Manuscript prepared for Clim. Past
with version 5.0 of the LaTeX class copernicus.cls.
Date: 29 November 2013

Similarity estimators for irregular and age uncertain time series

Kira Rehfeld[1,2] **and Jürgen Kurths**[1,2,3]

[1]Potsdam Institute for Climate Impact Research, P.O. Box 601203, D-14412 Potsdam, Germany
[2]Department of Physics, Humboldt-Universität zu Berlin, Newtonstr. 15, D-12489 Berlin, Germany
[3]Institute for Complex Systems and Mathematical Biology, University of Aberdeen, Aberdeen AB243UE, UK

Correspondence to: K. Rehfeld
(krehfeld@awi.de)

Abstract. Paleoclimate time series are often irregularly sampled and age uncertain, which is an important technical challenge to overcome for successful reconstruction of past climate variability and dynamics. Visual comparison and interpolation-based linear correlation approaches have been used to infer dependencies from such proxy time series. While the first is subjective, not measurable and not suitable for the comparison of many datasets at a time, the latter introduces interpolation bias, and both face difficulties if the underlying dependencies are nonlinear.

In this paper we investigate similarity estimators that could be suitable for the quantitative investigation of dependencies in irregular and age uncertain time series. We compare the Gaussian-kernel based cross correlation (*gXCF,* Rehfeld et al., 2011) and mutual information (*gMI*, Rehfeld et al., 2013) against their interpolation-based counterparts and the new event synchronization function (*ESF*). We test the efficiency of the methods in estimating coupling strength and coupling lag numerically, using ensembles of synthetic stalagmites with short, autocorrelated, linear and nonlinearly coupled proxy time series, and in the application to real stalagmite time series.

In the linear test case coupling strength increases are identified consistently for all estimators, while in the nonlinear test case the correlation-based approaches fail. The lag at which the time series are coupled is identified correctly as the maximum of the similarity functions in around 60-55% (in the linear case) to 53-42% (for the nonlinear processes) of the cases when the dating of the synthetic stalagmite is perfectly precise. If the age uncertainty increases beyond 5% of the time series length, however, the true coupling lag is not identified more often than the others for which the similarity function was estimated. Age uncertainty contributes up to half of the uncertainty in the similarity estimation process. Time series irregularity contributes less, particularly for the adapted Gaussian-kernel based estimators and the event synchronization function. The introduced link strength concept summarizes the hypothesis test results and balances the individual strengths of the estimators: while *gXCF* is particularly suitable for short and irregular time series, *gMI* and the *ESF* can identify nonlinear dependencies. *ESF* could, in particular, be suitable to study extreme event dynamics in paleoclimate records. Programs to analyze paleoclimatic time series for significant dependencies are included in a freely available software toolbox.

1 Introduction

Time series are often used to assess the properties of the processes that generated them, in climate science (Rehfeld et al., 2011) but also in many other scientific fields ranging from ecology (Lhermitte et al., 2011) to astrophysics (Scargle, 1989). Time series similarity measures quantify the degree of statistical association and are, particularly in the geoscientific context, often equated with Pearson correlation (Chatfield, 2004). They help to identify the strength of dependencies between climate processes and potential lead/lag relationships. For modern-day weather stations, both daily temperature and the time of observations are logged precisely. To identify relationships between distant weather evolution, time series of temperature anomalies can be compared. Paleoclimate data are crucial to investigate climate inter-relationships beyond the instrumental record. Paleoclimate time series are, however, more challenging than the data sources in other disciplines: Neither observation time nor the climatic variable are known precisely. Both have to be reconstructed, resulting in irregular and age uncertain time series, because variability in the growth of the archive impacts on the temporal resolution of the resulting proxy time series (Fig. 1). The dependency

of reconstructed paleoclimate time series, and its relationship to global or external forcing, is often inferred from similarities, coinciding maxima/minima or trends, between graphical visualizations of the time series (for example in Zhang et al., 2008; Cheng et al., 2012; Sinha et al., 2011; Zhang et al., 2011). Visual comparison is, however, inherently subjective, can not be quantified and tested in a hypothesis test and will not suffice with the growing number of paleoclimatic datasets available.

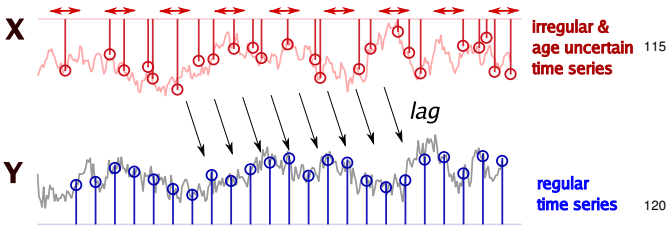

Fig. 1: Illustration: Assume that the climatic processes Y is driven by process X at a given lag. They are sampled by a paleoclimate proxy archive (X) and an automatized measurement device (Y), resulting in corresponding time series. A typical task in paleoclimate data analysis is to estimate the strength of statistical association between such time series, the delay time can hint towards physical driving mechanisms.

Standard statistical techniques, such as estimating the Pearson correlation (XC), can not readily be applied when the sampling of the time series is irregular. XC is, in principle, computed by taking the arithmetic mean over the products of coeval, centralized and standardized observations and reflects the goodness of a linear fit to the scatterplot of the data. If the two time series to be correlated are irregular, coeval observations are only given in the special case that both time series have the same timescale. In practice, this would arise only if, for example, two proxies were measured on the same samples. In the general case the irregularity precludes the direct computation. Interpolating the time series to a regular coinciding timescale, however, results in a loss of high-frequency variability and a spectral bias towards low frequencies (Schulz and Stattegger, 1997). In a comparison of correlation analysis techniques the Gaussian-kernel-based Pearson correlation was identified as a reliable and robust estimator for irregular time series (Rehfeld et al., 2011). However, relationships in the climate system are not always linear, and therefore not necessarily identifiable by linear techniques such as Pearson correlation. This is not a problem in the geosciences alone, and similarity measures that can capture nonlinear inter-relationships exist. Mutual Information (MI), an entropy-based measure, has been used to investigate nonlinear dependencies of processes from observations (Donges et al., 2009; Runge et al., 2012; Hlinka et al., 2013). In this measure, the joint and marginal distributions of processes X and Y are evaluated. Its advantage is that it is model-free and able to quantify nonlinear dependencies, but it is symmetric, MI(X,Y)=MI(-X,Y), and more difficult to quantify as the quantification bias changes considerably for different sample sizes and estimator techniques (Khan et al., 2007; Kraskov et al., 2004). It has been adapted and tested for irregular and autocorrelated time series (Rehfeld et al., 2013) in a Gaussian-kernel-based variant. Both MI and XC depend on the notion of a scatterplot between the data. An alternative, especially in the analysis of extreme events, could be found in the measure of Event Synchronization (*ES,* Quian Quiroga et al., 2002), which is not based on the available time series, but the relative timing of distinguished events in two time series. Originally conceived for neurophysiological signals, it has become a popular measure to investigate dependencies in precipitation time series (Malik et al., 2010, 2011; Rheinwalt et al., 2012), but it has not been tested for its suitability on short and autocorrelated time series. In its original form it provides a measure for the strength of synchronization and for the direction of a potential coupling between the processes generating the events, but not for the lag of the potential coupling. Although stated differently in the original paper, ES does not require regular observation intervals.

A number for an individual correlation coefficient can be interpreted, when its level of significance is determined as well. For the usually short and autocorrelated paleoclimatic time series, this can be done by bootstrapping the result (Mudelsee, 2002), or by testing the similarity for mutually uncorrelated surrogate time series with similar autocorrelation properties (Rehfeld et al., 2011, 2013). The values of the different estimators, however, can not be compared directly, as they vary on different scales. In this paper we evaluate the impact of age uncertainty and time series irregularity on the accuracy of the estimators.

Furthermore we propose the concept of a *link strength*, to summarize the hypothesis test results of different estimators. If no outcome is significant, it is zero, if three out of five employed estimators yield a significant similarity, the link strength is $\frac{3}{5}$ and if all tests for null correlation were rejected the link strength is equal to unity. The advantage of this approach lies in its robustness due to the different estimators, and in the easy consideration of uncertain datasets. If the uncertainty of the time series can be modeled, for example using the Monte Carlo techniques in age modeling software such as StalAge (Scholz and Hoffmann, 2011) or COPRA (Breitenbach et al., 2012), it can be incorporated in the link strength considerations in a straightforward manner.

In this paper we will investigate how well each of these estimators identify the strength and the delay time of actual coupling between paleoclimatic processes from irregular and age uncertain time series. First we review the similarity measures (XC, MI), and develop a Event Synchronization Func-

tion (ESF) based on the concept of ES. We simulate artificial stalagmites with linearly and nonlinearly coupled proxy time series based on autoregressive (AR) and threshold-autoregressive (TAR) models. Using these and the stalagmite time series from Dandak (Sinha et al., 2007; Berkelhammer et al., 2010) and Wanxiang (Zhang et al., 2008) cave we investigate how the similarity estimators perform for irregular, age uncertain and autocorrelated time series, and how they are impacted by age uncertainty.

2 Methods

In this section we first give necessary definitions for time series and similarity measures, and derive the ESF and the link strength concept.

2.1 Time series

Time series are a collection of measurements of specific properties of an dynamical process, together with the time when the observation (or measurement) took place. The individual data points of the series are often regarded as observations of processes, which may be deterministic, stochastic, or a combination of both. In classical time series analysis the observation times of the process X_t are expected to be regular and certain, and the observation values to be measured exactly.
In contrast to this, for irregular time series no unique sampling rate can be defined, and the observation times cannot be directly related to an index anymore, but have to be given explicitly for each measurement.

Definition 1 (Irregular time series) *An irregular time series $\boldsymbol{x}(t) = (\boldsymbol{t}_i, \boldsymbol{x}_i)$ is defined by its observation times $\boldsymbol{t}_i$ and the respective observations $\boldsymbol{x}_i$, where $i = 1, \ldots, N$. The two vectors have a common length N_x, with $\boldsymbol{t}^x_1 < \boldsymbol{t}^x_2 < \cdots < \boldsymbol{t}^x_{N^x}$ as observation times.*

In the following we focus on the age uncertain paleoclimate proxy time series for which a growth model of the archive has been combined with point-wise age information, for example from uranium/thorium measurements. Input data to this age modeling are (i) a dating table with its entries containing depths, associated age estimates and their uncertainties, usually given as standard deviations, and (ii) the proxy observations.

Definition 2 (Dating table) *A dating table $\mathbb{D} = (\boldsymbol{D}_i, \boldsymbol{T}_i, \boldsymbol{\sigma}_{T_i})_{i=1,\ldots,N_{dat}}$ contains N_{dat} point-wise age estimates $\boldsymbol{T}_i$ taken at depths $\boldsymbol{D}_i$ and their corresponding age standard deviations $\boldsymbol{\sigma}_{T_i}$.*

Definition 3 (Proxy observation series) *Proxy observation series $\mathbf{X}^d = (\boldsymbol{d}_j, \boldsymbol{x}_j)$ are given for $j = 1, \ldots, N_{obs}$ measurement depths $\boldsymbol{d}_j$ and proxy measurements $\boldsymbol{x}_j$.*

For paleoclimate archives, the ages at few depths are estimated, with some uncertainty. Age models are then created to interpolate from these few dates to a time axis for the proxy time series, which is sampled much more densely in depth than the dating table. Thus, an age model is defined here as one potential depth-age relationship $\boldsymbol{t}_i(\boldsymbol{z}_i)$ out of the possible ensemble of age models $\mathbb{T}$. For Monte Carlo (MC) age modeling, whole *ensembles* of age models, $\mathbb{T}$ are created, sampling the probability space inherent in the dating table (cf. def. 2). By convention, usually the *most likely* age model is selected as the time axis for proxy time series (Breitenbach et al., 2012; Scholz and Hoffmann, 2011). Finally the dating table is combined with the proxy observation series using a single age model to form a time-uncertain time series.

