Dear Editor,

With this letter we would like to resubmit our manuscript entitled “Detecting dynamical anomalies in time series from different palaeoclimate proxy archives using windowed recurrence network analysis” for publication in *Nonlinear Processes in Geophysics.*

In the revision, we implemented all changes that we mentioned in our comments during the discussion phase. In the following, we reproduce our answers to the Referee’s comments to our manuscript (corresponding to the answers provided during the discussion phase) and highlight the associated changes in the revised manuscript. A difference of the originally submitted and resubmitted manuscript can be found at the end of this document.

Yours sincerely,
Jaqueline Lekscha and Reik Donner

Answers to Referee 1

Overall:
In this manuscript the authors test the suitability of the earlier developed windowed recurrence network analysis (wRNA) for detecting dynamical anomalies in paleoclimate proxy times series. The method skills are tested on the suite of stationary and nonstationary timeseries of forward modelled pseudoproxies with known dynamical properties. This work is a natural continuation/extension of the earlier studies of the group on the application of networks to the analysis of (paleo)climatic data

The paper is clearly written and results are well presented. I therefore consider the manuscript deserves to be published after some very minor modifications /additions to the content if the authors/editor finds them relevant.

We thank the Referee for this positive assessment of our work. Our replies to the suggested modifications and additions are given below.

Minor comments
Page 3 Line 59: “for estimating embedding delay...autocorrelation function” is it a global or a windowed estimate? Please be specific

It is a global estimate. We have specified this in the revised version of the manuscript (line 64).

Page 3 Line 66: why namely the maximum norm is used? Is it possible to justify the choice? Did the authors check the sensitivity of the results to the use of other norms?

We use the maximum norm because there is a direct analytical relationship between the network transitivity and the dimensionality of the dynamics of the system for it, as shown in Donner et al., EPJB, 2011. Thus, using the maximum norm is particularly useful when assessing the network transitivity to study the system’s dynamics. Moreover, due to its specific form, the calculation of the maximum norm of a given vector in Euclidean space is particularly simple. This is also why this specific norm has been particularly widely used in previous works on recurrence network analysis and other recurrence plot based techniques. We have clarified this in the revised manuscript (lines 72/73 and 83/84).

In our present work, we did not explicitly check the sensitivity of the results with respect to the use of other norms for the sake of conciseness. However, we performed similar sensitivity tests for other (similar) time series and typically found the results to be qualitatively robust when, for example, using the Euclidean norm instead of the maximum norm. We expect this to also hold for the present work.

Page 3 Line 71: "...such that a fraction \rho of all possible links in the network is realized": is the threshold global or window-based?

For each window, we construct a recurrence network and thus, set the threshold \epsilon for every window separately. That is, in each window, a fraction \rho of all possible links are realized. We have clarified this in the revised manuscript (lines 76/77).

Page 3 Eq. 4 please indicate that |i-j|=|v-i|=1

In this equation, the summation is performed over all i, j and v and not only over neighboring indices. In fact, the network transitivity as given by Eq. 4 can be interpreted as the probability that two random neighbors of a randomly chosen node are mutually connected. Assuming that the randomly chosen node has index v, the denominator of Eq. 4 counts how many combinations of all i and j are neighbors of v. By summing over all possible v, the denominator equals the number of all possible triangles in the network. The nominator then counts how many of the possible triangles are actually realised.

Page 5-6: forward proxy model for tree rings. One should not that the juvenile growth is not modelled/accounted for in the model used. Hence an effect of its subtraction, which can be substantial, depending on the species used, technique applied and the entire age structure of the tree-ring network (archive) is also discarded. It is worth mentioning in a context of results demonstrated for tree rings.

We thank the Referee for drawing our attention to this important point. Indeed, the model does not take into account juvenile growth and it would be very interesting to compare the results to those of a model that does account for this effect in future work. We have added a corresponding comment in the revised manuscript (lines 149-151 and 412/413).

Page 8 table 2: Please check if the amount of measured foraminifera is correct (number of species? Sample weight?) please indicate units

For this model, we used the default values for the amount of measured foraminifera, mixed layer thickness, and abundance of species of the TURBO2 model which are unitless (compare Trauth, Comp. Geosci., 2013). Of course, it would be very interesting to compare the results for varying values of these parameters. Unfortunately, such a systematic study on the effects of different model parameter settings was beyond the scope of our present manuscript and is thus left as a topic for future work. Generally, we expect here more quantitative rather than qualitative changes if the mentioned parameters are varied within some reasonable ranges.

Page 11: Use of nonstationary Røssler system: How realistic this model actually is for climate applications? Are there any larger-scale climatic processes that can potentially be associated with this model?

Unlike the Lorenz system, which has been originally introduced as a simplified toy model for atmospheric convection processes, the Rössler system has no such close climatological interpretation to the best of our knowledge. One of the main motivations for introducing this model

(inspired by chemical reaction kinetics in Rössler's (Phys. Lett. A, 1976) original work) was to provide a chaotic dynamical system model with somewhat simpler behavior than the Lorenz system (i.e., without a double-scroll structure). This simpler type of attractor topology, along with the still non-trivial and rich cascade of bifurcations, was the main motivation for us to use the Rössler system as a generator for complex input dynamics to our proxy system models. Notably, we expect that this well studied model is known to a vast majority of the readership of Nonlinear Processes in Geophysics. By contrast, we did not attempt here to reach any dedicated level of realism in making the input variable particularly similar to real world climate dynamics. We have clarified this point in our revised manuscript (lines 262/263 and 268).

Page 12 Line 309: "...respond to temperature rather than to precipitation..." mind that compared with a temperature, precipitation is not reproduced in the models that well, though for this particular case (boreal forest), temperature indeed will be a stronger limiting factor.

As we chose the model parameters for the tree ring width model to simulate tree growth in Eastern Canada, we were indeed expecting that the temperature and not the precipitation will be the limiting factor for the model in this case. However, we fully agree with this comment, and added a brief explicit statement on this fact in the revised supplement (lines 9-13).

Page 13 Line 322: "...closely follow the respective temperature input", note my comment on the used forward proxy model for tree rings. Such a good concistency can partly be attributed to a lack of juvenile growth effect in the model.

We agree with the Referee that the good correspondence between the temperature input and the tree ring width model output can partly be attributed to the lack of correction for juvenile growth in the tree ring width model. We have commented on this in the revised manuscript (lines 149-151 and 412/413).

Page 17 Line 368: "....lower-dimensional dynamics during the MCA.....higher-dimensional... during the LIA" Can the authors elaborate a bit more on this result? What are the actual features in the analyzed timeseries manifested in wRNA as lower and higher network \Tau?

The MCA has often be attributed to more stable climate conditions while the LIA has often been attributed to more variable climate conditions even though this imprint has varied locally and has been mainly discussed for Europe. As described, for the LMR data from Eastern Canada, we do not find any significant periods in the model output data meaning that we cannot detect any dynamical anomalies in the model output. Still, for the tree ring width model output, we observe higher values of the network transitivity roughly coinciding with the MCA and lower values of the network transitivity during the LIA. As the network transitivity (\mathcal{T}) has been shown to be related to some proxy for the dimensionality of the system's dynamics (m) via the relation m=\log(\ mathcal{T}) / \log(3/4) (cf. Donner et al., EPJB, 2011), we tentatively conclude that the higher values of the network transitivity during the MCA and the lower values of the network transitivity during the LIA reflect a lower/higher dimensional dynamics of the system at this particular location (in terms of complexity of temporal variations rather than just a change in variance). In terms of the time series properties for the different periods, we indeed additionally observe an increase in variance of the time series from the MCA to the LIA, which is very likely also reflected in the different recurrence networks and, thus, the resulting network transitivity. We note that this non-stationarity in variance along with the MCA-LIA transition in the European/North Atlantic sector has also been reported as being reflected in other nonlinear characteristics, which have been previously interpreted as a hallmark of some dynamical anomaly (Schleussner et al., Clim. Dyn., 2015; Franke & Donner, Clim. Change, 2017).

We have added a corresponding more detailed discussion of the results to the revised manuscript (lines 355-366).

Page 19 Lines 403-404: Did the authors consider block shuffling of surrogates (same as in block bootstrapping) as a possible method to tackle this problem?

We did not consider block shuffling of surrogates in this work, but rather addressed this issue by using iterative amplitude adjusted Fourier transform surrogates (that is, surrogates that exactly preserve the probability distribution and linear correlation structure of the data) for the areawise significance test.

We thank the Referee for suggesting using block shuffling as a possibly less computationally demanding alternative and will keep this option in mind for further work on significance tests. For the present work, we think that adding an additional type of surrogate data would primarily increase the already large amount of material presented without providing topically relevant results markedly deviating from those reported. Still, we agree that this is a very important topic, and suggest that apart from block shuffling surrogates, also other surrogate routines should be studied more systematically for time series with different autocorrelation properties.

References

Donner, R. V., Heitzig, J., Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: The geometry of chaotic dynamics — a complex network perspective, The European Physical Journal B, 84, 653–672, 2011.

Franke, J. G., Donner, R. V.: Dynamical anomalies in terrestrial proxies of North Atlantic climate variability during the last 2ka, Climatic Change, 143(1-2), 87-100, 2017.

Rössler, O.: An equation for continuous chaos, Physics Letters A, 57, 397 – 398, 1976.

Schleussner, C.-F., Divine, D., Donges, J. F., Miettinen, A., Donner, R. V.: Indications for a North Atlantic ocean circulation regime shift at the onset of the Little Ice Age, Climate Dynamics, 45, 3623-3633, 2015.

Trauth, M. H.: TURBO2: A MATLAB simulation to study the effects of bioturbation on paleoceanographic time series, Computers & Geosciences, 61, 1 – 10, 2013.

Answers to Referee 2

The purpose of the manuscript is to apply a method developed and documented by the authors elsewhere (the windowed recurrence network analysis (wRNA)) on artificial time series simulating the output of speleothem, lake, tree archives and isotopic water concentration in ice cores. The purpose is to test to what extent the 'proxy process' (transformation of the climate signal by the natural archive) could mask anomalies that would have been otherwise detected in the original time series. Different time series have been tested, including Gaussian white noise, an autoregressive process of order 1, output non-linear dynamical system model, and data from the reanalysis project of Hakim et al.

The main diagnostic is 'network transitivity', and the application of the method to the

different input datasets generates Figures 2 to 6. The reader is invited to concentrate on the 'area-wise' significance test, which is supposed to indicate a signal of significant change in network transitivity, to be interpreted as a change in the effective dimensionality of the system.
Some results appear a bit disconcerting, especially the test on the AR(1) signal, because it shows, on the one hand, a large patch of area-wise significant anomalies in network transitivity which was — if I understood correctly — a priori not expected in this signal. In addition, this large patch disappears in the natural archives simulated with this model, while another patch emerges in the speleothem simulation. Simulations with other inputs tend to confirm that the speleothem model is prone to create or destroy areas of significant network transitivity anomalies seen in the original time series.

Indeed, the AR(1) signal shows a patch of areawise significant anomalies that is a priori not expected. As argued in the manuscript (original manuscript, lines 398-400), such artefacts appearing in single realisations of the stochastic input time series should be excluded in future work by considering ensembles of realisations of the processes. Unfortunately, this was not possible within the scope of the original manuscript.

With respect to the results of the different proxy model output time series, we generally found that the tree ring and the lake sediment model tend to miss, while the speleothem model both misses and creates additional areawise significant patches of the network transitivity, which is in agreement with the above observations. We have further disentangled the description of these results in the revised manuscript in order to make these points clearer (lines 313-366).

The study is quite systematic in its approach, to the point of being slightly repetitive, and yet, one might argue that it falls short convincing the reader about the robustness of the conclusions. Basically, up to p. 12 the manuscript consists in an exposition of the methods, which for their greatest part have been described elsewhere (the significant area test is in press, and the proxy models have been published elsewhere). The output of the wRNA analysis follows a show-and-tell description running until p. 18, and even though some main conclusions are correctly outlined, the discussion does not really help identify mechanisms or key conclusions that would actually help to 'improve the interpretation of windowed recurrence network analysis' as announced in the abstract. For example, the authors have observed that the speleothem model modifies the significant patches, but we do not really which process, in the speleothem model, is responsible for this behaviour. Do we expect this to be an idiosyncrasy of the particular speleothem model used here, or do we expect it to be a general result? Which aspects of the 'nonlinear filtering' should be incriminated? The presence of a large significant area patch on the AR(1) time series, along with its the quasi-absence of significant changes in network transitivity in the last-millennium reanalysis data is also disconcerting, because we no longer know how to reasonably interpret the output of the wRNA for understanding climate dynamics. Is the last millennium actually the right test bed for this study?

We thank the Referee for pointing out the weaknesses of the presentation of the material and have worked on the mentioned points in the revised manuscript. This particularly concerns the presentation of the results and the corresponding discussion and conclusions (lines 313-366 and 387-424). We also decided to move the description of the time series properties of the model output time series to the supplementary information document. For the (admittedly long) description of the methods and proxy system models, we consider it useful to be included in the manuscript to provide

a complete picture and make the contribution self-consistent. Still, we shortened the detailed description of the areawise significance test (lines 109-112).

Also, we want to stress that this study has been meant to improve the interpretation of wRNA results in terms of highlighting not only the potentials, but also the limitations of the method when applied to palaeoclimate data. That is, in particular for the speleothem model, our results show the need of further studying the effects of different filtering mechanisms on the results of the wRNA in order to draw reliable conclusions when analysing real-world data. (Note that there have been quite a few papers reporting the results of windowed recurrence analyses of real-world speleothem data, e.g., Donges et al., Clim. Past, 2015; Eroglu et al., Nat. Comm., 2016.) In this regard, we argue that the last millennium is a good test bed because for this period, highly resolved proxy data are available from various different archives. We have clarified the direct conclusions from the obtained results with respect to the interpretation of the wRNA and outline further work that will help to increase the robustness of the conclusions (lines 397-403). For the speleothem model results, for example, we expect the conclusions to hold, independently of the actual model used, but potentially depend on the choice of the model parameters.

To sound hopefully a bit more constructive, I would suggest the authors to seek for more general aspects of the filtering process which may destroy or generate spurious changes in network transitivity. Is this caused by non-linearity in the instantaneous response (what would an 'exp(x)' filtering generate?) Is this the temporal smoothing process? Is it the amount of noise? What would this analysis tell us about how to find proxies that would preserve the wRNA signal, beyond the particular example chosen here? Which are the desirable characteristics for such proxies? Answering these questions would provide some more general and perhaps valuable hints for the interpretation of the wRNA, which could then be summarised in the abstract.

