https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions

A statistical approach to the phasing of atmospheric reorganization and sea ice retreat at the onset of Dansgaard-Oeschger events under rigorous treatment of uncertainties

Keno Riechers[1,2] and Niklas Boers[1,2,3]

[1]Department of Mathematics and Computer Science, Free University Berlin, Berlin, Germany
[2]Potsdam Institute for Climate Impact Research, Potsdam, Germany
[3]Department of Mathematics and Global Systems Institute, University of Exeter, Exeter, UK

Correspondence: Keno Riechers (riechers@pik-potsdam.de)

Abstract. For previous glacial intervals, concomitant shifts in different proxy records from Greenland ice cores indicate repeated abrupt climate transitions comprising - among others - abrupt warming, a sudden reorganization of the atmospheric circulation, and a retreat of perannial sea ice. The physical mechanism underlying these so-called Dansgaard-Oeschger (DO) events remains debated. Former studies have made an effort to deduce the progression of temperature, circulation, and sea-ice changes at the onset of DO events from paleoclimate proxy records to constrain potential triggering mechanisms. In this context, recent research reports on systematically delayed transitions in Na^+ concentrations and $\delta^{18}O$ values compared to Ca^{2+} concentrations and the annual layer thickness by about one decade. This is interpreted as a temporal lag of sea ice retreat and Greenland warming with respect to atmospheric reorganization at the onset of DO-events. Here, we present a comprehensive statistical analysis of the relative phasing of DO transitions in Ca^{2+} and Na^+ concentration records from the NGRIP ice core for the period 60 - $10\,\mathrm{kyr\,BP}$. Regarding the time lags identified in this period as a sample generated from an unknown population, we derive probability density functions for the sample and population mean and test the null-hypothesis of a simultaneous transition. Special attention was paid to the uncertainties inherent to the transition onset detection in noisy data. Their rigorous propagation changes the test results from significant to non-significant and therefore a purely stochastic origin of the observed tendency for Ca^{2+} to lead the transition cannot be ruled out. In fact, we show that the data is very likely to comprise both: DO events that were led by a Ca^{2+} transition, as well as events led by a Na^+ transition. Together, these findings clearly contradict a systematic lead or lag between the DO transitions in the two proxies, and the apparent Ca^{2+} lead should therefore not be interpreted as indication of a causal relationship. Under the assumption that all DO events followed the same physical mechanism and that the proxy interpretation holds true, the we conclude that at DO transition onsets, neither was the atmospheric reorganization caused by sea ice retreat, nor was the sea ice retreat triggered by atmospheric reorganization.

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions
Open Access
EGU

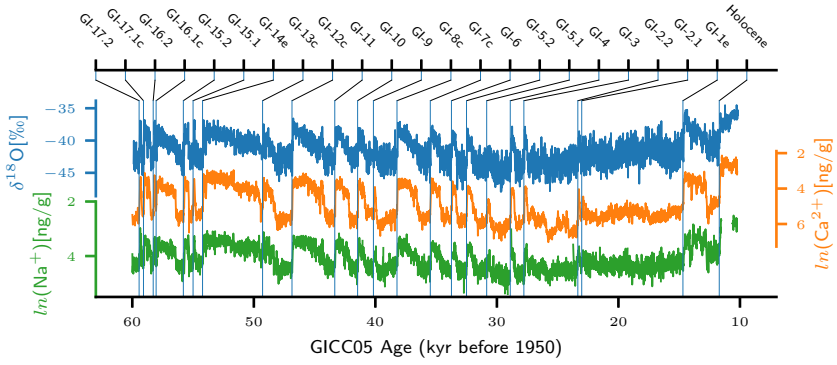

Figure 1. Time series of δ^{18}O (blue), Ca^{2+} (orange) and Na^{+} (green) from the NGRIP ice core on the GICC05 timescale in ky BP, at 10-year resolution. Light blue vertical lines mark the timings DO events. Data retrieved from Erhardt et al. (2018) and DO event timings are according to Rasmussen et al. (2014).

1 Introduction

In view of anthropogenic global warming, concerns have been raised that several subsystems of the earth's climate system may undergo abrupt and fundamental changes of state if temperatures are to exceed corresponding critical thresholds (Lenton and Schellnhuber, 2007; Lenton et al., 2008, 2019). Under sustained warming, the Atlantic Meridional Overturning Circulation (AMOC), the Amazon rainforest, or the Greenland ice sheet are possible candidates among others to abruptly transition to new equilibrium states that may differ strongly from their current states (Lenton et al., 2008). Understanding the physical mechanisms behind abrupt shifts in climatic subsystems is crucial for assessing the associated risks and for defining safe operating spaces in terms of cumulative greenhouse gas emissions. To date, empirical evidence for abrupt climate transitions only comes from paleoclimate proxy records encoding climate variability in the long-term past. First discovered in the δ^{18}O records from Greenland ice cores, the so-called Dansgaard-Oeschger (DO) events are considered the archetype of past abrupt climate changes (see Fig. 1) (Johnsen et al., 1992; Dansgaard et al., 1993; Bond et al., 1993; Andersen et al., 2004). These events constitute a series of abrupt regional warming transitions that punctuated the last and previous glacial intervals at millennial recurrence periods. Amplitudes of the decadal-scale temperature increases reach from 5 °C to 16.5 °C over Greenland (Kindler et al., 2014; Huber et al., 2006; Landais et al., 2005). The abrupt warming is followed by gradual cooling over centuries or even a few millennia, before the climate abruptly transitions back to cold conditions. The relatively cold (warm) intervals within the glacial episodes have been termed Greenland stadials (GS) (Greenland interstadials (GI)). GS typically persist over millennial time scale, before another abrupt warming starts a new cycle (Rasmussen et al., 2014). Despite being less pronounced, a global impact of DO events on climate and ecosystems is evident in manifold proxy records (e.g., Moseley et al., 2020; Buizert et al., 2015; Lynch-Stieglitz, 2017; Kim et al., 2012; Fleitmann et al., 2009; Voelker, 2002; Cheng et al., 2013). Apart from δ^{18}O,

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020


Climate of the Past Discussions

other Greenland ice core proxy variables such as Ca^{2+} and Na^{+} also bear the signature of DO cycles as can be seen in Fig. 1 (e.g., Erhardt et al., 2019; Fuhrer et al., 1999; Ruth et al., 2007). While $\delta^{18}O$ is interpreted as a qualitative proxy for ice core site temperatures (Gkinis et al., 2014, and references therein), changes in Ca^{2+} – or equivalently dust – are believed to reflect changes in global atmospheric circulations (Ruth et al., 2007; Erhardt et al., 2019). Na^{+} proxy records indicate past sea-salt aerosol concentrations and are thought to negatively correlate with the North Atlantic sea ice cover (Erhardt et al., 2019; Schüpbach et al., 2018). With the trigger mechanism for DO events still under debate, there have been several attempts to investigate the relative temporal order of the Greenland warming, the potential reorganization of the atmospheric circulation, sea-ice retreat, and increasing precipitation across the DO transitions, by analyzing the phasing of abrupt shifts detected in corresponding multi-proxy time series from Greenland ice cores (Erhardt et al., 2019; Thomas et al., 2009; Steffensen et al., 2008; Ruth et al., 2007). While Thomas et al. (2009) and Steffensen et al. (2008) report delayed Greenland warming with respect to atmospheric changes for the onsets of GI-8 and GI-1 and the Holocene, Ruth et al. (2007) find no systematic lead or lag between NGRIP dust concentration changes and $\delta^{18}O$ changes across the onsets of GI-1 to GI-24. However, the comprehensive study conducted by Erhardt et al. (2019) concludes that on average, changes in both terrestrial dust aerosol concentrations (Ca^{2+}) and local precipitation have preceded the changes in local temperatures ($\delta^{18}O$) and sea salt aerosol concentrations (Na^{+}) across the DO events during the last glacial cycle. These observation-based studies are complemented by numerous conceptual theories and modeling studies that explore a variety of mechanisms to explain the DO events. Many authors emphasize the role of the AMOC in the emergence of DO events (Broecker et al., 1985; Clark et al., 2002; Ganopolski and Rahmstorf, 2001; Henry et al., 2016). In that context, Vettoretti and Peltier (2018) observed a self-sustained sea-salt oscillation mechanism to initiate the transitions between stadials and interstadials in a GCM model run, while Boers et al. (2018) proposed a coupling between sea-ice growth, subsurface-warming, and AMOC changes to explain the DO cycles. Moreover, Li and Born (2019) draw attention to the subpolar gyre, a sensitive region that features strong interactions between atmosphere, ocean and sea ice. In line with the empirical studies that suggest a delayed Greenland warming with respect to atmospheric changes, Kleppin et al. (2015) and Zhang et al. (2014) find DO like transitions in model studies triggered by an abrupt reorganization of atmospheric circulation patterns.

