Analyzing Granger Causality in Climate Data with Time Series Classification Methods

Christina Papagiannopoulou[1]([✉]), Stijn Decubber[1], Diego G. Miralles[2,3],
Matthias Demuzere[2], Niko E. C. Verhoest[2], and Willem Waegeman[1]

[1] Department of Mathematical Modelling, Statistics and Bioinformatics,
Ghent University, Ghent, Belgium
{christina.papagiannopoulou,stijn.decubber,willem.waegeman}@ugent.be
[2] Laboratory of Hydrology and Water Management,
Ghent University, Ghent, Belgium
{matthias.demuzere,niko.verhoest,diego.miralles}@ugent.be
[3] Department of Earth Sciences, VU University Amsterdam,
Amsterdam, The Netherlands

Abstract. Attribution studies in climate science aim for scientifically ascertaining the influence of climatic variations on natural or anthropogenic factors. Many of those studies adopt the concept of Granger causality to infer statistical cause-effect relationships, while utilizing traditional autoregressive models. In this article, we investigate the potential of state-of-the-art time series classification techniques to enhance causal inference in climate science. We conduct a comparative experimental study of different types of algorithms on a large test suite that comprises a unique collection of datasets from the area of climate-vegetation dynamics. The results indicate that specialized time series classification methods are able to improve existing inference procedures. Substantial differences are observed among the methods that were tested.

Keywords: Climate science · Attribution studies · Causal inference
Granger causality · Time series classification

1 Introduction

Research questions in climate change research are mostly related to either *climate projection* or to *climate change attribution*. Climate projection or forecasting aims at predicting the future state of the climatic system, typically over the next decades. The goal of climatic attribution on the other hand is to identify and quantify cause-effect relationships between climate variables and natural or anthropogenic factors. A well-studied example, both for projection and attribution, is the effect of human greenhouse gas emissions on global temperature.

The standard approach in the field of climate science is based on simulation studies with mechanistic climate models, which have been developed, expanded and extensively studied over the last decades. Data-driven models, in contrast to mechanistic models, assume no underlying physical representation of reality

but directly model the phenomenon of interest by learning a more or less flexible function of some set of input data. Climate science is one of the most data-rich research domains. With global observations on ever finer spatial and temporal resolutions from both satellite and *in-situ* measurements, the amount of (publicly available) climatic data sets has vastly grown over the last decades. It goes without any doubt that there is a big potential for making progress in climate science with advanced machine learning models.

The most common data-driven approach for identifying causal relationships in climate science consists of Granger causality modelling [17]. Analyses of this kind have been applied to investigate the influence of one climatic variable on another, e.g., the Granger causal effect of CO_2 on global temperature [1,20], of vegetation and snow coverage on temperature [19], of sea surface temperatures on the North Atlantic Oscillation [26], or of the El Niño Southern Oscillation on the Indian monsoon [25]. In Granger causality studies, one assumes that a time series A Granger-causes a time series B, if the past of A is helpful in predicting the future of B. The underlying predictive model that is commonly considered in such a context is a linear vector autoregressive model [8,32]. Similar to other statistical inference procedures, conclusions are only valid as long as all potential confounders are incorporated in the analysis. The concept of Granger-causality will be reviewed in Sect. 2.

In recent work, we have shown that causal inference in climate science can be substantially improved by replacing traditional statistical models with non-linear autoregressive methods that incorporate hand-crafted higher-level features of raw time series [27]. However, approaches of that kind require a lot of domain knowledge about the working of our planet. Moreover, higher-level features that are included in the models often originate from rather arbitrary decisions. In this article, we postulate that causal inference in climate science can be further improved by using automated feature construction methods for time series. In recent years, methods of that kind have shown to yield substantial performance gains in the area of time series classification. However, we believe that some of those methods also have a lot of potential to improve causal inference in climate science, and the goal of this paper is to provide experimental evidence for that. We experimentally compare a large number of time series classification methods – an overview and more discussion of these methods will be given in Sect. 3.