2.2 Estimating similarity of irregular time series

Similarity measures reflect statistical properties of time series, which may not reflect the same climatic parameters. Different estimators focus on different characteristic properties related to the distributions of the observations, we summarize them in Table 1.

Assume that the processes X and Y generated time series $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t)$. These processes, and the time series, are similar if, for example, coeval minima or maxima were observed. Comparison can then give information about functional relationships between processes underlying time series: Given that two processes X and Y are not independent, there may either be a causal relationship or they are both driven by a global *common driver*, or there are unobservable intermediate processes, as illustrated in Fig. 2. A significant similarity estimate may therefore arise for such physical reasons - or as a false positive of the statistical test. If a transfer function between these two processes exists in a form of $Y_t = \mathcal{F}(X_{t+\ell})$, this results in a repetition of a pattern, though maybe distorted, that occurs in X_t at t_0 and in Y_t at a time $t = t_0 + \ell$ later. A similarity estimator can help identify $\mathcal{F}$ and quantifies the similarities in the contemporary evolution of two time series:

Definition 4 (Similarity estimator) *A similarity estimator $S = \mathcal{F}((\boldsymbol{t}^x, \boldsymbol{x})(\boldsymbol{t}^y, \boldsymbol{y}))$ reflects the similarity between $\boldsymbol{x}(\boldsymbol{t})$ and $\boldsymbol{y}(\boldsymbol{t})$ to a numeric value in an interval $[a, b]$, $S : \boldsymbol{x}(t) \times \boldsymbol{y}(\boldsymbol{t}) \to [a, b]$.*

For most similarity measures $a = -1, b = 1$ is considered, but for different estimators different bounds exist. Here we only require that the relationship between true dependency and estimated similarity is monotonically increasing, which is what we test for artificially generated time series. If the delay time ℓ in the transfer function is nonzero, a similarity function gives the similarity between two time series for increasing delay:

Table 1: Properties, parameters and references of the similarity estimator algorithms for irregularly sampled time series developed and tested in this paper.

Estimator (Abbr.)	Quantif. property	Parameter choice	References
1 (*gXCF*)	Gaussian-kernel-based XCF (goodness of linear fit to scatterplot)	$h = 0.25$	Rehfeld et al. (2011); Babu and Stoica (2010)
2 (*iXCF*)	interpolation + Pearson correlation (goodness of linear fit to scatterplot)	$\Delta t = \max(\Delta t^x, \Delta t^y)$	e.g. Rehfeld et al. (2011), basics for example in Chatfield (2004)
3 (*gMI*)	Gaussian-kernel-based MI (rel. non-randomness in joint vs. marginal distribution)	$h = 0.5$, $\tau = 3$	Rehfeld et al. (2013), basics for example in Cover and Thomas (2006)
4 (*iMI*)	interpolation + MI (rel. non-randomness in joint vs. marginal distribution)	$\Delta t = \max(\Delta t^x, \Delta t^y)$, $n_{bins} = 10$	Rehfeld et al. (2013), basics for example in Cover and Thomas (2006)
5 (*ESF*)	Relative timing of extreme events	$q = 0.8$	here, based on Quian Quiroga et al. (2002); Malik et al. (2010)

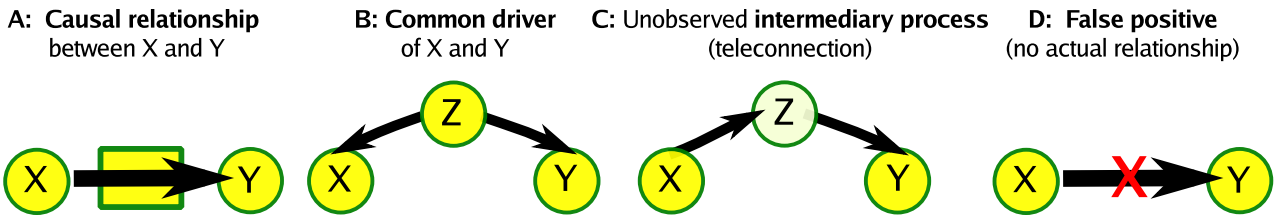

Fig. 2: Significant similarities between the time series at two locations, X and Y, can arise from a) direct physical coupling, b) a teleconnection, c) a common driving mechanism or d) by chance as false positives.

Definition 5 (Similarity function) *A similarity function $S(\ell)$ gives the estimated similarity over different lag times ℓ:*

$$S(\ell) = S(\ell \cdot \Delta t) = f\left((\boldsymbol{t}^x, \boldsymbol{x}), (\boldsymbol{t}^y + \ell \cdot \Delta t, \boldsymbol{y})\right) \quad (1)$$

The spacing of the lag vector is uniform and depends on the mean time resolution of the time series: $\Delta t = \max(\Delta t_x, \Delta t_y)$. To indicate that we are focusing on bivariate similarity we also use the alternative notation $S(X,Y)$ which does not explicitly refer to the possible lags.

Similarity measures as required in this context should be symmetric, reflexive, translation and scale invariant (Batyrshin et al., 2012). The estimators presented here fulfill these requirements.

2.2.1 Kernel-based estimators for Pearson correlation

Pearson correlation is defined as the mean over co-eval products of standardized observations (Chatfield, 2004). For irregular time series the inter-sampling time intervals vary and the classical definition cannot be applied. Rehfeld et al. (2011) tested different correlation estimators for irregular time series and found that a Gaussian-kernel based estimator performed best. In the definition of the correlation function $\hat{\rho}(\boldsymbol{k}\Delta t)$ at the lag $\boldsymbol{k}\Delta t$,

$$\hat{\rho}(\boldsymbol{k} \cdot \Delta t) = \frac{\sum_{i=1}^{N^x} \sum_{j=1}^{N^y} \boldsymbol{x}_i \boldsymbol{y}_j b_k(\boldsymbol{t}_j^y - \boldsymbol{t}_i^x)}{\sum_{i=1}^{N} \sum_{j=1}^{N} b_k(\boldsymbol{t}_j^y - \boldsymbol{t}_i^x)} \ , \quad (2)$$

the *kernel* $b_k(\boldsymbol{t}_j^y - \boldsymbol{t}_i^x)$ weights those products higher whose time lag lies closer to $\boldsymbol{k}\Delta t$:

$$b_k(\boldsymbol{d}) = \frac{1}{\sqrt{2\pi h}} e^{-|\boldsymbol{d}|^2/2h^2} \ , \quad (3)$$

where $h = \Delta t/4$ or 0.25 for the rescaled time axis, $\boldsymbol{t}_i^x = \boldsymbol{t}_i^{\text{orig}}/\Delta_t^x$, and $\boldsymbol{d}$ denotes the distance between the product inter-observation time and the desired lag, $\boldsymbol{d} = \boldsymbol{t}_j^y - \boldsymbol{t}_i^x - \boldsymbol{k}\Delta t$,

$\boldsymbol{k}$ denotes the lag index. The standard width parameter h is chosen to result in a main lobe width of Δt, the mean sampling interval or common sampling period in the bivariate case. Note that the observations have to be standardized to zero mean and unit variance before the analysis.

2.2.2 Kernel-based estimators for mutual information

Mutual information $I(X,Y) = I_{xy}$ is a measure of the dependency (linear or nonlinear) between two random variables, X and Y. This measure from information theory can be interpreted as the uncertainty reduction in variable X, given that Y was observed. It is symmetric, i.e. relationships of opposite sign but the same association strength, correlation and anti-correlation, give the same MI. By definition, the measure yields a null result if, and only if, the two random variables, in this case time series of observations, are independent (Kraskov et al., 2004; Cover and Thomas, 2006).

While more complex estimators exist (e.g. Kraskov et al., 2004), the simplest estimator is

$$\hat{I}_{xy} = \sum_{x,y} p_{x,y} \log \frac{p_{x,y}}{p_x p_y} \ , \tag{4}$$

where $p_{x,y}$ is the two-dimensional joint probability density function of the variables X and Y and p_x resp. p_y are the one-dimensional probability distributions of X resp. Y. The unit of measurement of MI depends on the *logarithm* chosen in the estimator: it is measured in *bits*, if the logarithmic base 2 is chosen, and in *nats* for the natural logarithm.
In case of irregular sampling, however, the bivariate observation set (X_t, Y_t) at regular observation points t that is required for a scatterplot are not available. In standard interpolation procedures, both $(\boldsymbol{t}_x, \boldsymbol{x})$ and $(\boldsymbol{t}_y, \boldsymbol{y})$ would be resampled to obtain a bivariate set of observations with regular observation time intervals, $(\boldsymbol{t}_r, \boldsymbol{x}_r, \boldsymbol{y}_r)$. This is undesirable for paleoclimate records a) because every interpolation routine involves an assumption on the dynamics of the underlying process, and this is difficult to justify for climate data and b) it reduces the observable variability in the process (Schulz and Stattegger, 1997; Stoica and Sandgren, 2006; Babu and Stoica, 2010).

There are two main points where this problem can be addressed: Either by reconstructing bivariate observations while avoiding variance reduction as much as possible or by a modification of the joint distribution, for example by introducing weights proportional to the sampling time-distance similar to the Gaussian-kernel based XC (Rehfeld et al., 2011). For MI the latter is difficult to achieve. But following the former solution, the probabilities required for Eq. 4 are straightforward to derive from relative frequencies.

Algorithmically, this can be described as follows:

1. A local reconstruction of the signal is performed by estimating for each point i in the time series $X = (\boldsymbol{t}^x, \boldsymbol{x})$ a corresponding observation from $Y = (\boldsymbol{t}^y, \boldsymbol{y})$, by estimating a local, observation-time weighted mean $\boldsymbol{y}_j^{lr}$ around a time point $\boldsymbol{t}_i^x$ in Y,

$$\boldsymbol{y}_j^{lr} = \sum_{i=1}^{N_y} b_k(\boldsymbol{d}) \boldsymbol{y}_i \ , \tag{5}$$

with the Gaussian-kernel based local weight $b_k(\boldsymbol{d})$ defined as in Eq. 3. For MI the standard deviation of the Gaussian weight function is set to $h = 0.5$. If there are less than five observations $\boldsymbol{y}_i$ available in a time window $\pm 3\Delta t$ around $\boldsymbol{t}_i^x$ this reconstruction is not performed. Repeating this for each time point $j = 1, \ldots, N^x$ in X one obtains a new, bivariate set of observations

$$Y^x = (\boldsymbol{t}_i^x, \boldsymbol{x}_i, \boldsymbol{y}_i^{lr}) \ .$$

2. Afterwards the procedure is repeated by stepping through $\boldsymbol{t}_j^y$, which yields

$$X^y = (\boldsymbol{t}_j^y, \boldsymbol{x}_j^{lr}, \boldsymbol{y}_j) \ .$$

3. The local reconstruction Y^x and the original observations Y are then concatenated into one series $Y^r = \{Y \cup Y^x\}$ combining locally reconstructed and original observations. Similarly, a time series $X^r = (X \cup X^y)$ is obtained.

