Noname manuscript No.
(will be inserted by the editor)

Detection of early warning signals in paleoclimate data using a genetic time series segmentation algorithm

Nikolaou, Athanasia · Gutiérrez, Pedro Antonio · Durán, Antonio. · Dicaire, Isabelle · Fernández-Navarro, Francisco · Hervás-Martínez, César.

Received: date / Accepted: date

Abstract This paper proposes a time series segmentation algorithm combining a clustering technique and a genetic algorithm to automatically find segments sharing common statistical characteristics in paleoclimate time series. The segments are transformed into a six-dimensional space composed of six statistical measures,

Athanasia Nikolaou and Isabelle Dicaire

Advanced Concepts Team, European Space Research and Technology Centre (ESTEC), European Space Agency (ESA), Noordwijk, Netherlands

E-mail: athanasia.nikolaou@esa.int; isabelle.dicaire@esa.int

Pedro A. Gutiérrez, Antonio Durán and C. Hervás-Martínez,

Department of Computer Science and Numerical Analysis, University of Córdoba, Spain

E-mail: pagutierrez@uco.es; i92duroa@uco.es; chervas@uco.es

Francisco Fernández-Navarro

Department of Mathematics and Engineering, Universidad Loyola Andalucia, Spain

E-mail: i22fenaf@uco.es; fafernandez@uloyola.es

most of which have been previously considered in the detection of warning signals of critical transitions. Experimental results show that the proposed approach applied to paleoclimate data could effectively analyse Dansgaard-Oeschger (DO) events and uncover commonalities and differences in their statistical and possibly their dynamical characterisation. In particular, warning signals were robustly detected in the GISP2 and NGRIP δ^{18}O ice core data for several DO events (e.g. DO 1, 4, 8 & 12) in the form of an order of magnitude increase in variance, autocorrelation and mean square distance from a linear approximation, the mean square error. The increase in mean square error, suggesting nonlinear behaviour, has been found to correspond with an increase in variance prior to several DO events for ∼90% of the algorithm runs for the GISP2 δ^{18}O dataset and for ∼100% of the algorithm runs for the NGRIP δ^{18}O dataset. The proposed approach applied on both model output and paleoclimate datasets provides a novel visualisation tool of climate time series analysis.

Keywords Warning Signals · Time series Segmentation · Tipping Points · Abrupt Climate Change · Genetic Algorithms · Clustering

1 Introduction

The statistical tools used to extract knowledge from time series analysis have undergone considerable development during the past decade (see Livina and Lenton, 2007; Livina et al., 2011; Lenton et al., 2012; Scheffer et al., 2009; Dakos et al., 2008; Held and Kleinen, 2004; Cimatoribus et al., 2013). Driven by the ultimate aim of understanding past climate variability, the above studies focused on statistical analysis of time series that demonstrate *threshold behaviour* as used in Alley

et al. (2003). Candidate explanations for transitions of a system over thresholds link to dynamical systems analysis, which is used for gaining insight into internal variability modes and response to external forcing on both simple and complex systems (Saltzman, 2001). Adopting the notation from Ashwin et al. (2012) the abrupt shift from a stable state to another stable state could be e.g. due to *B-tipping* or *N-tipping*. In B-tipping the system is driven past bifurcation points, where equilibrium solutions loose their stability past a critical value with respect to a control parameter (Saltzman, 2001). In N-tipping, noise fluctuations of a fast component affect a slower variable pushing the system away from the neighbourhood of an attractor to that of another one (Hasselmann, 1976). Those are only two of the candidate explanations of tipping points; combined with the fact that in open systems subject to internal and external forcings such as the climate system both dynamical behaviours can take place (Ashwin et al., 2012), attributing causal relations to abrupt transitions with certainty is a challenging task. Therefore, statistical analysis is required to complement the mapping of the variability before tipping points. To reveal additional properties of the system, sophisticated tools have been developed for statistical monitoring of its evolution.

One of the most studied paleoclimate proxy data series that demonstrates tipping behaviour is the oxygen isotope δ^{18}O of the Greenland ice cores ranging to more than 100,000 years before today (Svensson et al., 2008). Abrupt transitions can be seen in this time series, that correspond to Dansgaard-Oeschger (DO) events. DOs consist of sudden warming in Greenland followed by gradual cooling. The last of those warmings is called the Younger Dryas event, and marked the transition to today's warm climate (Alley et al., 2003).

By focusing on a single transition event, the work of Dakos et al. (2008) drew attention on bifurcation points and their precursors as they could identify slowing down as an early warning signal (EWS) before the Younger Dryas period. In particular, the increase in autocorrelation within a fixed sliding window as wide as half the record length was found to precede the transition. Ditlevsen and Johnsen (2010) brought forward that the fluctuation-dissipation theorem from Kubo (Kubo, 1966) imposes both increasing autocorrelation and variance before crossing a bifurcation. Cimatoribus et al. (2013) suggested preprocessing or filtering the raw proxy data in order to infer novel diagnostics, such as the Detrended Fluctuation Analysis (DFA) coefficient. Based on that indicator Livina and Lenton (2007) have measured the proximity of a system to a tipping point not in units of critical parameter difference, but by monitoring the evolution of the DFA propagator. Held and Kleinen (2004) proposed the method of degenerate fingerprinting and hypothesized the system as an auto-regressive process with lag-1 autocorrelation, considering only exponentially decaying modes and white noise in the dynamics of the system. Rahmstorf (2003) tried to automatize the characterization of the DO events by introducing an event detection algorithm based on the slope of the data curve within time intervals of 200 years. The study failed to detect certain DO events because of the different time scales within which each DO event developed. From the above studies the statistical quantities of variance and lag-1 autocorrelation were pointed out as early warning indicators of critical transitions and the slope within a fixed time interval was found to characterise most of the DO events.

In this paper, a time series segmentation algorithm combining a genetic algorithm and a clustering technique is proposed to address the existence of EWSs before tipping points in climate time series. Starting from a random division pattern

the different segments within the time series are classified according to similarities in their statistical parameters (including variance, lag-1 autocorrelation and slope; as considered in previous works). The algorithm is then evolved by renewing its segmentation pattern at each iteration in a process called genetic evolution. to optimize the class label assigned to each segment, where algorithm operators are applied and the best-fitted individuals are selected. By keeping intact the segments that group well in statistical similarity and rearranging those that were far from any of the groups, the iteration converges to a final pattern of division of the time series, providing as output the data regions that have common characteristics. No prior knowledge of the tipping point locations are supplied to the algorithm. The goal of the algorithm is to provide a more compact representation of time series while keeping high the number of statistical parameters employed. The time series segmentation problem has been widely studied within various disciplines. For example, such algorithms have been successfully applied in phoneme recognition (Xiong et al., 1994; Prandom et al., 1997), paleoecological problems (Bennett, 1996), telecommunication applications (Himberg et al., 2001) or financial problems (Tseng et al., 2009). For an excellent review of the field see Keogh et al. (2001). Furthermore the interest in GAs applied to climate tipping points is rising, e.g. Lenton et al. (2009) used a GA to tune 12 physical parameters of an Earth System Model to study the tipping of the Atlantic thermohaline circulation following a multi-objective optimization method.

The layout of the paper is as follows. Section 2 introduces the segmentation algorithm with a detailed description of the embedded genetic algorithm, the six statistical metrics and the clustering process. Section 3 presents the synthetic and paleoclimate datasets used in this study and the algorithm parameters. Section

4 presents the main results of the segmentation algorithm on the paleoclimate and synthetic datasets, the latter used to analyse the algorithm's capabilities in a controlled environment. Section 5 discusses the results from the point of view of dynamical system theory in addition to possible limitations of the algorithm. Finally section 6 reviews the main findings of this paper.

2 Segmentation Algorithm

2.1 General overview of the segmentation algorithm

This paper proposes a novel Genetic Algorithm (GA) from the field of time series segmentation (see Sclove, 1983; Himberg et al., 2001; Keogh et al., 2001; Chung et al., 2004). The general objective of the GA is to identify segments with common characteristics by applying a label to these segments. In practice this means finding the cutpoints of the different segments to be discovered together with their class labelling. As in traditional GAs, the proposed approach considers a population of candidate solutions representing different possible segmentations which are evolved towards better segmentation solutions. Each individual is represented by an array of integer values (chromosome representation) which can be mutated and recombined. The evolution starts from a population of randomly generated segmentations. After that, every segment in every chromosome is categorized using six statistical metrics, most of which were previously considered in climate tipping point research (including variance, autocorrelation, and skewness). The clustering technique is then applied over this six-dimensional space for every chromosome using the k-means clustering algorithm (MacQueen et al., 1967) and a fitness value is assigned to every chromosome according to the degree of homogeneity of the

segments with respect to their centroids (i.e. the mean value of each cluster). A class label is assigned during the clustering process. After that, different mutation and crossover operators are applied to explore the search space. This procedure is repeated during n generations. The final mathematical goal of the proposed GA is to minimize the distance of each segment to its centroid in the six-dimensional space where the six dimensions are statistical properties of each segment.