This study systematically refines the investigation of a potential Ca^{2+} lead with respect to Na^{+} at DO transition onsets. Our work is based on piece-wise high-resolution multi-proxy time series around 16 DO-transitions from the second half of the last glacial, retrieved from the NGRIP ice core, and a probabilistic transition detection method, both presented by Erhardt et al. (2019). We present a statistical framework for the stringent treatment of uncertainties that arise from the transition detection in paleoclimate time series and investigate the significance of the postulated lag between Ca^{2+} and Na^{+}. Generally, threefold uncertainty makes the study of abrupt transition in paleoclimate records challenging (Goswami et al., 2018). First, the proxy data is subject to ordinary measurement uncertainty. Second, dating uncertainties in each individual record (Boers et al., 2017; Breitenbach et al., 2012) mostly prevent any meaningful cross-archive intercomparison on the time scale relevant for abrupt events (and sometimes even within the same archives, for example when gas proxies are compared to solid proxies in ice cores). Third, high-frequency internal climate variability blurs abrupt transitions. Taking the perspective of stochastic differential equations, drift and diffusion can in most practical cases not be separated unequivocally and therefore, the onset

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions

Table 1. List of DO events (Greenland Interstadial onsets) for which high-resolution Ca^{2+} and Na^{+} concentration data for windows centered at the transition are available. Bold print indicates those events for which the stochastic, MCMC-based method successfully detected empirical density distributions for transition onset. $\langle \Delta t \rangle$ indicates the expected time lag between the Ca^{2+} and the Na^{+} transition onsets. $P(\Delta t > 0)$ gives the probability for events to exhibit a positive Ca^{2+}–Na^{+} lag, i.e., a sodium lead at the transition onset. Dates are according to Rasmussen et al. (2014).

Event	age [yr BP]	$\langle \Delta t \rangle$ [yr]	$P(\Delta t > 0)$
Holocene	11703	4.83	0.65
GI-1e	14692	-10.67	0.27
GI-2.1	23020		
GI-2.2	23340		
GI-3	27780	2.02	0.56
GI-4	28900		
GI-5.1	30840		
GI-5.2	32500	-5.76	0.37
GI-6	33740		
GI-7c	35480	-4.09	0.34
GI-8c	38220	-0.53	0.48
GI-9	40160		
GI-10	41460	-2.64	0.38
GI-11	43340	32.50	0.85
GI-12c	46860	-12.88	0.21
GI-13c	49280	-4.45	0.38
GI-14e	54220	-2.49	0.39
GI-15.1	55000	-18.11	0.23
GI-15.2	55800	-16.1	0.29
GI-16.1c	58040		
GI-16.2	58280	-15.89	0.10
GI-17.1c	59080	-21.30	0.03
GI-17.2	59440	-14.36	0.18

of any abrupt change in the data can be determined only with limited precision (see Fig. 2(a)). Besides these quantitative uncertainties, interpretation of proxy data is always subject to qualitative uncertainty, since there is no exact mapping from the proxy variable to the climate variable that the proxy is supposed to represent.

We follow the lines of (Erhardt et al., 2019) and use a Bayesian Markov-Chain-Monte-Carlo (MCMC)-based algorithm to infer posterior probability distributions for the DO transition onset time t_0 in the set of Ca^{2+} and Na^{+} time series. This approach captures the uncertainties in the transition onset determination most conveniently and allows for the derivation of a stochastic sample of Ca^{2+}–Na^{+} lags. Given this stochastic sample, we propose a method to estimate the mean of an underlying population that is assumed to have generated the small observed sample of lags. Additionally, we use standard tests to assess the significance of leads or lags between Ca^{2+} and Na^{+} transition onsets. A comparison is given between the application

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020


Climate of the Past Discussions

of the tests to the expected sample, where uncertainties in the representation of lags have been averaged out, and a rigorous propagation of the uncertainties to the resulting p-values. Finally, we derive the probabilities for $n_{Ca^{2+}}$ out of the 16 events under study to feature a leading Ca^{2+} transition. Together, these methods form a comprehensive statistical analysis and allow to interpret the data with regard to the question of relative phasing and potential causality between atmospheric reorganization and sea ice retreat.

This article is structured as follows: First, the data used for the study is briefly described. Second, we generically introduce our methods, in order to facilitate potential adaptation to structurally similar problems. Attention has been paid to their sound mathematical formulation. This is followed by the presentation of our results together with their statistical interpretation. We underpin and extend our interpretation in the subsequent discussion. The last section consicely summarizes the key conclusion that can be drawn from our analysis.

2 Data

Recently, Erhardt et al. (2018) published 24 highly resolved Ca^{2+} and Na^{+} time series from the NGRIP ice core for time intervals of 250 to 500 years centered around DO events from the period 60-10 kyr BP. The data set covers all interstadial onsets from GI-17.2 to the Holocene, as determined by Rasmussen et al. (2014) (see Tab. 1). The time resolution decreases from 2 to 4 years with increasing depth in the NGRIP ice core due to the compression of the core. With Ca^{2+} deposition rates reflecting the state of the atmospheric circulation and Na^{+} concentrations being related to the sea ice extent in the North Atlantic, the comparison of these two proxies from the same archive gives insights into the sequence of events leading to DO events (Erhardt et al., 2019, and references therein). Exact co-registration of the ion concentrations guarantees the absence of any relative dating uncertainties that would impede the approach. For details on the measurement process we refer to Erhardt et al. (2019).

3 Methods

Here, we present a generic yet detailed description of the methods we use to infer characteristics of a generating population from a sample which is subject to uncertainty. First, we introduce the stochastic transition onset detection algorithm that by design returns an uncertain transition onset t_0 in form of a posterior probability density distribution. Application of this algorithm to the DO events present in the data from Sec. 2 yields an uncertain sample of transition onset lags $\mathbf{\Delta t} = (\Delta t_1, ..., \Delta t_{16})$ between the calcium and the sodium transition - the starting point for our statistical analysis. Second, a statistical perspective on the series of DO events is presented, with special emphasize on the particularity of two levels of randomness being involved. This is followed by some remarks on the numerical treatment of high dimensional probability densities. Subsequently, inference of the population mean and application of hypothesis tests are discussed under rigorous propagation of the samples' uncertainties. We highlight that situations similar to ours, where a small-to-medium-size sample generated from an underlying population is subject to uncertainty, may occur in many different disciplines and the content of this section may be adapted to such