4. Based on this set of bivariate observations (X^r, Y^r) the joint density of X and Y can be estimated using standard binning estimators for MI.

The reconstructed set of bivariate observations can also be used to construct Gaussian-weighted scatterplots, where the size of the marker reflects the amount of weight placed on the reconstructed observation (cf. Figs. 4b and 5b). MI is difficult to estimate in practice, first and foremost because of the large bias effects produced in the inference of the joint and marginal probabilities. Elaborate algorithms have been devised to improve this (described, for example, in Kraskov et al., 2004; Papana and Kugiumtzis, 2009; Roulston, 1999), but no straightforward solution to this has been found yet. We have tested several algorithms and finally resorted to the most simple equidistant *binning estimator* (Kraskov et al., 2004), due to its computational efficiency and simplicity. Bias effects are predominantly tied to the temporal sampling and length of the time series due to the occurrence of empty bins. Thus, if necessary, we can estimate and subtract the bias using uncorrelated processes with the same observation times as in X and Y. However, for the use as a similarity measure comparable to *XCF* and *ES* in the context of paleoclimate networks we only require that the estimated MI be proportional to the actual association strength. For bivariate normally distributed and linearly correlated X and Y *MI* is by

definition proportional to their estimated correlation coefficient r_{xy}^2,

$$I_{xy} = -\frac{1}{2}\log(1 - r_{xy}^2)\ , \tag{6}$$

and can, by inversion of this equation, be scaled to the positive range of the correlation coefficient so that $\hat{I} \in [0,1]$ (Nazareth et al., 2007). The expected value for mutual information of these processes at the lag of coupling is then given by $MI(X(t), Y(t+l)) = -0.5\log(1 - r_{xy}^2)$. For the evaluation of the joint and marginal distributions, $n_{bins} = 10$ equidistant bins were employed. In principle, the number of bins should be adapted to the respective length of the time series involved, to reduce bias effects from empty bins.

2.2.3 Event Synchronization function

The concept of event synchronization (ES) was introduced by Quian Quiroga et al. (2002). The motivation behind the development was to obtain a simple, fast method that quantifies the synchronization between time series where certain *events* can be distinguished. The primary application was focused on neurophysiological signals (Quian Quiroga et al., 2002; Kreuz et al., 2009), but it later was also applied for the investigation of rainfall patterns in the Asian Monsoon domain (Malik et al., 2010, 2011; Rheinwalt et al., 2012).

The main idea behind ES is that two time series are synchronized, if events in time series $\boldsymbol{x}$ occur close in time to events in time series $\boldsymbol{y}$. Considering the temporal order of the events, e.g. if an event in $\boldsymbol{y}$ occurred *before* one in $\boldsymbol{x}$, it is also possible to infer which process is *leading*. In the following we will define the Event Synchronization Function, ESF, further developing the ES concept (Quian Quiroga et al., 2002; Malik et al., 2010).

Given two time series $(\boldsymbol{t}^x, \boldsymbol{x})$ and $(\boldsymbol{t}^y, \boldsymbol{y})$ that represent observations of autocorrelated stochastic processes, *events* are given by the set of observations that are considered *extreme*, in that their observation value lies above or below the $q/2$ resp. $(1-q/2)$ percentiles of the distributions of $\boldsymbol{x}$ and $\boldsymbol{y}$. The actual *value* of the observation at the event points is not relevant for the further analysis. Once the events are defined, only the observation *times* are considered in the event time vectors $\boldsymbol{t}_x^*$ and $\boldsymbol{t}_x^*$. Next a temporal threshold τ is defined to evaluate the relationship between the events in X and Y with a maximum separation time:

$$\tau = \max\left(\Delta t^x, \min(\Delta t_x^*, \Delta t_y^*)/2\right)\ . \tag{7}$$

Here, Δt^x is the mean sampling rate of X, and Δt_x^* and Δt_y^* are the inter-event times in X and Y, respectively.

Subsequently, the co-occurrence of events in X and Y is counted and summed for all events as

$$C(X|Y) = \sum_{l=1}^{N_x}\sum_{m=1}^{N_y} \mathbf{J}_{lm}^{xy}\ , \tag{8}$$

where N_x and N_y, respectively, give the total numbers of events in X and Y. The counter variable $\mathbf{J}_{lm}^{xy}$ is defined as

$$\mathbf{J}_{lm}^{xy} = \begin{cases} 1 & \text{if } 0 < t_l^x - t_m^y < +\tau \\ 1/2 & \text{if } t_l^x - t_m^y = 0 \\ 0 & otherwise\,. \end{cases} \tag{9}$$

$C(Y|X)$ is obtained by exchanging X vs. Y in the above expression, and combining both,

$$Q_{xy} = Q_{xy}(X,Y) = \frac{C(X|Y) + C(Y|X)}{\sqrt{N_x, N_y}} \tag{10}$$

gives the *strength* of the event synchronization and

$$q_{xy} = \frac{C(X|Y) - C(Y|X)}{\sqrt{N_x, N_y}} \tag{11}$$

the *direction* of the association. Unless double-counting of events occurs, these are normalized to $0 \le Q \le 1$ resp. $-1 \le q \le 1$. $Q = 1$ corresponds to completely synchronous occurrence of events in X and Y, and $q = 1$ implies that all events in Y *precede* those in X.

For the previous studies (Quian Quiroga et al., 2002; Malik et al., 2010, 2011) local definitions of the temporal threshold τ were used, preventing, in most cases, events from being double-counted, and adapting it to the local inter-event rate. The chosen definition of τ is motivated by the fact that, to be able to compare the results for ES to those obtained from MI and XCF, a similarity function over the *delay* is needed. Thus, the delay τ cannot allowed to be arbitrarily large or small, as in Malik et al. (2010); Quian Quiroga et al. (2002).

The ESF is obtained by shifting the observation times of time series X according to the desired lag,

$$ES(\boldsymbol{k}\Delta t) = Q_{xy}((\boldsymbol{t}_x - \boldsymbol{k}\Delta t, \boldsymbol{x}), (\boldsymbol{t}_y, \boldsymbol{y})). \tag{12}$$

which, using the delay time τ from Eq. 7, makes it possible to use the ESF as a similarity function.

2.3 An approach to similarity assessment of time-uncertain time series

Age uncertainty is a key obstacle to be overcome for a comprehensive understanding of past Earth system dynamics. To investigate the potential dependency structure of paleoclimate processes X and Y as they are reflected in natural archives, the contribution of age uncertainty to the uncertainty of the similarity $S(X,Y)$ is important.

Thus the aim is to estimate the distribution **p(S(X,Y))** of similarity for given datasets X and Y, where

$$X = \left[\mathbb{D}^x = \{\boldsymbol{D}^x, \boldsymbol{T}^x, \boldsymbol{\sigma}_{T^x}\}, Y^d = \{\boldsymbol{d}^x, \boldsymbol{x}\}\right] \text{ and} \tag{13}$$

$$Y = \left[\mathbb{D}^y = \{\boldsymbol{D}^y, \boldsymbol{T}^y, \boldsymbol{\sigma}_{T^y}\}, X^d = \{\boldsymbol{d}^y, \boldsymbol{y}\}\right]\ , \tag{14}$$

both input datasets consist of a dating table (Def. 2) $\mathbb{D}$ with dating depth vector $\boldsymbol{D}$, the corresponding estimated ages $\boldsymbol{T}$ and their uncertainties $\boldsymbol{\sigma}_{T^y}$ and a set of proxy measurements $\boldsymbol{X}^d$ resp. $\boldsymbol{Y}^d$ (Def. 3), visualized as **step 1** in Fig. 3. The smoothing resulting from the size of the samples in depth direction, $\boldsymbol{\sigma}_D$, is assumed to be negligible here. The input proxy measurements are mapped to observation times in the *age modeling* process. In general, algorithms to assess similarity between time series are not capable of processing *probability distributions* or *confidence intervals* instead of singleton values, neither for the observation times nor for the measurement values.

For Pearson correlation, an analytical approach to propagate the uncertainty around the input data into the correlation estimate is possible. However, Pearson correlation alone is insufficient to characterize similarity between paleoclimate time series in general and in the context of paleoclimate networks. Therefore a Monte-Carlo based approach based on time series ensembles which are obtained via age modeling is used here, to keep the flexibility regarding similarity estimators:

1. In a first step the input datasets X and Y are processed. The monotonicity of the depth control variables, d and D is checked.

2. A Monte-Carlo simulation for the uncertain age estimates in the dating table is performed: N_{ens} ages are drawn from $\mathcal{N}(\boldsymbol{T}_i^X, \boldsymbol{\sigma}_{T_i^X})$ and $\mathcal{N}(\boldsymbol{T}_j^Y, \boldsymbol{\sigma}_{T_j^Y})$, respectively, for all $i = 1, \ldots, N_{dtg}^X$ pointwise age estimates corresponding to $j = 1, \ldots, N_{dtg}^Y$ entries in the dating table. This results in dating matrices $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ with N_{ens} columns containing the sampled ages. If no distribution of ages is otherwise given the ages are expected to be Gaussian distributed with the given standard deviation.

3. The age estimates in each column and $\hat{\mathbf{X}}$ ($\hat{\mathbf{Y}}$) are interpolated to the depths of the proxy observations: $\mathbf{T} = \texttt{interp}(\boldsymbol{D}, \hat{\mathbf{X}}, \boldsymbol{d})$ which results in a matrix of reconstruction observation times $\mathbf{T}$. We used conventional linear interpolation of the ages in COPRA. Thus we obtain an ensemble of possible age-depth relationships $\{\mathbf{T}, \boldsymbol{d}\}$ and an ensemble of proxy time series $\{\mathbf{T}, \boldsymbol{x}\}$.

4. Each of the members of the ensemble of proxy time series is used as an input to the similarity statistic $S(X, Y)$. This results in a distribution of estimates $p(S(\hat{\mathbf{X}}, \hat{\mathbf{Y}}))$.

5. Analysis of distribution $S(\hat{\mathbf{X}}, \hat{\mathbf{Y}})$: Apart from inspection of mean, variance and skewness of this distribution, a hypothesis test can be conducted, comparing $S(\hat{\mathbf{X}}, \hat{\mathbf{Y}})$ with a distribution obtained from suitable surrogate time series $S(\hat{\mathbf{X}}^*, \hat{\mathbf{Y}}^*)$.

This approach is general in the sense that it is independent of the specific function $\mathcal{F}([\hat{\mathbf{X}}, \hat{\mathbf{Y}}])$ that maps the uncertain input to some output estimate. Apart from $\mathcal{F} = S$, $\mathcal{F}$ may represent any bivariate statistic, and with minor modification is also applicable to calculate the influence of sampling uncertainty on univariate statistics, like the autocorrelation coefficients or persistence times (Rehfeld et al., 2011; Mudelsee, 2002). Bivariate similarity assessment is often concerned with estimation of a potential *coupling strength* α (hinting towards the same process of origin) and/or the *lag of coupling* ℓ for model-building. For Pearson correlation, the ratio of shared vs. total variance between two linearly correlated processes at a given lag ℓ, $S(\ell)$, is given in the maximum of the cross-correlation function. While the relation to the overall variance of the processes does not necessarily hold by definition for other similarity measures, they, too, will observe the maximum of their similarity function $\max(\hat{S})$, at the lag of coupling ℓ.