Most attribution studies in climate science infer causal relationships between time series of continuous measurements, leading to regression settings. However, classification settings arise when targeting extreme events, such as heatwaves, droughts or floods. We will conduct an experimental study in the area of investigating climate-vegetation dynamics, where such a classification setting naturally arises. This is an interesting application domain for testing time series classification methods, due to the availability of large and complex datasets with worldwide coverage. It is also a practically-relevant setting, because extremes in vegetation can reveal the vulnerability of ecosystems w.r.t. climate change [23]. A more precise description of this application domain and the experimental setup will be provided in Sect. 4. In Sect. 5, we will present the main results, which will allow us to

formulate conclusions concerning which methods are more appropriate in the area of climate sciences.

2   Granger Causality for Attribution in Climate Science

Granger causality [17] can be seen as a predictive notion of causality between time series. In the bivariate case, when two time series are considered, one compares the forecasts of two models; a baseline model that includes only information from the target time series (which resembles the effect) and a so-called full model that includes also the history of the second time series (which resembles the cause). Given two time series $\boldsymbol{x} = [x_1, x_2, ..., x_N]$ and $\boldsymbol{y} = [y_1, y_2, ..., y_N]$, with N being the length of the time series, one says that the time series $\boldsymbol{x}$ Granger-causes the time series $\boldsymbol{y}$ if the forecast of y at a specific time stamp t improves when information of the history of $\boldsymbol{x}$ is included in the model.

In this paper we will limit our analysis to situations where the target time series $\boldsymbol{y}$ consists of $\{0, 1\}$-measurements that denote the presence or absence of an extreme event at time stamp t. As such, one ends up with solving two classification problems, one for the baseline and one for the full model. We will work with the Area Under the Curve (AUC) as performance measure, because the class distribution will be heavily imbalanced, as a natural result of modelling extreme events. Let $\hat{\boldsymbol{y}}$ denote the new time series that originates as the one-step ahead forecast of $\boldsymbol{y}$ using either the baseline or the full model, then Granger causality can be formally formulated as follows:

Definition 1. *A time series $\boldsymbol{x}$ Granger-causes $\boldsymbol{y}$ if $AUC(\boldsymbol{y}, \hat{\boldsymbol{y}})$ increases when $x_{t-1}, x_{t-2}, ..., x_{t-P}$ are considered for predicting y_t, in contrast to considering $y_{t-1}, y_{t-2}, ..., y_{t-P}$ only, where P is the lag-time moving window.*

Granger causality studies might yield incorrect conclusions when additional (confounding) effects exerted by other climatic or environmental variables are not taken into account [13]. The problem can be mitigated by considering time series of additional variables. For example, let us assume one has observed a third variable $\boldsymbol{w}$, which might act as a confounder in deciding whether $\boldsymbol{x}$ Granger-causes $\boldsymbol{y}$. The above definition then naturally extends as follows.

Definition 2. *We say that time series $\boldsymbol{x}$ Granger-causes $\boldsymbol{y}$ conditioned on time series $\boldsymbol{w}$ if $AUC(\boldsymbol{y}, \hat{\boldsymbol{y}})$ increases when $x_{t-1}, x_{t-2}, ..., x_{t-P}$ are included in the prediction of y_t, in contrast to considering $y_{t-1}, y_{t-2}, ..., y_{t-P}$ and $w_{t-1}, w_{t-2}, ..., w_{t-P}$ only, where P is the lag-time moving window.*

An extension to more than three time series is straightforward. In our experiments, $\boldsymbol{y}$ will represent the vegetation extremes at a given location, whereas $\boldsymbol{x}$ and $\boldsymbol{w}$ can be the time series of any climatic variable at that location (e.g., temperature, precipitation or radiation).

Generally, the null hypothesis (H_0) of Granger causality is that the baseline model has equal prediction error as the full model. Alternatively, if the full model predicts the target variable $\boldsymbol{y}$ significantly better than the baseline

model, H_0 is rejected. In most applications, inference is drawn in vector autoregressive models by testing for significance of individual model parameters. Other studies have used likelihood-ratio tests, in which the full and baseline models are nested models [26]. Those procedures have a number of important shortcomings: (1) existing statistical tests only apply to stationary time series, which is an unrealistic assumption for attribution studies in climate science, (2) most tests are based on linear models, whereas cause-effect relationships can be non-linear, and (3) the models used for such tests are trained and evaluated on in-sample data, which will typically result in overfitting when the dimensionality or the model complexity increases.