The time series segmentation algorithm presented in this paper features the following characteristics:

- A class label is assigned to the different segments via the combination of the GA with the clustering technique; traditional approaches would only provide the segmentation points (Sclove, 1983; Himberg et al., 2001; Keogh et al., 2001). This is specially useful for finding common patterns arising in the climate datasets.
- Apart from the determination of the cutpoints ($t_i, i = 1, \ldots, m-1$), the main idea is that of *transitions* between classes. The analysis of these transitions is crucial to the detection of EWSs, as it indicates changes in the statistical parameters.
- Each segment is represented by a six-dimensional vector where the dimensions are the six statistical metrics, some of which have been previously considered in the detection of EWSs of critical transitions (Cimatoribus et al., 2013; Dakos et al., 2008, 2012).
- Instead of representing the time series evolution by plotting one of its metrics as done in previous works, the approach proposed in this paper allows to visualise

several metrics simultaneously and to compare several sections of the time series to find common patterns.

- The proposed algorithm finds automatically the length of the data segment of interest, in contrast to a fixed sliding window or fixed time interval when calculating the statistics without providing prior information nor following the trial and error approach.
- This algorithm is a significant extension of the one presented in Tseng et al. (2009) as it enables the clustering of segments of various lengths without performing advanced signal processing such as wavelet analysis, which would add bias to the signal by assuming arbitrary mother wavelet shapes.

The different GA characteristics are defined in the following subsections.

2.2 Chromosome representation and initial population

A direct encoding of the final segmentation solution is adopted where each individual chromosome consists of an array of integer values (Michalewicz, 1996). Each position stores a cutting point of the time series. A chromosome of m segments is represented by $\{t_1, \ldots, t_{m-1}\}$, where the value t_i is the index of the i-th cutting point. In this way, the first segment is delimited by the cutting points 1 and t_1, the second by the cutting points t_1 and t_2 and so on. An example of this chromosome representation is given in Figure 1.

A GA requires a population of feasible solutions to be initialized and updated during the evolutionary process. As mentioned above, each individual within a population is a possible segmentation result for the time series considered. An initial set of chromosomes is thus generated with some constraints to form feasi-

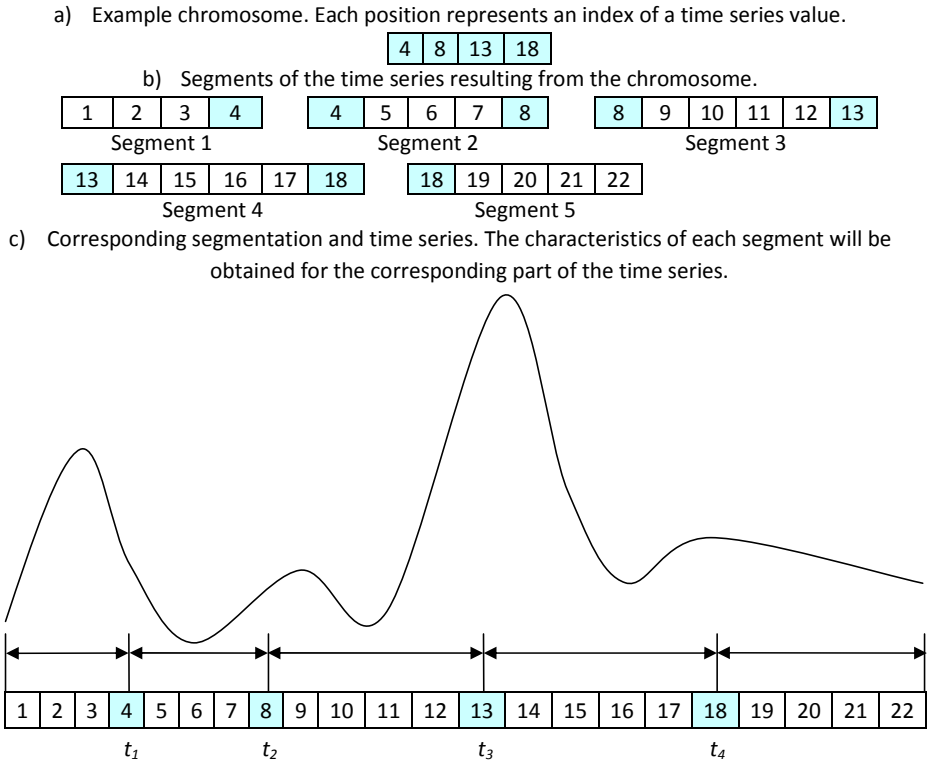


Fig. 1: Chromosome representation (Online version in colour)

ble segments. This initial population of t individuals is randomly generated. The number of individuals will be kept constant during the evolution. Further information of the creation of each initial individual can be found in in Pérez-Ortiz et al. (2014).

2.3 Segment characteristics

As a result of the genetic operators, the segments in a chromosome may have different lengths. Thus, an approach had to be designed to transform all the segments to the same dimensional space. In this paper, six statistical metrics are

measured for all the segments included in a chromosome, allowing the GA to calculate similarities between segments using the same dimensional space. For the sake of simplicity, the following characteristics are referred to the segment S_s which is mathematically defined as $S_s = \{y_{t_{s-1}}, \ldots, y_{t_s}\}$:

1. Variance (S_s^2): It is a measure of variability that indicates the degree of homogeneity of a group of observations. The mathematical expression of this metric is:

$$S_s^2 = \frac{1}{t_s - t_{s-1}} \sum_{i=t_{s-1}}^{t_s} (y_i - \bar{y_s})^2 , \tag{1}$$

where $(t_s - t_{s-1})$ is the number of points of the segment, t_{s-1} is the index of the first point in the s-th segment, t_s is the index of the last point in the segment, y_i are the time series values of the segment, and $\bar{y_s}$ is the average value of the segment.

2. Skewness (γ_{1s}): The skewness represents the asymmetry of a distribution of values within a segment. Segments can be skewed either up or down with respect to the arithmetic mean. The skewness is defined as:

$$\gamma_{1s} = \frac{\frac{1}{t_s - t_{s-1}} \sum_{i=t_{s-1}}^{t_s} (y_i - \bar{y_s})^3}{S_s^3}, \tag{2}$$

where S_s is the standard deviation of the s-th segment.

3. Kurtosis (γ_{2s}): It measures the degree of concentration that the values present around the mean of the distribution. Positive kurtosis (i.e. long tails) indicate large excursions away from the arithmetic mean. Kurtosis is defined as:

$$\gamma_{2s} = \frac{\frac{1}{t_s - t_{s-1}} \sum_{i=t_{s-1}}^{t_s} (y_i - \bar{y_s})^4}{S_s^4} - 3. \tag{3}$$

4. Slope of a linear regression over the points of the segment (a_s): A linear model is constructed for every segment, trying to achieve the best linear approximation of the points in the evaluated segment. The slope of the linear model

is a measure of the general tendency of the segment. The slope parameter is obtained as:

$$a_s = \frac{S_s^{yt}}{(S_s^t)^2}, \tag{4}$$

where S_s^{yt} is the covariance of the time indexes, t, and the time series values, y, for the s-th segment; and where S_s^t is the standard deviation of the time values. The mathematical expression for the covariance is:

$$S_s^{yt} = \frac{1}{t_s - t_{s-1}} \sum_{i=t_{s-1}}^{t_s} (i - \bar{t_s}) \cdot (y_i - \bar{y_s}). \tag{5}$$

5. Mean Squared Error (MSE_s): This metric measures the degree of nonlinearity of the segment. As for the slope, a linear model is fitted and used to obtain the MSE_s:

$$MSE_s = S_s^2 \cdot (1 - r_s^2), \tag{6}$$

where:

$$r_s^2 = \frac{S_s^{yt}}{S_s^2 \cdot (S_s^t)^2}. \tag{7}$$

6. Autocorrelation coefficient (AC_s): It measures the dependence of a time series with itself shifted by some time delay, i.e. the degree of correlation between the current values of the time series and the values in the previous time stamp. The AC_s is defined as:

$$AC_s = \frac{\sum_{i=t_{s-1}}^{t_s-1} (y_i - \bar{y_s}) \cdot (y_{i+1} - \bar{y_s})}{S_s^2}. \tag{8}$$

Once the six statistical metrics have been calculated for each segment in a chromosome, a clustering technique is applied over this six-dimensional space.

2.4 Clustering: *K-Means* Algorithm

A clustering process has to be applied in order to obtain the value of the fitness function for each segment. The algorithm chosen, *K-Means*, is applied to the time-series segments For further information on K-Means and the initialization procedure see Pérez-Ortiz et al. (2014).

Before applying the clustering algorithm one should normalize the values of the segment metrics, as the distance of each segment to its centroid strongly depends on the range of values of each metric (e.g. variance can have a much broader range of variation than skewness). Thus, distances from each metric with larger ranges would disrupt others with smaller ranges. Scaling is used to avoid this problem: for a given segmentation, the segment metrics are normalized to the range $[0, 1]$ using the min-max normalization:

$$v^* = \frac{v - v_{\min}}{v_{\max} - v_{\min}}, \tag{9}$$

where v is the value of the metric for a given segment, v^* is the normalized value, $v_{\min}$ is the minimum value for this metric when considering all the segments and $v_{\max}$ is the maximum one. A constant value of $v^* = 0.5$ is assigned whenever the metric is constant for all segments.