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions
Open Access
EGU

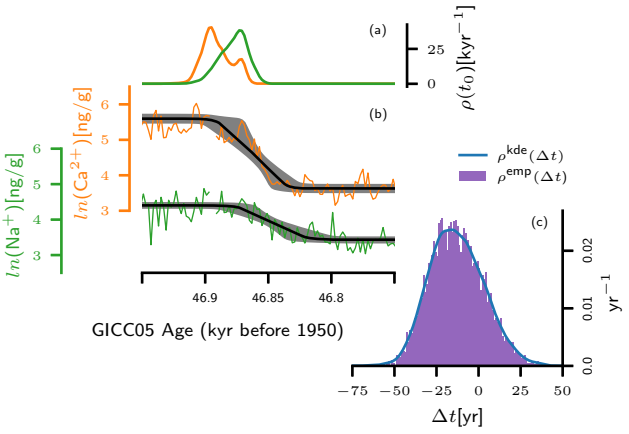

Figure 2. (a) Posterior probability distribution $\rho(t_0)$ for the onset of Ca^{2+} and Na^+ transitions associated with the onset of GI-12c, derived from Ca^{2+} (orange) and Na^+ (green) values around the GI-12c onset at 2-year resolution, using the probabilistic ramp-fitting shown in (b). The black lines in (b) indicates the mean of the ramps sampled in the MCMC process and the gray shaded area indicates the 5th-95th percentiles of these ramps. (c) Histogram associated with the empirical density of transition onset lags Δt between the two proxies (violet), together with the corresponding Gaussian kernel density estimate (blue).

problems. For example, consider a drug study that is meant to test the effect of a medicine against fever. The difference in the body temperature of the probands before and after the treatment can however be measured only with finite precision. Hence, the sample of temperature differences is uncertain in the very same sense as the sample of time lags between DO transition onsets of different proxies. In most practical cases the relative sample uncertainty may be small and correspondingly omitted. Also, the same shape of uncertainty on each sample member typically allows for application of hierarchical models. It is the individuality in the sample members uncertainty that prevents the application of hierarchical methods in our case.

3.1 Transition onset detection

Given intervals of Ca^{2+} and Na^+ concentration time series centered around DO events, application of an MCMC-based Bayesian ramp-fitting method developed by (Erhardt et al., 2019) yields posterior probability densities $\rho_{t_0}(t_0)$ for the transition onset time t_0. The approach is based on modeling the transition present in the data as a linear ramp between two constant levels (see Fig. 2 a and b). Tab. 1 lists the 24 DO transitions for which data is available, with bold font indicating the 16 events for which $\rho_{t_0}(t_0)$ was successfully derived for both proxies simultaneously. The probability densities for the transition onsets of both proxies during individual events allow to compute corresponding probability densities for the lag between them:

$$\rho(\Delta t) = \int\int \delta(t_0^{Ca^{2+}} - t_0^{Na^+} - \Delta t)\rho_{t_0}^{Ca^{2+}}(t_0^{Ca^{2+}})\rho_{t_0}^{Na^+}(t_0^{Na^+})dt_0^{Ca^{2+}}dt_0^{Na^+}. \quad (1)$$

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020


Climate of the Past Discussions
Open Access
EGU

In the following, $\rho_i(\Delta t_i)$ denotes the probability distribution for the lag of the i-th event under study. The index skips those events for which the transition detection method failed. Fig. 2 shows the transitions to GI-12c for Ca^{2+} and Na^{+}, together with the derived probability distributions for the transition onsets (panel a and b) and the corresponding distribution of lags between the two (panel c). An overview of all probability density functions $\rho_i(\Delta t_i)$ for the transition onset lags of the 16 DO events is given in Fig. 4. In Tab. 1 the probability for each transition to be lead by Ca^{+} is listed. Importantly, there is a clear tendency for Ca^{2+} to lead the transition for many of the events, as already reported by Erhardt et al. (2019). The sample of lags $\mathbf{\Delta t} = (\Delta t_1, \Delta t_2, ..., \Delta t_{16})$ constitutes the data basis for this study. Due to its stochastic character, we refer to the sample as an 'uncertain sample'. In the statistical analysis all of these events are treated equally even though some transitions might be more pronounced than others and the quality of the fit differs between the events. The ramp-fitting method does of course not return continuous probability densities but rather empirical densities

$$\rho_i^{\text{emp}}(\Delta t_i) = \frac{1}{6000} \sum_{j=1}^{6000} \delta(\Delta t_i - \Delta t_{i,j}^{\text{emp}}) \tag{2}$$

based on a representative set of 6000 values $\{\Delta t_{i,j}^{\text{emp}}\}$ sampled by the MCMC algorithm for each event. In general, empirical densities approximate their continuous counterparts such that

$$\int f(x)\rho_X(x)dx \simeq \int f(x)\rho_X^{\text{emp}}(x)dx = \frac{1}{m}\sum_{j=1}^{m} f(x_j) \tag{3}$$

for large enough m and any well behaved function f. The densities $\rho_i(\Delta t_i)$ are derived from the empirical densities by means of a Gaussian kernel density estimator.

We consider the empirical densities generated with the Bayesian MCMC-based ramp fit algorithm most convenient to capture the uncertainty in the detection of an abrupt transition onset. Further details of this method can be found in Erhardt et al. (2019). In view of the different signal-to-noise levels in the calcium and sodium records, we carried out a performance test by applying the algorithm to synthetically generated time series of abrupt transitions perturbed by Gaussian noise of different amplitude. As expected, the larger the noise amplitude, the less precise is the transition onset detection. Importantly, however, there is no systematic bias of the detected transition onset in the one or the other direction. For details, see Appendix A.

3.2 Statistical setting

Despite their diversity in terms of temperature amplitude, duration, and frequency across the last glacial, the reoccurring patterns and their common manifestation in different proxies suggest that the DO events follow a common physical mechanism. Provided this is the case, the relative phasing between Ca^{2+} and Na^{+} transitions still involves randomness due to climate variability. From a frequentist perspective, each DO warming is drawn from a hypothesized, infinitely large population of reoccurring DO events and the realization of a Ca^{2+}–Na^{+} lag can thus be regarded as a random experiment $(\Omega, \mathcal{F}, \mathcal{P}_{\Delta t})$ on the sample space $\Omega = \mathbb{R}$. Here, $\mathcal{F}$ is a σ-algebra defined on Ω and may be taken as the Borel algebra with $\mathcal{P}_{\Delta t}$ denoting a probability measure with respect to $\mathcal{F}$ that is called population. This said, the reasoning behind our statistical analysis is as follows: If the Ca^{2+} transition was to trigger the Na^{+} transition, the population $\mathcal{P}_{\Delta t}$ would not ascribe any probability to

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions
Open Access
EGU

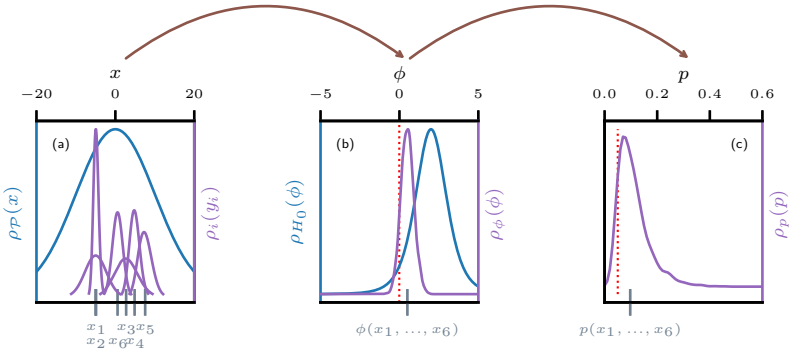

Figure 3. (a) Schematic representation of the two-level randomness inherent to the measurement of a continuous random variable. The blue line indicates the probability density associated with the generating population $\mathcal{P}$. A certain sample realized from $\mathcal{P}$ is illustrated by the gray lines. Measuring the sample necessarily introduces a new level of uncertainty, such that true values can at best be approximated by probability density functions depicted in purple. Depending on the measurement process, the uncertainty distributions of the sample members may all exhibit individual shapes or they may share a common one. (b) Distribution of the test statistic ϕ derived from the uncertain sample (purple) together with the corresponding value derived from the certain sample (gray). In blue, the distribution of ϕ under the null hypothesis is shown. The dotted red line separates the rejection region (left) from the acceptance region in the one-sided test setup. (c) Distribution of the uncertain p-value corresponding to the uncertain sample. In gray, the p-value of the certain sample is marked. The red line indicates the significance threshold η.

lags $\Delta t > 0$. In particular the population mean $\mu_{\Delta t}$ would necessarily be less than zero. If we fail to exclude a population mean $\mu_{\Delta t} \geq 0$ based on the sample $\mathbf{\Delta t}$, it cannot be guaranteed that this necessary criterion is fulfilled and, therefore, the data cannot serve as evidence for a Ca^{2+} transition lead, let alone a causality between the two transitions. Thus, we focus on deriving qualified statistical statements on the population mean $\mu_{\Delta t}$ in view of the sample $\mathbf{\Delta t}$.