We thank the Referee for these suggestions and fully agree that including the results for a set of general filtering functions and noises on the wRNA results will help interpreting the obtained results by providing insights into the mechanisms that are responsible for them in the different proxy system models. We have performed some further analyses and briefly elaborated on these more general aspects in the revised version of the manuscript (lines 304-312).

Perhaps the reader will also better appreciate the interest of the wRNA if more clues are given about how to interpret it: can one get a more or less adequate intuition of what a change in wRNA implies about the dynamics of climate. What, physically, does an increase or a decrease in network transitivity mean? Would this be associated with a form of 'global synchronisation' ? Are we expecting it when we approach a form of bifurcation (a "tipping point")? What is the wRNA telling us that is not obvious from visual inspection of the time series?

We agree with the Referee that including a paragraph on the interpretation of wRNA with respect to the climate system is a very good idea. Still, general statements will hardly be possible as climate-related interpretations vary depending on the location and thus, local boundary conditions have to be taken into account. Also, the network transitivity calculated from a single time series cannot be associated with a 'global synchronisation' as only information from a single location is taken into account. Instead, the network transitivity has rather been related to the dynamical regularity of the variations in the analysed time series (e.g., Donges et al., PNAS, 2011) with higher values of the network transitivity corresponding to less irregular variability and vice versa. This is in accordance with the interpretation of the network transitivity as an indicator of the dimensionality of the system's dynamics. In this regard, detected anomalies in the network transitivity could be related to some tipping point, but do not have to be. We have attempted to provide a more concise description

of possible interpretations of the wRNA results in the (palaeo)climate context in the revised version of the manuscript (lines 404-410).

Finally, the choice of an embedding dimension m = 3 was, to this reader, difficult to reconcile the quote that "The embedding theorem of Takens guarantees that, when choosing the embedding dimension larger than twice the box-counting dimension of the original attractor, the reconstructed and the original system's attractor are related by a smooth one-to-one coordinate transformation with smooth inverse, independent of the choice of the delay". Wouldn't we have expected, on this basis, a much larger embedding dimension? This may invite some discussion, perhaps available in Lekscha and Donner, (in press). In 1984 (Nature, vol. 311, p. 311), Nicolis and Nicolis published an estimate of the 'climate attractor dimension' but subsequent authors (including Grassberger, 1986, Nature, 1996, vol. 323, p. 609, and Vautard and Ghil, 1989, Physica D, vol. 35, p.395) pointed the difficulty of actually getting a meaningful estimate of "a climate dimension" from a 1-dimensional, finite record. Could the authors clarify their position in this respect?

We are well aware of the problem of choosing an appropriate embedding dimension when the available data are univariate, finite, and subject to noise. The embedding theorem of Takens is a sufficient condition and not a necessary one; thus, embedding dimensions smaller than twice the box-counting dimension of the original system's attractor may lead (at least approximately) to an appropriate embedding, which might possibly, at least partly, reconcile the quote and the choice of the embedding dimension m = 3. Also, it should be noted that the theorem is strictly valid only for perfect data. Thus, for finite and noisy data, the estimation of the embedding dimension has to rely on some more or less heuristic approaches. Many of these approaches however either have problems of distinguishing deterministic chaos and noise or systematically underestimate the embedding dimension.

In this spirit, we fully agree with the critiques of the cited papers that it is difficult to get a meaningful estimate of an embedding dimension from univariate, finite and noisy data. For the particular case of a 'climate attractor dimension', we agree that the climate system is not a low-dimensional dynamical system. Still, we think that lower dimensional embeddings can be used to obtain meaningful information about a system. Furthermore, in the palaeoclimate context where available time series are often rather short, we do not think that high-dimensional embeddings are useful as the limited amount of data points will in most cases not be sufficient to sample the attractor in a high-dimensional embedding space.

In the majority of the paper, we do not directly analyse climate data but rather synthetic data representing different kinds of underlying processes. For convenience, comparability, and with respect to the time series length and the increasing computational effort for larger embedding dimensions, we chose m = 3 for all analysed time series, thus, also for the last millennium reanalysis data which indeed represents more closely the actual dynamics of the climate system. In general, when analysing real-world data, we think that the problem of choosing an appropriate embedding dimension is best tackled by employing one of the estimation methods that can distinguish deterministic and stochastic signals and do not require too many subjective parameter choices (such as the one presented in Cao (1997), Physica D, 110, 43-50), and additionally slightly varying the obtained embedding dimension when analysing the data to check the robustness of the results. In the revised manuscript, we added a more detailed comment on our choice of m = 3 (lines 55-64).

Detecting dynamical anomalies in time series from different palaeoclimate proxy archives using windowed recurrence network analysis

Jaqueline Lekscha[1,2] and Reik V. Donner[1,3]

[1]Potsdam Institute for Climate Impact Research (PIK) – Member of the Leibniz Association, 14473 Potsdam, Germany
[2]Department of Physics, Humboldt University, 12489 Berlin, Germany
[3]Department of Water, Environment, Construction and Safety, Magdeburg–Stendal University of Applied Sciences, 39114 Magdeburg, Germany

Correspondence: lekscha@pik-potsdam.de

Abstract. Analysing palaeoclimate proxy time series using windowed recurrence network analysis (wRNA) has been shown to provide valuable information on past climate variability. In turn, it has also been found that the robustness of the obtained results differs among proxies from different palaeoclimate archives. To systematically test the suitability of wRNA for studying different types of palaeoclimate proxy time series, we use the framework of forward proxy modelling. For this, we create artificial input time series with different properties and ~~, in a first step, compare the time series properties~~ compare the areawise significant anomalies detected using wRNA of the input and the model output time series. ~~In a second step, we compare the areawise significant anomalies detected using wRNA. For~~ Also, taking into account results for general filtering of different time series, we find that the variability of the network transitivity is altered for stochastic input time series while being rather robust for deterministic input. In terms of significant anomalies of the network transitivity, we observe that these anomalies may be missed by proxies from tree and lake archives ~~, we find that significant anomalies present in the input time series are sometimes missed in the input time series after the nonlinear~~ after the non-linear filtering by the corresponding proxy system models. For proxies from speleothems, we additionally observe falsely identified significant anomalies that are not present in the input time series. Finally, for proxies from ice cores, the wRNA results show the best correspondence with those for the input data. Our results contribute to improve the interpretation of windowed recurrence network analysis results obtained from real-world palaeoclimate time series.

1 Introduction

Palaeoclimate proxy time series from archives such as trees, lakes, speleothems, or ice cores play an important role in past climate reconstructions (Bradley, 2015). Apart from information on climatic boundary conditions such as local and global mean temperatures and precipitation sums, the analysis of the proxy time series using non-linear methods offers the possibility to study dynamical anomalies in past climate variability (Marwan et al., 2002, 2003; Trauth et al., 2003; Marwan et al., 2009; Donner et al., 2010b; Franke and Donner, 2017). Due to the limitations of palaeoclimate proxy time series in terms of non-uniform sampling of the data points, uncertainties in dating and measurement, the contamination with noise, and the often

non-unique interpretation of the climate signal within the proxy, not all methods are suitable to gain reliable information from all data (Goswami et al., 2018; Lekscha and Donner, 2018).

Windowed recurrence network analysis (wRNA) has already been successfully used to detect dynamical anomalies in time series from different palaeoclimate archives (Donges et al., 2011a, b; Eroglu et al., 2016; Lekscha and Donner, 2018). But it has also been observed that this method often yields high numbers of false positive significant points, and, that not all palaeoclimate archives give equally robust results. In order to improve the interpretation of results obtained with wRNA for real-world time series, we here systematically test the suitability of the approach for analysing palaeoclimate data from different types of archives by employing proxy system models.

Proxy system models are forward models that simulate the formation of a palaeoclimate proxy based on the systemic understanding of that proxy (Evans et al., 2013). That is, given the climate input variables, the proxy system model outputs a proxy time series. We here use intermediate complexity models for tree ring width, branched glycerol dialkyl glycerol tetraethers in lake sediments, and the isotopic composition of speleothems and ice cores (Tolwinski-Ward et al., 2011; Dee et al., 2015, 2018). We first create artificial climate input time series with different properties. In particular, we consider two stochastic processes, Gaussian white noise and an autoregressive process of order 1, and two non-stationary time series from the paradigmatic Rössler and Lorenz systems. Additionally, we use climate input from the last millennium reanalysis project as an estimate of realistic past climate variability (Hakim et al., 2016; Tardif et al., 2019). We then compare the input and the model output with respect to the properties of the time series and the results of wRNA. To quantify significant dynamical anomalies, we use an areawise significance test that was recently introduced for wRNA (Lekscha and Donner, 2019).

With this work, we want to contribute to a better understanding and an improved interpretation of results obtained with wRNA for palaeoclimate applications. We first introduce our analysis framework in Sec. 2, the proxy system models in Sec. 3, and the input time series in Sec. 4. Then, we show and discuss the ~~time series properties and the~~ results of the wRNA for the different input and model output time series in Sec. 5 before we present our main conclusions in Sec. 6. Additional information and figures are included in the Supplementary Material accompanying this paper.

2 Analysis framework

2.1 Phase space reconstruction

The first step when analysing data using recurrence based approaches is to reconstruct the higher-dimensional phase space of the system from the measured univariate time series $x(t)$ with observations made at times $\{t_i\}_{i=1}^{N}$. For this, we ~~here~~ use uniform time delay embedding (Takens, 1980; Packard et al., 1980) ~~. Time delay embedding~~ which relates the higher-dimensional coordinates of the system's phase space to delayed versions of the measured, univariate time series

$$x(t_i) \to \boldsymbol{x}(t_i) = \{x(t_i), x(t_i - \tau), \ldots, x(t_i - (m-1)\tau)\}\,. \tag{1}$$

Here, m denotes the embedding dimension and τ is the embedding delay. The embedding theorem of Takens guarantees that, when choosing the embedding dimension larger than twice the box-counting dimension of the original attractor, the

reconstructed and the original system's attractor are related by a smooth one-to-one coordinate transformation with smooth inverse, independent of the choice of the delay (Takens, 1980). That is, analysing the dynamics in the reconstructed phase space can give information on the original system dynamics. Still, this theorem only provides a sufficient condition for the embedding dimension, while also smaller-dimensional embeddings may give useful information about the system's dynamics.

~~To determine the embedding parameters in~~ In practical applications, however, the box-counting dimension of the original attractor is usually unknown and the data are finite and subject to noise such that the choice of the embedding parameters plays an important role for the quality of the phase space reconstruction. In particular, the embedding dimension can be estimated using the method of false nearest neighbours (Kennel et al., 1992)~~can be used for estimating the embedding dimension~~, while the first zero of the autocorrelation function or the first minimum of the mutual information can serve as an estimate for ~~the~~ a suitable embedding delay (Abarbanel et al., 1993; Fraser and Swinney, 1986; Kantz and Schreiber, 2004). We here use ~~embedding dimension~~ an embedding dimension of $m = 3$ as a compromise between the number of available data points and the minimum number of independent coordinates as indicated by the false nearest neighbour criterion. For estimating the embedding delay, we use the first zero of the global lagged autocorrelation function.

2.2 Windowed recurrence network analysis

We analyse the embedded time series $\boldsymbol{x}(t)$ by means of windowed recurrence network analysis (wRNA), using a sliding window approach and dividing the embedded time series into windows of width W with mutual offset dW. For each of those $N' = (N-(m-1)\tau - W)/dW$ windows, we construct a network from the time series by identifying the different $\boldsymbol{x}_i = \boldsymbol{x}(t_i)$ as nodes of the network and drawing a link between two nodes $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ if they are mutually closer in phase space than some threshold ϵ. For this purpose, we measure the distances in phase space using the maximum norm,

$$\|\boldsymbol{x}\|_\infty = \max_{k=1,\ldots,m} \left\{ x^{(k)} \right\}, \tag{2}$$

which has, due to its particularly simple form when calculated in Euclidean space, been widely used in recurrence-based analyses. Then, the resulting recurrence network can be described by its adjacency matrix $\mathbf{A}$ with entries

$$A_{i,j}(\epsilon) = \theta\left(\epsilon - \|\boldsymbol{x}_i - \boldsymbol{x}_j\|\right) - \delta_{i,j}. \tag{3}$$

where $\theta(\cdot)$ is the Heaviside function and the delta function $\delta_{i,j}$ excludes self-loops in the network. Here, we select the threshold adaptively by choosing a fixed edge density $\rho = 0.05$. This means, for every window, we choose the threshold ϵ such that a fraction ρ of all possible links in the network are realised (corresponding to the $100\rho\%$ mutually closest pairs of state vectors).

From the adjacency matrix, we can estimate various network properties. In the course of this work, we will restrict ourselves to the network transitivity

$$\mathcal{T} = \frac{\sum_{v,i,j} A_{v,i} A_{i,j} A_{j,v}}{\sum_{v,i,j} A_{v,i} A_{j,v}}, \tag{4}$$

which gives the probability that two randomly chosen neighbours of a randomly chosen node are mutually connected. Network transitivity is particularly well suited for detecting dynamical anomalies in non-stationary time series, since this network measure has been shown to be closely related to the dimensionality of the system dynamics, in particular, when using the maximum norm to measure distances in phase space (Donner et al., 2010a, 2011b). Specifically, the quantity $\log \mathcal{T} / \log(3/4)$ then corresponds to a generalised fractal dimension (Donner et al., 2011a). Thus, low values of the network transitivity can be related to higher-dimensional dynamics and vice versa.

For the analysis performed here, we repeat the wRNA for different values of the window width $W \in [100, 300]$ with step size $\Delta W = 1$ in order to check the robustness of the results for this analysis parameter. Thus, the resulting values of the network transitivity can be stored in some matrix $\mathbf{Q}$ which is of size $N_W \times N'$ with N_W being the number of window widths for which the analysis has been performed and N' being the number of windows into which the original time series is divided. We choose the offset of the windowed analysis to be always $dW = 1$.

2.3 Significance testing and confidence levels

In order to identify dynamical anomalies from the resulting network transitivity, we first perform a pointwise significance test using random shuffling surrogates. That is, for every window width W, we randomly draw $N_s = 1,000$ times W embedded state vectors from $\boldsymbol{x}(t)$ and calculate the network transitivity of this set of W state vectors. We then take the empirical 2.5th and 97.5th percentiles from the N_s realisations to obtain a two-sided confidence interval of $s_{pw} = 95\%$. All transitivity values of the original analysis results that fall outside this interval are considered to show pointwise significant anomalies, that is, the null hypothesis of the corresponding values of transitivity resulting from random data with the same amplitude distribution as the original data is rejected. The result of this pointwise significance test can be summarised in the binary significance matrix $\mathbf{S}^{pw}$ of the same size as the matrix of the transitivity results $\mathbf{Q}$. The entries of $\mathbf{S}^{pw}$ equal one if the corresponding value of the transitivity has been found to be pointwise significant, and zero otherwise.