In recent work, we have introduced an alternative way of assessing Granger-causality, by focussing on quantitative instead of qualitative differences in performance between baseline and full models [27]. In this way, traditional linear models can be replaced by more accurate machine learning models. If both the baseline and the full model give evidence of better predictions, one can draw stronger conclusions w.r.t. cause-effect relationships. To this end, no statistical tests are computed, but the differences between the two types of models is visualized and interpreted in a quantitative way.

3   From Granger Causality to Time Series Classification

In the general framework that we presented in [27] we constructed hand-crafted features based on knowledge that has been described in the climate literature [12]. These features include lagged variables, cumulative variables as well as extreme indices. Therefore, we ended up with in total ~360 features extracted from one time series. Our previous study has shown that incorporating those features in any classical regression or classification algorithm might lead to a substantial increase in performance (for both the baseline and the full model).

In this article, we investigate whether this feature construction process can be automated using time series classification methods. Due to the increased public availability of datasets from various domains, many novel time series classification algorithms have been proposed in recent years. All those methods either try to find higher-level features that represent discriminative patterns or similarity measures that define an appropriate notion of relatedness between two time series [2,11,21]. The following categories can be distinguished:

(a) Algorithms that use the whole series or the raw data for classification. To this family of algorithms belongs the one nearest neighbour (1-NN) classifier with different distance measures such as the dynamic time warping (DTW) [29], which is usually the standard benchmark measure, and variations of it, the complexity invariant distance (CID) [3], the derivative DTW [14], the derivative transform distance (DTD) [15] and the Move-split-merge (MSM) [33] distance.

(b) Algorithms that are based on sub-intervals of the original time series. They usually use summary measures of these intervals as features. Typical algorithms in this category are the time series forest (TSF) [10], the time series bag of features (TSBF) [5] and the learned pattern similarity (LPS) [4].

(c) Algorithms that are attempting to find informative patterns, called shapelets, in the data. An informative shapelet is a pattern that helps in distinguishing the classes by its presence or absence. Representative algorithms of this class are the Fast shapelets (FS) [28], the Shapelet transform (ST) [18] and the Learned shapelets (LS) [16].

(d) Algorithms that are based on the frequency of the patterns in a time series. These algorithms build a vocabulary of patterns and form a histogram for each observation by using this vocabulary. Algorithms such as the Bag of patterns (BOP) [22], the Symbolic aggregate approximation-vector space model (SAXVSM) [31] and the Bag of SFA symbols (BOSS) [30] are based on the idea of a pattern vocabulary.

(e) Finally, there are approaches that combine more than one from the above techniques, forming ensemble models. A recently proposed algorithm named Collection of transformation ensembles (COTE) combines a large number of classifiers constructed in the time, frequency, and shapelet transformation domains.

In our comparative study, we run algorithms from the first four different groups. The main criteria for including a particular algorithm in our analysis are (1) availability of source code, (2) running time for the datasets that we consider, and (3) interpretability of the extracted features. Since we have collected multiple time series for a large part of the world (3,536 locations in total), the algorithms should run in a reasonable amount of time. Several algorithms had problems to finish within 3 days.

4   Experimental Setup

In order to evaluate the above-mentioned time series classification methods for causal inference, we adopt an experimental setup that is similar to [27]. The non-linear Granger causality framework is adopted to explore the influence of past-time climate variability on vegetation dynamics. To this end, data sets of observational nature were collected to construct climatic time series that are then used to predict vegetation extremes. Data sets have been selected on the basis of meeting the following requirements: (a) an expected relevance of the variable for vegetation dynamics, (b) the availability of multi-decadal records, and (c) the availability of an adequate spatial and daily temporal resolution. In our previous work, we collected in this way in total 21 datasets. For the present study, we retained three of them, while covering the three basic climatic variables: water availability, temperature, and radiation. The main reason for making this restriction was that in that way the running time of the different time series classification algorithms could be substantially reduced. Specifically, we collected one precipitation dataset, which is coming from a combination of *in-situ*, satellite data, and reanalysis outputs, called Multi-Source Weighted-Ensemble Precipitation (MSWEP) [7]. We include one temperature dataset, which is a reanalysis data set, and one radiation dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim [9].