A population $\mathcal{P}$ is by itself not observable, but random samples $\mathbf{X}^n = (X_1, X_2, ..., X_n)$ of independently and identically distributed random variables may be generated from $\mathcal{P}$, where the randomness of each component of $\mathbf{X}^n$ is determined by the populations probability density $\rho_{\mathcal{P}}$:

$$\mathcal{P}(X_i \in [a, b]) = \int_a^b \rho_{\mathcal{P}}(x_i) dx_i. \tag{4}$$

Inferring properties of $\mathcal{P}$ from a finite-size sample $\mathbf{X}^n$ (which we call an n-sample in the following) is a well-studied problem in statistics and a broad range of methods is available. The sample of lags $\mathbf{\Delta t} = (\Delta t_1, ...\Delta t_{16})$ introduced above can be regarded as such a sample generated from $\mathcal{P}_{\Delta t}$. However, this description does not account for the uncertainty inherent to each individual sample member Δt_i. Whenever a random sample is realized on a continuous sample space, its measurement necessarily involves uncertainty and adds a second level of randomness. In our case, this is due to the uncertainty quantified in the MCMC-based transition detection. Instead of a sample $\mathbf{x}^n = (x_1, x_2, ..., x_n)$ comprised of scalar entries, one obtains what we call an uncertain sample $\mathbf{Y}^n = (Y_1, Y_2, ...Y_n)$ of random variables Y_i. The generation of a sample from a population and

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020


Climate of the Past Discussions
Open Access
EGU

the corresponding two-fold randomness is illustrated in Fig. 3. We emphasize that the randomness of $\mathbf{Y}^n$ is not the same as the randomness of the stochastic process under study, but instead originates from the imprecise measurement process. In this section we use capital letters to denote random variables and lower-case letters to denote realizations. A realized sample $\mathbf{x}^n$ free of measurement uncertainty will be referred to as a certain sample in the following to avoid confusion. The components of $\mathbf{Y}^n$ are independently, but not necessarily identically distributed. Therefore, all sample members require an individual representation $\rho_i(y_i)$, giving rise to the distribution of $\mathbf{Y}^n$ in $\mathbb{R}^n$:

$$\rho_{\mathbf{Y}}(\mathbf{y}) = \prod_{i=1}^{n} \rho_i(y_i). \tag{5}$$

The sample $\mathbf{\Delta t}$ falls into the class of uncertain samples, with each component Δt_i being a realization of the same random process, yet carrying its individual measurement uncertainty represented by $\rho_i(\Delta t_i)$. Surprisingly there seems to be a lack of literature on how to treat these uncertainties when inferring the properties of $\mathcal{P}$, as soon as the uncertainty of each sample member has its individual shape and hierarchical distributional models cannot be invoked. An obvious approach is to work with expected values $\langle Y_i \rangle = \int y_i \rho(y_i) dy_i$ to obtain a scalar valued sample, which allows for immediate application of the typical statistical toolkit. However, we will show in the following that this approach is too simplistic and important information may be hidden by the averaging process. Before we discuss how the uncertainties may be treated and inference on the population from uncertain samples can be realized, we need to clarify how the high-dimensional probability distribution can be treated numerically.

3.3 Numerical treatment of high dimensional probability densities

Given empirical densities $\rho_i^{\text{emp}}(y_i) = \frac{1}{m}\sum_{j=1}^{m} \delta(y_i - y_{i,j}^{\text{emp}})$ of the members Y_i of an uncertain n-sample based on representative sets $\{y_{i,j}^{\text{emp}}\}$ of size m, formally the empirical density in $\mathbb{R}^n$ reads

$$\rho_{\mathbf{Y}}^{\text{emp}}(\mathbf{y}) = \prod_{i=1}^{n} \rho_i^{\text{emp}}(y_i) = \prod_{i=1}^{n} \frac{1}{m} \sum_{j=1}^{m} \delta(y_i - y_{i,j}^{\text{emp}}). \tag{6}$$

However, this results in a sum with m^n terms or in a vectorial representation $\rho_{\mathbf{Y}}^{\text{emp}}(\mathbf{y}) = \frac{1}{m^n}\sum_{k=1}^{m^n} \delta(\mathbf{y} - \mathbf{y}_k^{\text{emp}})$ with a representative set $\{\mathbf{y}_k^{\text{emp}}\}$ with m^n members. Since in our case $m = 6000$ and $n = 16$, m^n exceeds the computational memory by far. Therefore, we reduced the size of the representative set from 6000^{16} to 6000 values by setting

$$\rho_{\mathbf{Y}}^{\text{emp}}(\mathbf{y}) = \frac{1}{m} \sum_{j=1}^{m} \prod_{i=1}^{n} \delta(y_i - y_{i,j}^{\text{emp}}). \tag{7}$$

To check that Eq. 3 is not violated, we have carried out our analysis with a control group of 10 alternative realizations of $\rho_{\mathbf{Y}}^{\text{emp}}(\mathbf{y})$. The representative sets for the control group densities have been realized by randomly selecting 6000 vectors $\{\mathbf{y}_1, ..., \mathbf{y}_{6000}\}$ from the total possible 6000^{16} vectors. Since all final results reported in the main text of this article fall inside the two standard deviation uncertainty environment of the control group, we consider our chosen approximation of $\rho_{\mathbf{Y}}(\mathbf{y})$ to be justified. The results for the control group are presented in Appendix C.

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions
Open Access
EGU

3.4 Inference of the population mean from an uncertain sample

As explained above, we are interested in the mean $\mu_{\Delta t}$ of the population $\mathcal{P}_{\Delta t}$ that presumably has generated the given sample $\mathbf{\Delta t}$. In this section, we derive a way to directly infer a probability distribution for a population mean from an uncertain sample. First, consider an n-sample $\mathbf{x}^n$ generated from a population $\mathcal{P}$ with the probability density $\rho_{\mathcal{P}}(x)$. Without any further knowledge, the mean $u = \frac{1}{n}\sum_{i=1}^{n} x_i$ of the sample is the best point estimate for the mean of its generating population. Let

$$U = \frac{1}{n}\sum_i Y_i \tag{8}$$

denote the mean of a corresponding uncertain sample. Being a function of random variables makes $U(\mathbf{Y})$ a random variable itself, with a probability density function given by

$$\rho_U(u) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^{n} \rho_i(y_i)\delta(u - \frac{1}{n}\sum_{i=1}^{n} y_i)\, dy_1\, dy_2 \dots dy_n\,. \tag{9}$$