2.3.1 Synthetic data

'True' growth histories for two synthetic stalagmites $SS1$ and $SS2$ and according climate histories are obtained via simulation. These pseudo-archives are then 'dated', and correlated pseudo-proxy for the climate histories are 'sampled'. Then the age modeling procedure is performed and its output is fed into similarity estimation. Finally, we assess how much of the similarity that was originally present in the climate history is still recognizable significantly, considering the uncertainties. The test strategy is illustrated in Fig. 3.

2.3.2 The synthetic stalagmite

A synthetic (or: virtual) stalagmite is grown for the sensitivity analysis. The main parameters controlled are

- the growth rate λ in $\frac{\text{mm}}{\text{year}}$,
- the total length of the stalagmite (in mm),
- the type of accumulation (linear growth, or growth modeled via randomly distributed accumulation rates).

A growth rate of $\mu(\lambda(\boldsymbol{z})) = 1$mm/yr is chosen. Linear growth may be a reasonable first order approximation (Telford et al., 2004), but microscopically, the growth rates of natural archives archive vary. Therefore, Gamma-distributed accumulation times are drawn for each depth $\boldsymbol{z}_i = \{0, ..., \text{Z}\}$mm of the stalagmite, with the sampling time step mean $\mu(\lambda(\boldsymbol{z}))$ determined by the desired growth rate and shape and scale parameters α and β as $\Gamma(\alpha, \beta) = \Gamma(\alpha, \mu(\lambda(\boldsymbol{z}))/\alpha)$. This way, the mean sampling rate can be kept constant, even when the irregularity of the sampling distribution is changed (Rehfeld et al., 2011). The cumulative sum of the accumulation times then give the 'true' ages of the archive at the depths $\boldsymbol{z}_i$: $\boldsymbol{t}_i^{true}(\boldsymbol{z}_i) = \sum_{j=1}^{i} \lambda_i$.

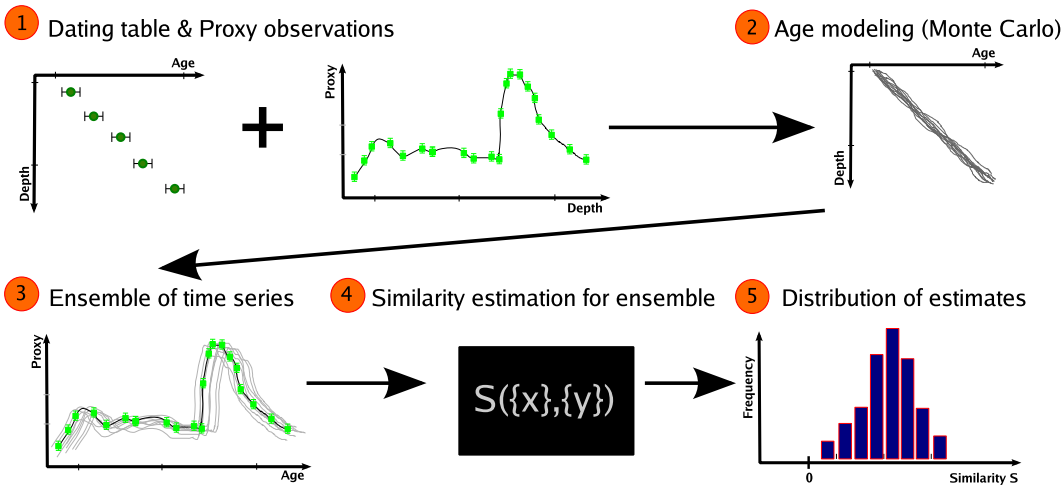

Fig. 3: How much age uncertainty is allowed to still enable reliable similarity estimation? Artificial stalagmites with increasing standard deviations of the ages are evaluated.

2.3.3 The simulated climate history

We attach each synthetic stalagmite $SS1$ and $SS2$ to a climate history. The climate/pseudo-proxy simulation is based on the assumption that $SS1$ lies in an area whose climate is controlling that around $SS2$, through a teleconnection or, for example, by being situated downstream of the same monsoon branch (cf. 2. We simulate climate variability using two different coupling schemes, one linear, one nonlinear, to investigate how the proposed methods perform.

Linearly coupled AR(1) processes

Assuming that the archive $SS2$ samples the same climate variability as $SS1$, in the same way though at a later time, we model such a causal sequence using coupled AR(1) processes. Then, the *true* proxy history of climate as recorded in $SS1$ is given by

$$X(\boldsymbol{t}_i^{true}, \boldsymbol{z}_i) = \phi X(\boldsymbol{t}_{i-1}^{true}) + \sigma_\varepsilon \boldsymbol{\varepsilon}_i, \tag{15}$$

and it determines part of the proxy history of $SS2$:

$$Y(\boldsymbol{t}_i^{true}, \boldsymbol{z}_i) = \alpha X(\boldsymbol{t}_{i-\ell}^{true}) + \sigma_\xi \boldsymbol{\xi}_i. \tag{16}$$

Here, $\boldsymbol{\varepsilon}$ and $\boldsymbol{\xi}$ are additional Gaussian white noise whose variances σ_ε and σ_ξ are scaled such that the variances of X and Y are equal to unity. $\alpha \in [-1,1]$ is the coupling strength between $SS1$ and $SS2$ and ϕ the autocorrelation of $SS1$. Since there is no autocorrelative term in Y_t the true similarity $S(X,Y)$ is equal to the cross correlation: $S(X,Y) = \rho_{xy} = \alpha$ (Rehfeld et al., 2011).

Nonlinear Threshold-AR(1) processes

Let us assume that $SS1$ samples climate variability in a certain place, and this can be modeled as in Eq. 15. Then the climate variability in another place, where $SS2$ is located, could be controlled in a nonlinear manner: The processes are negatively correlated, similar to Eq. 16 with $\alpha < 0$. If, however, a threshold in the climate system is exceeded, $X(\boldsymbol{t}) > \tau$, the correlation changes and might even become positive. Such a multi-scale behavior can be modeled using Threshold-AR-processes (TAR, Tsay, 1989), which are similar to the regime-dependent AR models Zwiers and Storch (1990) used to model the behavior of the Southern Oscillation. Assume that the negative coupling α below the thresholdτ, here $\tau = 0$, for $X(\boldsymbol{t}-1) \leqslant \tau$ turns into a positive correlation, with the same magnitude, for $X(\boldsymbol{t}-1) > \tau$. Then the proxy history of $SS2$ can be modeled as

$$Y(\boldsymbol{t}_i^{true}, \boldsymbol{z}_i) = \alpha \boldsymbol{\kappa} X(\boldsymbol{t}_{i-\ell}^{true}) + \sigma(\boldsymbol{t}^{true}) \boldsymbol{\xi}_i \ , \tag{17}$$

where the $\boldsymbol{\kappa} = -1$ if $X(t-1) \leqslant \tau$ and $\boldsymbol{\kappa} = 1$ when $X(\boldsymbol{t}-1) > \tau$. For convenience, the variance of the innovation term $\boldsymbol{\xi}$ is scaled such that the overall variance of Y is equal to unity in both cases.

2.3.4 'Dating' of the synthetic stalagmite

Mimicking the real life situation, the *true* growth history of the synthetic stalagmite, $\boldsymbol{z}(\boldsymbol{t}_{\text{true}})$ is, in the following, inaccessible. The stalagmite is subjected to *dating* along its depth. The dating table contains the for the dating depths $\boldsymbol{D}$, the estimated age at these depths, $\boldsymbol{T}_j$, the proxy measurement sample width $\boldsymbol{\sigma_D}$ and the age uncertainty $\boldsymbol{\sigma}_T$.

In real life, the stalagmite would be dated using radiometric dating techniques based on Uranium-Thorium (Sinha et al., 2007; Dykoski et al., 2005; Breitenbach et al., 2012) or radiocarbon (Yadava et al., 2004; Webster et al., 2007), yielding an estimate of $\boldsymbol{T}(\boldsymbol{z_j})$ at a few points. The corresponding dating uncertainty, in reality dependent on many

factors from initial isotope concentrations, overall age of the core, dating technique to lab and contamination (Fairchild and Baker, 2012), often lies between 0.1 to 0.5% of the age for stalagmites, but may be considerably higher.

For the synthetic stalagmites, dating 'samples' are taken at equidistant depths $\boldsymbol{D}_j$ and the center points of the assumed age distribution are taken directly from the *true* age-depth relationship. The age uncertainty, however, is modeled as increasing proportionally with age, as $p \cdot \boldsymbol{T}_j$. p here denotes the (im-)precision of the dating and is varied in the following numerical experiments.

2.3.5 Age modeling for SS1 and SS2

Age modeling aims at reconstructing the 'true' depth-age relationship that is inaccessible in real paleoclimate archives.

Based on the synthetic stalagmite dating tables $\mathbf{D}^x$ and $\mathbf{D}^y$ for $SS1$ and $SS2$, the 'observation times' for the proxy observations $\boldsymbol{X}^d$ and $\boldsymbol{Y}^d$, $\boldsymbol{t}^x$ and $\boldsymbol{t}^y$, are constructed by interpolation from the known ages (see Eq. 13). In Monte-Carlo based numerical frameworks such as StalAge (Scholz and Hoffmann, 2011) or COPRA (Breitenbach et al., 2012), an ensemble of age models $\mathbf{T} = \{\boldsymbol{t}_k, \boldsymbol{z}_k\}^{k=1,\ldots,N_{ens}}$ is created, which, in their entirety, reflect the age uncertainty of the estimated depth-age relationship. Based on this ensemble of age models, the uncertainty in the similarity estimates can be inferred, as is visible in Fig. 3.

In summary, the test plan is thus as follows:

1. Simulate a growth history $\boldsymbol{z}(t)$ of a synthetic stalagmite of length Z mm, corresponding to a 'true' age-depth relationship $\boldsymbol{t}_i^{true}(\boldsymbol{z}_i)$, resp. $\boldsymbol{z}_i(\boldsymbol{t}^{true})$. For this, assume gamma-distributed growth and an accumulation rate $\lambda = 1$mm/year. Z can be varied to study the influence of changing time series length.

2. Simulate proxy histories $\{\boldsymbol{T}, \boldsymbol{x}\}^{SS1}$ and $\{\boldsymbol{T}, \boldsymbol{y}\}^{SS2}$ according to the *true* growth history using coupled autoregressive processes (cf. Eqs. 16 and 17). Forget the true growth history.

3. Sample the true growth history at the dating depths and infer corresponding uncertainties.

4. Create N_{ens} surrogate dating tables for $SS1$ and $SS2$ with increasing uncertainty of the ages according to the (im)precision p, i.e. an ensemble of dating tables.

5. Assess if the estimates $S(\hat{\mathbf{X}}, \hat{\mathbf{Y}})$ are statistically significant for the given uncertainty, and how they are influenced by sampling heterogeneity and time uncertainty.

The core of the COPRA algorithm is used for MC simulations. $N_{ens} = 2000$ MC iterations are used to sample the probability space and linear interpolation is employed to infer ages between point estimates of the age at depth.

3 Tests on synthetic stalagmites

We evaluate the performance of the different estimators described in Sect. 2, for which parameter choices and references are given in Table 1.