In addition to those three climatic datasets, we also collected a vegetation dataset. We use the satellite-based Normalized Difference Vegetation Index (NDVI) [34], which is a commonly used monthly long-term global indicator of vegetation [6]. Roughly speaking, NDVI is a graphical greenness indicator which measures how green is a specific point on the Earth at a specific time stamp. The study period starting from 1981–2010 is set by the length of the NDVI record. The dataset is converted to a 1° spatial resolution to match with the climatic datasets.

For most locations on Earth, NDVI time series exhibit a clear seasonal cycle and trend – see top panel of Fig. 1 for a representative example. However, in climate science, the interesting part of such a time series is the residual component, usually referred to as seasonal anomalies. In a statistical sense, climatic data can only be useful to predict this residual component, as both the seasonal cycle and the trend can be modelled with pure autoregressive features. Similarly as in [27], we isolate the residual component using time series decomposition methods, and we work further with this residual component – see bottom panel of Fig. 1 for an illustration. In a next step, extremes are obtained from the residuals, while taking the spatial distribution of those extremes into account. The most straightforward way is setting a fixed threshold per location, such as the 10% percentile of the residuals. However, this leads to spatial distributions that are physically not plausible, because one cannot expect that the same number of vegetation extremes is observed in all locations on Earth. At some locations, vegetation extremes are more probable to happen. For this reason, we group the location pixels into areas with the same vegetation type, by using the global vegetation classification scheme of the International Geosphere-Biosphere Program (IGBP) [24], which is generically used throughout a range of communities. We selected the map of the year 2001 (closer to the middle of our period of interest). In order to end up with coherent regions that have similar climatic and vegetation characteristics, we further divided the vegetation groups into areas in which only neighboring pixels can belong to the same group. That way, we create 27 different pixel groups in America, see Fig. 2. We limit the study to America because some of the time series classification methods that we analyse have a long running time. Once we know which of those methods perform well, the study can of course be further extended to other regions, under the assumption that the same methods are favored for those regions.

The vegetation extremes are then defined by applying a 10th percentile threshold on the seasonal anomalies of each region. This is a common threshold in defining extremes in vegetation [35]. Applying a lower threshold would result in extreme events that are extremely rare, making it impossible to train predictive models. In this way, we produce the target variable of our time series classification task. The presence of an extreme is denoted with a '1' and the absence with a '0'. Unsurprisingly, the distribution of the vegetation extremes in time indicates that many more extremes occur in recent years, which means that a clear trend appears again in the time series of extreme events, even though the initial time series was detrended. This makes the time series highly
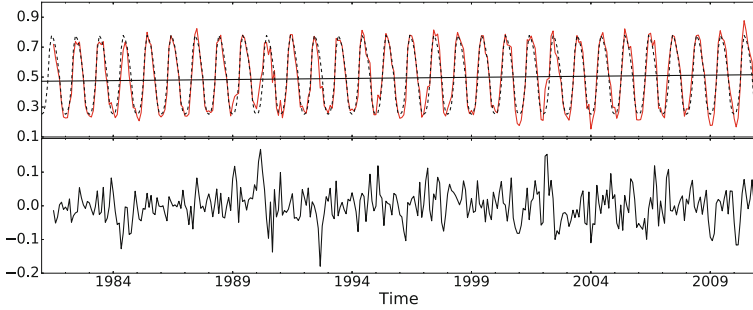
Fig. 1. The three components of an NDVI time series visualized for one particular location. The top panel shows the linear trend (black continuous line) and the seasonal cycle (dashed black line) that are obtained from the raw time series (red). The bottom panel visualizes the residuals, which are obtained by subtracting the seasonal cycle and the linear trend from the raw data. Only the residuals are further used to define extreme events. (Color figure online)
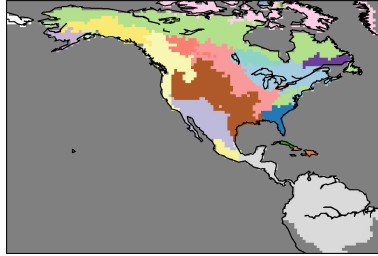
Fig. 2. Groups of pixels that are regions with similar climatic and vegetation characteristics. Based on the time series of each region we calculate the vegetation extremes for the pixels of that region.

non-stationary. Moreover, also a seasonal cycle typically re-appears, as one observes more extremes in certain months. Correctly identifying those two components (trend and seasonality) is essential when inferring causal relationships between vegetation extremes and climate.