In the above equation the δ-function selects the probability weight allocated to those $\mathbf{y}$ that are mapped to u. If $\rho_{\mathcal{P}}$ is justifiably assumed or known to be normal, the estimation of the population mean can be refined one step further by drawing on the t-distribution. First, consider an n-sample $\mathbf{X}^n$ with sample mean $U = \frac{1}{n}\sum_{i=1}^{n} X_i$ and sample standard deviation $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - U)^2}$, to be realized from a normal population $\mathcal{N}(\mu,\sigma)$ with population mean μ and population standard deviation σ. The quantity $Z = \frac{U-\mu}{S/\sqrt{n}}$ is then distributed according to the t-distribution with $n-1$ degrees of freedom (the proof is originally due to Student (1908) and can for example be found in Rice (2007)):

$$P(a < Z < b) = \int_a^b t_{n-1}(z)dz\,, \tag{10}$$

irrespective of the precise values for μ and σ. Here, $P(A)$ is an informal - though widely used - way to denote the probability for the outcome A of a random experiment without previous definition of the sample space and the sigma-algebra. For a given pair (u,s) derived from a sample $\mathbf{x}^n$ we can therefore interpret $t_{n-1}(z)$ as the probability density that the n-sample originates from a normal population with mean μ. Substituting $z = \frac{u-\mu}{s/\sqrt{n}}$ and $dz = -\frac{1}{s/\sqrt{n}}d\mu$ in Eq. (10) yields

$$P(a < z < b) = P(a < \frac{u-\mu}{s/\sqrt{n}} < b) = P(a\,s/\sqrt{(}n) < (u-\mu) < b\,s/\sqrt{n}) = P(u - b\,s/\sqrt{n} < \mu < u - a\,s/\sqrt{n}) \tag{11}$$

and

$$\int_a^b t_{n-1}(z)dz = \int_{u-bs/\sqrt{n}}^{u-as/\sqrt{n}} \underbrace{t_{n-1}(\frac{u-\mu}{s/\sqrt{n}})\frac{\sqrt{n}}{s}}_{=\rho_\mu(\mu,u,s)} d\mu. \tag{12}$$

The integrand on the right constitutes a probability density function for μ for any given (u,s). Since a and b can be freely chosen, one can replace $u - as/\sqrt{n}$ and $u - bs/\sqrt{n}$ with β and α, respectively. Combining these results we obtain

$$P(\alpha < \mu < \beta | u,s) = \int_\alpha^\beta \rho_\mu(\mu,u,s)\, d\mu\,, \tag{13}$$

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions

with $\rho_\mu(\mu,u,s) = t_{n-1}\left(\frac{u-\mu}{s/\sqrt{n}}\right)\frac{\sqrt{n}}{s}$. Eq. (13) relies on an uncertainty-free tuple (u,s) and is in this form not suited to capture uncertainties in the sample. Just as the mean $U(\mathbf{Y})$, the standard deviation $S(\mathbf{Y})$ of a random vector $\mathbf{Y}$ is a random variable itself. In general, for a given probability density $\rho_{\mathbf{Y}}(\mathbf{y})$, the statistics U and S are not independently distributed and their joint probability density reads

$$\rho_{(U,S)}(u,s) = \int \rho_{\mathbf{Y}}(\mathbf{y})\,\delta\left(u - \sum_{i=1}^{n} y_i/n\right)\delta\left(s - \frac{1}{n-1}\sum_{i=1}^{n}(y_i-u)^2\right)d\mathbf{y}. \tag{14}$$

Based on $\rho_{(U,S)}$ an expected distribution $\langle\rho\rangle_\mu(\mu)$ can be computed, such that any pair (u,s) contributes to $\langle\rho\rangle_\mu(\mu)$ according to its probability density:

$$\langle\rho\rangle_\mu(\mu) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\rho_\mu(\mu,u,s)\rho_{(U,S)}(u,s)\,du\,ds. \tag{15}$$

Eq. (15) takes account of all uncertainties inherent to $\mathbf{Y}$. Replacing the continuous density with the empirical density defined in Eq. 7 leads to

$$\rho^{\text{emp}}_{(U,S)}(u,s) = \int \frac{1}{m}\sum_{j=1}^{m}\prod_{i=1}^{n}\delta(y_i - y^{\text{emp}}_{i,j})\,\delta\left(u - \sum_{i=1}^{n} y_i/n\right)\delta\left(s - \frac{1}{n-1}\sum_{i=1}^{n}(y_i-u)^2\right)d\mathbf{y} = \frac{1}{m}\sum_{j=1}^{m}\delta\left(u - u^{\text{emp}}_j\right)\delta\left(s - s^{\text{emp}}_j\right), \tag{16}$$

with $u^{\text{emp}}_j = \frac{1}{n}\sum_{i=1}^{n} y^{\text{emp}}_{i,j}$ and $s^{\text{emp}}_j = \frac{1}{n-1}\sum_{i=1}^{n}(y^{\text{emp}}_{i,j} - u^{\text{emp}}_j)^2$ and correspondingly:

$$\langle\rho\rangle^{\text{emp}}_\mu(\mu) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\rho_\mu(\mu,u,s)\rho^{\text{emp}}_{(U,S)}(u,s)\,du\,ds = \frac{1}{m}\sum_{j=1}^{m}\rho_\mu(\mu,u_j,s_j) = \frac{1}{m}\sum_{j=1}^{m} t_{n-1}\left(\frac{u_j-\mu}{s_j/\sqrt{n}}\right)\frac{\sqrt{n}}{s_j}. \tag{17}$$

3.5 Bootstrapping the distribution of the population mean

Given a certain n-sample $\mathbf{x}^n$, bootstrapping constitutes a complementary approach to derive a distribution for the population mean μ (e.g., Lehmann and Romano, 2006). The great advantage of the non-parametric bootstrapping method is that it does not rely on any assumptions on $\mathcal{P}$ (such as normality). Generally speaking, the method assesses the uncertainty of statistics such as the sample mean computed from a certain sample. Here, we make use of the fact that the sample mean is the best point estimate of the population mean and hence an uncertainty distribution of the sample mean can equivalently be interpreted as an uncertain estimate for the population mean. In short, bootstrapping works as follows: a set of κ synthetic samples $\{\mathbf{x}^{\text{bs}}_k |\, 1 \le k \le \kappa\}$ is generated from the original sample $\mathbf{x}^n$ by drawing n values with replacement from $\mathbf{x}^n$. Next, the mean $u^{\text{bs}}_k = \frac{1}{n}\sum_{i=1}^{n} x^{\text{bs}}_{k,i}$ of each synthetic sample is calculated. The set of the κ means u^{bs}_k generates an empirical bootstrapped probability density for the population mean

$$\rho^{\text{bs}}_\mu(\mu) = \frac{1}{\kappa}\sum_{k=1}^{\kappa}\delta(u^{\text{bs}}_k - \mu). \tag{18}$$

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions

Given an uncertain sample $\mathbf{Y}$ represented by the empirical density $\rho_{\mathbf{Y}}^{\text{emp}}(\mathbf{y}) = \frac{1}{m}\sum_{j=1}^{m}\delta(\mathbf{y}-\mathbf{y}_j^{\text{emp}})$ instead of a certain sample, the bootstrapping technique can be applied to all $\mathbf{y}_j^{\text{emp}}$ comprised in the empirical density, giving rise to m synthetic bootstrapped samples $\{u_{k,j}^{\text{bs}}|1 \leq j \leq m\}$, each inducing a distribution $\rho_{\mu,j}^{\text{bs}}(\mu)$ according to Eq. 18. Averaging over these yields the general expression

$$\langle\rho\rangle_{\mu}^{\text{bs}} = \frac{1}{m}\sum_{j=1}^{m}\rho_{\mu,j}^{\text{bs}} = \frac{1}{m\kappa}\sum_{j,k}^{m,\kappa}\delta(u_{j,k}^{\text{bs}}-\mu) \tag{19}$$

for the population mean distribution, where $u_{j,k}^{\text{bs}}$ denotes the the k-th bootstrapped mean from the j-th member of the empirical density's representative set $\{\mathbf{y}_i\}$.