3.1 Characterization of linear proxy dependency

We first consider the linear dependency case, where the proxy history of $SS1$ is linearly correlated with that of $SS2$ a lag time ℓ later. We chose a length for the stalagmite of $L = 100mm$ for which we expect the time series to be roughly 100 years long (cf. Sect. 2.3.2) and linearly correlated, as in Fig. 4a. For each test 100 time series were generated from AR1 processes (cf. subsection 2.3.3), where process Y is coupled to process X at an intrinsic lag ℓ and with a coupling strength α. The autocorrelation parameter was set to $\phi = 0.8$, the coupling lag to $\ell = 5$ and the coupling parameter to $\alpha = 0.6$. For such stochastic processes, the true similarity function is single-peaked, with its peak height determined by α, and its location on the lag-axis by the coupling lag ℓ. The time series are irregular, therefore a direct scatterplot of the data is not possible. Fig. 4b shows a weighted scatterplot where the time series have been reconstructed using Gaussian weights, as for the MI estimation in Sect. 2.2.2.

The tests were guided by two questions: Do the similarity estimators reflect the actual similarity (here: the coupling strength at lag ℓ, α) truthfully and monotonically? And, how well do they identify the lag of coupling ℓ as the maximum of the similarity function?

To answer the first question we fix the imprecision at zero (at the dating points) and vary the coupling strength by setting the parameter α in Eq. 16 to values from 0.1 to 1. The results are given in Fig. 4c. The expected value of the similarity, α_{est}, and the variance of the estimate are computed from the mean and standard deviations of the estimated $\alpha_{\text{est},i}$ for 100 realizations for each value of the coupling parameter. Each of the similarity measures returns estimates whose expectation values increase monotonically with the actual similarity, α_{true} in Eq.16, except for the ESF, which has a single reversal which may be due to the low number of MC realizations (100) for each point in this diagram.

In practical data analysis, the potential lag and strength of (primary) coupling, identified as the maximum of the similarity function is of interest (e.g. for model-building). If no age uncertainty exists at the dating points, the maximum of the similarity function correctly identified in 50-60% of the ensemble cases. When time-scale uncertainty exists in the time series, this becomes difficult quickly (Fig. 4d). When the percentage has dropped to $\frac{1}{n_\ell} \approx 0.05$, where n_ℓ is the number of lags for which $S(\ell)$ has been estimated, the maxima of the similarity functions are perfectly uncorrelated. This limit is

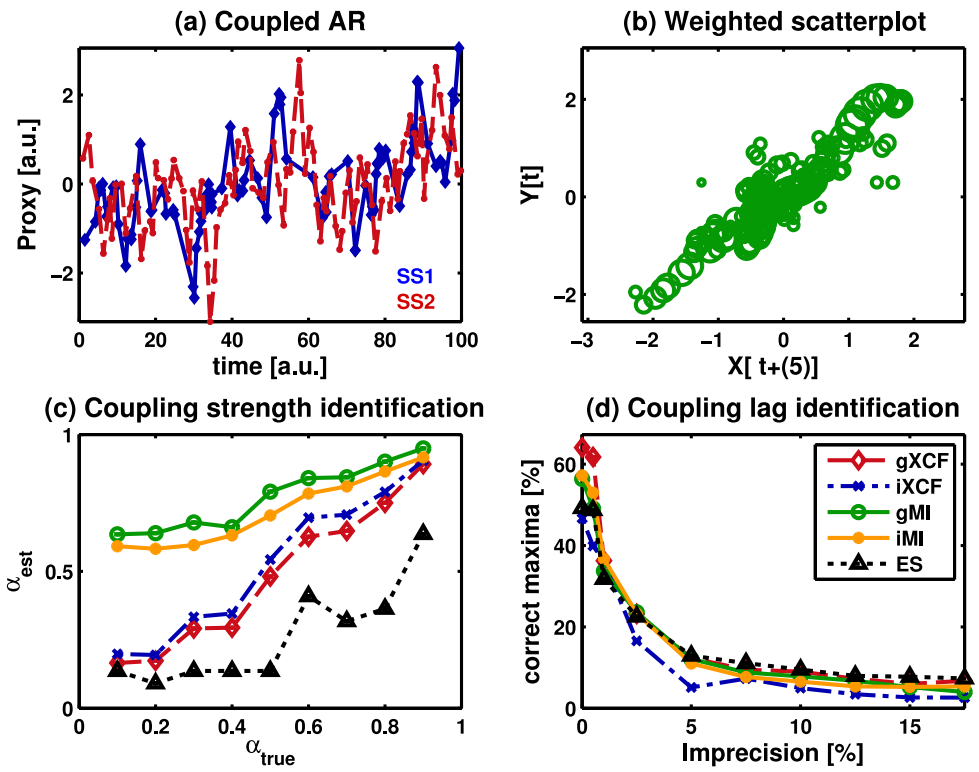

Fig. 4: Testing the similarity measures: for linearly coupled AR time series (cf. Eq. 16) from two synthetic stalagmites, SS1 and SS2, we give the sample time series **(a)** and the Gaussian weighted scatterplot **(b)**. We check the monotonicity of the estimators with increasing coupling strength **(c)** and how often the maximum of the similarity function correctly coincides with the lag of coupling **(d)**.

approached as an imprecision of more than 10% is reached. Increasing imprecision contained in the time series also results in increasing estimation error (i.e. RMSE, root mean square error) for the similarity at the lag of coupling, $S(\ell)$ (results not shown). When the stalagmite length is increased the time series length increases and both the RMSE and the false identification rate decrease for all estimators.

3.2 Nonlinear dependencies

For the nonlinear TAR model, the time series in Fig. 5a are not straightforward to compare visually as the linearly coupled ones in Fig. 4a. The weighted scatterplot for these time series in Fig. 5b shows the two different slopes of the positive and negative correlation regimes above and below the threshold value of zero.

The comparison of true vs. estimated coupling strength α in Fig. 5c shows no monotonous behavior for the linear correlation measures and no overall increase of their expected similarity estimates with the coupling strength. The MI estimators retain a monotonic increase, starting from a considerable bias value, while the ESF increases monotonically, but does not show consistent similarity estimate increases until the coupling strength is rather large. The monotonicity and linearity of the response for *gMI*, *iMI* and *ESF* improve considerably when the time series are chosen longer, i.e. with a length of 200 or more (results not shown).

In the identification of the maximum lag the Gaussian MI succeeds most often for imprecisions up to 2.5%. For more imprecise datasets the ESF remains stable, while the other measures perform worse and worse. The linear estimators, *gXCF* and *iXCF* do not identify the maxima correctly, neither the coupling strength, nor the lag of coupling.

3.3 Error source attribution

Age uncertainty has a considerable impact on the accuracy of similarity estimates, as we have shown in the previous section. But to what extent can this impact be attributed to short length of the time series, or the time series irregularity that results from the increasing age uncertainty? The uncertainty around the ages in the dating table is, in Monte-Carlo-based age-depth modeling, reflected by drawing different 'dates' from distributions around these ages for each MC realization. These realizations will therefore have different partial slopes between any date $\boldsymbol{D}_i$ and $\boldsymbol{D}_{i+1}$. This corresponds to

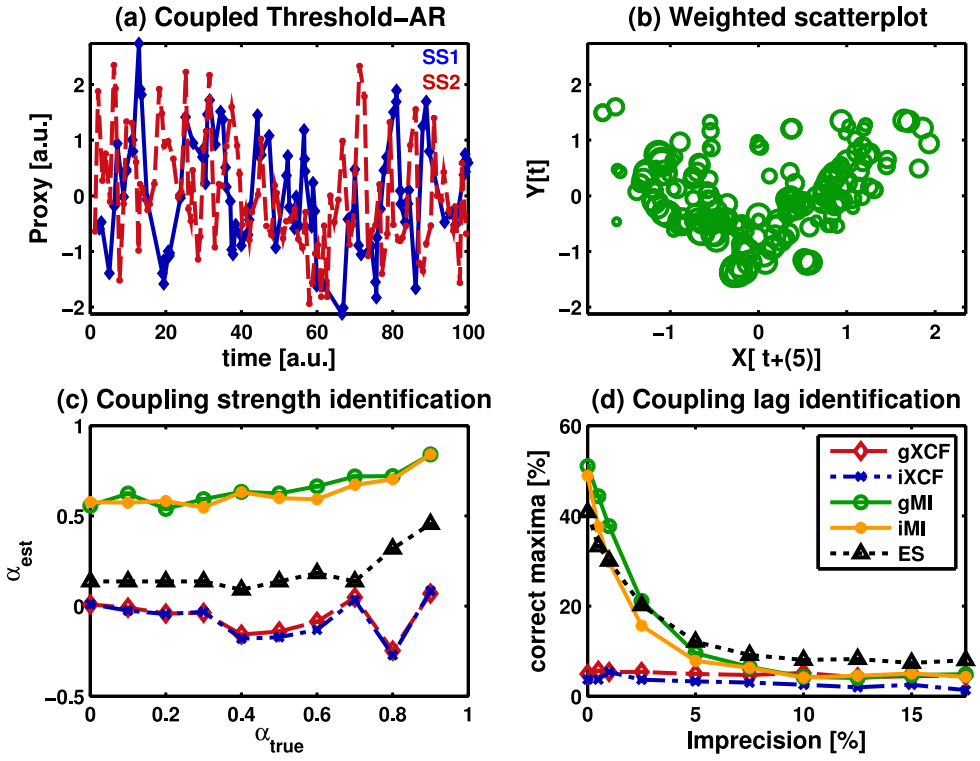

Fig. 5: Testing the similarity measures for nonlinear threshold-AR time series (cf. Eq. 17). For caption please refer to Fig. 4.

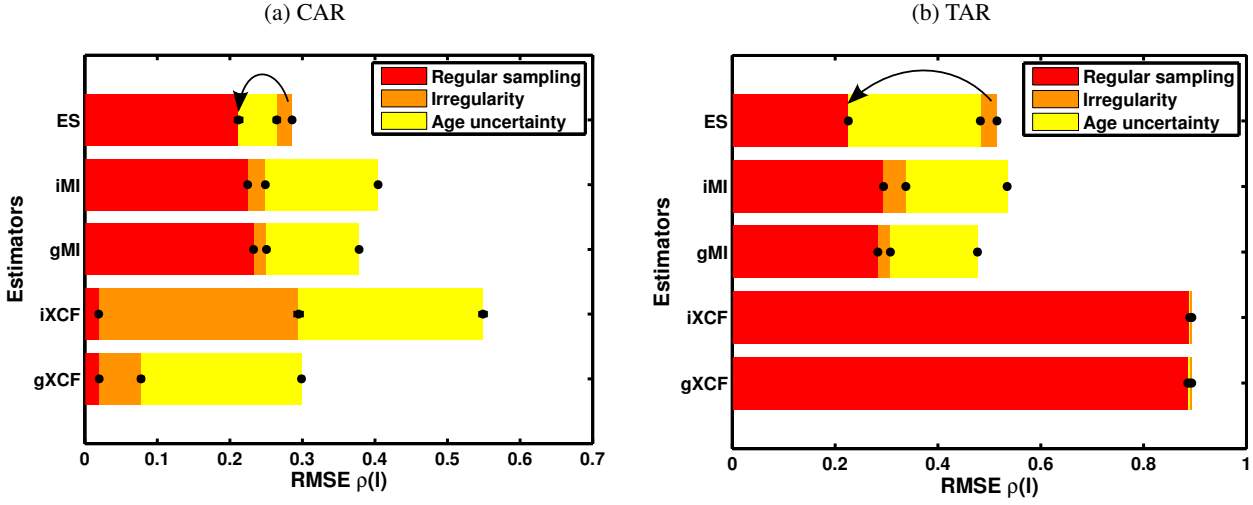

Fig. 6: Attribution of the uncertainty to its sources for (a) the linear CAR model and (b) the nonlinear TAR model: General (estimator) error in red, error introduced via irregular sampling (orange) and additional error due to the age uncertainty (yellow). The source-dependent RMSE was averaged over the second through to fifth imprecision levels given in Fig. 4d and 5d, as these correspond to the error levels most likely found in real world studies. Errorbars indicate the associated standard deviation. For event synchronization the RMSE is lower for irregular than regular sampling, folding the irregular part of the bar backwards.

different estimated growth rates for the individual segments of the synthetic core. At a proxy sampling rate over depth that is constant, this will lead to uneven observation times for the time series which correspond to the MC realizations, and this irregularity increases with the age uncertainty. The RMSE of $S(\ell)$ is, however, also dependent on the irregularity of the time series, as it was shown for both XCF and MI previously (Rehfeld et al., 2011, 2013).