As intrinsic correlations of the analysis results in both, the time domain (due to the short offset dW) and the window width domain can lead to patches of false positives in the pointwise significance matrix, we additionally perform an areawise significance test (Lekscha and Donner, 2019; Maraun et al., 2007). A pointwise significant point is defined to be areawise significant, if all neighbouring points within a given rectangle are also pointwise significant, i. e., if the pointwise significant point lies within a patch of pointwise significant points that is larger than this rectangle. The side lengths of the rectangle depend on the intrinsic correlations of a chosen null model that are estimated as ~~detailed in the following~~described in detail in Lekscha and Donner (2019).

~~First, we choose a null model for which we want to perform the areawise test. We here~~ Here, we employ a data-adaptive null model using iterative amplitude-adjusted Fourier transform surrogates of the original time series (Schreiber and Schmitz, 2000). That is, we test whether the same analysis results could have been obtained from data with the same amplitude distribution and linear correlation structure as the original data but that are otherwise random, i. e., originate from linear stochastic processes with prescribed correlations. ~~We thus create $N_{da} = 100$ such surrogate data sets and perform the same analysis on them as on the original data. We then quantify the intrinsic correlations by calculating correlation functions in the time~~

and the window width domain for the different values of the window width W. The corresponding preliminary decorrelation lengths are obtained as the values of the time lag or window width where the correlation function falls below a threshold that we set to $1/e$ for the time and $2/e$ for the window width domain. Those different values of the threshold can be explained by the different behaviour of the correlation function in the different domains and are set in accordance with the findings in Lekscha and Donner (2019). We then estimate the final decorrelation lengths l_t and l_W of the two domains by applying a linear fit to the preliminary decorrelation length as a function of the window width

$$l_{t/W} = m_{t/W} W + n_{t/W}.$$

Finally, we define the areawise significance matrix $\mathbf{S}^{aw}$ to have entries $S^{aw}_{i,j} = 1$ if the relation

$$\sum_{p=i-(l_t/2-1)}^{i+(l_t/2-1)} \sum_{q=j-(l_W/2-1)}^{j+(l_W/2-1)} S^{pw}_{p,q} = (l_t - 1)(l_W - 1)$$

holds, and to have entries $S^{aw}_{i,j} = 0$ otherwise. A detailed and more general description of the areawise significance test can be found in Lekscha and Donner (2019).

3 Proxy system models

Forward modelling of palaeoclimate proxies offers the possibility to gain insights into the underlying processes that influence the sensitivity of a given proxy to climate variations and can thus be used to investigate characteristic properties of time series of different palaeoclimate archives and their implications for further analyses. We here use four models for typical proxies from tree rings, lake sediments, ice cores and speleothems, respectively, in order to test how well dynamical anomalies can be identified when applying wRNA to time series originating from those archives.

Generally, a proxy system model can be divided into an environment, a sensor, an archive and an observation sub-model (Evans et al., 2013). The environment model can be used to model the environmental factors that the archive is sensitive to using the climatic input variables. The sensor model then describes how the archive reacts to the environment, and the archive model accounts for how this reaction is written into the archive. Finally, an observation model can be used to simulate uncertainties in dating or measurements. Here, depending on the archive, we use different combinations of the environment, sensor and archive sub-models, while neglecting possible dating and measurement uncertainties. In the following, we will give a brief overview over the model structures and parameters. A full description of the models can be found in the corresponding references.

3.1 Tree rings

Tree rings are one of the most important archives for palaeoclimate reconstructions of the last millennium (Bradley, 2015; St. George, 2014; St. George and Esper, 2019). To model the tree ring width as a function of time at a particular site, we use the intermediate complexity model Vaganov-Shashkin-Lite (VS-Lite) (Tolwinski-Ward et al., 2011). This is a reduced version of the full Vaganov-Shashkin model (Vaganov et al., 2006) and requires monthly input data of temperature T and

either precipitation P or soil moisture M. Additional model parameters are thresholds for temperature and soil moisture below which growth is not possible and above which growth is optimal (T_1, M_1, T_2, M_2), the latitude of the site Φ, and integration start and end months I_0 and I_f that set the period over which climate is responsible for growth in a given year.

If precipitation is given, the Leaky Bucket model (Huang et al., 1996) is used as environment model to calculate the soil moisture $M(t)$ based on the water balance in soil. This model requires additional parameters such as the initial moisture content of the soil M_0, minimum and maximum soil moisture content $M_{\text{min,max}}$, runoff parameters m, α and μ and the root depth d_r. The sensor model for the tree ring width then basically relies on the principle of limiting factors (Fritts, 1976), that is, it assumes that tree ring growth is limited by the resource that is the scarcest for optimal growing conditions, i. e., in this case, the temperature or the soil moisture. A temperature-based growth response is calculated as

$$g_T(t) = \begin{cases} 0 & \text{if } T(t) < T_1 \\ \dfrac{T(t) - T_1}{T_2 - T_1} & \text{if } T_1 \leq T(t) \leq T_2 \\ 1 & \text{if } T_2 > T(t) \end{cases} \tag{5}$$

and similarly, a growth response $g_M(t)$ is calculated for soil moisture. In addition, a third insolation-based growth response $g_E(t)$ is calculated based on the mean of the normalised lengths of the day for each month. The total growth response $g(t)$ of the tree to the climatic input is then given as the minimum of the temperature and moisture based growth responses modulated by the insolation based growth response,

$$g(t) = g_E(t) \min\{g_T(t), g_M(t)\}. \tag{6}$$

To obtain the annual growth response from those monthly data, the model integrates the growth response $g(t)$ over those months that are specified as the start and end months I_0 and I_f. Finally, the annually resolved time series of tree ring width anomalies is obtained by normalising the annual growth response to zero mean and unit variance.

It should be noted that this model does not take into account juvenile tree growth. Real tree ring width data are of course subject to juvenile growth and the effect is usually subtracted from the measured data. Problems arising from this are thus disregarded in the model which we will further address when discussing the results for the model.

To set the model parameters to realistic values, we use an exemplary real-world data set of a local tree ring width index chronology from eastern Canada ($54.2°$ N, $70.3°$ W) that was previously used for regional summer temperature reconstruction (Gennaretti et al., 2014). The quality of this data set with respect to data homogeneity, sample replication, growth coherence, chronology development, and climate signal has been ranked high (Esper et al., 2016). Regional average annual temperature is about $-3.8°$C with average maximum temperatures of $16°$C in July and minimum temperatures of $-23°$C in January. The average annual precipitation sum is 693 mm with a monthly minimum of about 30 mm in February/March and a maximum of about 100 mm in September. This information has been used in order to choose the mean and standard deviations of the climatic input variables and their annual cycles. An overview of the climate input and also the model parameters is given in Table 1. To determine the threshold parameters, we used the Bayesian parameter estimation as suggested in Tolwinski-Ward et al. (2013). The parameters for the Leaky Bucket model are chosen as recommended in Tolwinski-Ward et al. (2011).

Table 1. Climatic input variables and model parameters for the tree ring width model as derived from the eastern Canada data (Gennaretti et al., 2014).

variable	description	value
T_m	mean temperature at site	-3.8° C
P_m	mean annual precipitation sum	693 mm
Φ	latitude of site	54.2°N
$I_{0,f}$	integration period influencing growth	$[1, 12]$
$T_{1,2}$	temperature thresholds for growth	$[5.8, 17]^\circ$C
$M_{1,2}$	soil moisture thresholds for growth	$[0.032, 0.24]$
$M_{\text{min,max}}$	minimum/maximum soil moisture content	$[0.01, 0.76]$
M_0	soil moisture content at start of simulation	0.2
m	runoff parameter	4.886
α	runoff parameter	0.093 month^{-1}
μ	runoff parameter	5.8
d_r	root depth	1000 mm

3.2 Lake sediments

Records from lake sediments are available from many regions worldwide and can provide information about past temperatures and precipitation, depending on the regional boundary conditions and measured proxy (Cohen, 2003). We here model branched glycerol dialkyl glycerol tetraethers (brGDGTs) from lacustrine archives that have been related to air temperatures by using one of the sensor models provided in the PRYSM v2.0 framework (Dee et al., 2018). For this, as model input, a time series of mean annual air temperature T is required.

BrGDGTs are produced by bacteria and their degree of cyclisation and methylation has been related to soil temperatures, lake pH, and also to mean annual air temperatures (Weijers et al., 2007; De Jonge et al., 2014; Russell et al., 2018). The degree of methylation is quantified by a methylation of branched tetraether (MBT) index. The sensor model employs the MBT'_{5ME} index that only uses 5-methyl isomers. In particular, the calibration of Russell et al. (2018) is used, in which the mean annual air temperature is related to the MBT'_{5ME} index via the equation

$$MBT'_{\text{5ME}} = (T + 1.21)/32.42. \tag{7}$$

The archive model then accounts for bioturbation and mixing of the sediments using the TURBO2 model (Trauth, 2013). TURBO2 models the benthic mixing effects on individual sediment particles by assuming that in a mixed layer of a specified thickness on top of a sediment core, instantaneous mixing of the sediment particles occurs, while the rest of the core is not affected by the mixing. In addition to the time series of the sensor model output, the archive model requires three further input parameters, the abundance of the signal carrier over time (abu), the mixed layer thickness over time (mxl), and the number

Table 2. Climatic input variables and model parameters for the lake sediment model as derived from the eastern Canada data (Gennaretti et al., 2014).

variable	description	value
T_m	mean temperature at site	-3.8° C
abu	abundances of input species	200
mxl	mixed layer thickness	4
numb	amount of measured foraminifera	10

of fossil foraminifera on which the proxy signal is measured (numb). The model then returns time series of original and bioturbated abundances and corresponding proxy signatures for the original and a second virtual species that is required to keep the total abundance of all species constant over time. The bioturbated proxy signatures of the first species are then used as final proxy for the mean annual air temperature.

To tune the climatic input variables, we use the climatic setup corresponding to the one used for the tree ring archive in eastern Canada, while for the model parameters, we use typical values for lake sediments that are taken as default in the PRYSM implementation of the lake archive model (Dee et al., 2018). In particular, it should be noted that for the abundances of the input species and the mixed layer thickness, we use constant values over time. The climatic and model input parameters are specified in Table 2.

3.3 Speleothems

Oxygen isotope fractions of speleothems have been shown to provide valuable insights into past climate variability (Wong and Breecker, 2015). We here model the isotopic composition of speleothems by using the speleothem model presented within the PRYSM framework (Dee et al., 2015). This intermediate-complexity model is based on the model in Partin et al. (2013) and requires the mean annual temperature T and the mean of the precipitation-weighted annual isotopic composition of the precipitation $\delta^{18}\mathrm{O}_P$ as input. Additionally, the ground water residence time τ_{gw} has to be specified.

The sensor model covers processes in the karst and the cave, while processes in the soil such as evapotranspiration are neglected. The model filters the $\delta^{18}\mathrm{O}_P$ signal by applying an aquifer recharge model to simulate the storage and thus the mixing of water of different ages in the karst. This process is parameterised by the mean transit time τ_{gw}. The isotopic composition of the cave water is then given as the convolution ($*$) of $\delta^{18}\mathrm{O}_w$ with the impulse response of the aquifer recharge model $g(t) = \tau_{\mathrm{gw}}^{-1} \exp(-t\tau_{\mathrm{gw}})$ for $t > 0$:

$$\delta^{18}\mathrm{O}_d = g(t) * \delta^{18}\mathrm{O}_P. \tag{8}$$

Table 3. Climatic input variables and model parameters for the speleothem δ^{18}O model as derived from the Dongge cave data (Wang et al., 2005).

variable	description	value
T_m	mean temperature at site	15.6° C
$\delta^{18}\mathrm{O}_P$	mean isotopic composition of precipitation	$-8.05\,‰$
τ_{gw}	mean aquifer transit time	5 years

Finally, to obtain the isotopic composition of the flowstone calcite $\delta^{18}\mathrm{O}_c$, the model implements a temperature-dependent fractionation (Wackerbarth et al., 2010)

$$\delta^{18}\mathrm{O}_c = \frac{\delta^{18}\mathrm{O}_d + 1000}{1.03086} \exp\left(\frac{2780}{T_a^2} - \frac{2.89}{1000}\right) - 1000 \tag{9}$$

with the temperature T_a being the decadal average of T that is calculated using a Butterworth filter (Zumbahlen, 2008).

The parameters for the speleothem δ^{18}O model are tuned by using the data of stalagmite DA from Dongge cave (Wang et al., 2005) which is one of the most studied speleothem data sets. Dongge cave is located in Southern China (25.3° N, 108.1° E) at an elevation of 680 m and the data could be related to the history of the Asian monsoon of the last $9,000$ years. The proxy values have an average of $-8.05\,‰$ and the average temperature in the cave is 15.6°C. The mean of the precipitation-weighted isotopic composition of the input is chosen such that the average of the model output data equals the average of the Dongge cave proxy data. There is no information available about the mean transit time, i. e., the time the water has spent inside the karst before entering the cave. We here choose to use an average transit time of five years, which is slightly larger than the average sampling rate of the data (4.2 years). An overview of the input variables and model parameters is given in Table 3.

3.4 Ice cores

Proxy time series from ice cores have been used in a variety of contexts to study past climate variability on different time scales (Jouzel, 2013; Thompson et al., 2005). As for the speleothem δ^{18}O, we use the model for ice core δ^{18}O that is implemented and presented within the PRYSM framework (Dee et al., 2015). The model requires the precipitation-weighted mean annual isotopic composition of the precipitation $\delta^{18}\mathrm{O}_P$ as input. Additional parameters are the mean temperature at the site T_m, the altitude of the glacier z, the mean surface pressure p, the mean accumulation rate at the site A, and the total depth of the core $h_{\max}$ which is given by the time span of the observations times the average accumulation rate.

The sensor model corrects the isotopic composition of the precipitation for the altitude of the glacier by using the relation

$$\delta^{18}\mathrm{O}_{\mathrm{ice}} = \delta^{18}\mathrm{O}_P + \frac{z}{100}a, \tag{10}$$

with a describing the altitude effect. The archive model then accounts for compaction and diffusion within the ice core. First, the density of the core has to be calculated as a function of the depth of the core. For this, an adapted version of the firn densification model by Herron and Langway is used (Herron and Langway, 1980). From the density and the mean temperature

Table 4. Climatic input variables and model parameters for the ice core δ^{18}O model as derived from the Quelccaya ice cap data (Thompson et al., 2013; Yarleque et al., 2018).

variable	description	value
$\delta^{18}O_P$	mean isotopic composition of precipitation	$-3.75\,‰$
T_m	mean temperature at site	-3.99° C
A	average accumulation rate at site	1.15 m w.e./a
p	mean surface pressure at site	1 Atm
z	altitude of site	5670 m
ρ_0	surface density of snow at site	300 kg/m^3
a	altitude effect	$-0.25\,‰/100$ m

T_m, we can then compute the diffusion length σ within the core as a function of the depth h (Johnsen et al., 2000). This, in turn, is used to calculate the final proxy time series $\delta^{18}O_d$ by convolving the isotopic signal of the ice $\delta^{18}O_{\text{ice}}$ with a Gaussian kernel function with standard deviation equalling the diffusion length at a given depth,

$$\delta^{18}O_d = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-h^2}{2\sigma^2}\right) * \delta^{18}O_{\text{ice}}, \tag{11}$$

where the convolution is again denoted by the asterisk ($*$).