As discussed in Sect. 2, a baseline model only includes information from the target time series (i.e. previous time stamps). We both consider the residuals as well as their binarized extreme counterparts as features for the baseline model. However, due to the existence of seasonal cycles and trends when considering binary time series of extreme vegetation, we also include 12 dummy variables which indicate the month of the observation and a variable for the year of this observation. These last two components are necessary because the baseline model should tackle as good as possible the seasonality and the trend that exists in the time series of NDVI extremes. In this paper, we perform a general test for causal relationships between climatic time series and vegetation. As such, the full model extends the baseline model with the above-mentioned climatic variables.

5   Results and Discussion

We present two types of experimental results. First, we analyze the predictive performance of various time series classification methods as representatives for the full model in a Granger-causality context. Subsequently, we select the best-performing algorithm for a Granger causality test, in which a baseline and a full model are compared.

5.1   Comparison of Time Series Classification Methods

For the first step we performed a straightforward comparison of the performance of the following algorithms: CID [3], LPS [4], TSF [10], SAXVSM [31], BOP [22], BOSS [30], FS [28] and hand-crafted features in combination with a classification algorithm [27]. In this setting, our dataset consists of monthly observations (there are in total 360 observations per pixel), and the input time series for each observation includes the 365 past daily values of precipitation time series before the month of interest (excluding the daily values of the current month). Only the precipitation time series is used, as some of the methods are unable to handle multivariate time series as input. We train the models per region by concatenating the observations of the pixels. The evaluation is performed per pixel by using random 3-fold cross-validation and AUC as performance measure.

Figure 3 shows the results. The vocabulary-based algorithms outperform the other representations, which implies that the frequency of the patterns makes the two classes of our dataset more distinguishable. Algorithms which distinguish the observations according to a presence or an absence of a shapelet perform poor, probably because observations originating from consecutive time windows have similar shapelets (the daily values of the next month is added for the next observation). In addition, the shapelet-based FS algorithm is also not very efficient in terms of memory space for large datasets. For this reason, we could not obtain results for the 4 largest regions of our dataset – see Table 1. For the algorithms that compare the whole raw time series by using a distance measure (i.e., CID) one can observe that the performance is also very low, probably also due to the strong similarity between consecutive observations. Similarly, algorithms that attempt to form a characteristic vector for each class fail since the patterns in both classes are very similar (i.e., SAXVSM). On the other hand, from the algorithms that use sub-intervals of time series, LPS has a similar performance as the vocabulary-based algorithms, because it takes local patterns and their relationships into account and forms a histogram out of them, while TSF fails in capturing useful information. We note that the LPS algorithm includes randomness so in each run it extracts different patterns from the data and also it is more time and space inefficient by comparison with the vocabulary-based algorithms. Finally, the hand-crafted features are not able to extract local patterns of the raw daily time series and are mostly based on statistic measurements. Table 1 presents the arithmetical results for the 9 largest regions. As one can observe, the results of BOP and BOSS are very similar. In most regions they give rise to substantially better results than the other methods that were tested.