To separate these sources of uncertainty, $M = 2000$ realizations of coupled climate histories, as defined in 2.3.2, were generated in three different ways: *age uncertain*, *irregularly* and *regularly sampled*. The age uncertain ensembles were the direct product of the age modeling efforts, as in the previous sections and with same parameter settings ($\phi = 0.8$, $\alpha = 0.9$, $\ell = 5$) For the irregular dataset the proxy histories were re-generated with the true coupling strength on the irregular timescales of the age modeling output. To assess the impact of regular sampling, regular time series of the same length, average temporal spacing and coupling scheme were also simulated. We evaluated the performance of the different estimators for the different sampling schemes at increasing dating imprecision using the *root mean square error* (RMSE) of the estimators for the target coupling parameter α:

$$\mathrm{RMSE}(\alpha_{\mathrm{est}}) = \sqrt{\mathrm{var}\left(\alpha_{\mathrm{est}}\right) + \mathrm{bias}\left(\alpha_{\mathrm{est}}\right)^2}\ , \tag{18}$$

where $\mathrm{bias}(\alpha_e st) = \alpha_{\mathrm{true}} - \alpha_{\mathrm{est}}$.

We did this separately for each sampling scheme to obtain the $\mathrm{RMSE}_{\mathrm{reg}}$, the 'baseline' RMSE for each estimator under regular sampling, $\mathrm{RMSE}_{\mathrm{irreg}}$ for the irregularly sampled ensembles and the $\mathrm{RMSE}_{\mathrm{au}}$ for the age uncertain ensemble. Coupling strength, autocorrelation and time series length were fixed to the same values for the three different sampling schemes. To improve the comparability for the MI estimators the bias offset was estimated from mutually uncorrelated time series with the same autocorrelation and length and subtracted prior to the conversion to the XCF scale.

Based on the assumption that the RMSE should increase from regular to irregular to age uncertain time series,

$$\mathrm{RMSE}_{\mathrm{reg}} < \mathrm{RMSE}_{\mathrm{irreg}} < \mathrm{RMSE}_{\mathrm{au}}\ ,$$

the 'baseline' contribution is estimated from regular time series as $\mathrm{RMSE}_{\mathrm{reg}}$, the additional contribution from time scale irregularity as $\mathrm{RMSE}_{\mathrm{irreg}} - \mathrm{RMSE}_{\mathrm{reg}}$ and the additional RMSE of the age uncertain time series' similarity as $\mathrm{RMSE}_{\mathrm{au}} - \mathrm{RMSE}_{\mathrm{irreg}}$.

The results, averaged over the realistic imprecision values (the 2nd-5fth points in Fig. 4d and 5d), are given in Fig. 6.

Ideally the RMSE should of course be as small as possible. For the linear (CAR) case in Fig.6a the smallest RMSE is observed for the ESF and the gXCF, the largest - by far - for the interpolation-based iXCF. While the regular (estimator) bias is low for the correlation estimators, the contribution of increasing irregularity of the time series sampling (due to the uncertain inputs) is non-negligible particularly for the interpolation-based cases. The age uncertainty alone accounts for additional, but generally smaller, error. While a large amount of the uncertainty of the interpolation-based estimators, *iMI* and *iXCF*, is due to sampling irregularity, ES has a large RMSE for regular time series, which is even higher than that for regular to slightly irregular time series. Therefore the contribution of irregular sampling to the cumulative uncertainty, as depicted in Fig. 6a, is negative, thus improving the estimation efficiency!

In the nonlinear (TAR) case the picture is quite different. The correlation-based estimators are not able to tell the coupling strength, regardless of the sampling scheme. The gMI estimator ranks lowest, with a lower uncertainty contribution from irregular sampling compared to the iMI estimator. The ESF, again, improves its accuracy when the time series are irregular. The overall error level is higher than for the linear case.

3.4 The link strength concept

Each of the tested similarity estimators comes with different underlying assumptions, estimator bias and variance, and they refer to different properties of the time series: the goodness of a linear fit to the joint distribution (XCF), the sharpness of the joint vs. the marginal distributions (MI) or the relative positions of extreme points, or events, in the time series (ES).

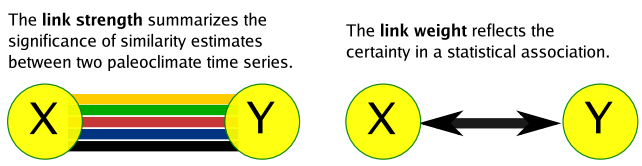

Fig. 7: The link strength concept: For each similarity estimator, significant results result in a link between the time series. The sum of these links determine the strength, or weight, of the link.

Therefore direct results obtained from the different estimators are difficult to compare, and they respond to coupling strength increases differently (Figs. 4c and 5c). The MI estimates, to this end, have to be converted to the XCF scale and thus are bound to the interval $[0,1]$, not $[-1,1]$ as for XC. This, together with the substantial and non-negative bias, induces a different proportionality between the actual coupling and the inferred association strength. Inferred ES, on the other hand, increases nonlinearly, but monotonically, with the coupling.

The main use of similarity measures is to assess the association strength between dynamics of processes. This can only be interpreted properly, if the significance of this estimate is known. To unify the results obtained from different similarity

estimators we propose to use a *link strength* $p(X,Y)$, to homogenize and summarize the results obtained for individual similarity measures.

The *link strength* $p(X,Y)$ for two observed time series X and Y is defined as the relative frequency of significant estimates from the N_{sim} employed estimators S_i:

$$p^q_{sim}(X,Y) = \frac{\sum_{i=1}^{N_{sim}} P_i(X,Y)}{N_{\text{sim}}} \,, \tag{19}$$

as illustrated in Fig. 7. The link strength of the individual estimators, $P^q_i(X,Y)$ is recorded on a binary scale:

$$P^q_i(X,Y) = \begin{cases} 1 & \text{if } S_i \text{ symmetric and } S_i^{\text{xy}} > S_i^{\text{hi,xy}} \\ 1 & \text{if } S_i \text{ asymmetric and} \\ & \left(S_i^{\text{xy}} > S_i^{\text{hi,xy}}\right) | \left(S_i^{xy} < S_i^{\text{lo,xy}}\right), \\ 0 & \text{otherwise} \,, \end{cases} \tag{20}$$

and here $S^{\text{hi/lo}}$ refer to the critical values of a hypothesis test, the null hypothesis being that both X and Y are autocorrelated, but mutually uncorrelated, Gaussian distributed stochastic processes. The significance q determines the critical values $S_i^{\text{hi,xy}}$ and $S_i^{\text{lo,xy}}$ which are obtained from the $q_{\text{hi}} = 1-0.5q$ and $q_{\text{lo}} = 0.5q$ quantiles of surrogate similarity estimates $S_i(X^*,Y^*)$.

Independent AR(1) surrogate time series X^* and Y^* are generated on the same time axes as X and Y according to Eq.15. The individual AR(1) persistence time for actual paleoclimate data can be obtained using an efficient least-squares fitting algorithm (Rehfeld et al., 2011; Mudelsee, 2002). The link strength can be extended to incorporate age uncertainties by computing the similarities for N_{mc} realizations of an age model and adding a second summation over these in Eq. 19.

4 Application to real stalagmite data

Now after having ensured the efficacy of the estimators using synthetic datasets, we apply the estimators to real-world stalagmite datasets from India, (the Dandak cave δ^{18}O record originally published in Sinha et al., 2007), and China (the Wanxiang record, Zhang et al., 2008). Comparisons of these datasets have been performed by Berkelhammer et al. (2010) and Rehfeld et al. (2011). Thirteen U/Th dates constrain the age model of the Dandak cave record, 19 are available for the Wanxiang cave record. Age modeling was performed on the full proxy datasets, comprising of 1875 and 703 oxygen isotope measurements over depth and using the COPRA algorithm with 1000 realizations (Breitenbach et al., 2012). The time series were cut to the overlapping time period from AD 600-1550 and detrended by subtracting the long-term mean. Berkelhammer et al. (2010) determined an averaged correlation of 0.27 for 50-year overlapping time windows, while Rehfeld et al. (2011) found a lag zero correlation coefficient of 0.290 and 0.295 for iXCF and gXCF, respectively. This

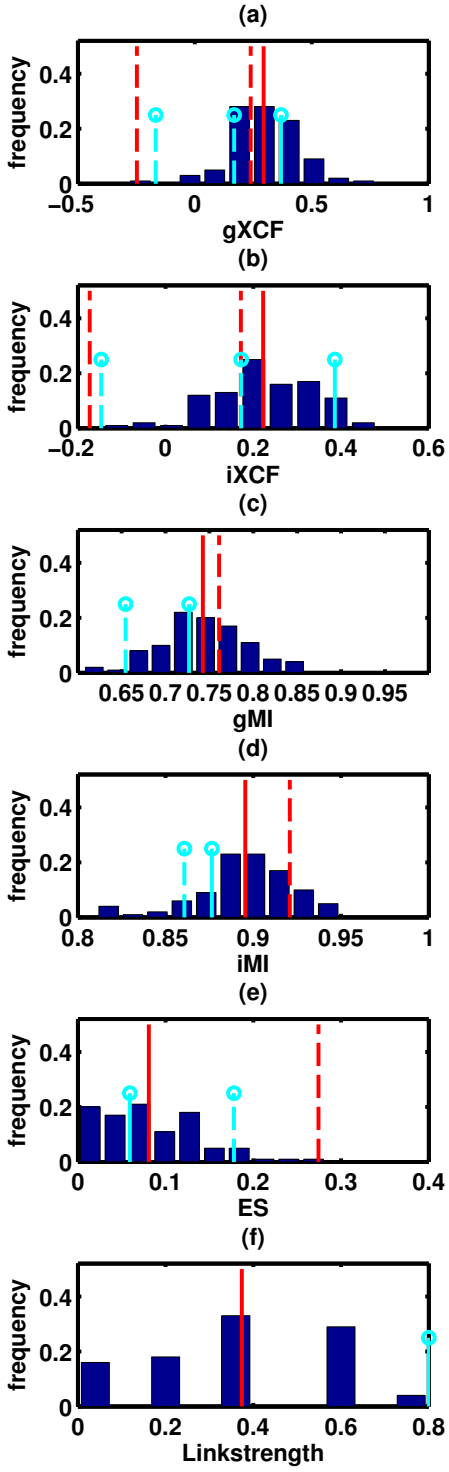

Fig. 8: Estimated lag zero similarities and link strength between the Dandak and Wanxiang cave records for the overlapping time period. The results for the age uncertain ensembles are given in the dark blue histograms. The red solid line refers to the mean of these estimates, the light blue stem to the results for the mean timescale. The dashed lines refer to the respective confidence intervals.

correlation was found to be significant to the 95%-level in the two-sided test for zero correlation, the null hypothesis being that the time series are autocorrelated but mutually uncorrelated.