To tune the climatic input variables and model parameters of the ice core δ^{18}O model, we use an exemplary real world data set of the Quelccaya ice cap (Thompson et al., 2013) which is one of the most studied ice core data sets outside the polar regions. The Quelccaya ice cap is located in the Peruvian Andes ($13^\circ 56'$ S, $70^\circ 50'$ W) at an altitude of 5670 m above sea level. The average accumulation rate is 1.15 m water equivalent per year and the mean $\delta^{18}O_{\text{ice}}$ is $-17.9\,‰$. The average annual temperature over the last decade at the Quelccaya ice cap is given as $T_m = -3.99^\circ$ C (Yarleque et al., 2018). An overview of the climatic input variables and the model parameters can be found in Table 4.

4 Input data

We now introduce the data sets that we use as input for the proxy system models. We first consider two stationary stochastic processes, namely Gaussian white noise (GWN) and an autoregressive process of order 1 (AR(1) process) to evaluate whether such input can lead to the detection of dynamical anomalies in the proxy time series. Then, we consider non-stationary versions of the two well-known Rössler (ROS) and Lorenz (LOR) systems. For all those processes, time series of length $N = 1{,}000$ are independently created to describe temperature, precipitation and precipitation-weighted oxygen isotope fractions as detailed below, where the precipitation is proportional to negative temperature. Additionally, we use data from the last millennium reanalysis project (Hakim et al., 2016; Tardif et al., 2019). Then, as for the tree ring width model monthly input is required, an annual cycle is added to the temperature and precipitation data. The amplitude of the annual cycle is chosen according to the climatic boundary conditions of the tree ring width index time series from eastern Canada presented in Sec. 3. Finally, yearly

means for temperature and sums for precipitation are calculated for those models that require yearly input. The input time series are normalised to zero mean and unit standard deviation, and for each model, the mean is adjusted to the corresponding climatic boundary conditions.

4.1 Gaussian white noise

For the case of GWN, we draw N data points independently at random from the probability distribution

$$p_G(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{12}$$

where $\sigma = 1$ is the standard deviation and $\mu = 0$ is the mean of the distribution. We do not expect to detect significant dynamical anomalies from time series created in this way as the process is stationary.

4.2 Autoregressive process of order 1

For the AR(1) process, we create a time series of length N by using the relation

$$x(t) = \alpha x(t-1) + \epsilon_t \tag{13}$$

with ϵ_t a Gaussian random variable with zero mean and constant standard deviation $\sigma_\epsilon = 0.5$. The scaling factor is given as $\alpha = 0.7$ corresponding to the approximate value that we obtained when fitting an AR(1) process to the tree ring width data from eastern Canada (Gennaretti et al., 2014). As initial condition, we use $x(0) = 0.3$. The resulting time series is normalised to zero mean and unit standard deviation. As for GWN, we do not expect to detect significant dynamical anomalies in the corresponding time series.

4.3 Non-stationary Rössler system

In a next step, we use data from a non-stationary version of the Rössler system which exhibits non-trivial and rich cascades of bifurcations despite its rather simple attractor topology. The Rössler system is defined by the set of ordinary differential equations (ODEs) (Rössler, 1976)

$$\begin{aligned} \dot{x}(t) &= -y(t) - z(t) \\ \dot{y}(t) &= x(t) + a y(t) \\ \dot{z}(t) &= b(t) + z(t)(x(t) - c). \end{aligned} \tag{14}$$

It should be noted that this model is not meant to simulate real world climate dynamics.

We here use the two fixed parameters $a = 0.2$ and $c = 5.7$ and a time varying parameter $b(t) = b_0 + \Delta b(t - t_0)$ with $b_0 = 0.02$ and $\Delta b = 0.001$. We numerically solve this system of ODEs with a temporal resolution of $\Delta t = 0.1$ for times in the range $[0, 730]$, discard the first 300 points and then use every seventh point of the remaining time series of the x-component to end up with a time series of length $N = 1,000$. As initial conditions we use $x(0) = 0.5$, $y(0) = 0$ and $z(0) = 0$. Again, we normalise

the time series to have zero mean and unit standard deviation. The resulting time series and the corresponding Feigenbaum diagram of the stationary system (i. e., $b = \text{const.}$) are shown in the Supplementary Material in Fig. S1.

From the Feigenbaum diagram, it becomes clear that we expect to detect alternating periods of lower and higher dimensional dynamics in the time series. In particular, we stress that we do not expect to detect the bifurcation points but periods of outstandingly high- or low-dimensional dynamics in between them as we use random shuffling surrogates for the pointwise significance test.

4.4 Non-stationary Lorenz system

The Lorenz system shows a more complicated attractor topology than the Rössler system and has been originally introduced as a simple toy model for atmospheric convection processes (Lorenz, 1963). It is given by the following set of ODEs (Lorenz, 1963):

$$\begin{aligned} \dot{x}(t) &= a(y(t) - x(t)) \\ \dot{y}(t) &= x(t)(b(t) - z(t)) - y(t) \\ \dot{z}(t) &= x(t)y(t) - cz(t). \end{aligned} \tag{15}$$

We here use the setting studied in Donges et al. (2011a) and correspondingly fix the parameters $a = 10.0$ and $c = 8/3$, while the parameter b is again varied over time as $b(t) = b_0 + \Delta b(t - t_0)$ with $b_0 = 160.0$ and $\Delta b = 0.02$. We numerically solve this system of ODEs with a temporal resolution of $\Delta t = 0.05$ for times in the range $[0, 500]$ and use every fifth point of the x-component of the system to end up with a univariate time series of length $N = 1,000$. As initial conditions we use $x(0) = 10.0$, $y(0) = 10.0$ and $z(0) = 10.0$. Again, we normalise the time series to have zero mean and unit standard deviation.

The stationary Lorenz system has been found to exhibit a shift from periodic to chaotic dynamics at $b = 166.0$ (Barrio and Serrano, 2007), while for the transient system as described above, transitions could be detected at $b = 161.0$ and $b = 166.5$ (Donges et al., 2011a). Thus, we expect to detect regimes of more periodic dynamics for $b < 166.5$ and of more chaotic dynamics for $b \geq 166.5$ in terms of significantly high and low values of the network transitivity, respectively.

4.5 Last millennium reanalysis data

Finally, we consider reconstructed temperature and precipitation data of the years $501 - 2000$ AD from the last millennium reanalysis project version 2 (Hakim et al., 2016; Tardif et al., 2019), which combines information from general circulation models and from proxy measurements using palaeoclimate data assimilation. As no information about the isotopic composition of the precipitation is available, we only consider the models for tree ring width and lake sediments with this input. From the available global gridded data, we use the standardised, thus, unit-less, ensemble average time series of temperature and precipitation from the grid points with the coordinates closest to those for which we have calibrated the tree and the lake model, that is, at ~~(~~54° N, 70° W~~)~~.

5 Results

5.1 ~~Time series properties~~

~~As a first step to evaluate the results, we take a look at the properties of the time series generated by~~ Before studying the different proxy system models~~and compare them to the~~ , we note that applying some general filtering techniques such as moving average filtering or exponential smoothing to the described stochastic input time series ~~. Supplementary Figures S2 to S4 display the annually sampled input time series for temperature, precipitation and isotopic compositions and the corresponding output time series of the four proxy system models for the five input scenarios of GWN, the AR(1)process, the non-stationary Rössler system, the non-stationary Lorenz system, and the last millennium reanalysis data. The expected low-pass filter effects of the speleothem, ice and lake models due to the cave residence time, diffusion, and bioturbation, respectively, are directly visible in the~~ on the one hand and the deterministic input time series on the other hand changes the results of the windowed recurrence network analysis in different ways (not shown here for brevity). For the stochastic time series, ~~while for the tree model, such an effect is neither expected nor visible. Also, it should be noted that the tree ring model with the parameters as specified in Table 1 seems to primarily respond to temperature rather than to precipitation, meaning that the limiting factor for tree growth in eastern Canada is temperature, which is ecologically reasonable.~~

~~Normalised histograms, autocorrelation functions (acf) and estimated power spectral densities (psd) of the different input and proxy system model output time series for GWN, AR(1), ROS, LOR, and LMR (top to bottom). The input variables are denoted in grey, the tree model output in green, the lake model output in blue, the speleothem model output in red, and the ice model output in magenta. For further evaluation, we standardise all time series to zero mean and unit variance and examine some properties of the different input and output series. The left panels of Fig. **??** show the normalised histograms of the input and output variables. To quantify differences in the histograms, we consider the skewness of the distributions of the different time series (see Supplementary Table S1). We observe that the resulting time seriesfrom the tree model all show significant deviations in skewness from the input data, which is probably due to the thresholding of the growth response functions for temperature and soil moisture. For the other proxy models, we observe a significant influence of the model on the skewness of the output for those inputs that are not stochastic. This might either be related to the smaller confidence bounds due~~ the filtering mostly changes the variability of the calculated network transitivity. In particular, filtering an AR(1) process with moving averages produces extended and additional areawise significant patches of high values of the network transitivity. In turn, for deterministic input time series, the variability of the network transitivity remains rather robust under filtering. Furthermore, the results for the network transitivity remain similar when adding white noise to the different ~~surrogate generation, or hint to a different reaction of the models to different distributions of the input data~~input time series up to signal-to-noise ratios of 10 and show increasingly different variability for lower signal-to-noise ratios.

~~Finally, we have a look at the autocorrelation functions and the power spectral densities of the different input and proxy model~~ We then turn to the output time series ~~. The middle and right panels of Fig. **??** show the resulting autocorrelation functions and estimates of the power spectral density obtained using the Welch method (Welch, 1967). As before, we observe that the autocorrelation function and power spectral density of the tree ring model output closely follow the respective temperature~~

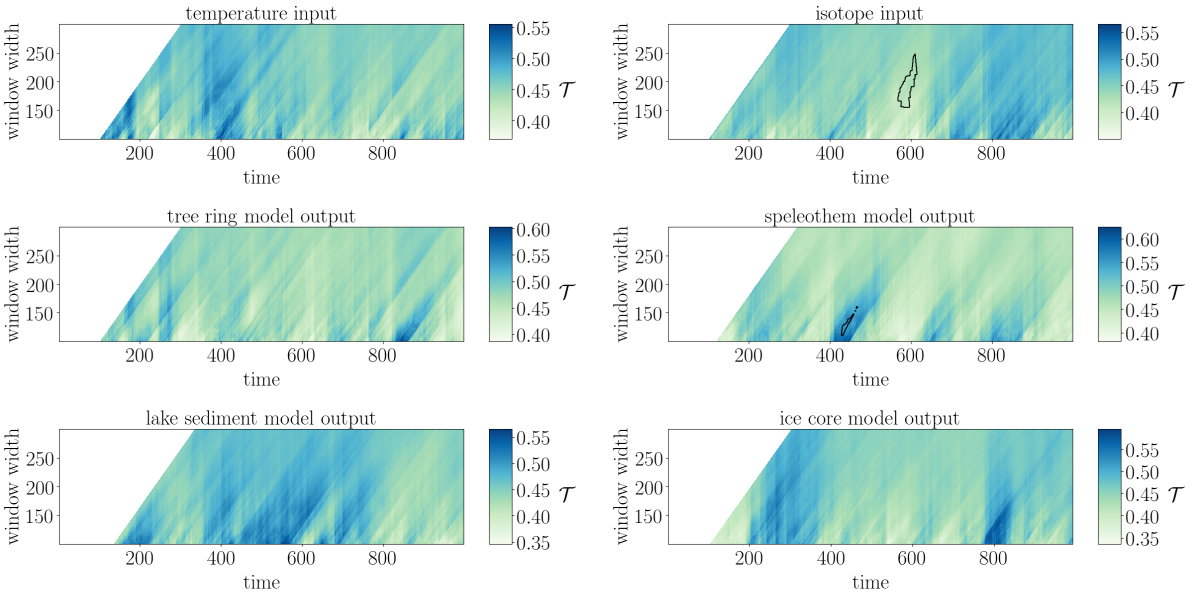

Figure 1. Network transitivity (colour-coded) for GWN model in- and output with areawise significance test (contours).