(a) Hand-crafted features   (b) LPS   (c) BOP

(d) BOSS   (e) SAXVSM, TSF, FS   (f) CID

0.5 .55 .6 .65 .7 ≥ .8
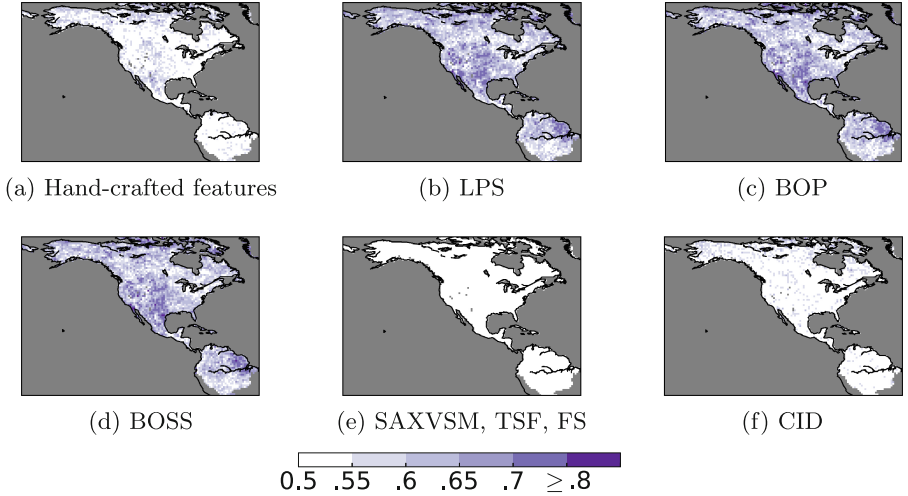
Fig. 3. Performance comparison in terms of AUC of the time series classification algorithms in the univariate time series classification setting on climate data.

Table 1. Mean and standard deviation of the AUC for areas which include more than 100 pixels. The vocabulary-based algorithms as well as the LPS algorithm perform very similar. Results of the algorithms SAXVSM and TSF are omitted due to their low performance.

Algorithm	Reg 1	Reg 2	Reg 3	Reg 4	Reg 5	Reg 6	Reg 7	Reg 8	Reg 9
Hand-crafted	0.50±0.01	0.50±0.00	0.54±0.05	0.52±0.03	0.51±0.02	0.50±0.00	0.50±0.00	0.50±0.01	0.51±0.01
LPS	0.59±0.06	**0.56**±0.04	**0.65**±0.09	**0.65**±0.07	**0.61**±0.06	**0.62**±0.05	0.60±0.05	0.65±0.07	0.59±0.05
BOP	**0.60**±0.07	**0.56**±0.05	**0.65**±0.08	0.64±0.07	0.60±0.06	0.61±0.05	**0.61**±0.06	0.66±0.07	**0.60**±0.05
BOSS	**0.60**±0.06	**0.56**±0.04	0.64±0.08	**0.65**±0.07	**0.61**±0.05	0.61±0.05	**0.61**±0.05	**0.67**±0.07	0.59±0.05
CID	0.50±0.03	0.50±0.02	0.51±0.05	0.51±0.04	0.50±0.03	0.54±0.04	0.53±0.03	0.55±0.05	0.51±0.03
FS	–	0.50±0.00	0.50±0.00	–	0.50±0.00	–	0.50±0.00	–	0.50±0.00

5.2   Quantification of Granger Causality

In a second step, we combine the best representation coming from the time series classification algorithms and we apply it to the non-linear Granger causality framework in order to test causal effects of climate on vegetation extremes. Our main goal is to replace the hand-crafted features constructed in [27]. As the BOSS algorithm has the best performance compared to the other time series algorithms, we use the vocabulary of patterns that BOSS automatically extracts from the climatic time series as features. To evaluate Granger causality, the baseline model includes information from the NDVI extremes, while the full model includes also the automatically-extracted features from the climatic time series. In contrast to the previous set of experiments, we now include three climatic time series instead of only the precipitation time series.
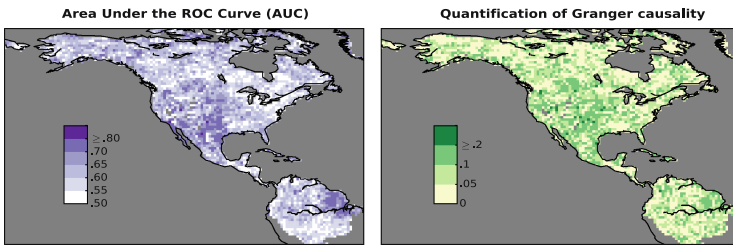
Fig. 4. On the left, the performance of the full model that uses the patterns extracted by the BOSS algorithm as predictors. On the right, a quantification of Granger causality; positive values indicate regions with Granger-causal effects of climate on vegetation extremes.