Does this correlation persist, when the age uncertainties are considered in the analysis? We estimated the similarities for the two records considering all five estimators of Tab. 1 and for the original records as well as the results from age modeling, and give the results in Fig. 8. The histograms of similarity estimates for 100 realizations of the age models show a considerable spread. The mean similarity (indicated by the solid red line in Figs. 8a and 8b) for the correlation estimators is higher than that of the 95% quantile of the surrogate distribution, and the mean gMI estimate is close to the critical value. The median link strength is equal to 0.4. In contrast, the original age models published by Berkelhammer et al. (2010) and Zhang et al. (2008) yield significant results for all estimators except the ESF, resulting in an overall link strength of 0.8.

When we compute the similarities using the COPRA ensembles for the more sparse Dandak δ^{18}O time series published earlier (Sinha et al., 2007) the outcome is quite different - the link strength is only 0.2.

5 Discussion

Age uncertainty clearly affects all estimators of similarity for time series, and it is an illusion that it would be possible to mitigate the effects of uncertainty on the time axis for any type of analysis depending on observation times. Even if the observation – or accumulation – time of a grown archive is known precisely at some depths, an observation time reconstruction from age modeling requires an assumption on the accumulation behavior which, necessarily, will be wrong to some extent, as stochasticity and irregularity in the growth will always be present. This is a fact not challenged by the choice of a different interpolation routine, e.g. to a continuous cubic spline, which is often preferred by geoscientists (Breitenbach et al., 2012; Scholz and Hoffmann, 2011). On the positive side, and although counter-intuitive, incorporating (small) age uncertainty in the analysis might even improve the estimate when a deterministic (thus: necessarily wrong) assumption on the growth of the archive is made.

A low imprecision of 0-0.5% or an age uncertainty of approximately 1-2 years over a period of 200 years results in minimal relative estimation error and maximal confidence on the similarity peak position for the time series similarity functions $\hat{S}$. If a similarity analysis for real-world datasets covering a time span of 100 000 years was desired, this would amount to an 'allowed' age error of 500 years at a mean time series resolution of 500 years, which is a lower than what is usually found (Taylor et al., 2004). Thus, the resolution desired in the analysis is necessarily dependent on age uncertainty – only if that is lower, or comparable, an analysis of such short time series with full consideration of age uncertainties is feasible. One way to achieve higher certainty could be the incorporation of layer-counted ada in the agemodeling process, for example for annually laminated archives (Breitenbach et al., 2012).

The similarity estimators tested show different behavior, dependent on the signal type. The correlation-based estimators perform better for the linear coupling scheme, but fail for the nonlinear processes.

The **gXCF and iXCF** error split is dominated by the age uncertainty as the largest source of error in the linear CAR case. Both have small baseline bias for regular sampling. gXCF estimates coupling strength more effectively, however, for both age uncertainty and irregular sampling contributions of iXCF are significantly larger due to interpolation effects. In the nonlinear coupling scheme there is little difference whether the time series is regular, irregular or age uncertain – the correlation-based methods can not capture such type of dependencies.

gMI and iMI perform badly on the first glance in the linear CAR case, as their baseline bias for regular sampling RMSE is large. However, one needs to take into account that the RMSE is determined by both variance and bias – and that MI estimation, especially using binning estimators, is always associated with a significant positive bias, particularly for short time series. This bias, however, decreases with increasing time series length. If a direct comparison of MI and XC estimates is desired, this bias should be subtracted from the MI estimate prior to scaling it to the correlation scale. In the nonlinear TAR case the Gaussian-kernel-based version has the lowest overall RMSE.

The **ESF**, originally intended for the analysis of event series, performs well and has the lowest total RMSE, followed closely by gXCF, in the linear test case. There, its baseline RMSE dominates the RMSE split, and the RMSE for irregular sampling is *lower* than that for regular sampling. This is similar for the nonlinear processes. One reason for this might be that, for irregularly sampled time series of the same mean observation time distance, the number of observations spaced *closely* together is higher, which might increase the chances to find multiple events spaced closely together, resulting in effective *double-counting* of events. The comparably small contribution from age uncertainty in the linear test indicates, that neither the relative nor the absolute observation time distance between the time series are crucially important to the measure. Thus, it is quite a robust similarity measure with respect to age uncertainty and comparable to gXCF for linear coupling and gMI for nonlinear coupling, which both ultimately depend on the notion of simultaneous observations.

Although the irregularity of the time series is rather low (the inter-sampling-time distribution is narrow and close to normally distributed) the estimators that do not require the time series to be sampled regularly perform better than the interpolation-based records, which confirms the previous finding (Rehfeld et al., 2011, 2013) that large

sampling irregularity (i.e. the presence of gaps) leads to large interpolation bias, where the adapted estimators gXCF and gMI are particularly suitable. We have applied the similarity estimators to investigate the similarities between the Dandak and Wanxiang cave records. We find that the link strength aptly summarizes the results of the similarity significance tests: The time series are quite likely to be correlated, but age uncertainty blurs the results. There are several other parameters which can have a critical impact on the analysis: The choice of the significance level for link strength estimation, the detrending width and the respective resolution of the time series. The dependence of the results on the detrending parameter (Fig. 9) illustrates the time-scale dependence of the analysis: A small detrending width W results in a high-pass-filter and very low linkstrengths, large W yields high similarity on larger timescales. This indicates that the paleoclimatic records are more clearly associated at centennial to multi-centennial timescales than at decadal timescales, which are more impacted by age uncertainty. A higher temporal resolution of proxy measurements improves the accuracy of the estimators, particularly for the data-demanding MI estimators. Bootstrapping of the time series to successively lower lengths could be used to test the robustness of the estimators against such effects.

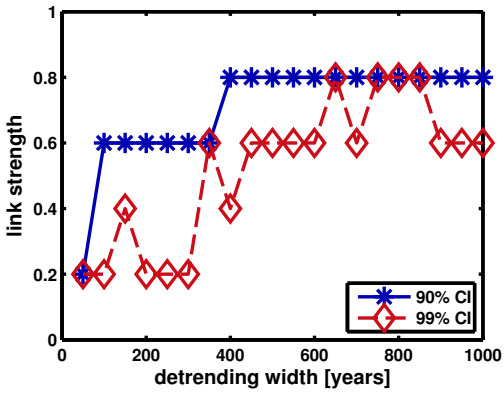

Fig. 9: Sensitivity of the link strength result for the original records of Berkelhammer et al. (2010) and Zhang et al. (2008) to changes in the detrending parameter W of a Gaussian-kernel detrending and the significance level in the hypothesis test.

We have only considered five similarity estimators, *gXCF*, *iXCF*, *gMI*, *iMI* and *ESF*, here, but this could be expanded for other concepts, for example based on (cross-)recurrence plots (Romano et al., 2005; Marwan et al., 2007; Marwan, 2002; Lange, 2011), recurrence networks (Feldhoff et al., 2012), convergent cross-mapping (Sugihara et al., 2012) or distance measures (Lhermitte et al., 2011). The notion of a link strength, instead of XC, MI or ES values, makes it straightforward to extend the analysis to a whole ensemble of time series, be it from age modeling or out of a database of paleoclimate records. If age uncertainty does not impact the cross-similarity, the link strength will not drop substantially. The actual value of the link strength can be interpreted in terms of a "degree of confidence": If the value is close to the significance level, a relationship cannot be concluded with confidence. If the link strength is close to one, all the estimators return significant similarity estimates and a similarity can be deduced with certainty.

In the future it could be evaluated whether p-values from the the surrogate tests can replace the binary thresholding for the link strength metric to improve the sensitivity of the link strength estimate. The ESF alone, however, could be particularly suitable for the analysis of extreme events since it does not place strong restrictions on the time series beyond stationarity, and performs particularly well for irregular time series.

The NESToolbox containing scripts and programs for the similarity analysis of age uncertain time series in Matlab and the open source software Octave are available with this paper. We also include a function to simulate age uncertainties that arise for archives for which the chronology is based on layer counting, trees, ice cores or laminated sediments, so that these, too, can be investigated using the methods presented in this paper.

6 Conclusions

In this paper we have investigated similarity estimators that do not require regular sampling in time and can capture linear (*gXCF*) and nonlinear (*gMI* and *ESF*) relationships. We found that interpolation to regular spacing of the observation times results in worse estimates. By contrast, the adapted estimators are more efficient in the presence of sampling time irregularity and cope with age uncertainty better. Table 1 gives a comprehensive overview over the similarity estimators, parameter choices and further references. *gXCF* and *ESF* perform particularly well if the relationship is linear, but the correlation estimator fails in the presence of nonlinear coupling, where the *ESF* and *gMI* are better suited to infer dependences. The significance of results from different estimators and under varying time series length and sampling can be unified using the concept of a link strength. It combines similarity estimators and significance tests and is given by the relative frequency of positive significance tests and could be especially useful in the analysis of large paleoclimatic datasets where it is infeasible to check each pair of time series for similarity individually. We have shown that age uncertainty is the largest contributor to estimation error for time series similarity, and for a reliable of similarity function shape and coupling structure, the time scale imprecision should be as low as possible. When it exceeds 5% of the time series length coupling phenomena on time-scales close to the sampling resolution can not be deduced any longer.

While time series irregularity can be well addressed by the use of the adapted estimators, age uncertainty can not, and should therefore be reduced as much as possible by measuring more ages, improved dating techniques or the use of additional temporal information from layer-counting (Breitenbach et al., 2012) where possible. These are, in essence, good news, because the irregular growth of the archives cannot be reversed, but measurement devices can be optimized.

Acknowledgements. The authors thank Norbert Marwan, Jobst Heitzig, Bedartha Goswami and Sebastian Breitenbach for helpful comments and discussion, and Franziska Lechleitner for assistance with data pre-processing. We thank Ashish Sinha for providing us with the depth data for the Dandak cave stalagmite and Richard Telford and an anonymous reviewer for their constructive feedback. This work has been financially supported by the Federal Ministry for Education and Research (BMBF) via the Potsdam Research Cluster for Georisk Analysis, Environmental Change and Sustainability (PROGRESS). The NESToolbox containing software tools to handle irregularly sampled datasets can be found on *tocsy.pik-potsdam.de/nest.php*.

References

Babu, P. and Stoica, P.: Spectral analysis of nonuniformly sampled data – a review, Digital Signal Processing, 20, 359–378, doi:10.1016/j.dsp.2009.06.019, 2010.

Batyrshin, I., Sheremetov, L., and Velasco-Hernandez, J. X.: On axiomatic definition of time series shape association measures, in: Operations Research and Data Mining ORADM 2012 workshop proceedings, edited by Villa-Vargas, U., Sheremetov, L., and Haasis, H.-D., pp. 1–12, National Polytechnic Institute, Mexico City, 2012.