2.5 Fitness

All GAs need an evaluation mechanism to assign a quality index to each population individual. For *clustering* processes, one way to evaluate the obtained groups is to consider the *Sum of Squared Errors* (SSE), which consists of the sum of the

243 squared distances between each segment and its cluster centroid:

$$SSE = \sum_{i=1}^{m} d_i^2, \tag{10}$$

244 where i is the segment being evaluated, m is the total number of segments, and
245 d_i is the distance between the segment i and its closest centroid.

246 Our goal is to minimize this SSE in order to obtain more compact clusters
247 (where each point is as closer as possible to its centroid, but the centroids are as far
248 as possible from each other). However, when the GA tries to minimize the SSE,
249 it tends to minimize the number of segments as much as possible and could even
250 produce a partition where each cluster is a single segment. For instance, assuming
251 that the number of clusters considered is five and that a chromosome includes only
252 five segments, the SSE would be minimum in this case, $SSE = 0$, because each
253 segment would constitute a cluster. Since this is not an acceptable solution, the
254 fitness function is redefined considering also the number of segments:

$$fitness = \frac{m}{SSE}. \tag{11}$$

255 In this way, the algorithm tries to find partitions of the time series where
256 the number of segments is sufficiently high to ensure the acquisition of valuable
257 information from the clustering process.

258 2.6 Selection and replacement processes

259 In each generation, all individuals within the population are selected for reproduc-
260 tion and generation of offspring to promote a greater diversity since the parents
261 are not selected based on their fitness.

The replacement process has been performed by roulette wheel selection, i.e. a selection probability for each individual chromosome is calculated from its fitness value. The number of individuals selected is the population size minus one, and the vacant place is occupied by the best segmentation (with the highest fitness) of the previous generation, thus being an elitist algorithm.

As can be seen, the selection process promotes diversity, while the replacement process promotes elitism.

2.7 Mutation Operator

The algorithm has been endowed with four mutation operators whose principal function is to perform a better random exploration of the search space, aiming to reduce the dependency to the initial population and escape from local optima. The probability p_{m} of performing any mutation is decided by the user. Once a mutation is decided, the kind of perturbation applied to the chromosome is randomly selected from the following list: 1) add a cutpoint, 2) remove a cutpoint, 3) move half of the cutpoints to the left, and 4) move half of the cutpoints to the right.

The number of cutpoints to be added or removed is determined randomly. The number of points to move is approximately half of the available points and they are randomly selected and randomly pushed to the previous or the following point, with the constraint that it never reaches the previous or the next point. An example of the four mutation operations is included in Figure 2, where two cutpoints are removed, one is added and half of the them are moved to the left and to the right.

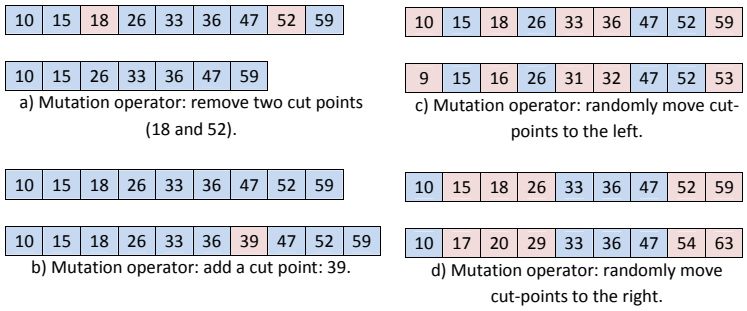


Fig. 2: Mutation operator (Online version in colour)

2.8 Crossover Operator

The algorithm includes a crossover operator, whose main function is to perform a better exploitation of the existing solutions. For each parent individual, the crossover operator is applied with a given probability p_c. The operator randomly selects the other parent, a random index of the time series, and it interchanges the left and right parts of the chromosomes with respect to this point. An illustration of the crossover operator can be seen in Figure 3.

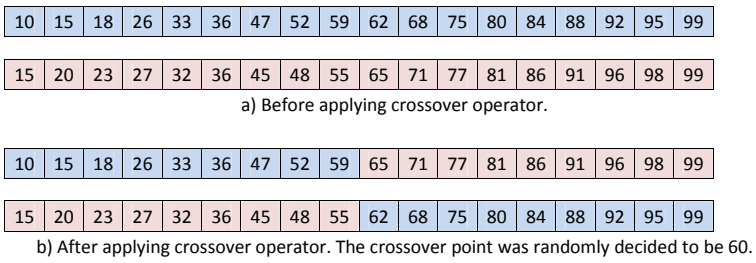


Fig. 3: Crossover operator (Online version in colour)

3 Experiments

3.1 Climate datasets

The paleoclimate datasets chosen for this study are the GISP2 Greenland Ice Sheet Project Two and the NGRIP North Greenland Ice Core Project δ^{18}O time series (Grootes and Stuiver, 1997; Stuiver and Grootes, 2000; Andersen et al., 2004; Svensson et al., 2008). The δ^{18}O isotope record is a proxy for past surface temperature variations (Bradley, 2015). In this study we focus on the 20-yr resolution δ^{18}O isotope records from both drilling sites. Pre-processing the datasets in the form of a 5-point average was found to help reduce short-term fluctuations within the datasets and improve the analysis of time series segmentations. If $\{y_n\}_{n=1}^{N}$ is the original time series, then the considered time series is $\{y_n^*\}_{n=1}^{N/5}$ with $y_i^* = \frac{1}{5}\sum_{j=5i}^{5i+4} y_i$.

In addition to the paleoclimate records, synthetic datasets obtained from well-known dynamical systems are also studied here to better understand the algorithm behaviour and as a preliminary attempt to reject or reinforce hypotheses related to underlying dynamical mechanisms for the DO events. Synthetic time series were produced using two simple mathematical models demonstrating noise-induced transitions as described in Benzi et al. (1981). We name *Benzi-A* the time series x(t) of a Langevin equation evolution with a gaussian noise component, a Wiener process dW:

$$dx = [x(a - x^2)]dt + \epsilon dW, \tag{12}$$

where $\epsilon = 0.5$ is the noise level and a is a constant. For $a > 0$ the system has two equilibrium solutions and is able to alternate between them because of the noise fluctuations on x. We name *Benzi-B* the time series with an additional periodic

315 forcing with frequency Ω as described in Benzi et al. (1981):

$$dx = [x(a - x^2) + Acos(\Omega t)]dt + \epsilon dW. \quad (13)$$

316 The Benzi-B time series demonstrates stochastic resonance for the noise level cho-
317 sen. The following values were used for the parameters: $a = 1.0, dt = 1.0, A =$
318 $0.12, \epsilon = [0.08, 0.10]$, and $\Omega = 10^{-3.0}$ In addition to the noise-induced transitions,
319 time series of a simplified 2D model of the Atlantic thermohaline Meridional Over-
320 turning circulation demonstrating critical as well as other types of bifurcations
321 were also examined in this paper. A detailed description of the Thermohaline
322 Ocean Model (T.O.M.) and its parameterizations can be found in Artale et al.
323 (2002). Two experiments (TOM-A and TOM-B) were conducted by time varying
324 the hydrological cycle strength F_{NS} to push the system through bifurcation points.
325 The F_{NS} was varied with a rate of $10 \cdot 10^{-11} psu/sec$ units within a total period of
326 700,000 years, which is slower than the time scale of the slowest process (diffusion)
327 included in the system, in order to ensure resilience of the system to the exter-
328 nal forcing (Saltzman, 2001). Experiments TOM-A and TOM-B were performed
329 by increasing and decreasing the control variable F_{NS} following the sequences of
330 states (A,F,B,C,D,C,E,F,A) and (D,C,E,F,A,G,H) as shown in the state diagram
331 of Fig. 4, respectively.

332 3.2 Algorithm parameters

333 GAs usually involve adjusting a notable set of parameters. However, their search
334 dynamics can adapt to the problem under study, resulting in a performance which
335 is negligibly affected by minor changes in the initial parameters. In this paper, the
336 following parameters were initially set by trial and error and then used for every

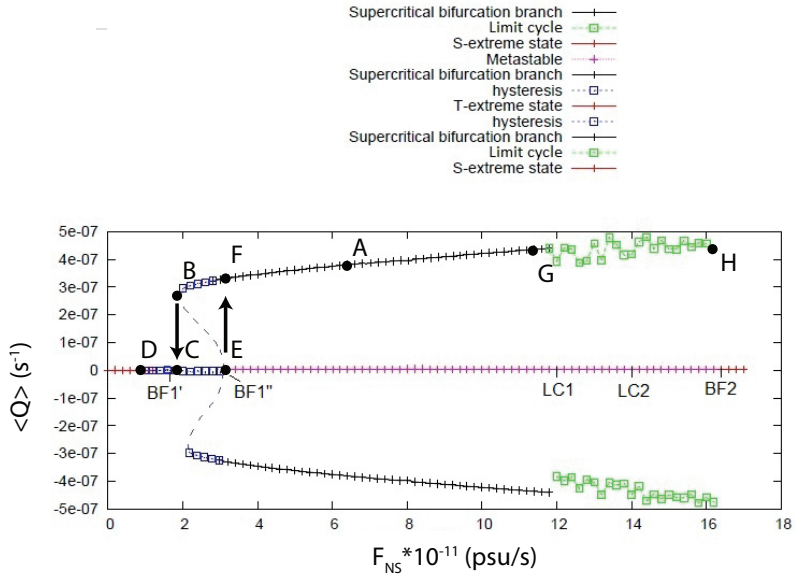


Fig. 4: Bifurcation diagram of the T.O.M system: average vorticity ($< Q >$) of the circulation cell as a function of F_{NS}: the strength of the evaporation minus precipitation at the surface of the ocean, or hydrological cycle, in practical salinity units psu/s. Notice the various bifurcation points $BF1'$, $BF1''$, BF2 and the limit cycles LC1, LC2 that mark certain values of F_{NS} where transitions in the stability of the equilibrium solutions take place. Each cross point is a stable solution. The dashed branches represent unstable solutions where the system cannot be found and are included for the shake of completeness. The two experiments TOM-A and TOM-B were conducted by increasing and decreasing F_{NS} to follow the state sequences: TOM-A: (A,F,B,C,D,C,E,F,A) and TOM-B: (D,C,E,F,A,G,H). Notice the hysteresis branches involved along the track of the experiments (Online version in colour)