3.6 Hypothesis testing with uncertain samples

In order to identify a causal relationship between the transitions in the atmospheric circulation as indicated by Ca^{2+}, and transitions in the sea-ice extent represented by Na^{+}, we must be able to discriminate stochastic from systematic features in the data. In our case, this translates into ruling out that the given sample of time lags between the two proxy variables was generated from a population with mean equal to zero, corresponding to a simultaneous transition. In view of the apparent tendency of Ca^{2+} to lead the transition across most of the analyzed DO events (see Fig. 4 and previous work by Erhardt et al. (2019)) one-sided hypothesis tests addressing the population mean constitute a convenient method to provide evidence for a population mean less than zero and therefore a systematic time lag. The other way round, not being able to preclude population means equal to or greater than zero implies that causality cannot be evidenced in the data sample, though it does not prove the absence of causality between the two proxy variables.

Numerous statistical hypothesis tests have been designed to systematically rule out distributions which are unlikely to have generated a given sample. As noted above, the classic literature deals with certain samples, while we discuss the application of three different tests – all targeting the population mean – to an uncertain sample of the previously described kind. First, we present our approach how hypothesis tests can in general be applied to uncertain data samples. In short, we propose a propagation of the uncertainties to the p-value and introduce two potential decision criteria to map the resulting p-value distribution $\rho_p(p(\mathbf{Y}))$ to a binary decision between rejection and acceptance of the null hypothesis. Subsequently, we discuss our choice of tests. Generally speaking, we wish to test how the populations of the transition onsets in Ca^{2+} and Na^{+} compare and, in particular, whether the transition onset of Na^{+} lags the onset of Ca^{2+}. Since we are given only relative data, that is, the difference Δt in the transition onset times of the two proxies, we must rely on tests that can be applied to what are called paired samples or paired comparisons. This said, we find the paired t-test, the Wilcoxon-signed-rank (WSR) test, and a bootstrap test suited for our purpose.

3.7 Testing uncertain data

Typically, a statistical hypothesis test draws on a test statistic $\phi(\mathbf{X})$:

$$\phi : \mathbb{R}^n \to \mathbb{R}; \qquad \mathbf{X} \mapsto \phi(\mathbf{X}),$$

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions
Open Access
EGU

whose distribution $\rho_{H_0}(\phi)$ is known under the so-called null-hypothesis H_0. The null-hypothesis constitutes a set of assumptions on the population $\mathcal{P}$. Given a statistic of a certain sample realization $\phi(\mathbf{x_0}) = \phi_0$, a p-value is calculated which indicates the cumulative probability for obtaining a more extreme $\phi_1(\mathbf{X_1})$ from a second sample $\mathbf{X_1}$, provided H_0 was true:

$$p(\phi_0) = \int_{\omega} \rho_{H_0}(\phi) d\phi, \quad \omega = \{\phi : \rho_\phi(\phi) < \rho_\phi(\phi_0)\}. \tag{20}$$

The hypothesis H_0 is then rejected at a significance level α if the p-value is below the predefined significance threshold α. In that case one also speaks of a statistically significant result. For a comprehensive description of hypothesis testing see for example Lehmann and Romano (2006). While the application of hypothesis tests to a certain sample $\mathbf{x}$ is in general straightforward, for an uncertain sample it is not. A distribution of sample values $\rho_{\mathbf{Y}}(\mathbf{y})$ first translates into a distribution for the test statistic $\rho_\phi(\phi)$ and from there to a distribution of p-values $\rho_p(p)$ as depicted in Fig. 3 with $\rho_p(p)$ featuring probabilities for both, $p < \alpha$ and $p > \alpha$. One might be tempted to boil down the uncertain sample to its expected value $\langle \mathbf{Y} \rangle$ before the application of the test to obtain a scalar p-value. In general, however, the p-value of the samples expectation does not equal the expected p-value,

$$p\left(\int \mathbf{y} \rho_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}\right) \neq \int p(\mathbf{y}) \rho_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}, \tag{21}$$

due to the non-linearities in both the statistic ϕ and in the mapping of ϕ to a p-value. In other words, the significance of a sample may change if its uncertainties are propagated into the space of the test statistic or the p-value. So what is a convenient and sound way to project a distribution of p-values derived from an uncertain sample $\mathbf{Y}$ to a scalar in order to take a meaningful decision on the significance by comparing it to α?

We argue that the uncertainties should be propagated to the p-values, before any projection to a scalar is applied. The reason for this is that a distribution $\rho_{\mathbf{Y}}(\mathbf{y})$ might assign finite probabilities for $\mathbf{y}$-values that strongly differ from the expectation. Such values do not find any consideration if a test was applied to the expectation of $\mathbf{Y}$. Consider the case of a bimodal distribution of $\mathbf{Y}$ with both bulks of the distribution lying in the rejection region of the hypothesis H_0. The expected $\langle \mathbf{Y} \rangle$ might yield a non-significant test result as it is located somewhere between the two bulks while a propagation of the uncertainty would potentially indicate the significance of the uncertain sample. Given a distribution of p-values $\rho_p(p)$ that incorporates the original uncertainties, the following two criteria can be invoked to decide between acceptance and rejection of the Hypothesis:

- *The hypothesis shall be rejected at the significance level α if and only if the expected p-value is less than α, that is*

$$\int_0^1 \rho_p(p) dp < \alpha\,. \tag{22}$$

- *The hypothesis shall be rejected at the significance level α if and only if the probability for p to be less than α is greater than a predefined threshold η (we propose $\eta = 90\%$), that is*

$$P(p < \alpha) = \int_0^{\alpha} \rho_p(p) > \eta\,. \tag{23}$$

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020


Climate of the Past Discussions

As the expected p-value constitutes a convenient measure of the overall extremeness of an uncertain sample $\mathbf{Y}$ with respect to H_0, its comparison to the significance threshold is meaningful. However, in some situations one might want to guarantee that in fact the probability to achieve a significant test result is high – e.g. when mistakenly attested significance is associated with high costs. In these cases the second criterion is more appropriate even though it does not imply that the first criterion is fulfilled.

In the following we draw on the empirical density representation of $\mathbf{Y}$ in order to rigorously propagate the uncertainties to the space of p-values. A corresponding empirical density for the p-values may be obtained by

$$\rho_p^{\text{emp}}(p) = \int \frac{1}{m}\sum_{j=1}^{m} p(\mathbf{y})\delta(\mathbf{y}-\mathbf{y}_j)d\mathbf{y} = \frac{1}{m}\sum_{j=1}^{m}\delta(p-p_j), \tag{24}$$

with $p_j = p(\mathbf{y}_j)$.

3.7.1 t-test

The t-test can be applied to a sample of differences $\{d_i\}$ derived from a paired sample $\{(x_i, y_i)\}$ (Rice, 2007). The hypothesis H_0 in this case states that both x_i and y_i were generated from normal populations with the same mean $\mu_x = \mu_y$. This implies that the population of differences is normally distributed as well, but with mean $\mu_0 = 0$. Note that the standard deviations σ_x and σ_y may differ. For large samples, deviations from normality in the original populations can be tolerated and the hypothesis reduces to the equality of means. With $n = 16$ our sample constitutes an intermediate case. This is one reason why we use the WSR and bootstrap tests in addition to the t-test.

Recall the definition of z, which constitutes the test statistic for the t-test:

$$z(\mathbf{d}) = \frac{u(\mathbf{d})-\mu_0}{s(\mathbf{d})/\sqrt{n}}, \quad \mathbf{d} = (d_1, ..., d_n). \tag{25}$$

The distribution of z under the null-hypothesis is given by the t-distribution $t_{n-1}(z)$ of $n-1$ degrees of freedom. Setting $\mu_0 = 0$ in the above equation and calculating the p-value p_z from the left hand side only allows us to effectively test the hypothesis $H_0 : \mu_0 \geq 0$. The p-value

$$p_z = \int_{-\infty}^{z} t_{n-1}(z')dz' \tag{26}$$

may be compared to the chosen significance threshold.