~~input, while the speleothem model output shows the expected loss of power in the higher frequencies. Similarly, a loss of power in the higher frequencies can be observed in the ice core model output but not as pronounced as for the speleothem model output. The same holds for the lake sediment model output except for the Rössler and Lorenz scenarios, where the lake model output shows more power in the higher frequencies than the corresponding input . This is most likely related to the different time-scales of variability present in those time seriesin comparison with the others and the particular choice of mixed layer thickness in the lake sediment model, mxl = 4.~~

5.1 ~~Windowed recurrence network analysis~~

~~In a second step, we analyse the different proxy~~ of the different proxy system models which combine linear and non-linear transformations and filtering of the input time series. The different input and model output time series ~~using wRNA as introduced in Sec. 2 and compare the results to those of the input time series~~and some remarks on their properties can be found in the Supplementary Material. Figures 1 to 5 show the resulting network transitivity over time for the different ~~input~~ time series. ~~The analysis is combined with the areawise significance test and areawise~~ Areawise significant points indicating anomalously low or high values of the network transitivity are highlighted by ~~the contours.~~

contours. For Gaussian white noise (Fig. 1), we ~~do not get any areawise significant points in the input temperature and also the output of the tree ring width and the lacustrine archive models do not exhibit any areawise significant points~~observe the expected changes in the variability of the network transitivity for stochastic input when processing the time series through different filters. For the isotopic composition of the precipitation, we find a patch of areawise significant anomalously low

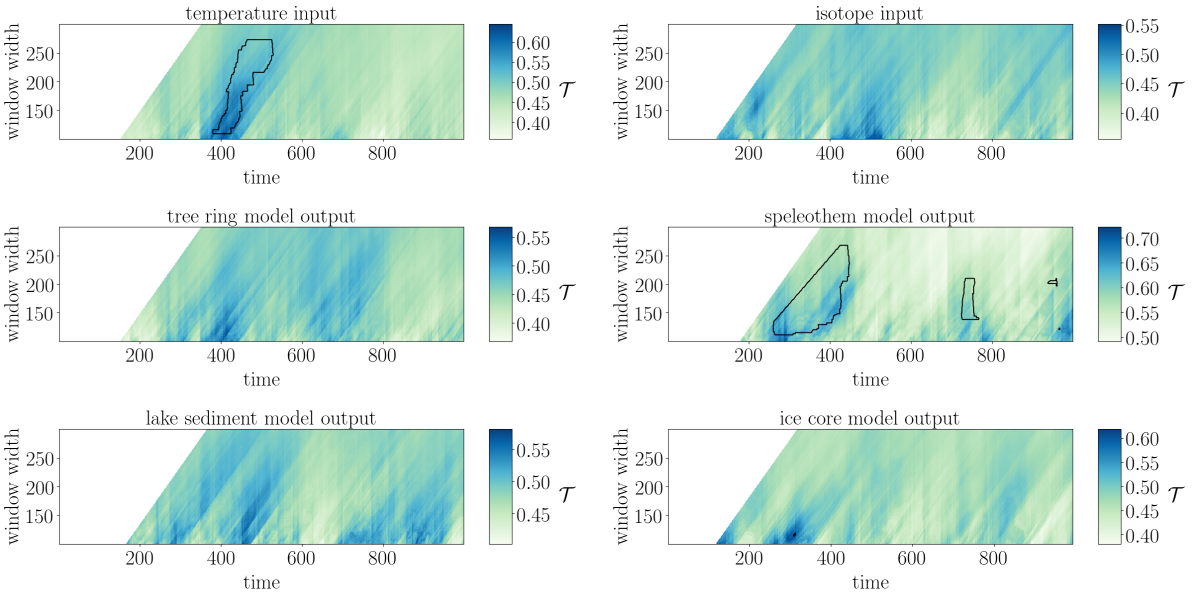

Figure 2. Same as in Fig. 1, but for AR(1) model input.

values of the network transitivity, which is a random ~~artefact~~ artifact in this particular realisation of GWN. The output of the speleothem and the ice core model do not show this anomaly, i. e., the processing through the model causes the initially significant (false positive) points to be missed ~~. Additionally, the speleothem output shows a small patch of falsely identified areawise significant high values of the network transitivity for small window width after $t = 400$~~and creates an additional patch of areawise significant network transitivity in the case of the speleothem model.

In the case of the AR(1) process (Fig. 2), we observe an areawise significant patch of anomalously high values of network transitivity for the input temperature that is not apparent in the tree ring width and lake sediment model output. For the isotope input, we do not find any areawise significant points, while the speleothem model shows two large and two small falsely identified significant patches of network transitivity. That is, when recalling the results for the filtering of the AR(1) process, we find that the speleothem model seems to particularly involve a way of filtering that produces patches of areawise significant high values of the network transitivity while the other models rather change the variability of the stochastic input.

For the non-stationary Rössler system (Fig. 3), the input time series show areawise significant patches of low network transitivity values in the parameter range $b \in [0.30, 0.35]$ and of high network transitivity for $b \in [0.55, 0.57]$. The isotope input has an additional small areawise significant patch around $b = 0.15$. ~~The~~ As expected for deterministic input, we observe a much better agreement of the variability of the network transitivity for the input and model output time series than for the two stochastic input and corresponding model output time series. Still, for the tree ring and lake sediment model~~outputs again show no areawise significant points, i. e., the model processing~~ , the processing through the model prevents the significant points to

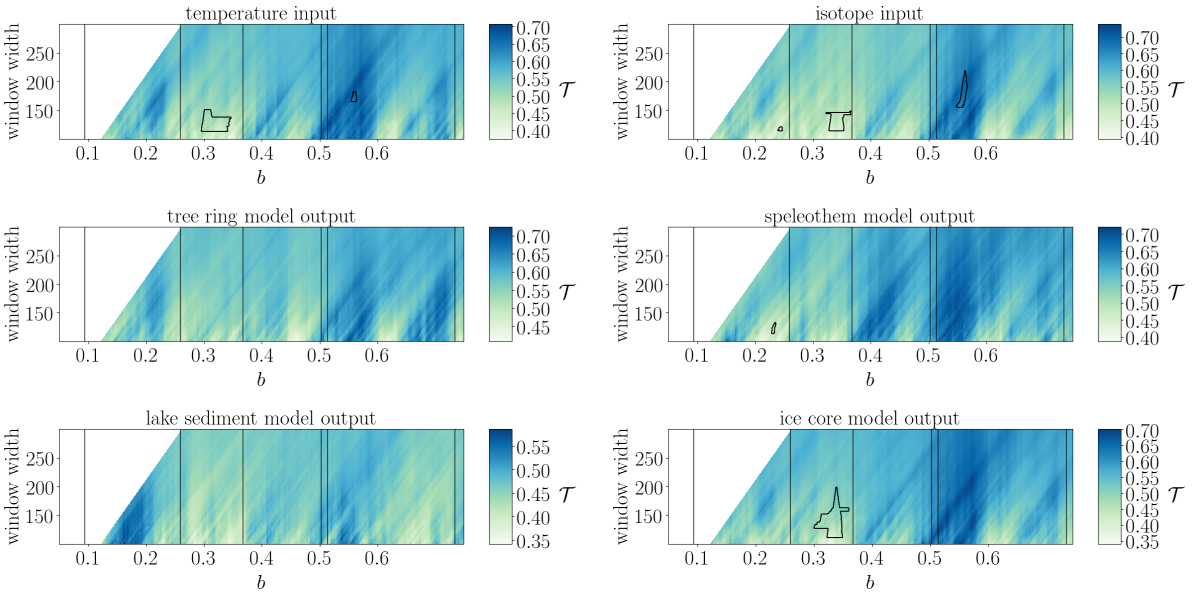

Figure 3. Same as in Fig. 1, but for non-stationary Rössler system input. The vertical lines mark the expected bifurcation points identified in the Feigenbaum diagram in Supplementary Fig. S1.

be detected. The speleothem model output only shows the small areawise significant patch at $b = 0.15$ but not the others, while the output from the ice core model displays the areawise significant patch for $b \in [0.30, 0.35]$.

The input time series of the non-stationary Lorenz system (Fig. 4) exhibits a large patch of areawise significant low values of the network transitivity for high values of b and a small patch of areawise significant high values of the network transitivity around $b = 163.2$. The isotope input shows another small significant patch around $b = 166.6$. The resulting network transitivity for all proxy system model output time series except the one from the lake model reproduce the shift in network transitivity from higher to lower values after $b = 166.5$ as expected for this deterministic input. The tree ring width model output reproduces the large and the small patch of significant values of the network transitivity and additionally exhibits another areawise significant patch of high transitivity values close to the small patch. ~~The lake sediment output does not show any areawise significant points. The speleothem output shows a large patch of areawise significant high transitivity values for~~ In the speleothem model output, not the low values of the network transitivity at the end of the time series but rather the high values for $b \in [162.5, 166.5]$ ~~and, like this, captures the small patch from the isotope input, but also exhibits a large number of falsely identified significant points, while the large patch of low transitivity values from the input is not found to be areawise significantfor the model output except for a very small part of it~~are classified as areawise significant, which might again be related to the particular filtering in the speleothem model favouring areawise significant high values of the network transitivity. The ice core model output only shows a small patch for high values of b and low values of the window width and misses most of the areawise significant points apparent in the isotope input.

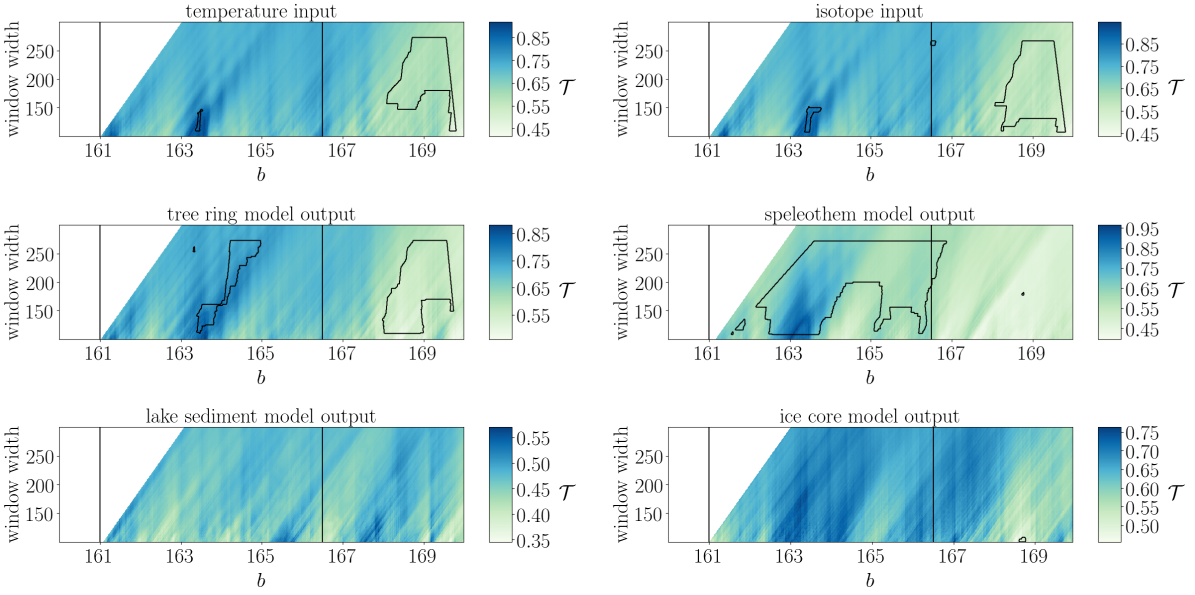

Figure 4. Same as in Fig. 1, but for non-stationary Lorenz system input. The vertical lines denote the transitions at $b = 161.0$ and $b = 166.5$ as detected in Donges et al. (2011a).

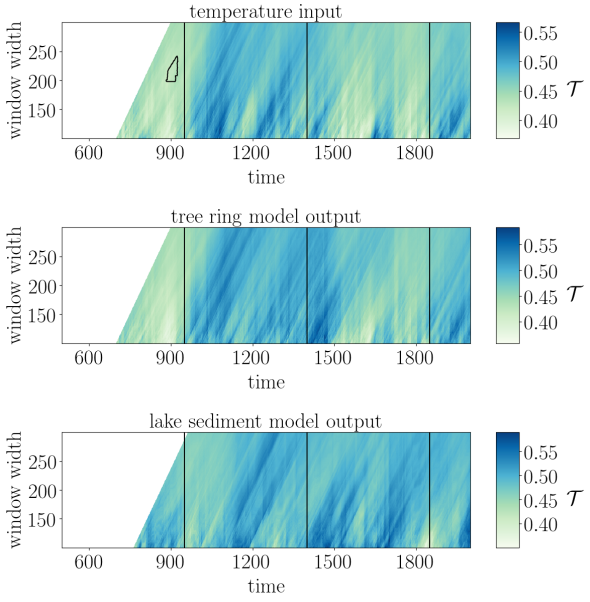

Figure 5. Network transitivity (colour-coded) for last millennium reanalysis input at (54° N, 70° W) and corresponding output with areawise significance test (contours). The vertical lines denote the approximate onset dates of the Medieval Climate Anomaly, the Little Ice Age, and the industrial age, respectively.

Table 5. Fraction of missed and falsely identified significant points in the different proxy models with respect to the corresponding reference input variables.

archive	reference		GWN	AR(1)	ROS	LOR	LMR
tree rings	temperature	missed	0.000	0.056	0.012	0.003	0.004
		falsely identified	0.000	0.000	0.000	0.064	0.000
lake sediments	temperature	missed	0.000	0.056	0.011	0.083	0.004
		falsely identified	0.000	0.000	0.000	0.000	0.000
speleothems	isotopes	missed	0.010	0.000	0.011	0.106	–
		falsely identified	0.001	0.105	0.001	0.289	–
ice cores	isotopes	missed	0.010	0.000	0.006	0.109	–
		falsely identified	0.000	0.000	0.011	0.001	–

For the last millennium reanalysis data temperature input (Fig. 5), we find a patch of areawise significant values of anomalously low values of the network transitivity around the year 900 AD, roughly coinciding with the onset of the European Medieval Climate Anomaly (MCA, denoted by a vertical line at 950 AD). The other vertical lines depict the approximate onset of the European Little Ice Age (LIA) at 1400 AD, and the onset of the industrial age at 1850 AD. It should be noted that the timings and imprints of these episodes are known to exhibit substantial regional differences (Franke and Donner, 2017), thus, the reference lines are only for orientation. The model output time series do not show any areawise significant points even though the network transitivity of the ~~tree ring model output~~ model output (and particularly the tree ring width model output) varies similarly as the network transitivity of the temperature input with higher values ~~corresponding to lower-dimensional dynamics during the Medieval Climate Anomaly~~ during the MCA and lower values ~~corresponding to higher-dimensional dynamicsduring the Little Ice Age~~during the LIA. The MCA has often be attributed to more stable climate conditions while the LIA has often been associated with more variable climate conditions even though this imprint has varied locally and has been mainly discussed for Europe. Thus, given the theoretical relation between network transitivity and the dimensionality of the system's dynamics (Donner et al., 2011a), we tentatively conclude that the higher values of the network transitivity during the MCA and the lower values of the network transitivity during the LIA reflect a lower/higher dimensional dynamics of the system at this particular location in terms of the complexity of temporal variations rather than just a change in variance. In terms of the time series properties for the different periods, we indeed additionally observe an increase in variance of the time series from the MCA to the LIA, which is very likely also reflected in the different recurrence networks and, thus, the resulting network transitivity. We note that this non-stationarity in variance along with the MCA-LIA transition in the European/North Atlantic sector has also been reported as being reflected in other non-linear characteristics, which have been previously interpreted as a hallmark of some dynamical anomaly (Franke and Donner, 2017; Schleussner et al., 2015).

In order to quantify the effect of the proxy system models as ~~nonlinear~~ non-linear filters of the input signal on the detection of areawise significant points, Table 5 displays the fraction of falsely identified and missed significant points in the different

proxy models for the different input time series. We observe that for the tree ring width model, the lacustrine sediment model, and the ice core model missed significant points are more common than falsely identified significant points, with the exception of the Lorenz input for the tree ring width model. For the speleothem model, both falsely identified and missed significant points occur. In fact, this model shows particularly high fractions of falsely identified and missed significant points, while overall, the ice core model seems to best reproduce the results of the input in terms of the classification of areawise significant points.