dataset under study. The number of individuals (segmentation possibilities) of the population is $t = 80$. The crossover probability is $p_c = 0.9$ and the mutation probability $p_m = 0.075$. The number of clusters to be discovered from each candidate segmentation is $k = [4, 5]$; such values are high enough to discover new information among the derived clusters, but not so high as to threaten the interpretation and reproducibility of the results. The maximum number of generations is set to 2000, and the k-means clustering process is allowed a maximum of 20 iterations.

Finally a GA is a stochastic optimization algorithm with an embedded random number generator. Given that the results can be different depending on the seed value, the algorithm is run several times with different seeds. For each dataset, the GA was run 10 times, with seeds in the set $\{10, 20, \ldots, 100\}$ to evaluate and remove the dependence of the results on the seed value. It is also a step for evaluating the accuracy of the algorithm.

4 Results

This section presents the main results of the segmentation algorithm for the synthetic and paleoclimate datasets under study. The segmentation returned by the GA in the final generation was analyzed using the following approach: First it was verified whether the abrupt transitions were belonging to different classes or if they were grouped to the same class according to some common characteristics. Second the behaviour of each metric in the six-dimensional parameter space was observed on the onset of the transitions to find common patterns in the form of certain class sequences, that would be indicative of EWSs, e.g. increasing variance and

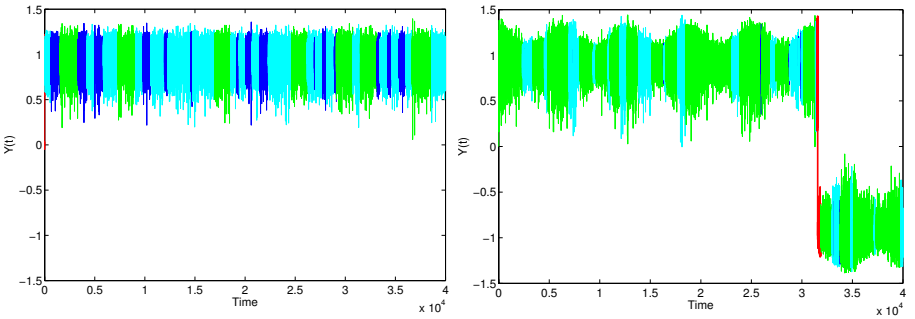
autocorrelation coefficient. This was done for the two independent paleoclimate datasets, the synthetic datasets, and for the ten seed values.

4.1 Synthetic Datasets

Based on the approach described above the following results have been obtained for the synthetic datasets to evaluate the performance of the proposed algorithm on well-described dynamical systems. Figures 5 and 6 present the segmentation results for the Benzi-A and -B models where transitions are caused by gaussian noise and stochastic resonance, respectively. The main results are listed below.

1. The algorithm constantly attributes classes with high variance, autocorrelation, and MSE for the abrupt transitions (shown in red, magenta, and green). Variance, MSE, and autocorrelation increase by one order of magnitude close to the transition.
2. There are no false alarms for the N-tippings in the Benzi-A -B models , i.e. the above classes are never found outside abrupt transitions.
3. The algorithm performs much less robustly for transitions that are closely spaced to each other. Detection of transitions occuring less than 700 time steps after the previous transition goes down to 40% compared to transitions occurring after longer time intervals (e.g. $\Delta t \geq 1200$).
4. The algorithm cannot identify whether the system under study is governed by stochastic resonance or only white noise-induced transitions since the statistical parameters observed by the algorithm follow a similar behaviour in both cases.
5. A spurious class with segments of kurtosis value equal to -1 (i.e. broad and flat distribution) can easily be identified from the segmentation results, otherwise

the kurtosis value is usually positive ($\gamma \in (0, 2)$), corresponding to narrower distributions.



	Variance	Skewness	Kurtosis	Slope LR	MSE	Autocorrelation
$\mathcal{C}_1$ (red)	0.216	0.376	-1.493	0.068	0.045	16.419
$\mathcal{C}_2$ (green)	0.020	-0.850	1.436	2.8E-07	0.020	-956.558
$\mathcal{C}_3$ (blue)	0.018	-0.774	0.978	2.0E-05	0.018	-137.964
$\mathcal{C}_4$ (cyan)	0.019	-0.781	1.016	2.8E-06	0.019	-425.634

(a) Benzi-A model: Gaussian noise

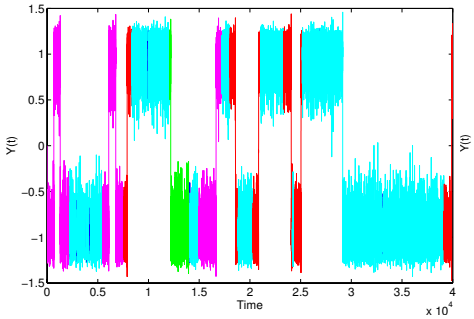
	Variance	Skewness	Kurtosis	Slope LR	MSE	Autocorrelation
$\mathcal{C}_1$ (red)	0.424	2.167	3.141	-0.003	0.279	309.046
$\mathcal{C}_2$ (green)	0.025	-0.396	0.965	-2.1E-06	0.025	-589.891
$\mathcal{C}_3$ (blue)	0.025	-0.212	-0.742	-0.003	0.024	-11.440
$\mathcal{C}_4$ (cyan)	0.024	-0.470	0.860	-2.1E-05	0.024	-117.835

(b) Benzi-B model: stochastic resonance

Fig. 5: Results of the segmentation algorithm for the Benzi-A and B models governed by Gaussian noise (left) and stochastic resonance (right), $\epsilon = 0.08$. The transition belongs to a single class (shown in red) with high autocorrelation, variance, MSE, skewness and kurtosis compared to the segments preceding the transition (Online version in colour).

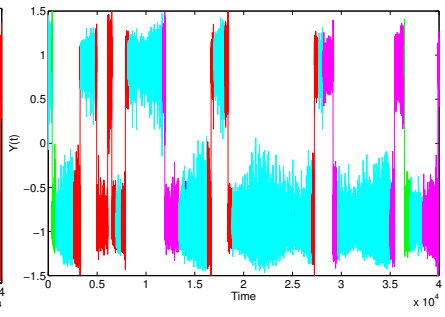
Figures 7 and 8 present the segmentation results for the Thermohaline Ocean Model (T.O.M.), where the system undergoes critical bifurcations in experiment TOM-A and both critical and limit cycle transitions in experiment TOM-B. The main results are listed below, employing the maximum flow strength Ψ as a state variable second to the average vorticity Q.

1. The TOM-A experiment pushes consecutively the system across two bifurcation points $BF1'$ and $BF1''$ that mark the limits of a hysteresis on the bifur-



	Variance	Skewness	Kurtosis	Slope LR	MSE	Autocorrelation
$\mathcal{C}_1$ (red)	0.665	0.238	0.264	0.001	0.296	756.749
$\mathcal{C}_2$ (green)	0.202	3.310	11.073	-2.7E-04	0.178	1536.851
$\mathcal{C}_3$ (blue)	0.035	0.016	-0.784	-0.005	0.034	-8.032
$\mathcal{C}_4$ (magenta)	0.770	0.905	-0.978	2.0E-04	0.638	2066.442
$\mathcal{C}_5$ (cyan)	0.029	0.152	1.298	4.8E-07	0.029	-161.242

(a) Benzi-A model: Gaussian noise



	Variance	Skewness	Kurtosis	Slope LR	MSE	Autocorrelation
$\mathcal{C}_1$ (red)	0.751	0.697	-0.614	0.001	0.290	664.215
$\mathcal{C}_2$ (green)	0.152	3.951	18.010	-0.001	0.139	349.409
$\mathcal{C}_3$ (blue)	0.026	0.291	-0.549	-0.019	0.024	-2.576
$\mathcal{C}_4$ (magenta)	0.553	0.905	2.397	-5.4E-05	0.271	1407.011
$\mathcal{C}_5$ (cyan)	0.031	0.386	0.962	1.75E-06	0.031	-116.209

