menne, M.J.; Gleason, B.E.; Williams, C.N.; Wuertz, D.B.; Vose, R.S.; Rennie, J. An overview of the Global Historical Climatology Network monthly mean temperature data set, version 3. *J. Geophys. Res. Atmos.* **2011**, *116*. [[CrossRef](#)]
- Jones, P.; Lister, D.; Osborn, T.; Harpham, C.; Salmon, M.; Morice, C. Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *J. Geophys. Res. Atmos.* **2012**, *117*. [[CrossRef](#)]
- Capilla, C. Time series analysis and identification of trends in a Mediterranean urban area. *Glob. Planet. Chang.* **2008**, *63*, 275–281. [[CrossRef](#)]

21. Grieser, J.; Trömel, S.; Schönwiese, C.D. Statistical time series decomposition into significant components and application to European temperature. *Theor. Appl. Climatol.* **2002**, *71*, 171–183. [[CrossRef](#)]
22. Hughes, G.L.; Rao, S.S.; Rao, T.S. Statistical analysis and time-series models for minimum/maximum temperatures in the Antarctic Peninsula. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **2007**, *463*, 241–259. [[CrossRef](#)]
23. Viola, F.M.; Paiva, S.L.; Savi, M.A. Analysis of the global warming dynamics from temperature time series. *Ecol. Model.* **2010**, *221*, 1964–1978. [[CrossRef](#)]
24. Founda, D.; Papadopoulos, K.; Petrakis, M.; Giannakopoulos, C.; Good, P. Analysis of mean, maximum, and minimum temperature in Athens from 1897 to 2001 with emphasis on the last decade: Trends, warm events, and cold events. *Glob. Planet. Chang.* **2004**, *44*, 27–38. [[CrossRef](#)]
25. Ge, Q.S.; Zheng, J.Y.; Hao, Z.X.; Shao, X.M.; Wang, W.C.; Luterbacher, J. Temperature variation through 2000 years in China: An uncertainty analysis of reconstruction and regional difference. *Geophys. Res. Lett.* **2010**, *37*. [[CrossRef](#)]
26. Deser, C.; Phillips, A.S.; Alexander, M.A. Twentieth century tropical sea surface temperature trends revisited. *Geophys. Res. Lett.* **2010**, *37*. [[CrossRef](#)]
27. Oñate, J.J.; Pou, A. Temperature variations in Spain since 1901: A preliminary analysis. *Int. J. Climatol.* **1996**, *16*, 805–815. [[CrossRef](#)]
28. Stephenson, D.B.; Doblas-Reyes, F.J. Statistical methods for interpreting Monte Carlo ensemble forecasts. *Tellus A* **2000**, *52*, 300–322. [[CrossRef](#)]
29. Machado, J.A.T.; Lopes, A.M.; Galhano, A.M. Multidimensional scaling visualization using parametric similarity indices. *Entropy* **2015**, *17*, 1775–1794. [[CrossRef](#)]
30. Lempel, A.; Ziv, J. On the complexity of finite sequences. *IEEE Trans. Inf. Theory* **1976**, *22*, 75–81. [[CrossRef](#)]
31. Shao, Z.G. Contrasting the complexity of the climate of the past 122,000 years and recent 2000 years. *Sci. Rep.* **2017**, *7*, 4143. [[CrossRef](#)] [[PubMed](#)]
32. Zhang, Y.; Wei, S.; di Maria, C.; Liu, C. Using Lempel–Ziv complexity to assess ECG signal quality. *J. Med. Biol. Eng.* **2016**, *36*, 625–634. [[CrossRef](#)] [[PubMed](#)]
33. Abásolo, D.; Simons, S.; da Silva, R.M.; Tononi, G.; Vyazovskiy, V.V. Lempel–Ziv complexity of cortical activity during sleep and waking in rats. *J. Neurophysiol.* **2015**, *113*, 2742–2752. [[CrossRef](#)] [[PubMed](#)]