3.7.2 Wilcoxon-signed-rank

Given a set of pairs $\{(x_i, y_i)\}$ the Wilcoxon-signed-rank test (Wilcoxon, 1945) was originally introduced to test whether $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$ stem from the same population (Rice, 2007). The test relies on the set of differences

$$\{d_i = x_i - y_i\}\,, i \in \{1, 2, ..n\}\,, \tag{27}$$

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020


Climate of the Past Discussions

and the hypothesis H_0 states that the population of differences is symmetrically distributed around zero. We apply the test in a one-sided way, such that we can reject distributions centered around any value greater than zero in case of a significant outcome. The test statistic $w(\mathbf{d} = (d_1, ..., d_n))$ for the Wilcoxon-signed-rank test is defined as

$$w(\mathbf{d}) = \sum_{i=1}^{n} R(|d_i|)\, \Theta(d_i), \tag{28}$$

where $R(d_i)$ denotes the rank of $|d_i|$ within the sorted set of the absolute values of differences $\{|d_i|\}$. The Heaviside function Θ guarantees that exclusively $d_i > 0$ are summed. Note that a significant w-value does not preclude an asymmetric distribution with mean zero but higher portions of probability located below zero. The corresponding one-sided p-value p_w is given by the cumulative probability that the null hypothesis assigns to w'-values smaller than a given w:

$$p_w(w) = \sum_{i=1}^{n} P(w_i')\, \Theta(w - w'). \tag{29}$$

3.7.3 Bootstrap test

Given a sample of differences $\mathbf{d} = (d_1, ..., d_n)$, a bootstrap test constitutes a third option to test the compatibility of the sample with the hypothesis that the population of differences features a mean equal to or greater than zero: $H_0 := \{\mu_0 \geq 0\}$. The advantage of the bootstrap test lies in its independence from assumptions regarding the distributions shape. Guidance for the construction of a bootstrap hypothesis test can be found in Lehmann and Romano (2006) and Hall and Wilson (1991). For our test, we bootstrap from the set of differences $\{d_i - u(\mathbf{d})\}$ shifted by the sample mean $u(\mathbf{d}) = \frac{1}{n}\sum_{i=1}^{n} d_i$. We denote the arising empirical distribution by

$$\hat{\rho}_0^{\mathrm{bs}}(u) = \rho_U^{\mathrm{bs}}(u) - u(\mathbf{d}) = \frac{1}{\kappa}\sum_{k=1}^{\kappa} \delta(u - \hat{u}_k^{\mathrm{bs}}), \tag{30}$$

where $\hat{u}_k^{\mathrm{bs}}$ are the bootstrapped means shifted by the samples' mean value. This distribution can be regarded as a data-derived empirical distribution for the test statistic U compatible with the null-hypothesis. The bootstrap p-value p^{bs} for the sample $\mathbf{d}$ is then given by the fraction of shifted bootstrap means that are smaller than the mean of the sample $u(\mathbf{d})$:

$$p^{\mathrm{bs}}(\mathbf{d}) = \int_{-\infty}^{u(\mathbf{d})} \hat{\rho}_0^{\mathrm{bs}}(u)\, du = \frac{1}{\kappa}\sum_{k=1}^{\kappa} \Theta(u(\mathbf{d}) - \hat{u}_k^{\mathrm{bs}}). \tag{31}$$

In order to generalize this test to the case of an uncertain sample $\mathbf{Y}$ the uncertain quantity $U(\mathbf{Y})$ should be compared to a bootstrapped distribution that is consistent with the null hypothesis and simultaneously captures the uncertainties. The application of the shifted bootstrapping scheme to all samples $\mathbf{y}_i$ comprised in the empirical density $\rho_{\mathbf{Y}}^{\mathrm{emp}}(\mathbf{y}) = \frac{1}{m}\sum_{j=1}^{m} \delta(\mathbf{y} - \mathbf{y}_j^{\mathrm{emp}})$ yields a convenient empirical distribution of the test statistic compatible with the null hypothesis:

$$\langle \hat{\rho} \rangle_0^{\mathrm{bs}}(u) = \frac{1}{m}\sum_{j=1}^{m} \underbrace{\rho_{u,j}^{\mathrm{bs}}(u) - u_j^{\mathrm{emp}}}_{=\hat{\rho}_{u,j}^{\mathrm{bs}}(u)} = \frac{1}{\kappa m}\sum_{k,j=1}^{\kappa,m} \delta(u - \hat{u}_{k,j}^{\mathrm{bs}}), \tag{32}$$

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions

with $u_j^{\text{emp}} = u(\mathbf{y}_j)$ as before. Subsequently, each u_j shall be compared to the distribution of shifted bootstrap means to compute a corresponding p-value:

$$p^{\text{bs}}(u_j) = \int_{-\infty}^{u_j} \langle\hat{\rho}\rangle_0^{\text{bs}}(u)du = \frac{1}{\kappa m}\sum_{k,j}^{\kappa,m}\Theta(u_j - \hat{u}_{k,j}^{\text{bs}}), \tag{33}$$

giving rise to the empirical p-value distribution

$$\rho_p^{\text{bs}}(p) = \frac{1}{m}\sum_j \delta(p - p_j^{\text{bs}}) \tag{34}$$

with $p_j^{\text{bs}} = p^{\text{bs}}(u_j)$. The test scheme for the certain sample could also be applied to all $\mathbf{y}_i$ individually. However, combining the shifted bootstrapped distributions $\hat{\rho}_{u,j}^{\text{bs}}(u)$ to one overarching null hypothesis distribution for the test statistic $\langle\hat{\rho}\rangle_0^{\text{bs}}(u)$ constitutes the more conservative approach for uncertainty propagation.

The three tests are applied in combination in order to compensate their individual deficits. If the population $\mathcal{P}_{\Delta t}$ was truly Gaussian, the t-test would be the most powerful test, i.e., its rejection region would be the largest across all tests on the population mean (Lehmann and Romano, 2006). Since normality of $\mathcal{P}_{\Delta t}$ cannot be guaranteed, the less powerful Wilcoxon-signed-rank test constitutes a meaningful supplement to the t-test, relying on the somewhat weaker assumption that $\mathcal{P}_{\Delta t}$ is symmetric around zero. Finally, the bootstrap test is non-parametric and in view of its independence from any assumptions adds a valuable contribution.

4 Results

The statistical methods formulated in Sec. 3 provide a sound framework for the statistical analysis of the data described in Sec. 2. From application of the Bayesian MCMC ramp-fit method introduced by Erhardt et al. (2019), we obtained the empirical densities $\rho_i^{\text{emp}}(\Delta t_i) = \frac{1}{6000}\sum_{j=1}^{6000}\delta(\Delta t_i - \Delta t_{i,j}^{\text{emp}})$ for the uncertain transition onset lag between Ca^{2+} and Na^{+} of the 16 DO events indicated in Tab. 1. Each empirical density is based on a representative set of size $m = 6000$. Additionally, continuous densities $\rho_i(\Delta t_i)$ were generated with a Gaussian kernel density estimator. We derived a 16-dimensional joint empirical density $\rho_{\mathbf{\Delta t}}(\mathbf{\Delta t} = (\Delta t_1, ..., \Delta t_{16})) = \frac{1}{6000}\sum_{j=1}^{6000}(\mathbf{\Delta t} - \mathbf{\Delta t}_j^{\text{emp}})$ according to Sec. 3.3, drawing on representative sets of 6000 vectors $\mathbf{\Delta t}_j^{\text{emp}}$. Hence, the sample vector $\mathbf{\Delta t}$ can be regarded as an uncertain sample as introduced in Sec. 3.2, generated from a population $\mathcal{P}_{\Delta t}$ with mean $\mu_{\Delta t}$ which is assumed to describe an hypothesized infinitely large pool of DO onset lags. The correspondence between $\mathbf{\Delta t}$ and the generic uncertain sample $\mathbf{Y} = (Y_1, ..., Y_n)$ allows for the immediate application the statistical framework developed in Sec. 3.