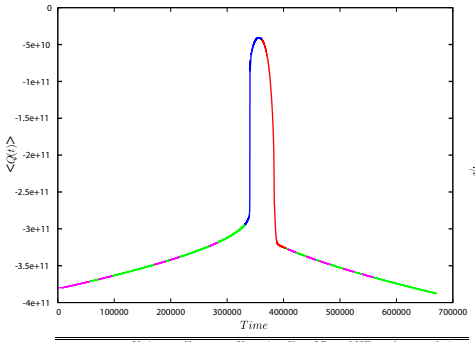
(b) Benzi-B model: Stochastic resonance

Fig. 6: Results of the segmentation algorithm for the Benzi-A and B models where transitions are governed by Gaussian noise (left) and stochastic resonance (right), $\epsilon = 0.10$. In both cases, the transitions are attributed classes (shown in red, magenta, and green) with high autocorrelation, variance, and MSE compared to the class of the preceding segments (Online version in colour).

cation diagram (see Fig. 4). The B-tipping associated with the $BF1'$ point is always detected by the algorithm in the form of increased autocorrelation, MSE and variance in the sequence of classes until the transition. This is encountered in both state variables Q and Ψ (see Fig. 7(a): green, blue, red, green, and Fig. 7(b): blue, red sequence)

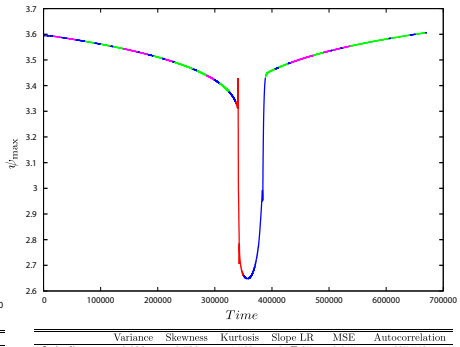
2. The detection of the second bifurcation tipping $BF1''$ was accompanied by false alarms for 40% of the cases for the variable Ψ, meaning that the sequence of classes at the B-tipping was repeated at parts of the time series where no bifurcation point was crossed (see Fig. 7(b): blue, green sequence), as consulted by the bifurcation graph (Fig. 4).

3. The experiment TOM-B pushes the system consecutively across the $BF1''$ and the limit cycles LC1 and LC2 where the system engages in internally excited oscillations, irrelevant to external periodic forcings therefore the use of the "cycle" term (see Fig. 4). The variability of Q and Ψ differ in amplitude, with Ψ having a larger amplitude.
4. The increase in variance, MSE and autocorrelation by two orders of magnitude at 100% of the runs while crossing the critical bifurcation point $BF1''$ is a robust indicator of B-tipping in the Q time series.
5. The algorithm divides the Q data series section that corresponds to the interval (F,A,G) (see Fig. 4) to consecutive segments of the same class. It successfully (90%) attributes the same class to the qualitatively same states of the system (see Fig. 8(a): multiple blue segments), without prior knowledge of their similarity on the bifurcation graph.
6. Failure to detect the B-tipping in the Ψ series at the $BF1''$ point as it falsely attributes to the B-tipping one of the two similar classes that are used for the interval (F,A,G) where no qualitative change is observed in the system's state (see Fig. 8(b): magenta, green, blue). In contrast, the algorithm captures the $BF1''$ successfully in the Q series.
7. Increase in variance, MSE but decrease in autocorrelation by one order of magnitude is diagnosed in Ψ time series following the transition to the limit cycle at point LC1, showing a different statistical profile for this transition. The signal is very robust, persisting during three consecutive classes for 80% of the algorithm seeds, instead of two classes for the previous statements, setting a stronger diagnostic of the tipping using these three metrics as robust indicators (see Fig. 8(b): green, red).



	Variance	Skewness	Kurtosis	Slope LR	MSE	Autocorrelation
$\mathcal{C}_1$ (red)	1.4E+22	0.159	-1.804	-9.0E+06	1.6E+21	2144.550
$\mathcal{C}_2$ (green)	7.0E+16	0.005	-1.201	1.3E+06	7.6E+11	163.973
$\mathcal{C}_3$ (blue)	1.1E+22	-0.915	-1.118	1.0E+07	3.6E+21	1517.360
$\mathcal{C}_4$ (magenta)	1.8E+18	0.048	-1.206	3414.640	6.1E+15	1041.230

(a) Experiment TOM-A ($< Q(t) >$), s=20



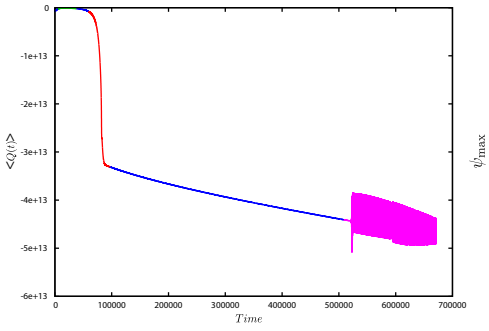
	Variance	Skewness	Kurtosis	Slope LR	MSE	Autocorrelation
$\mathcal{C}_1$ (red)	0.086	0.526	-1.569	-6.5E-05	0.015	695.542
$\mathcal{C}_2$ (green)	1.5E-05	-0.086	-1.090	-4.6E-08	3.2E-07	818.688
$\mathcal{C}_3$ (blue)	7.3E-04	0.066	-1.147	3.0E-06	4.7E-05	189.136
$\mathcal{C}_4$ (magenta)	6.3E-06	-0.023	-1.198	-3.4E-08	1.2E-09	668.998

(b) Experiment TOM-A (ψ_{max}),$s = 60$

Fig. 7: Results of the segmentation algorithm for the Thermohaline Ocean Model (T.O.M.) with the system undergoing critical bifurcations (Experiment TOM-A) for the average vorticity Q (left) and maximum flow strength Ψ_{max} (right). The $BF1'$ bifurcation is successfully detected through a unique class attribution with an increase in autocorrelation, MSE and variance. None of the datasets captures the intermediate stable state reached before the $BF1''$ transition, resulting in false alarms (decrease in autocorrelation, MSE and variance for the second $BF1''$ point, in contrast to the EWSs expected by the literature. The seed value s of the random initiation is noted below each image. (Online version in colour).

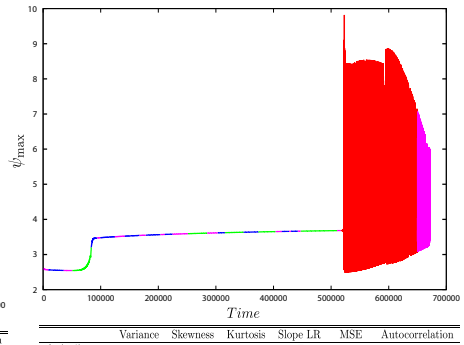
4.2 Paleoclimate Datasets

This subsection presents the main results of the segmentation algorithm for the paleoclimate datasets, following the same approach as for the synthetic datasets. They are listed below:



	Variance	Skewness	Kurtosis	Slope LR	MSE	Autocorrelation
$\mathcal{C}_1$ (red)	1.9E+26	-0.257	-1.779	-1.1E+09	2.3E+25	1.8E+04
$\mathcal{C}_2$ (green)	1.0E+21	-0.718	-0.676	8.5E+06	1.0E+20	794.300
$\mathcal{C}_3$ (blue)	1.4E+21	-0.009	-1.197	-2.2E+07	3.8E+17	231.636
$\mathcal{C}_4$ (magenta)	7.8E+24	-0.011	-0.889	-2.5E+07	6.3E+24	3.3E+04

(a) Experiment TOM-B ($< Q(t) >$), s=40



	Variance	Skewness	Kurtosis	Slope LR	MSE	Autocorrelation
$\mathcal{C}_1$ (red)	3.033	0.857	-0.548	-7.2E-06	3.030	16.603
$\mathcal{C}_2$ (green)	0.002	0.294	-0.432	2.1E-06	0.001	1.5E+04
$\mathcal{C}_3$ (blue)	2.5E-04	-0.124	-0.908	1.4E-06	9.4E-05	631.095
$\mathcal{C}_4$ (magenta)	0.041	0.072	-1.140	1.0E-07	0.041	250.311

(b) Experiment TOM-B (ψ_{max}),s=70

Fig. 8: Results of the segmentation algorithm for the Thermohaline Ocean Model (T.O.M.) with the system undergoing critical and limit-cycle bifurcations (Experiment TOM-B) for the average vorticity Q (left) and maximum flow strength Ψ_{max} (right). Transition through the $BF1''$ is captured in the Q series but is documented as a false alarm in the Ψ time series. The LC1 is detected in both runs with the difference that in (b) the autocorrelation decreases close to the transition instead of increasing as in (a). The seed value s of the random initiation is noted below each image. (Online version in colour)

.

1. The DO events are grouped into two main classes, sometimes three because the values of autocorrelation, variance, and MSE may differ significantly from one DO event to another. The high number of classes considered here (5 classes in total) allows for flexibility within the algorithm.
2. EWSs of DO events are found by the segmentation algorithm in the form of an order of magnitude increase in autocorrelation, variance, and mean square error (MSE) across a sequence of two classes. These EWSs are robustly found

(70%+) in the GISP2 δ^{18}O dataset for DO 0, 1, 2, 4, 7, 8, 11, 12 and for DO 0, 1, 4, 8, 10, 11, 12 for the NGRIP δ^{18}O dataset (see Fig. 9).