Taken together, the previous findings should raise awareness in the context of future applications of wRNA to palaeoclimate proxy time series, suggesting that interpretations of results obtained for individual records only may not be sufficiently robust for drawing substantiated conclusions. From a practical perspective, this calls for combining different time series from different proxies and/or archives from the same region to obtain further climatological knowledge from such kind of analysis (Donges et al., 2015a; Franke and Donner, 2017).

6 Discussion and conclusions

In this paper, we have studied the suitability of windowed recurrence network analysis (wRNA) for detecting dynamical anomalies in time series from different proxy archives. For this, we used proxy system models that simulate the formation of proxy archives, such as for example tree rings, lake sediments, speleothems, and ice cores, given some climatic input variables like temperature and precipitation. We created artificial input time series with different properties and additionally used temperature and precipitation data from the last millennium reanalysis project (Hakim et al., 2016; Tardif et al., 2019). We then processed the input time series through the different proxy models and ~~compared time series properties such as the autocorrelation function for the input and model output time series. Finally, we~~ contrasted the results of wRNA for the different input and model output time series.

We ~~observed the expected loss of power in the higher frequencies for the lake sediment, speleothem, and ice core model output time series (due to intrinsic smoothing /integration processes inherent to the physical system)and noted changes in the skewness of the model output time series for the deterministic and real-world inputscenarios. The tree ring width model showed changes in skewness also for the stochastic input~~first compared the results of wRNA for stochastic and deterministic input to corresponding results when filtering the time series using moving average filtering and exponential smoothing (results not shown). We found that filtering alters the variability of the network transitivity with a bias towards additional and extended patches of areawise significant high transitivity values for stochastic input. ~~For the wRNA~~The network transitivity of deterministic input seems to be rather robust under such filtering. When processing the input time series through the different proxy system models, these differences for stochastic and deterministic input were also apparent. In terms of areawise significant anomalies of the network transitivity, we found that time series of tree ring width and brGDGTs in lake sediments have problems with missing areawise significant points, while the isotopic composition of speleothems also exhibits falsely identified significant points of high values of the network transitivity probably related to the bias towards higher values of the network transitivity

due to the particular filtering in the model. Time series of the isotopic composition of ice yield comparable results to the corresponding input but also sometimes miss significant points.

~~For the stochastic input time series as for example the isotope input of GWN, we found some areawise significant artefacts in single realisations. To improve the reliability of the results for the these processes, more realisations should be considered to confirm~~ Taken together, our results show the need of further studying the effects of different filtering mechanisms on the results of the wRNA in order to interpret the results and ~~to exclude the influence of random artefacts. As we applied an areawise significance test to identify dynamical anomalies which reduces the number of false positives in the analysis results, this can also reduce the number of true positives and increase the number of false negatives independent of whether considering model input or output time series. In this regard, we also observed that time series with stronger autocorrelations in most cases show higher correlations for~~ draw reliable conclusions when analysing real-world data. This particularly concerns data from speleothems as they have been studied quite often using windowed recurrence analysis (e. g., Donges et al. (2015a); Eroglu et al. (2016)) and filtering effects seem to have a non-negligible influence on the results for this type of archive. Thus, the role of the ~~wRNA results in the different domains and, thus, have more restrictive bounds of the areawise significance test (not shown).~~ model parameters that control the filtering (here, the mean aquifer transit time) should be studied in more detail and in the context of the real-world applications, should be taken into account for estimates of these parameters for the particular data to better interpret corresponding results of the network transitivity.

~~Falsely identified significant points in the model output, particularly present in the results for the speleothem model, might for the stochastic processes be~~ Still, we want to stress that even then, providing a general recipe for interpreting the resulting network transitivity is hardly possible in the (palaeo)climate context. Climate-related interpretations always vary depending on the location and thus, local boundary conditions have to be taken into account. As motivated in Section 2, the network transitivity has been related to the ~~strong filtering within the model. For the Lorenz input, the processing through the speleothem model leads to large fractions of both,falsely identified and missed significant points as the high~~ dynamical regularity of the variations in the analysed time series (e.g., Donges et al. (2011b)) with higher values of the network transitivity ~~before the transition at $b = 166.5$ are classified as significant, while in the input, the low values~~ corresponding to less irregular variability and vice versa. This is in accordance with the interpretation of the network transitivity ~~after $b = 166.5$ are classified as significant. Missed significant points in the model output may also~~ as an indicator of the dimensionality of the system's dynamics. In this regard, detected anomalies in the network transitivity could be related to ~~the processing through the model destroying some structure within the data such as the thresholding process in the tree ring width model~~some tipping point, but do not have to be.

Future work should also include the study of alternative proxy system models within this framework. Results of proxy system models for both, the same proxies (but with more detailed systemic understanding of the formation of the proxies, as for example a tree ring width model accounting for juvenile growth of the trees) and different proxy variables (in particular, for other lacustrine proxy variables), will complement the improved understanding of the suitability of wRNA for these types of time series and will advance the interpretation of the corresponding results. Also, sensitivity studies for the different model pa-

rameters are of interest to better interpret results obtained with wRNA for a given real-world data set. This concerns particularly the mean aquifer transit time of the speleothem model.

For the stochastic input time series as for example the isotope input of GWN, we found some areawise significant artefacts in single realisations. To improve the reliability of the results for the these processes, more realisations should be considered to confirm the results and to exclude the influence of random artefacts. As we applied an areawise significance test to identify dynamical anomalies which reduces the number of false positives in the analysis results, this can also reduce the number of true positives and increase the number of false negatives independent of whether considering model input or output time series. In this regard, we also observed that time series with stronger autocorrelations in most cases show higher correlations for the wRNA results in the different domains and, thus, have more restrictive bounds of the areawise significance test (not shown).

Additionally, the study of properties of the analysed time series can serve as starting point to judge the suitability of wRNA for other data to be analysed. In particular, the effect of filtering the time series with different non-linear filters prior to the analysis as done within the different proxy system models can and should be studied more systematically. ~~In particular~~Also, the theory of non-linear observability might give an interesting new perspective on this as the filtering can be seen as creating a new observable, and the choice of observable has already been shown to influence results of recurrence quantification analysis and recurrence network analysis (Portes et al., 2014, 2019). Moreover, further systematically studying the relation between the autocorrelation of a time series and the resulting network properties might yield additional information on the role of the different archives for ~~wRNA.~~ the results of wRNA.

Code availability. Exemplary Python code for the windowed recurrence network analysis including the areawise significance test is available at https://gitlab.pik-potsdam.de/lekscha/awsig. A comprehensive implementation of recurrence network analysis can be found in the open-source Python package `pyunicorn` (Donges et al., 2015b), which can be found at https://github.com/pik-copan/pyunicorn.

Author contributions. JL and RVD designed the research. JL performed the numerical experiments and data analyses. JL and RVD discussed the results and wrote the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. This work has been financially supported by the German Federal Ministry for Education and Research (BMBF) via the BMBF Young Investigators Group *CoSy-CC*^{2} - *Complex Systems Approaches to Understanding Causes and Consequences of Past, Present and Future Climate Change* (grant no. 01LN1306A) and the Studienstiftung des deutschen Volkes. Calculations have been performed with the help of the Python package `pyunicorn` (Donges et al., 2015b).

References

Abarbanel, H. D. I., Brown, R., Sidorowich, J. J., and Tsimring, L. S.: The analysis of observed chaotic data in physical systems, Reviews of Modern Physics, 65, 1331–1392, https://doi.org/10.1103/revmodphys.65.1331, 1993.

Barrio, R. and Serrano, S.: A three-parametric study of the Lorenz model, Physica D: Nonlinear Phenomena, 229, 43 – 51, https://doi.org/10.1016/j.physd.2007.03.013, 2007.

Bradley, R. S.: Paleoclimatology (Third Edition), Academic Press, third edition edn., https://doi.org/doi.org/10.1016/C2009-0-18310-1, 2015.

Cohen, A.: Paleolimnology: The History and Evolution of Lake Systems, Oxford University Press, 2003.

De Jonge, C., Hopmans, E. C., Zell, C. I., Kim, J.-H., Schouten, S., and Damsté, J. S. S.: Occurrence and abundance of 6-methyl branched glycerol dialkyl glycerol tetraethers in soils: Implications for palaeoclimate reconstruction, Geochimica et Cosmochimica Acta, 141, 97 – 112, https://doi.org/10.1016/j.gca.2014.06.013, 2014.

Dee, S., Emile-Geay, J., Evans, M. N., Allam, A., Steig, E. J., and Thompson, D.: PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems, Journal of Advances in Modeling Earth Systems, 7, 1220–1247, https://doi.org/10.1002/2015MS000447, 2015.

Dee, S. G., Russell, J. M., Morrill, C., Chen, Z., and Neary, A.: PRYSM v2.0: A Proxy System Model for Lacustrine Archives, Paleoceanography and Paleoclimatology, 33, 1250–1269, https://doi.org/10.1029/2018PA003413, 2018.

Donges, J. F., Donner, R. V., Rehfeld, K., Marwan, N., Trauth, M. H., and Kurths, J.: Identification of dynamical transitions in marine palaeoclimate records by recurrence network analysis, Nonlinear Processes in Geophysics, 18, 545–562, https://doi.org/10.5194/npg-18-545-2011, 2011a.

Donges, J. F., Donner, R. V., Trauth, M. H., Marwan, N., Schellnhuber, H.-J., and Kurths, J.: Nonlinear detection of paleoclimate-variability transitions possibly related to human evolution, Proceedings of the National Academy of Sciences, 108, 20 422–20 427, https://doi.org/10.1073/pnas.1117052108, 2011b.

Donges, J. F., Donner, R. V., Marwan, N., Breitenbach, S. F. M., Rehfeld, K., and Kurths, J.: Non-linear regime shifts in Holocene Asian monsoon variability: potential impacts on cultural change and migratory patterns, Climate of the Past, 11, 709–741, https://doi.org/10.5194/cp-11-709-2015, https://www.clim-past.net/11/709/2015/, 2015a.

Donges, J. F., Heitzig, J., Beronov, B., Wiedermann, M., Runge, J., Feng, Q. Y., Tupikina, L., Stolbova, V., Donner, R. V., and Marwan, N., e. a.: Unified functional network and nonlinear time series analysis for complex systems science: The pyunicorn package, Chaos, 25, 113 101, https://doi.org/10.1063/1.4934554, 2015b.

Donner, R. V., Zou, Y., Donges, J. F., Marwan, N., and Kurths, J.: Ambiguities in recurrence-based complex network representations of time series, Physical Review E, 81, 015 101, https://doi.org/10.1103/physreve.81.015101, 2010a.

Donner, R. V., Zou, Y., Donges, J. F., Marwan, N., and Kurths, J.: Recurrence networks — a novel paradigm for nonlinear time series analysis, New Journal of Physics, 12, 033 025, https://doi.org/10.1088/1367-2630/12/3/033025, 2010b.

Donner, R. V., Heitzig, J., Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: The geometry of chaotic dynamics — a complex network perspective, The European Physical Journal B, 84, 653–672, https://doi.org/10.1140/epjb/e2011-10899-1, 2011a.

Donner, R. V., Small, M., Donges, J. F., Marwan, N., Zou, Y., Xiang, R., and Kurths, J.: Recurrence-based time series analysis by means of complex network methods, International Journal of Bifurcation and Chaos, 21, 1019–1046, https://doi.org/10.1142/s0218127411029021, 2011b.

Eroglu, D., McRobie, F. H., Ozken, I., Stemler, T., Wyrwoll, K.-H., Breitenbach, S. F. M., Marwan, N., and Kurths, J.: See-saw relationship of the Holocene East Asian-Australian summer monsoon, Nature Communications, 7, 12 929, https://doi.org/10.1038/ncomms12929, 2016.

Esper, J., Krusic, P. J., Ljungqvist, F. C., Luterbacher, J., Carrer, M., Cook, E., Davi, N. K., Hartl-Meier, C., Kirdyanov, A., Konter, O., Myglan, V., Timonen, M., Treydte, K., Trouet, V., Villalba, R., Yang, B., and Büntgen, U.: Ranking of tree-ring based temperature reconstructions of the past millennium, Quaternary Science Reviews, 145, 134 – 151, https://doi.org/https://doi.org/10.1016/j.quascirev.2016.05.009, 2016.

Evans, M., Tolwinski-Ward, S., Thompson, D., and Anchukaitis, K.: Applications of proxy system modeling in high resolution paleoclimatology, Quaternary Science Reviews, 76, 16 – 28, https://doi.org/https://doi.org/10.1016/j.quascirev.2013.05.024, 2013.

Franke, J. G. and Donner, R. V.: Dynamical anomalies in terrestrial proxies of North Atlantic climate variability during the last 2 ka, Climatic Change, 143, 87–100, https://doi.org/10.1007/s10584-017-1979-z, 2017.

Fraser, A. M. and Swinney, H. L.: Independent coordinates for strange attractors from mutual information, Physical Review A, 33, 1134–1140, https://doi.org/10.1103/PhysRevA.33.1134, 1986.

Fritts, H. C.: Tree rings and climate, Academic Press, https://doi.org/https://doi.org/10.1016/B978-0-12-268450-0.X5001-0, 1976.

Gennaretti, F., Arseneault, D., Nicault, A., Perreault, L., and Bégin, Y.: Volcano-induced regime shifts in millennial tree-ring chronologies from northeastern North America, Proceedings of the National Academy of Sciences, 111, 10 077–10 082, https://doi.org/10.1073/pnas.1324220111, 2014.

Goswami, B., Boers, N., Rheinwalt, A., Marwan, N., Heitzig, J., Breitenbach, S. F. M., and Kurths, J.: Abrupt transitions in time series with uncertainties, Nature Communications, 9, 48, https://doi.org/10.1038/s41467-017-02456-6, 2018.

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The last millennium climate reanalysis project: Framework and first results, Journal of Geophysical Research: Atmospheres, 121, 6745–6764, https://doi.org/10.1002/2016JD024751, 2016.

Herron, M. M. and Langway, C. C.: Firn Densification: An Empirical Model, Journal of Glaciology, 25, 373–385, https://doi.org/10.3189/S0022143000015239, 1980.

Huang, J., van den Dool, H. M., and Georgarakos, K. P.: Analysis of Model-Calculated Soil Moisture over the United States (1931–1993) and Applications to Long-Range Temperature Forecasts, Journal of Climate, 9, 1350–1362, https://doi.org/10.1175/1520-0442(1996)009<1350:AOMCSM>2.0.CO;2, 1996.

Johnsen, S. J., Clausen, H. B., Cuffey, K. M., Hoffmann, G., Schwander, J., and Creyts, T.: Diffusion of stable isotopes in polar firn and ice : the isotope effect in firn diffusion, Physics of Ice Core Records, 159, 121–140, https://doi.org/10.7916/D8KW5D4X, 2000.