Berkelhammer, M., Sinha, A., Mudelsee, M., Cheng, H., Edwards, R. L., and Cannariato, K.: Persistent multidecadal power of the Indian Summer Monsoon, Earth and Planetary Science Letters, 290, 166–172, doi:10.1016/j.epsl.2009.12.017, 2010.

Breitenbach, S. F. M., Rehfeld, K., Goswami, B., Baldini, J. U. L., Ridley, H. E., Kennett, D. J., Prufer, K. M., Aquino, V. V., Asmerom, Y., Polyak, V. J., Cheng, H., Kurths, J., and Marwan, N.: COnstructing Proxy Records from Age models (COPRA), Climate of the Past, 8, 1765–1779, doi:10.5194/cp-8-1765-2012, 2012.

Chatfield, C.: The analysis of time series: an introduction, CRC Press, Florida, US, 6th edn., 2004.

Cheng, H., Zhang, P. Z., Spötl, C., Edwards, R. L., Cai, Y. J., Zhang, D. Z., Sang, W. C., Tan, M., and An, Z. S.: The climatic cyclicity in semiarid-arid central Asia over the past 500,000 years, Geophysical Research Letters, 39, 1–5, doi:10.1029/2011GL050202, 2012.

Cover, T. and Thomas, J.: Elements of information theory, John Wiley & Sons, Inc., Hoboken, New Jersey, 2 edn., 2006.

Donges, J. F., Zou, Y., Marwan, N., and Kurths, J.: Complex networks in climate dynamics, The European Physical Journal Special Topics, 174, 157–179, doi:10.1140/epjst/e2009-01098-2, 2009.

Dykoski, C., Edwards, R., Cheng, H., Yuan, D., Cai, Y., Zhang, M., Lin, Y., Qing, J., An, Z., and Revenaugh, J.: A high-resolution, absolute-dated Holocene and deglacial Asian monsoon record from Dongge Cave, China, Earth and Planetary Science Letters, 233, 71–86, doi:10.1016/j.epsl.2005.01.036, 2005.

Fairchild, I. and Baker, A.: Speleothem Science: from process to past environments, Wiley-Blackwell, 2012.

Feldhoff, J. H., Donner, R. V., Donges, J. F., Marwan, N., and Kurths, J.: Geometric detection of coupling directions by means of inter-system recurrence networks, Physics Letters A, 376, 3504–3513, doi:10.1016/j.physleta.2012.10.008, 2012.

Hlinka, J., Hartman, D., Vejmelka, M., Runge, J., Marwan, N., Kurths, J., and Paluš, M.: Reliability of Inference of Directed Climate Networks Using Conditional Mutual Information, Entropy, 15, 2023–2045, doi:10.3390/e15062023, 2013.

Khan, S., Bandyopadhyay, S., Ganguly, A., Saigal, S., Erickson, D., Protopopescu, V., and Ostrouchov, G.: Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data, Physical Review E, 76, 1–15, doi:10.1103/PhysRevE.76.026209, 2007.

Kraskov, A., Stögbauer, H., and Grassberger, P.: Estimating mutual information, Physical Review E, 69, 1–16, doi:10.1103/PhysRevE.69.066138, 2004.

Kreuz, T., Chicharro, D., Andrzejak, R. G., Haas, J. S., and Abarbanel, H. D. I.: Measuring multiple spike train synchrony., Journal of neuroscience methods, 183, 287–99, doi:10.1016/j.jneumeth.2009.06.039, 2009.

Lange, H.: Recurrence Quantification Analysis in Watershed Ecosystem Research, International Journal of Bifurcation and Chaos, 21, 1113–1125, doi:10.1142/S0218127411028921, 2011.

Lhermitte, S., Verbesselt, J., Verstraeten, W., and Coppin, P.: A comparison of time series similarity measures for classification and change detection of ecosystem dynamics, Remote Sensing of Environment, 115, 3129–3152, doi:10.1016/j.rse.2011.06.020, 2011.

Malik, N., Marwan, N., and Kurths, J.: Spatial structures and directionalities in Monsoonal precipitation over South Asia, Nonlinear Processes in Geophysics, 17, 371–381, doi:10.5194/npg-17-371-2010, 2010.

Malik, N., Bookhagen, B., Marwan, N., and Kurths, J.: Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks, Climate Dynamics, 39, 971–987, doi:10.1007/s00382-011-1156-4, 2011.

Marwan, N.: Nonlinear analysis of bivariate data with cross recurrence plots, Physics Letters A, 302, 299–307, doi:10.1016/S0375-9601(02)01170-2, 2002.

Marwan, N., Romano, M. C., Thiel, M., and Kurths, J.: Recurrence plots for the analysis of complex systems, Physics Reports, 438, 237–329, doi:10.1016/j.physrep.2006.11.001, 2007.

Mudelsee, M.: TAUEST: a computer program for estimating persistence in unevenly spaced weather/climate time series, Computers & Geosciences, 28, 69–72, doi:10.1016/S0098-3004(01)00041-3, 2002.

Nazareth, D., Soofi, E., and Zhao, H.: Visualizing Attribute Interdependencies Using Mutual Information, Hierarchical Clustering, Multidimensional Scaling, and Self-organizing Maps, 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07), pp. 53–53, doi:10.1109/HICSS.2007.608, 2007.

Papana, A. and Kugiumtzis, D.: Evaluation of mutual information estimators for time series, International Journal of Bifurcation and Chaos, 19, 4197–4215, doi:10.1142/S0218127409025298, 2009.

Quian Quiroga, R., Kreuz, T., and Grassberger, P.: Event synchronization: A simple and fast method to measure synchronicity and time delay patterns, Physical Review E, 66, 041 904, doi:10.1103/PhysRevE.66.041904, 2002.

Rehfeld, K., Marwan, N., Heitzig, J., and Kurths, J.: Comparison of correlation analysis techniques for irregularly sampled time series, Nonlinear Processes in Geophysics, 18, 389–404, doi:10.5194/npg-18-389-2011, 2011.

Rehfeld, K., Marwan, N., Breitenbach, S. F. M., and Kurths, J.: Late Holocene Asian Summer Monsoon dynamics from small but complex networks of palaeoclimate data, Climate Dynamics, 41, 3–19, doi:10.1007/s00382-012-1448-3, 2013.

Rheinwalt, A., Marwan, N., Kurths, J., Werner, P., and Gerstengarbe, F.-W.: Boundary effects in network measures of spatially embedded networks, EPL (Europhysics Letters), 100, 28 002, doi:10.1209/0295-5075/100/28002, 2012.

Romano, M. C., Thiel, M., Kurths, J., Kiss, I. Z., and Hudson, J. L.: Detection of synchronization for non-phase-coherent and non-stationary data, Europhysics Letters (EPL), 71, 466–472, doi:10.1209/epl/i2005-10095-1, 2005.

Roulston, M.: Estimating the errors on measured entropy and mutual information, Physica D: Nonlinear Phenomena, 125, 285–294, 1999.

Runge, J., Heitzig, J., Marwan, N., and Kurths, J.: Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy, Physical Review E, 86, 061 121, doi:10.1103/PhysRevE.86.061121, 2012.

Scargle, J. D.: Studies in astronomical time series analysis. III - Fourier transforms, autocorrelation functions, and cross-correlation functions of unevenly spaced data, The Astrophysical Journal, 343, 874, doi:10.1086/167757, 1989.

Scholz, D. and Hoffmann, D. L.: StalAge – An algorithm designed for construction of speleothem age models, Quaternary Geochronology, 6, 369–382, doi:10.1016/j.quageo.2011.02.002, 2011.

Schulz, M. and Stattegger, K.: SPECTRUM: spectral analysis of unevenly spaced paleoclimatic time series, Computers & Geosciences, 23, 929–945, doi:10.1016/S0098-3004(97)00087-3, 1997.

Sinha, A., Cannariato, K. G., Stott, L. D., Cheng, H., Edwards, R. L., Yadava, M. G., Ramesh, R., and Singh, I. B.: A 900-year (600 to 1500 A.D.) record of the Indian summer monsoon precipitation from the core monsoon zone of India, Geophysical Research Letters, 34, 1–5, doi:10.1029/2007GL030431, 2007.

Sinha, A., Stott, L., Berkelhammer, M., Cheng, H., Edwards, R. L., Buckley, B., Aldenderfer, M., and Mudelsee, M.: A global context for megadroughts in monsoon Asia during the past millennium, Quaternary Science Reviews, 30, 47–62, doi:10.1016/j.quascirev.2010.10.005, 2011.

Stoica, P. and Sandgren, N.: Spectral analysis of irregularly-sampled data: Paralleling the regularly-sampled data approaches, Digital Signal Processing, 16, 712–734, doi:10.1016/j.dsp.2006.08.012, 2006.

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., and Munch, S.: Detecting causality in complex ecosystems, Science, 338, 496–500, doi:10.1126/science.1227079, 2012.

Taylor, K. C., Alley, R. B., Meese, D. A., Spencer, M. K., Brook, E. J., Dunbar, N. W., Finkel, R. C., Gow, A. J., Kurbatov, A. V., Lamorey, G. W., Mayewski, P. A., Meyerson, E. A., Nishiizumi, K., and Zielinski, G. A.: Dating the Siple Dome (Antarctica) ice core by manual and computer interpretation of annual layering, Journal of Glaciology, 50, 453–461, doi:10.3189/172756504781829864, 2004.

Telford, R., Heegaard, E., and Birks, H.: All age–depth models are wrong: but how badly?, Quaternary Science Reviews, 23, 1–5, doi:10.1016/j.quascirev.2003.11.003, 2004.

Tsay, R.: Testing and modeling threshold autoregressive processes, Journal of the American Statistical Association, 84, 231–240, 1989.

Webster, J., Brook, G., and Railsback, L.: Stalagmite evidence from Belize indicating significant droughts at the time of Preclassic Abandonment, the Maya Hiatus, and the Classic Maya collapse, Palaeogeography, Palaeoclimatology, Palaeoecology, 250, 1 – 17, doi:10.1016/j.palaeo.2007.02.022, 2007.

Yadava, M., Ramesh, R., and Pant, G.: Past monsoon rainfall variations in peninsular India recorded in a 331-year-old speleothem, The Holocene, 14, 517–524, doi:10.1191/0959683604hl728rp, 2004.

Zhang, J., Chen, F., Holmes, J. A., Li, H., Guo, X., Wang, J., Li, S., Lü, Y., Zhao, Y., and Qiang, M.: Holocene monsoon climate documented by oxygen and carbon isotopes from lake sediments and peat bogs in China: a review and synthesis, Quaternary Science Reviews, 30, 1973–1987, doi:10.1016/j.quascirev.2011.04.023, 2011.

Zhang, P., Cheng, H., Edwards, R. L., Chen, F., Wang, Y., Yang, X., Liu, J. J. J. J., Tan, M., Wang, X., An, C., Dai, Z., Zhou, J., Zhang, D., Jia, J., Jin, L., and Johnson, K. R.: A test of climate, sun, and culture relationships from an 1810-year Chinese cave record, Science, 322, 940–2, doi:10.1126/science.1163965, 2008.

Zwiers, F. and Storch, H. V.: Regime-dependent autoregressive time series modeling of the Southern Oscillation, Journal of climate, 3, 1347–1363, 1990.