3. The increase in mean square error (MSE) is suggested here as an additional indicator of abrupt climate change. The increase in MSE, which suggests non-linear behaviour, has been found to correspond with an increase in variance prior to DO events for ∼90% of the seed runs for the GISP2 δ^{18}O dataset (see Fig. 9) and for ∼100% of the seed runs for the NGRIP δ^{18}O dataset.
4. The increase in the autocorrelation coefficient cannot be solely used as indicator of climate change. The algorithm sometimes found an increase in MSE and variance but a decrease in autocorrelation coefficient on the onset of DO events. This signature was minor in the GISP2 δ^{18}O dataset (e.g. DO 2, 10) but much more present in the NGRIP δ^{18}O dataset (e.g. DO 0, 1, 5, 7, 8, 10). Hints of this behaviour could already be found for DO 1 by Lenton et al. (2012). We stress that the increase in variance and MSE is a much more robust EWS for the NGRIP dataset especially.
5. Analysis of paleoclimate records GISP2 and NGRIP did not find any consistent change in skewness nor kurtosis on the onset of DO events.

Considering algorithm runs with different seed values revealed minor differences such as DO events being attributed to other classes and the cutpoints between classes being not exactly at the very same location, but the main characteristics described here and in the five main points remained robust throughout the analysis. Finally, the average computational time of the 10 runs was 53.24 ± 10.36 and 65.45 ± 10.38 seconds for GISP2 δ^{18}O and NGRIP δ^{18}O datasets, respectively, using an Intel Core i7 (R) CPU 3610QM at 2.3GHz with 8GB of RAM. Taking

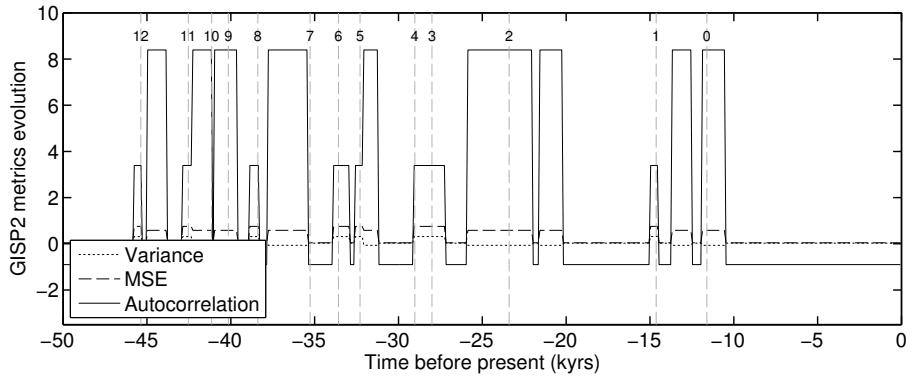
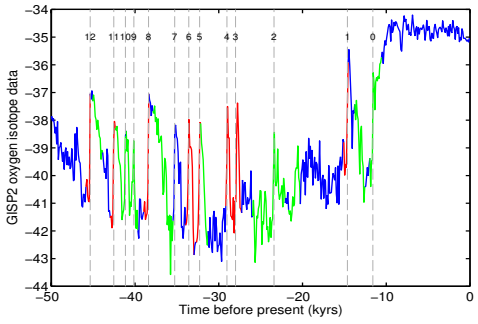


Fig. 9: Time series metrics after the clustering process (i.e. the segments found by the algorithm are replaced with their clusters centroids). The increase in MSE is associated with an order of magnitude increase in variance and autocorrelation on the onset of DO events. All DO events are represented for reference (GISP2 δ^{18}O ice core, seed = 10).

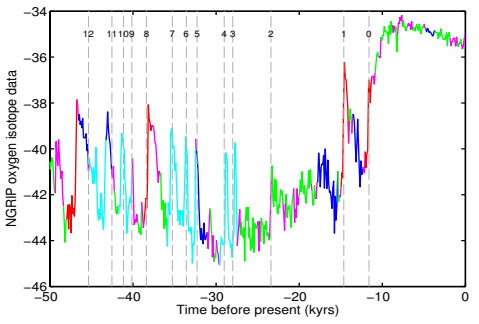
the length of the datasets into account, this computational cost is justified and affordable.

Figure 10 presents the detailed segmentation results for GISP2 and NGRIP δ^{18}O ice core data for a fixed seed value. The warning signals can be seen as the transition between consecutive segments and the label assigned to such segments, e.g. the upward trend in nonlinearity was seen via one or two consecutive increases in value, although this depends on the number of segment classes compared to the whole dataset length. The Dansgaard-Oeschger events are found grouped into two or three main classes with high autocorrelation, MSE, and variance corresponding to classes $\mathcal{C}_1$ and $\mathcal{C}_2$ for GISP2 and classes $\mathcal{C}_1$ and $\mathcal{C}_5$ for NGRIP for that run. An order of magnitude increase in these statistical parameters are found at the onset of the events; they decrease back to normal values as the dynamical system



	Variance	Skewness	Kurtosis	Slope LR	MSE	Autocorrelation
$\mathcal{C}_1$ (red)	2.734	0.313	-1.455	-0.004	0.753	3.391
$\mathcal{C}_2$ (green)	1.188	-0.068	-0.804	0.001	0.584	8.395
$\mathcal{C}_3$ (blue)	0.124	0.025	-1.716	0.001	0.045	-0.904
$\mathcal{C}_4$ (magenta)	2.189	0.445	-1.315	-0.001	2.159	-0.915
$\mathcal{C}_5$ (cyan)	0.843	-1.702	2.543	-9.5e-05	0.839	11.001

(a) Segmentation results on GISP2 dataset



	Variance	Skewness	Kurtosis	Slope LR	MSE	Autocorrelation
$\mathcal{C}_1$ (red)	2.952	0.072	-1.229	-0.002	1.734	6.219
$\mathcal{C}_2$ (green)	0.173	-0.043	-0.632	5.6E-05	0.137	-0.525
$\mathcal{C}_3$ (blue)	0.793	-0.023	-0.529	0.001	0.338	7.246
$\mathcal{C}_4$ (magenta)	0.079	-0.003	-1.915	2.2E-04	0.0169	-0.994
$\mathcal{C}_5$ (cyan)	2.143	0.491	-1.041	4.1E-04	2.016	12.685

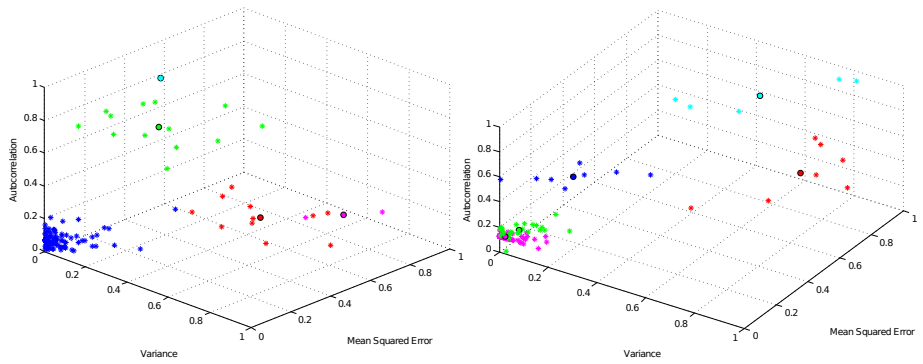
(b) Segmentation results on NGRIP dataset

Fig. 10: Results of the segmentation algorithm on δ^{18}O ice core data (seed = 10). The Dansgaard-Oeschger events are found grouped into two or three main classes with high autocorrelation, MSE, and variance. (a) GISP2: $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_5$. (b) NGRIP: $\mathcal{C}_1$ and $\mathcal{C}_5$. All Dansgaard-Oeschger events are numbered for reference. (Online version in colour)

slowly recovers to its stable state. This behaviour can also be seen in Fig. 9, which illustrates the evolution of the statistical metrics accross the whole GISP2 data series.

A detailed analysis of Fig. 10 reveals that class $\mathcal{C}_3$ for GISP2 dataset was the third main class, grouping segments with the lowest MSE, variance, and autocorrelation for that seed run and was found at the onset of several DO events (e.g. 1, 4, 8, 12) collocated with the Heinrich events H1, H3, H4, H5 as well as during the Holocene period (for an introduction to Heinrich events see Hemming, 2004). Classes $\mathcal{C}_4$ and $\mathcal{C}_5$ have been found outside the plotted area (in the -50ka,

-60ka range) and therefore do not appear in the graph. As for the NGRIP dataset classes $\mathcal{C}_2$ and $\mathcal{C}_4$ (with the lowest MSE, variance, and autocorrelation) have been found at the onset of several DO events as well (e.g. 4, 7, 8, 10 and 12) with an unusual behaviour in the autocorrelation coefficient for DO 1. A detailed analysis of their six-dimensional vector revealed that classes $\mathcal{C}_2$ and $\mathcal{C}_4$ differ only from the point of view of kurtosis. This is further discussed in the discussion section about the limitations of the algorithm. Class $\mathcal{C}_5$ (cyan curve in Fig. 10b) is considered the main DO class in NGRIP data for that particular run with a highly linear relationship (ratio of 1:1) between variance and MSE within that class and a constant high autocorrelation coefficient. This is illustrated in Fig. 11, where the 3D representation of the clustering results is shown for variance, autocorrelation, and MSE (normalised values), where each point is a segment within its own cluster, colour-coded according to the classes shown in the previous figure (Fig. 10).