Figure 4 shows the performance of the full model in terms of AUC, as well as the performance improvement of the full model compared to the baseline model. It is clear that by using information from climatic time series the prediction of vegetation extremes improves in most of the regions. Therefore, one can conclude that – while not bearing into consideration all potential control variables in our analysis – climate dynamics indeed Granger-cause vegetation extremes in most of the continental land surface of North and Central America.

As results of that kind could not be obtained with hand-crafted feature representations, we do conclude that more specialized time series classification methods such as BOSS have the potential of enhancing causal inference in climate science. While this paper presents particular results for the case of climate-vegetation dynamics, we believe that the approach might be useful in other causal inference studies, too.

References

1. Attanasio, A.: Testing for linear granger causality from natural/anthropogenic forcings to global temperature anomalies. Theor. Appl. Climatol. **110**(1–2), 281–289 (2012)
2. Bagnall, A., Lines, J., Bostrom, A., Large, J., Keogh, E.: The great time series classification bake off : a review and experimental evaluation of recent algorithmic advances. Data Min. Knowl. Disc. **31**(3), 606–660 (2017). https://doi.org/10.1007/s10618-016-0483-9. ISSN: 1573-756X
3. Batista, G.E.A.P.A., Keogh, E.J., Tataw, O.M., De Souza, V.M.A.: CID: an efficient complexity-invariant distance for time series. Data Min. Knowl. Discov. **28**(3), 634–669 (2014)

4. Baydogan, M.G., Runger, G.: Time series representation and similarity based on local autopatterns. Data Min. Knowl. Discov. **30**(2), 476–509 (2016)
5. Baydogan, M.G., Runger, G., Tuv, E.: A bag-of-features framework to classify time series. IEEE Trans. Patt. Anal. Mach. Intell. **35**(11), 2796–2802 (2013)
6. Beck, H.E., McVicar, T.R., van Dijk, A.I.J.M., Schellekens, J., de Jeu, R.A.M., Bruijnzeel, L.A.: Global evaluation of four AVHRR-NDVI data sets: intercomparison and assessment against landsat imagery. Remote Sens. Environ. **115**(10), 2547–2563 (2011)
7. Beck, H.E., van Dijk, A.I.J.M., Levizzani, V., Schellekens, J., Miralles, D.G., Martens, B., de Roo, A.: MSWEP: 3-hourly $0.25°$ global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. Hydrol. Earth Syst. Sci. Discuss. **2016**, 1–38 (2016)
8. Chapman, D., Cane, M.A., Henderson, N., Lee, D.E., Chen, C.: A vector autoregressive ENSO prediction model. J. Clim. **28**(21), 8511–8520 (2015)
9. Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. Royal Meteorol. Soc. **137**(656), 553–597 (2011)
10. Deng, H., Runger, G., Tuv, E., Vladimir, M.: A time series forest for classification and feature extraction. Inf. Sci. **239**, 142–153 (2013)
11. Ding, H., Trajcevski, G., Scheuermann, P., Wang, X., Keogh, E.: Querying and mining of time series data: experimental comparison of representations and distance measures. Proc. VLDB Endowment **1**(2), 1542–1552 (2008)
12. Donat, M.G., Alexander, L.V., Yang, H., Durre, I., Vose, R., Dunn, R.J.H., Willett, K.M., Aguilar, E., Brunet, M., Caesar, J., et al.: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset. J. Geophys. Res.: Atmos. **118**(5), 2098–2118 (2013)
13. Geiger, P., Zhang, K., Gong, M., Janzing, D., Schölkopf, B.: Causal inference by identification of vector autoregressive processes with hidden components. In Proceedings of 32th International Conference on Machine Learning (ICML 2015) (2015)
14. Górecki, T., Łuczak, M.: Using derivatives in time series classification. Data Min. Knowl. Disc. **26**(2), 310–331 (2013). https://doi.org/10.1007/s10618-012-0251-4. ISSN: 1573-756X
15. Górecki, T., Łuczak, M.: Non-isometric transforms in time series classification using DTW. Knowl.-Based Syst. **61**, 98–108 (2014)
16. Grabocka, J., Schilling, N., Wistuba, M., Schmidt-Thieme, L.: Learning time-series shapelets. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 392–401. ACM (2014)
17. Granger, C.W.J.: Investigating causal relations by econometric models and cross-spectral methods. Econometrica: J. Econ. Soc. **37**, 424–438 (1969)
18. Hills, J., Lines, J., Baranauskas, E., Mapp, J., Bagnall, A.: Classification of time series by shapelet transformation. Data Min. Knowl. Discov. **28**(4), 851–881 (2014)
19. Kaufmann, R.K., Zhou, L., Myneni, R.B., Tucker, C.J., Slayback, D., Shabanov, N.V., Pinzon, J.: The effect of vegetation on surface temperature: a statistical analysis of NDVI and climate data. Geophys. Res. Lett. **30**(22) (2003)
20. Kodra, E., Chatterjee, S., Ganguly, A.R.: Exploring granger causality between global average observed time series of carbon dioxide and temperature. Theor. Appl. Climatol. **104**(3–4), 325–335 (2011)
21. Liao, T.W.: Clustering of time series data a survey. Patt. Recogn. **38**(11), 1857–1874 (2005)