Jouzel, J.: A brief history of ice core science over the last 50 yr, Climate of the Past, 9, 2525–2547, https://doi.org/10.5194/cp-9-2525-2013, 2013.

Kantz, H. and Schreiber, T.: Nonlinear time series analysis, Cambridge University Press, 2 edn., 2004.

Kennel, M. B., Brown, R., and Abarbanel, H. D. I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction, Physical Review A, 45, 3403–3411, https://doi.org/10.1103/PhysRevA.45.3403, 1992.

Lekscha, J. and Donner, R. V.: Phase space reconstruction for non-uniformly sampled noisy time series, Chaos, 28, 085 702, https://doi.org/10.1063/1.5023860, 2018.

Lekscha, J. and Donner, R. V.: Areawise significance tests for windowed recurrence network analysis, Proceedings of the Royal Society A, 475, 20190 161, https://doi.org/10.1098/rspa.2019.0161, 2019.

Lorenz, E. N.: Deterministic Nonperiodic Flow, Journal of the Atmospheric Sciences, 20, 130–141, https://doi.org/10.1175/1520-0469(1963)020<0130:dnf>2.0.co;2, 1963.

Maraun, D., Kurths, J., and Holschneider, M.: Nonstationary Gaussian processes in wavelet domain: Synthesis, estimation, and significance testing, Phys. Rev. E, 75, 016 707, https://doi.org/10.1103/PhysRevE.75.016707, 2007.

Marwan, N., Thiel, M., and Nowaczyk, N. R.: Cross recurrence plot based synchronization of time series, Nonlinear Processes in Geophysics, 9, 325–331, https://doi.org/10.5194/npg-9-325-2002, 2002.

Marwan, N., Trauth, M. H., Vuille, M., and Kurths, J.: Comparing modern and Pleistocene ENSO-like influences in NW Argentina using nonlinear time series analysis methods, Climate Dynamics, 21, 317–326, https://doi.org/10.1007/s00382-003-0335-3, 2003.

Marwan, N., Donges, J. F., Zou, Y., Donner, R. V., and Kurths, J.: Complex network approach for recurrence analysis of time series, Physics Letters A, 373, 4246–4254, https://doi.org/10.1016/j.physleta.2009.09.042, 2009.

Packard, N. H., Crutchfield, J. P., Farmer, J. D., and Shaw, R. S.: Geometry from a Time Series, Physical Review Letters, 45, 712–716, https://doi.org/10.1103/physrevlett.45.712, 1980.

Partin, J. W., Quinn, T. M., Shen, C.-C., Emile-Geay, J., Taylor, F. W., Maupin, C. R., Lin, K., Jackson, C. S., Banner, J. L., Sinclair, D. J., and Huh, C.-A.: Multidecadal rainfall variability in South Pacific Convergence Zone as revealed by stalagmite geochemistry, Geology, 41, 1143–1146, https://doi.org/10.1130/G34718.1, 2013.

Portes, L. L., Benda, R. N., Ugrinowitsch, H., and Aguirre, L. A.: Impact of the recorded variable on recurrence quantification analysis of flows, Physics Letters A, 378, 2382 – 2388, https://doi.org/10.1016/j.physleta.2014.06.014, 2014.

Portes, L. L., Montanari, A. N., Correa, D. C., Small, M., and Aguirre, L. A.: The reliability of recurrence network analysis is influenced by the observability properties of the recorded time series, Chaos: An Interdisciplinary Journal of Nonlinear Science, 29, 083 101, https://doi.org/10.1063/1.5093197, 2019.

Rössler, O.: An equation for continuous chaos, Physics Letters A, 57, 397 – 398, https://doi.org/https://doi.org/10.1016/0375-9601(76)90101-8, 1976.

Russell, J. M., Hopmans, E. C., Loomis, S. E., Liang, J., and Damsté, J. S. S.: Distributions of 5- and 6-methyl branched glycerol dialkyl glycerol tetraethers (brGDGTs) in East African lake sediment: Effects of temperature, pH, and new lacustrine paleotemperature calibrations, Organic Geochemistry, 117, 56 – 69, https://doi.org/https://doi.org/10.1016/j.orggeochem.2017.12.003, 2018.

Schleussner, C.-F., Divine, D. V., Donges, J. F., Miettinen, A., and Donner, R. V.: Indications for a North Atlantic ocean circulation regime shift at the onset of the Little Ice Age, Climate Dynamics, 45, 3623–3633, https://doi.org/10.1007/s00382-015-2561-x, 2015.

Schreiber, T. and Schmitz, A.: Surrogate time series, Physica D: Nonlinear Phenomena, 142, 346 – 382, https://doi.org/https://doi.org/10.1016/S0167-2789(00)00043-9, 2000.

St. George, S.: An overview of tree-ring width records across the Northern Hemisphere, Quaternary Science Reviews, 95, 132 – 150, https://doi.org/10.1016/j.quascirev.2014.04.029, 2014.

St. George, S. and Esper, J.: Concord and discord among Northern Hemisphere paleotemperature reconstructions from tree rings, Quaternary Science Reviews, 203, 278 – 281, https://doi.org/10.1016/j.quascirev.2018.11.013, 2019.

Takens, F.: Detecting strange attractors in turbulence, in: Lecture Notes in Mathematics, pp. 366–381, Springer Science and Business Media, https://doi.org/10.1007/bfb0091924, 1980.

Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D.: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling, Climate of the Past, 15, 1251–1273, https://doi.org/10.5194/cp-15-1251-2019, 2019.

Thompson, L. G., Davis, M. E., Mosley-Thompson, E., Lin, P.-N., Henderson, K. A., and Mashiotta, T. A.: Tropical ice core records: evidence for asynchronous glaciation on Milankovitch timescales, Journal of Quaternary Science, 20, 723–733, https://doi.org/10.1002/jqs.972, 2005.

Thompson, L. G., Mosley-Thompson, E., Davis, M. E., Zagorodnov, V. S., Howat, I. M., Mikhalenko, V. N., and Lin, P.-N.: Annually Resolved Ice Core Records of Tropical Climate Variability over the Past ∼1800 Years, Science, 340, 945–950, https://doi.org/10.1126/science.1234210, 2013.

Tolwinski-Ward, S. E., Evans, M. N., Hughes, M. K., and Anchukaitis, K. J.: An efficient forward model of the climate controls on interannual variation in tree-ring width, Climate Dynamics, 36, 2419–2439, https://doi.org/10.1007/s00382-010-0945-5, 2011.

Tolwinski-Ward, S. E., Anchukaitis, K. J., and Evans, M. N.: Bayesian parameter estimation and interpretation for an intermediate model of tree-ring width, Climate of the Past, 9, 1481–1493, https://doi.org/10.5194/cp-9-1481-2013, 2013.

Trauth, M. H.: TURBO2: A MATLAB simulation to study the effects of bioturbation on paleoceanographic time series, Computers & Geosciences, 61, 1 – 10, https://doi.org/https://doi.org/10.1016/j.cageo.2013.05.003, 2013.

Trauth, M. H., Bookhagen, B., Marwan, N., and Strecker, M. R.: Multiple landslide clusters record Quaternary climate changes in the northwestern Argentine Andes, Palaeogeography, Palaeoclimatology, Palaeoecology, 194, 109–121, https://doi.org/10.1016/s0031-0182(03)00273-6, 2003.

Vaganov, E. A., Hughes, M. K., and Shashkin, A. V.: Growth Dynamics of Conifer Tree Rings, Springer-Verlag Berlin Heidelberg, 1 edn., https://doi.org/10.1007/3-540-31298-6, 2006.

Wackerbarth, A., Scholz, D., Fohlmeister, J., and Mangini, A.: Modelling the δ^{18}O value of cave drip water and speleothem calcite, Earth and Planetary Science Letters, 299, 387 – 397, https://doi.org/10.1016/j.epsl.2010.09.019, 2010.

Wang, Y., Cheng, H., Edwards, R. L., He, Y., Kong, X., An, Z., Wu, J., Kelly, M. J., Dykoski, C. A., and Li, X.: The Holocene Asian Monsoon: Links to Solar Changes and North Atlantic Climate, Science, 308, 854–857, https://doi.org/10.1126/science.1106296, 2005.

Weijers, J. W., Schouten, S., van den Donker, J. C., Hopmans, E. C., and Damsté, J. S. S.: Environmental controls on bacterial tetraether membrane lipid distribution in soils, Geochimica et Cosmochimica Acta, 71, 703 – 713, https://doi.org/10.1016/j.gca.2006.10.003, 2007.

Welch, P.: The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Transactions on Audio and Electroacoustics, 15, 70–73, https://doi.org/10.1109/TAU.1967.1161901, 1967.

Wong, C. I. and Breecker, D. O.: Advancements in the use of speleothems as climate archives, Quaternary Science Reviews, 127, 1 – 18, https://doi.org/10.1016/j.quascirev.2015.07.019, 2015.

Yarleque, C., Vuille, M., Hardy, D. R., Timm, O. E., De la Cruz, J., Ramos, H., and Rabatel, A.: Projections of the future disappearance of the Quelccaya Ice Cap in the Central Andes, Scientific reports, 8, 15 564–15 564, https://doi.org/10.1038/s41598-018-33698-z, 2018.

Zumbahlen, H.: CHAPTER 8 - Analog Filters, in: Linear Circuit Design Handbook, edited by Zumbahlen, H., Newnes, https://doi.org/10.1016/B978-0-7506-8703-4.00008-0, 2008.

Supplementary Material for: Detecting dynamical anomalies in time series from different palaeoclimate proxy archives using windowed recurrence network analysis

Jaqueline Lekscha[1,2] and Reik V. Donner[1,3]

[1]Potsdam Institute for Climate Impact Research (PIK) – Member of the Leibniz Association, 14473 Potsdam, Germany
[2]Department of Physics, Humboldt University, 12489 Berlin, Germany
[3]Department of Water, Environment, Construction and Safety, Magdeburg-Stendal University of Applied Sciences, 39114 Magdeburg, Germany

Correspondence: lekscha@pik-potsdam.de

Additional figures

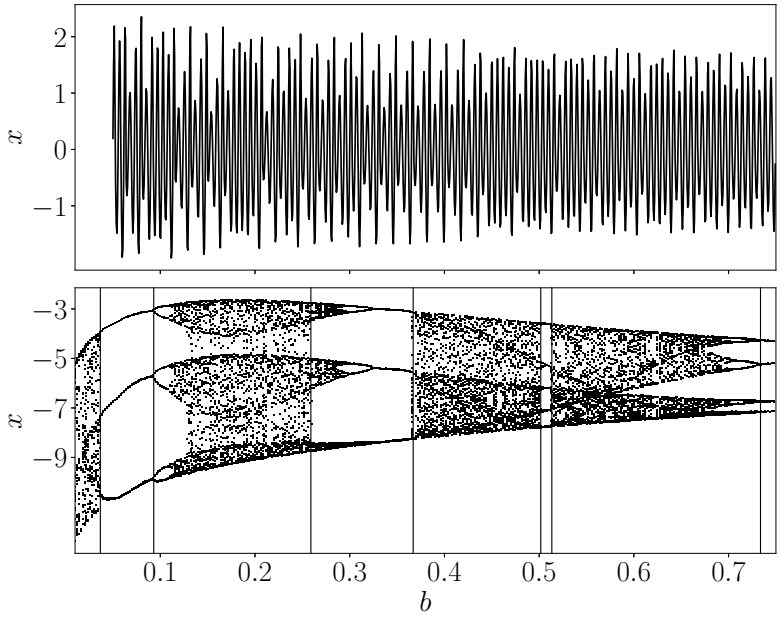

Figure S1. Transient time series of the x-component of the Rössler system with parameter b varying in every integration step (top) and Feigenbaum diagram for corresponding stationary system (bottom).

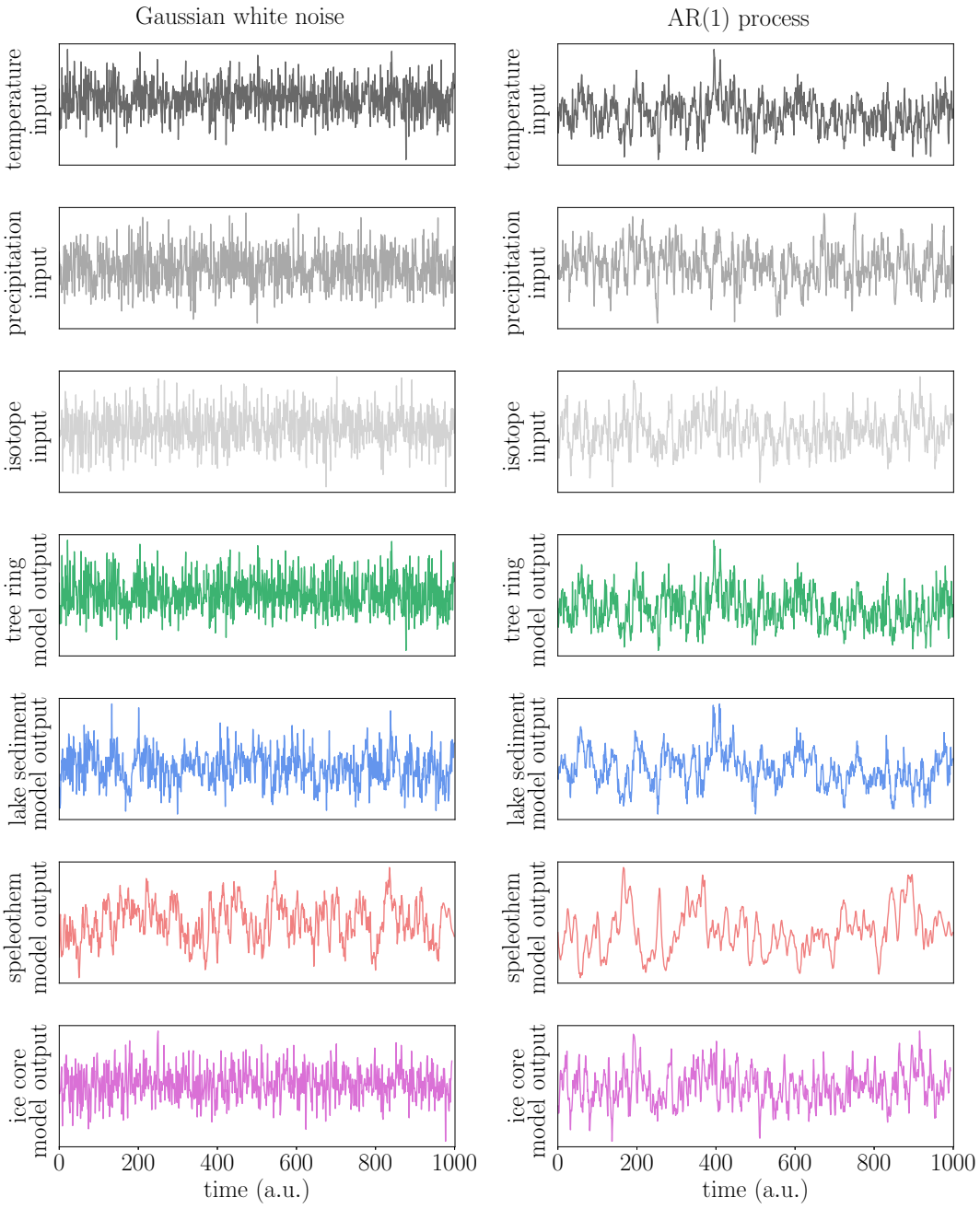

Figure S2. Annually resolved input time series for temperature, precipitation and isotopic composition and corresponding model output time series (top to bottom) for the two noise processes (left: Gaussian white noise, right: AR(1) process).