(a) GISP2 (b) NGRIP

Fig. 11: 3D representation of the clustering results for variance, autocorrelation and MSE (normalized values), where each point is a segment within its own cluster. The centroids are represented by black circles (Online version in colour).

5 Discussion

Before diagnosing the high variability of the ice core data series for EWSs, the algorithm is tested on synthetic data, which include transitions associated to N- and B-tipping (Ashwin et al., 2012). Patterns preceding such transitions can be studied in the controlled environment of simulated data for possible existence of EWSs, with the advantage of prior knowledge of their underlying dynamic mechanisms. It was found that testing the proposed algorithm on the simulated time series was successfull in detecting the N-tippings (100%) and the B-tippings (100%) in the time series. EWSs of those tippings were revealed in the form of sequence of classes with pronounced differences in their autocorrelation, variance and MSE. Therefore, before a N-transition point in the data series we observe at least an order of magnitude increase in autocorrelation, variance, and MSE via their changes of class through two or three consecutive segments. This sequence is not seen throughout the dataset in the absence of N-tipping event. Frequently in the runs the last segment includes the transition point itself. This occurs due to the randomised initial segmentation. However, the sequence of classes attributed still compactly describes the statistical evolution of the time series before and after the onset of the tipping and serves in classifying the type according to differences in the attributed class sequence.

In the case of B-tipping, the algorithm always captures the critical bifurcations using the same EWS as in the N-tipping in absence of other B-tippings in the dataset. In agreement with previous studies (see Introduction) an increase in variance and autocorrelation is found and an increase to the MSE metric is added to the tipping precursors. The existence of a second critical bifurcation, as seen

in the TOM-A experiment can undermine the detection of multiple B-tippings if they occur within a short time interval, compared to the time series length. It could also be that the availability of warning signals or resilience indicators are absent as proposed in Batt et al. (2013) for stable to cycle transitions of phytoplankton blooms. For the systems whose state is described by multiple variables, the dynamical transitions occuring could be differently depicted on each variable evolution and therefore be elusive to the algorithm from variable to variable as seen in experiment TOM-B . An important aspect of the precursor signal at the LC1 transition is that both the variance and MSE increases, while the algorithm fails to see an increase in autocorrelation prior to the transition: it only sees a sharp decrease in autocorrelation following the transition. Such types of signals mark the transition of the system from one stable state to a state that alternates between stable points and have been encountered in the study of similar ecosystem dynamics already (Batt et al., 2013). This encouraging result was captured by the algorithm demonstrating its ability to distinguish between qualitatively different B-tippings. The algorithm was able to cluster segments with qualitatively same states into one or two classes with minor differences in their statistical metrics without prior knowlegde of the bifurcation diagram. Regarding technical limitations of the algorithm, the kurtosis and slope coefficients are inconclusive as classification criteria and the algorithm is unable to detect multiple tipping transitions that occur within a short time interval, which is an expected weakness for any statistically-based approach that relies on the size of the sample.

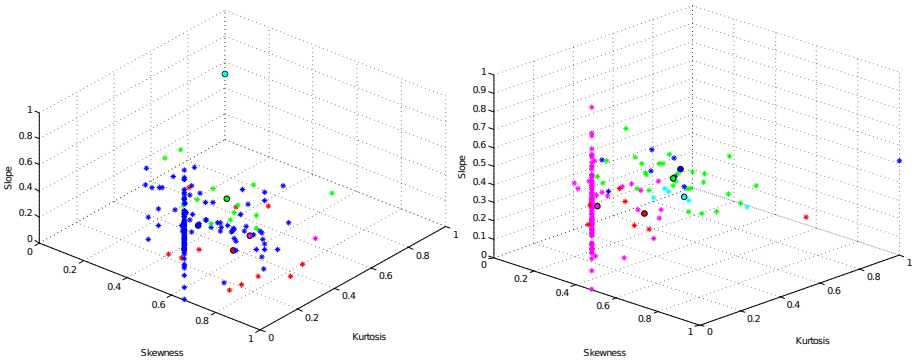
The comparative study of the two independent ice core datasets revealed that the algorithm could capture the same main characteristics in warning signals of DO events. For instance warning signals were robustly detected ($\geq$ 70% of the

seed runs) for DO events 0, 1, 4, 8, 11, and 12 for the two independent datasets in the form of an order of magnitude increase in autocorrelation, variance, and mean square error (MSE). Significant changes in these metrics were detected in the two ice cores for DO 2, 3, 5, 6, 7, and 10 for 20%–70% of the algorithm seeds, suggesting that these events possess weak warning signals. To analyse EWSs we considered a time span starting 100 years after the previous DO event until the one under study. Warning signals were detected at least 600 years in advance by the algorithm (e.g. DO 1, 2, 4, 8, GISP2 data) and up to 1.8 kyrs before the event (e.g. DO 12, NGRIP data). A minimum time period for investigating EWSs before the onset of DO events would therefore correspond to 600 years, which is comparable with values found in the literature (e.g. more than 700 years for the increase in variance as found by Cimatoribus et al. (2013). The starting point for considering this time span should be the start of the cold stadial state to discriminate from the dynamical behaviour of the previous event, e.g. a decrease in autocorrelation. The non-detection by the algorithm of any warning signals for DO event 9 supports this hypothesis: DO 9 starts 1 kyr after the onset of the previous event but only 500 years after the beginning of the cold, stable period. The case of DO 9 not being detected falls in the shortcomings of the algorithm, which underperforms for short temporal distance from the previous transition i.e. DO event. Detection of warning signals notably improves for DO events with previously long cold stadials. This is also consistent with the results of the synthetic datasets, e.g. the noise-driven system and the stochastic resonance system, $\epsilon = 0.1$: the performance of the algorithm in detecting warning signals for transitions occurring less than 700 years after the previous transition goes down to 40% compared to transitions following longer stable states (e.g. $\Delta t \geq 1200$).

Hypotheses of underlying dynamical mechanisms for DO events can be reinforced or rejected by additional comparison with the algorithm behaviour on simulated data produced with well-known dynamical models. In this reverse engineering approach more than a single model can reproduce the variability encountered in the ice core or any other paleoclimatic record (Crucifix, 2012), so any interpretation of dynamical systems should be used with caution and in combination with scientific insight. Distinct behaviours for the autocorrelation coefficient have been found according to the type of transition studied as revealed from the synthetic data analysis. The algorithm sometimes found an increase in MSE and variance but a decrease in the autocorrelation coefficient on the onset of DO events 1 and 10 in both the GISP and NGRIP datasets. It is therefore proposed to revisit the hypothesis that the DO events were caused by an uniform dynamical system, e.g. the stochastic resonance hypothesis (Ganopolski and Rahmstorf, 2001, 2002), as statistical indicators seem to point to certain DOs being attributed other dynamical causes. This aspect is left for investigation to the specialized paleoclimate scientist.

When analyzing the results of segmentation algorithms we also considered the segment lengths as a possible bias factor to the diagnostics. It can happen that the algorithm is not able to assign a proper class to a segment and prefers to divide the segments into smaller sections to reduce e.g. MSE and kurtosis values. The new smaller sections are likely to be grouped together in this parameter space, allowing the algorithm to perform the clustering process. Moreover, analyzing Eq. (11), fitness is directly proportional to the number of divisions so segmentations with a high number of cut points will be prefered. One signature of this effect is seen in the fact that all small segments are found in a single class with very low

kurtosis (γ=[-1.6,-1.9]), constant skewness (equal to 0), and a large range of slope coefficients. They are represented by a straight line in Fig. 12. Special care was taken to discard those small segments (e.g. containing 2 or 3 points) in the analysis of EWSs, otherwise they would have lead to false alarms, i.e. events of nonlinear behaviour that are not leading up to a DO event, e.g. at -20.66 kyrs (Fig. 9 & 10a, GISP2 data). Closer inspection revealed that the increase in nonlinearity was a spurious effect due to the small segments, which were discarded in order to avoid false alarms.



(a) GISP2 (b) NGRIP

Fig. 12: 3D representation of the clustering results for slope, skewness and kurtosis (normalized values) where each point is a segment within its own cluster. The centroids are represented by black circles (Online version in colour).

6 Conclusion

In this paper, a novel genetic algorithm (GA) from the field of time series segmentation is applied to paleoclimate data in order to identify common patterns

607 that would act as early warning signals for abrupt climate shifts. The algorithm
608 automatically finds the cutpoints, segment length, and the length of the window
609 for calculating six statistical metrics without providing climate knowledge or any
610 prior information. In addition to the variance and autocorrelation, the MSE met-
611 ric is found to respond robustly before a tipping point. A few datasets were used
612 to evaluate the algorithm behaviour, revealing similarities and differences amongst
613 the EWSs occurring before different qualitative tipping types. Experimental re-
614 sults show that warning signals of Dansgaard-Oeschger events could be robustly
615 found for several of these events in the form of an order of magnitude increase in
616 autocorrelation, variance, and mean square error in both GISP2 and NGRIP δ^{18}O
617 ice core data, but those that occurred within a short time interval were elusive.
618 The GA applied to NGRIP δ^{18}O ice core record showed that increasing autocor-
619 relation coefficient cannot be solely used as an indicator of climate shifts, as the
620 expected tendency of an increase in value is not seen for certain DO events. Com-
621 parison with synthetic datasets of well-known dynamical behaviour suggests that
622 different DO events might be triggered by different underlying dynamics. Finally
623 the proposed approach provides a novel visualisation tool in the field of climate
624 time series analysis and detection of warning signals of abrupt transitions.