22. Lin, J., Khade, R., Li, Y.: Rotation-invariant similarity in time series using bag-of-patterns representation. J. Intell. Inf. Syst. **39**(2), 287–315 (2012)
23. Liu, G., Liu, H., Yin, Y.: Global patterns of NDVI-indicated vegetation extremes and their sensitivity to climate extremes. Environ. Res. Lett. **8**(2), 025009 (2013)
24. Loveland, T.R., Belward, A.S.: The IGBP-DIS global 1km land cover data set, discover: first results. Int. J. Remote Sens. **18**(15), 3289–3295 (1997)
25. Mokhov, I.I., Smirnov, D.A., Nakonechny, P.I., Kozlenko, S.S., Seleznev, E.P., Kurths, J.: Alternating mutual influence of El-Niño/southern oscillation and Indian monsoon. Geophys. Res. Lett. **38**(8) (2011)
26. Mosedale, T.J., Stephenson, D.B., Collins, M., Mills, T.C.: Granger causality of coupled climate processes: ocean feedback on the North Atlantic Oscillation. J. Clim. **19**(7), 1182–1194 (2006)
27. Papagiannopoulou, C., Miralles, D.G., Verhoest, N.E.C., Dorigo, W.A., Waegeman, W.: A non-linear Granger causality framework to investigate climate-vegetation dynamics. Geosci. Model Dev. **10**, 1–24 (2017)
28. Rakthanmanon, T., Keogh, E.: Fast shapelets: a scalable algorithm for discovering time series shapelets. In: Proceedings of the 2013 SIAM International Conference on Data Mining, pp. 668–676. SIAM (2013)
29. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust. Speech Sig. Process. **26**(1), 43–49 (1978)
30. Schäfer, P.: The BOSS is concerned with time series classification in the presence of noise. Data Min. Knowl. Discov. **29**(6), 1505–1530 (2015)
31. Senin, P., Malinchik, S.: SAX-VSM: interpretable time series classification using sax and vector space model. In: 2013 IEEE 13th International Conference on Data Mining (ICDM), pp. 1175–1180. IEEE (2013)
32. Shahin, M.A., Ali, M.A., Ali, A.B.M.S.: Vector Autoregression (VAR) modeling and forecasting of temperature, humidity, and cloud coverage. In: Islam, T., Srivastava, P.K., Gupta, M., Zhu, X., Mukherjee, S. (eds.) Computational Intelligence Techniques in Earth and Environmental Sciences, pp. 29–51. Springer, Dordrecht (2014). https://doi.org/10.1007/978-94-017-8642-3_2
33. Stefan, A., Athitsos, V., Das, G.: The move-split-merge metric for time series. IEEE Trans. Knowl. Data Eng. **25**(6), 1425–1438 (2013)
34. Tucker, C.J., Pinzon, J.E., Brown, M.E., Slayback, D.A., Pak, E.W., Mahoney, R., Vermote, E.F., El Saleous, N.: An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. Int. J. Remote Sens. **26**(20), 4485–4498 (2005)
35. Zscheischler, J., Mahecha, M.D., Harmeling, S., Reichstein, M.: Detection and attribution of large spatiotemporal extreme events in Earth observation data. Ecol. Inform. **15**, 66–73 (2013)