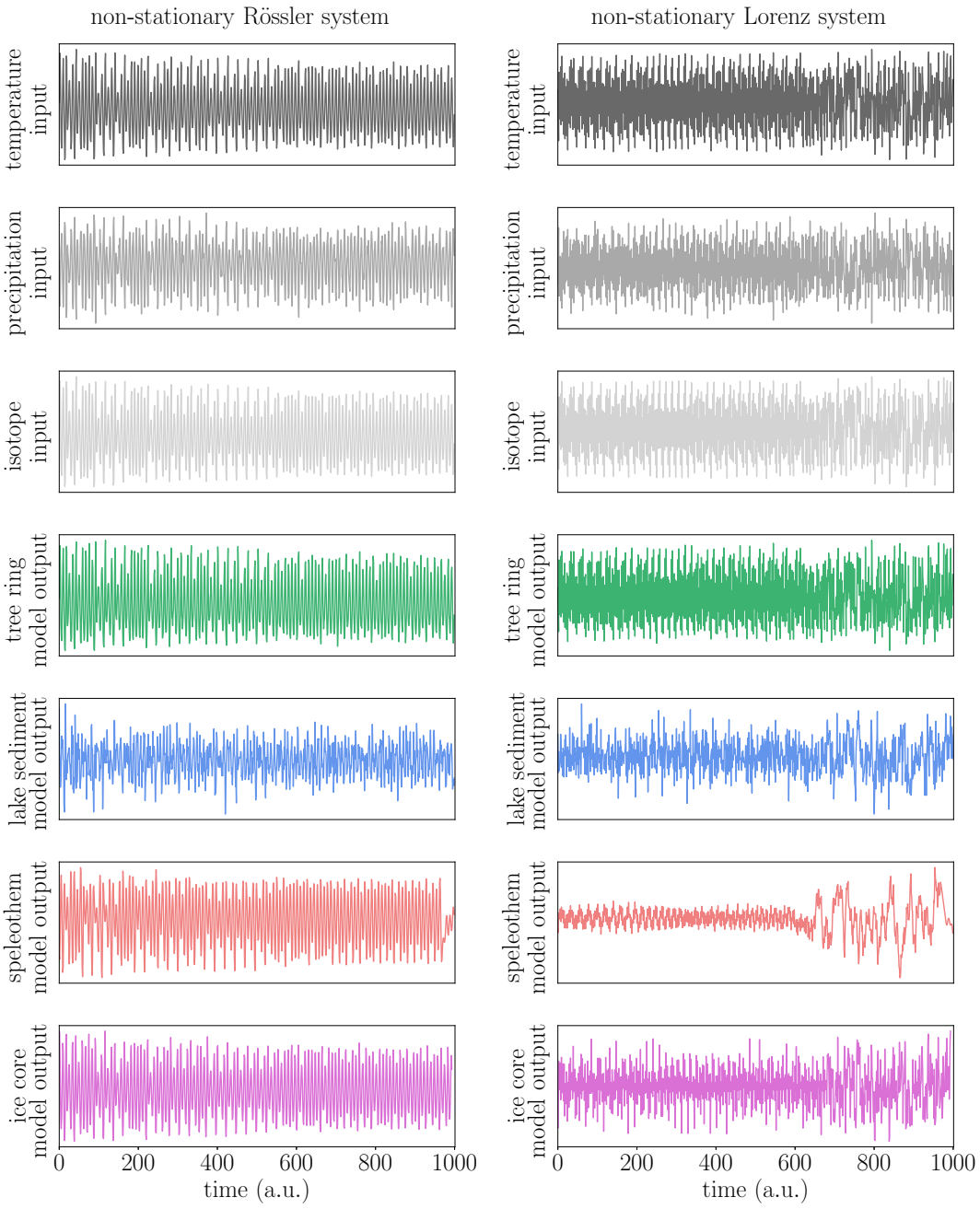

Figure S3. Same as in Fig. S2, but for the two deterministic non-stationary systems (left: non-stationary Rössler system, right: non-stationary Lorenz system).

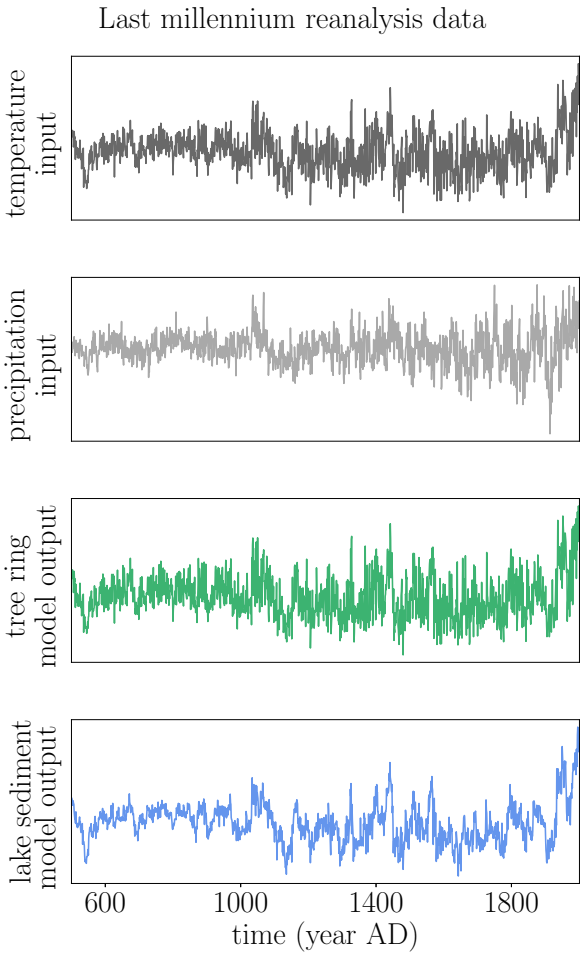

Figure S4. Annually resolved input time series for temperature and precipitation and corresponding model output time series (top to bottom) for the last millennium reanalysis data (Hakim et al., 2016; Tardif et al., 2019).

~~Skewness of time~~ Time series properties

In order to complement the results of recurrence network analysis and better understand their possible dynamical meaning, we take a look at the properties of the time series generated by the different proxy system models and compare them to those of the input time series. Figures S2 to S4 display the annually sampled input time series for temperature, precipitation and isotopic compositions and the corresponding output time series of the four proxy system models for the five input scenarios of GWN, the AR(1) process, the non-stationary Rössler system, the non-stationary Lorenz system, and the last millennium reanalysis data. The expected low-pass filter effects of the speleothem, ice and lake models due to the cave residence time, diffusion, and bioturbation, respectively, are directly visible in the time series, while for the tree model, such an effect is neither expected nor

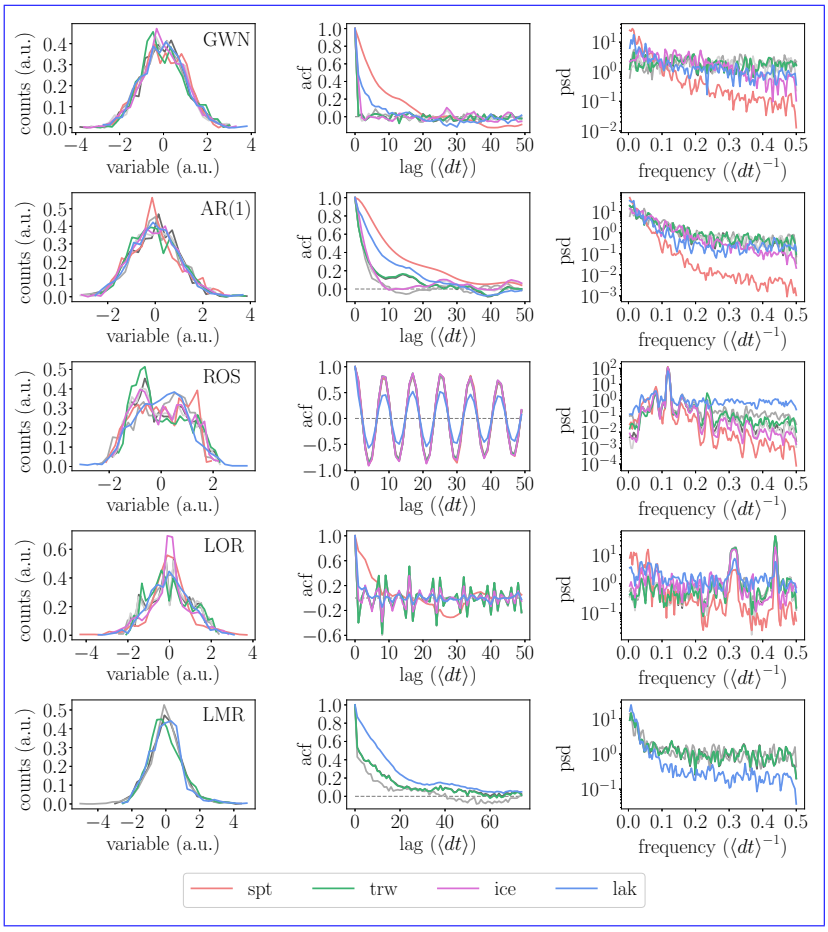

Figure S5. Normalised histograms, autocorrelation functions (acf) and estimated power spectral densities (psd) of the different input and proxy system model output time series for GWN, AR(1), ROS, LOR, and LMR (top to bottom). The input variables are denoted in grey, the tree model output in green, the lake model output in blue, the speleothem model output in red, and the ice model output in magenta.

visible. Also, it should be noted that the tree ring model with the parameters as specified in Table 1 seems to primarily respond to temperature rather than to precipitation, meaning that the limiting factor for tree growth in eastern Canada is temperature, which is ecologically reasonable and also be related to the fact that precipitation is only indirectly taken into account in the used model for tree ring width.

For further evaluation, we standardise all time series to zero mean and unit variance and examine some properties of the different input and output series. The left panels of Fig. S5 show the normalised histograms of the input and output variables. To quantify differences in the histograms, we consider the skewness of the distributions of the different time series (see Table S1). We observe that the resulting time series from the tree model all show significant deviations in skewness from the input data, which is probably due to the thresholding of the growth response functions for temperature and soil moisture. For the other proxy models, we observe a significant influence of the model on the skewness of the output for those inputs that are not

stochastic. This might either be related to the smaller confidence bounds due to the different surrogate generation, or hint to a different reaction of the models to different distributions of the input data.

Finally, we have a look at the autocorrelation functions and the power spectral densities of the different input and proxy model output time series. The middle and right panels of Fig. S5 show the resulting autocorrelation functions and estimates of the power spectral density obtained using the Welch method (Welch, 1967). As before, we observe that the autocorrelation function and power spectral density of the tree ring model output closely follow the respective temperature input. The good agreement between the temperature input and the tree ring width model output can be partially attributed to the model not taking into account juvenile growth of the trees. The speleothem model output shows the expected loss of power in the higher frequencies. Similarly, a loss of power in the higher frequencies can be observed in the ice core model output but not as pronounced as for the speleothem model output. The same holds for the lake sediment model output except for the Rössler and Lorenz scenarios, where the lake model output shows more power in the higher frequencies than the corresponding input. This is most likely related to the different time-scales of variability present in those time series in comparison with the others and the particular choice of mixed layer thickness in the lake sediment model, $\mathrm{mxl} = 4$.

Skewness of time series

Table S1 gives the skewness of the distributions of the different time series where those values showing a significant deviation from the skewness of the corresponding input are marked in bold. To test whether the skewness values of the model output time series differ significantly from that of the respective input, we create $N_{\mathrm{sk}} = 10,000$ surrogate data sets for each of the input time series. For GWN and the AR(1) process, this is done by creating different realisations of the corresponding process according to the descriptions in Sec. 4, while for the other time series, the surrogates are created by adding white noise with signal-to-noise ratios of 25 for the model systems and 100 for the last millennium reanalysis data. Then, for each surrogate realisation, we calculate the skewness of the distribution and take the 0.5th and 99.5th percentile as confidence bounds resulting in the 99% confidence intervals of $[-0.20, 0.20]$ for GWN, $[-0.29, 0.29]$ for the AR(1) process, $[0.20, 0.23]$ for the non-stationary Rössler system, $[-0.03, 0.00]$ for the non-stationary Lorenz system, $[0.32, 0.34]$ for the last millennium reanalysis temperature, and $[0.00, 0.02]$ for the last millennium reanalysis precipitation. As precipitation is proportional to the negative temperature and isotopes, the corresponding confidence interval for the Rössler system is $[-0.23, -0.20]$ and for the Lorenz system $[0.00, 0.03]$.

Table S1. Skewness of the different input and proxy system model output time series with significant deviations of the model output time series from the corresponding input marked in bold.

input/output	GWN	AR(1)	ROS	LOR	LMR
temperature	0.00	0.07	0.21	−0.02	0.33
precipitation	0.12	−0.03	−0.20	0.01	0.01
isotopes	−0.06	0.12	0.22	−0.02	–
trw	**0.34**	**0.35**	**0.40**	**0.26**	**0.63**
lak	0.11	0.23	**−0.05**	**0.01**	**0.71**
spt	−0.04	0.29	**−0.27**	**0.26**	–
ice	−0.01	0.15	**0.15**	**−0.05**	–

References

Hakim, G. J., Emile-Geay, J., Steig, E. J., Noone, D., Anderson, D. M., Tardif, R., Steiger, N., and Perkins, W. A.: The last millennium climate reanalysis project: Framework and first results, Journal of Geophysical Research: Atmospheres, 121, 6745–6764, https://doi.org/10.1002/2016JD024751, 2016.

Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D.: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling, Climate of the Past, 15, 1251–1273, https://doi.org/10.5194/cp-15-1251-2019, 2019.

Welch, P.: The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Transactions on Audio and Electroacoustics, 15, 70–73, https://doi.org/10.1109/TAU.1967.1161901, 1967.