625 As future steps, improvements of the algorithm are required to overcome
626 its limitations regarding consecutive fast transitions. We also suggest creating
627 a dataset comprising several tipping points and their statistical metrics to in-
628 vestigate the development of data-driven mathematical functions that would be
629 representative of abrupt transitions.

Acknowledgment

This work has been subsidized by the Ariadna project 13-9202 of the European Space Agency. The research work of P.A. Gutiérrez, A. Durán and C. Hervás-Martínez is partially funded by the TIN2011-22794 project of the Spanish Ministerial Commission of Science and Technology (MICYT), FEDER funds and the P11-TIC-7508 project of the "Junta de Andalucía" (Spain).

References

R.B. Alley, J. Marotzke, W.D. Nordhaus, J.T. Overpeck, D.M. Peteet, R.A. Pielke, R.T. Pierrehumbert, P.B. Rhines, T.F. Stocker, L.D. Talley, J.M. Wallace, Abrupt climate change. Science **299**(5615), 2005–2010 (2003)

K.K. Andersen, N. Azuma, J.-M. Barnola, M. Bigler, P. Biscaye, N. Caillon, J. Chappellaz, H.B. Clausen, D. Dahl-Jensen, H. Fischer, et al., High-resolution record of northern hemisphere climate extending into the last interglacial period. Nature **431**(7005), 147–151 (2004)

V. Artale, S. Calmanti, A. Sutera, Thermohaline circulation sensitivity to intermediate-level anomalies. Tellus A **54**(2), 159–174 (2002)

P. Ashwin, S. Wieczorek, R. Vitolo, P. Cox, Tipping points in open systems: bifurcation, noise-induced and rate-dependent examples in the climate system. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **370**(1962), 1166–1184 (2012)

R. Batt, W. Brock, S.R. Carpenter, J. Cole, M.L. Pace, D. Seekell, Asymmetric response of early warning indicators of phytoplankton transition to and from cycles. Theor. Ecol. **6**, 285–293 (2013)

653 K.D. Bennett, Determination of the number of zones in a biostratigraphical se-
654 quence. New Phytologist **132**(1), 155–170 (1996)
655 R. Benzi, A. Sutera, A. Vulpiani, The mechanism of stochastic resonance. J. Phys.
656 A: Math. Gen. **14**, 453–457 (1981)
657 R.S. Bradley, *Paleoclimatology*, 3rd edn. (Academic Press, San Diego, 2015), pp.
658 1–512
659 F.-L. Chung, T.-C. Fu, V. Ng, R.W. Luk, An evolutionary approach to pattern-
660 based time series segmentation. Evolutionary Computation, IEEE Transactions
661 on **8**(5), 471–489 (2004)
662 A.A. Cimatoribus, S.S. Drijfhout, V. Livina, G. van der Schrier, Dans-
663 gaard–oeschger events: bifurcation points in the climate system. Clim. Past **9**,
664 323–333 (2013)
665 M. Crucifix, Oscillators and relaxation phenomena in pleistocene climate theory.
666 Philosophical Transactions of the Royal Society A **370**(1962), 1140–1165 (2012)
667 V. Dakos, M. Scheffer, E.H. van Nes, V. Brovkin, V. Petoukhov, H. Held, Slowing
668 down as an early warning signal for abrupt climate change. Proceedings of the
669 National Academy of Sciences **105**(38), 14308–14312 (2008)
670 V. Dakos, S.R. Carpenter, W.A. Brock, A.M. Ellison, V. Guttal, A.R. Ives, S. Kefi,
671 V. Livina, D.A. Seekell, E.H. Van Nes, et al., Methods for detecting early warn-
672 ings of critical transitions in time series illustrated using simulated ecological
673 data. PLoS One **7**(7), 41010 (2012)
674 P.D. Ditlevsen, S.J. Johnsen, Tipping points: Early warning and wishful thinking.
675 Geophysical Research Letters **37**(19), 19703 (2010)
676 A. Ganopolski, S. Rahmstorf, Rapid changes of glacial climate simulated in a
677 coupled climate model. Nature **409** (2001)

678 A. Ganopolski, S. Rahmstorf, Abrupt glacial climate changes due to stochastic
679 resonance. Physical Review Letters **88**(3), 038501 (2002)

680 P.M. Grootes, M. Stuiver, Oxygen 18/16 variability in greenland snow and ice
681 with 10-3- to 105-year time resolution. Journal of Geophysical Research: Oceans
682 **102**(C12), 26455–26470 (1997)

683 K. Hasselmann, Stochastic climate models. i. theory. Tellus **28**, 473–485 (1976)

684 H. Held, T. Kleinen, Detection of climate system bifurcations by degenerate fin-
685 gerprinting. Geophysical Research Letters **31**(23), 23207 (2004)

686 S.R. Hemming, Heinrich events: Massive late pleistocene detritus layers of the
687 north atlantic and their global climate imprint. Reviews of Geophysics **42**(1),
688 (2004)

689 J. Himberg, K. Korpiaho, H. Mannila, J. Tikanmaki, H.T. Toivonen, Time series
690 segmentation for context recognition in mobile devices, in *Data Mining, 2001.*
691 *ICDM 2001, Proceedings IEEE International Conference on*, IEEE, 2001, pp.
692 203–210. IEEE

693 E. Keogh, S. Chu, D. Hart, M. Pazzani, An online algorithm for segmenting time
694 series, in *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Con-*
695 *ference on*, IEEE, 2001, pp. 289–296. IEEE

696 R. Kubo, The fluctuation - dissipation theorem. Reports on Progress in Physics
697 **29I** (1966)

698 T.M. Lenton, V.N. Livina, V. Dakos, M. Scheffer, Climate bifurcation during the
699 last deglaciation? Clim. Past **8**, 1127–1139 (2012)

700 T.M. Lenton, R.J. Myerscough, R. Marsh, V.N. Livina, A.R. Price, S.J. Cox, Us-
701 ing genie to study a tipping point in the climate system. Philosophical Transac-
702 tions of the Royal Society A: Mathematical, Physical and Engineering Sciences

367(1890), 871–884 (2009)

V.N. Livina, T.M. Lenton, A modified method for detecting incipient bifurcations in a dynamical system. Geophys. Res. Lett. **34**, 03712 (2007)

V. Livina, F. Kwasniok, G. Lohmann, J. Kantelhardt, T. Lenton, Changing climate states and stability: from pliocene to present. Climate dynamics **37**(11-12), 2437–2453 (2011)

J. MacQueen, et al., Some methods for classification and analysis of multivariate observations, in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, California, USA, 1967, p. 14. California, USA

Z. Michalewicz, *Genetic algorithms + data structures = evolution programs* (The Netherlands, HOOO, 1996)

P. Prandom, M. Goodwin, M. Vetterli, Optimal time segmentation for signal modeling and compression, in *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol. 3, IEEE, 1997, pp. 2029–2032. IEEE

M. Pérez-Ortiz, P.A. Gutiérrez, J. Sánchez-Monedero, C. Hervás-Martínez, A. Nikolaou, I. Dicaire, F. Fernández-Navarro, Time Series Segmentation of Paleoclimate Tipping Points by an Evolutionary Algorithm., in *HAIS*, 2014, pp. 318–329

S. Rahmstorf, Timing of abrupt climate change: A precise clock. Geophysical Research Letters **30**(10) (2003)

B. Saltzman, *Dynamical Paleoclimatology* (Academic Press, ???, 2001). ISBN 978-0-12-617331-4

M. Scheffer, J. Bascompte, W.A. Brock, V. Brovkin, S.R. Carpenter, V. Dakos, H. Held, E.H. Van Nes, M. Rietkerk, G. Sugihara, Early-warning signals for critical

728 transitions. Nature **461**(7260), 53–59 (2009)

729 S.L. Sclove, Time-series segmentation: A model and a method. Information Sci-
730 ences **29**(1), 7–25 (1983)

731 M. Stuiver, P.M. Grootes, Gisp2 oxygen isotope ratios. Quaternary Research
732 **53**(3), 277–284 (2000)

733 A. Svensson, K.K. Andersen, M. Bigler, H.B. Clausen, D. Dahl-Jensen, S. Davies,
734 S.J. Johnsen, R. Muscheler, F. Parrenin, S.O. Rasmussen, et al., A 60 000 year
735 greenland stratigraphic ice core chronology. Climate of the Past **4**(1), 47–57
736 (2008)

737 V.S. Tseng, C.-H. Chen, P.-C. Huang, T.-P. Hong, Cluster-based genetic segmen-
738 tation of time series with dwt. Pattern Recognition Letters **30**(13), 1190–1197
739 (2009)

740 Z. Xiong, C. Herley, K. Ramchandran, M.T. Orchard, Flexible time segmentations
741 for time-varying wavelet packets, in *Time-Frequency and Time-Scale Analysis,*
742 *1994., Proceedings of the IEEE-SP International Symposium on*, IEEE, 1994,
743 pp. 9–12. IEEE