34. Wu, X.; Xu, J. Complexity and brain function. *Acta Biophys. Sin.* **1991**, *7*, 103–106.
35. Radhakrishnan, N.; Gangadhar, B. Estimating regularity in epileptic seizure time-series data. *IEEE Eng. Med. Biol. Mag.* **1998**, *17*, 89–94. [[CrossRef](#)] [[PubMed](#)]
36. Li, C.; Li, Z.; Zheng, X.; Ma, H.; Yu, X. A generalization of Lempel–Ziv complexity and its application to the comparison of protein sequences. *J. Math. Chem.* **2010**, *48*, 330–338. [[CrossRef](#)]
37. Nagarajan, R. Quantifying physiological data with Lempel–Ziv complexity—certain issues. *IEEE Trans. Biomed. Eng.* **2002**, *49*, 1371–1373. [[CrossRef](#)] [[PubMed](#)]
38. Hu, J.; Gao, J.; Principe, J.C. Analysis of biomedical signals by the Lempel–Ziv complexity: The effect of finite data size. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 2606–2609. [[PubMed](#)]
39. Zhang, X.S.; Zhu, Y.S.; Thakor, N.V.; Wang, Z.Z. Detecting ventricular tachycardia and fibrillation by complexity measure. *IEEE Trans. Biomed. Eng.* **1999**, *46*, 548–555. [[CrossRef](#)] [[PubMed](#)]
40. Aboy, M.; Hornero, R.; Abásolo, D.; Álvarez, D. Interpretation of the Lempel–Ziv complexity measure in the context of biomedical signal analysis. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 2282–2288. [[CrossRef](#)] [[PubMed](#)]
41. Melchert, O.; Hartmann, A. Analysis of the phase transition in the two-dimensional Ising ferromagnet using a Lempel–Ziv string-parsing scheme and black-box data-compression utilities. *Phys. Rev. E* **2015**, *91*, 023306. [[CrossRef](#)] [[PubMed](#)]
42. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [[CrossRef](#)] [[PubMed](#)]
43. Pincus, S. Approximate entropy (*ApEn*) as a complexity measure. *Chaos Interdiscip. J. Nonlinear Sci.* **1995**, *5*, 110–117. [[CrossRef](#)] [[PubMed](#)]
44. Tang, L.; Lv, H.; Yang, F.; Yu, L. Complexity testing techniques for time series data: A comprehensive literature review. *Chaos Solitons Fractals* **2015**, *81*, 117–135. [[CrossRef](#)]
45. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [[CrossRef](#)] [[PubMed](#)]

46. Alcaraz, R.; Rieta, J. A novel application of sample entropy to the electrocardiogram of atrial fibrillation. *Nonlinear Anal. Real World Appl.* **2010**, *11*, 1026–1035. [[CrossRef](#)]
47. Eduardo Virgilio Silva, L.; Otavio Murta, L., Jr. Evaluation of physiologic complexity in time series using generalized sample entropy and surrogate data analysis. *Chaos Interdiscip. J. Nonlinear Sci.* **2012**, *22*, 043105. [[CrossRef](#)] [[PubMed](#)]
48. Mihailović, D.; Nikolić-Đorić, E.; Drešković, N.; Mimić, G. Complexity analysis of the turbulent environmental fluid flow time series. *Phys. Stat. Mech. Its Appl.* **2014**, *395*, 96–104. [[CrossRef](#)]
49. Stein, E.M.; Shakarchi, R. *Fourier Analysis: An Introduction*; Princeton University Press: Princeton, NJ, USA, 2003.
50. Dym, H.; McKean, H. *Fourier Series and Integrals*; Academic Press: San Diego, CA, USA, 1972.
51. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]
52. Pachori, R.B. Discrimination between ictal and seizure-free EEG signals using empirical mode decomposition. *Res. Lett. Signal Process.* **2008**, *2008*, 14. [[CrossRef](#)]
53. Wu, Z.; Huang, N.E.; Long, S.R.; Peng, C.K. On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 14889–14894. [[CrossRef](#)] [[PubMed](#)]
54. Berry, M.V. Diffractals. *J. Phys. Math. Gen.* **1979**, *12*, 781–797. [[CrossRef](#)]
55. Fleckinger-Pelle, J.; Lapidus, M. Tambour fractal: Vers une résolution de la conjecture de Weyl-Berry pour les valeurs propres du laplacien. *C. R. Acad. Sci. Sér. Math.* **1988**, *306*, 171–175.
56. Valério, D.; Lopes, A.M.; Tenreiro Machado, J.A.T. Entropy Analysis of a Railway Network's Complexity. *Entropy* **2016**, *18*, 388. [[CrossRef](#)]
57. Menne, M.J.; Durre, I.; Vose, R.S.; Gleason, B.E.; Houston, T.G. An overview of the global historical climatology network-daily database. *J. Atmos. Ocean. Technol.* **2012**, *29*, 897–910. [[CrossRef](#)]
58. Steffen, K.; Box, J.; Abdalati, W. *Greenland Climate Network: GC-Net*; US Army Cold Regions Reattach and Engineering (CRREL), CRREL Special Report; CRREL: Hanover, NH, USA, 1996; pp. 98–103.
59. Willmott, C.J.; Matsuura, K. *Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950–1996)*. 2000. Available online: http://climate.geog.udel.edu/~climate/html_pages/Global2014/README.GlobalTsT2014.html (accessed on 20 February 2018).
60. Huang, N.E.; Wu, M.L.C.; Long, S.R.; Shen, S.S.; Qu, W.; Gloersen, P.; Fan, K.L. A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proc. R. Soc. Math. Phys. Eng. Sci.* **2003**, *459*, 2317–2345. [[CrossRef](#)]
61. Polderman, J.W.; Willems, J.C. *Introduction to Mathematical Systems Theory: A Behavioral Approach*; Springer: Berlin/Heidelberg, Germany, 1998.
62. Lopes, A.M.; Machado, J.A.T. State space analysis of forest fires. *J. Vib. Control* **2016**, *22*, 2153–2164. [[CrossRef](#)]
63. Machado, J.A.T.; Lopes, A.M. Analysis and visualization of seismic data using mutual information. *Entropy* **2013**, *15*, 3892–3909. [[CrossRef](#)]
64. Murtagh, F. A survey of recent advances in hierarchical clustering algorithms. *Comput. J.* **1983**, *26*, 354–359. [[CrossRef](#)]
65. Cha, S.H. Comprehensive survey on distance/similarity measures between probability density functions. *Int. J. Math. Models Methods Appl. Sci.* **2007**, *4*, 300–307.
66. Ellenburg, W.; McNider, R.; Cruise, J.; Christy, J.R. Towards an understanding of the twentieth-century cooling trend in the southeastern United States: Biogeophysical impacts of land-use change. *Earth Interact.* **2016**, *20*, 1–31. [[CrossRef](#)]