4.1 Distribution of the population mean

With $U_{\mathbf{\Delta t}} = \frac{1}{16}\sum_{i=1}^{16}\Delta t_i$ denoting the mean of the uncertain sample $\mathbf{\Delta t}$ and $\mu_{\Delta t}$ denoting the mean of the population $\mathcal{P}_{\Delta t}$, we derived distributions for the stochastic sample mean $\rho_{U_{\mathbf{\Delta t}}}(u_{\mathbf{\Delta t}})$ and the population mean $\langle\rho\rangle_{\mu_{\Delta t}}^{\text{emp}}(\mu_{\Delta t})$ according to Eq. 9

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

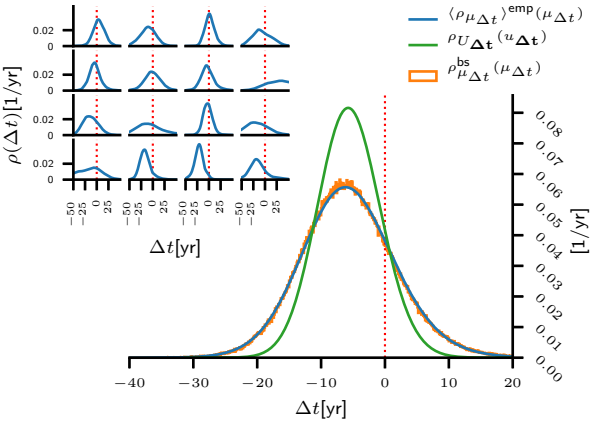

Figure 4. Distribution of the sample mean $\rho_{U_{\mathbf{\Delta t}}}(u_{\mathbf{\Delta t}})$ (green) derived from the Gaussian kernel density estimates of the individual lag distributions $\rho_i(\Delta t_i)$ (see inset) according to Eq. 9 together with the distribution for the population mean $\langle \rho \rangle^{\text{emp}}_{\mu_{\Delta t}}(\mu_{\Delta t})$ (blue) computed according to Eq. 15 and the bootstrapped distribution of the population mean $\langle \rho \rangle^{\text{bs}}_{\mu_{\Delta t}}(\mu_{\Delta t})$ (orange) according to Eq. 18. Clearly, the sample mean as well as the population are likely to be less than zero. The inset displays the probability densities $\rho_i(\Delta t_i)$ of the transition onset lags for all DO events under study, as listed in Tab. 1, with the index increasing along the rows. Most of the individual events feature elevated probabilities for a negative lag.

Table 2. Statistical features of the distributions $\rho_{U_{\mathbf{\Delta t}}}(u_{\mathbf{\Delta t}})$ for the sample mean of the uncertain sample $\mathbf{\Delta t}$, the population mean $\langle \rho \rangle^{\text{emp}}_{\mu_{\Delta t}}(\mu_{\Delta t})$ derived by means of the t-distribution and the bootstrapped distribution for the population mean $\langle \rho \rangle^{\text{bs}}_{\mu_{\Delta t}}(\mu_{\Delta t})$. q5,q50 and q95 denote the 5th, 50th and 95th percentiles of the distributions. The columns $\langle \cdot \rangle$ and $P(\cdot < 0)$ denote the expectation and the probability to be less than zero of the corresponding stochastic variable.

	q5	q50	q95	$\langle \cdot \rangle$	$P(\cdot < 0)$
$\rho_{U_{\mathbf{\Delta t}}}(u_{\mathbf{\Delta t}})$	-13.53	-5.63	2.63	-5.56	0.87
$\langle \rho \rangle^{\text{emp}}_{\mu_{\Delta t}}(\mu_{\Delta t})$	-17.50	-5.80	6.90	- 5.66	0.78
$\langle \rho \rangle^{\text{bs}}_{\mu_{\Delta t}}(\mu)$	-17.36	-5.91	6.95	-5.65	0.79

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions
Open Access
EGU

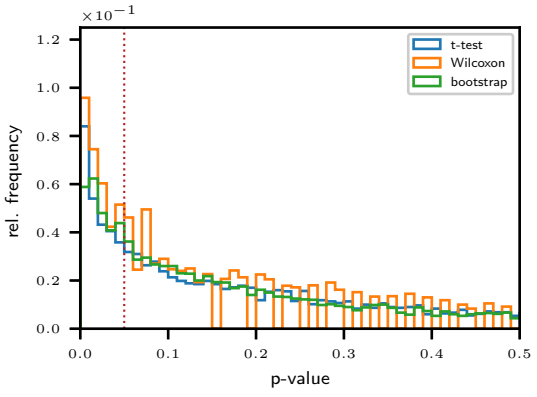

Figure 5. Histograms associated with the empirical densities $\rho_p^{\text{emp}}(p) = \frac{1}{6000}\sum_{j=1}^{6000}\delta(p-p_j)$ of p-values derived by application of the t-test (blue), the WSR-test (orange), and a bootstrap test (green) to the distribution $\rho_{\mathbf{\Delta t}}^{\text{emp}}(\mathbf{\Delta t})$ of the uncertain sample of transition onset lags $\mathbf{\Delta t}$ according to Eq. 24. The dotted red line indicates the significance threshold $\alpha = 0.05$.

and Eq. 17, respectively. The distributions are displayed in Fig. 4 together with the bootstrapped distribution $\langle\rho\rangle_{\mu_{\Delta t}}^{\text{bs}}(\mu_{\Delta t})$ according to Eq. 19, which reproduces the population mean distribution accurately. $\rho_{U_{\mathbf{\Delta t}}}$ appears to be a symmetrically squeezed version of the other two distributions. The expectations of these distributions are listed in Tab. 2 together with the 5th, 50th and 95th percentiles and the probabilities for $U_{\mathbf{\Delta t}}$ and $\mu_{\Delta t}$ to exceed zero. All three distributions exhibit almost the same mean value $\langle U_{\mathbf{\Delta t}}\rangle \simeq \langle\mu_{\Delta t}\rangle \simeq -5.6$ yr and a high degree of symmetry around this value. The equality of their expectations is consistent with the fact that the sample mean $U_{\mathbf{\Delta t}}$ is the best point estimate for the population mean $\mu_{\Delta t}$. The observation that $\langle\rho\rangle_{\mu_{\Delta t}}^{\text{emp}}(\mu_{\Delta t})$ is much broader than $\rho_{U_{\mathbf{\Delta t}}}(u_{\mathbf{\Delta t}})$ can be explained by the fact that any sample mean u_0 is associated with an interval (a,b) allowing a normal distribution $\mathcal{N}(\mu_0,\sigma)$ with $\mu_0 \in (a,b)$ to generate u_0 with finite probability. A distribution of U consequently must be associated with an even broader distribution of μ. The strong agreement between $\langle\rho\rangle_{\mu_{\Delta t}}^{\text{emp}}(\mu_{\Delta t})$ and $\langle\rho\rangle_{\mu_{\Delta t}}^{\text{bs}}(\mu_{\Delta t})$ makes us confident that our approach is justified and robust. The probabilities for $U_{\mathbf{\Delta t}}$ and $\mu_{\Delta t}$ to exceed 0 are $13\,\%$ and $22\,\%$, respectively.

4.2 Significance of the Ca^{2+} lead

In addition to estimating the population mean $\mu_{\Delta t}$, we applied the t-test, the Wilcoxon-signed-rank test, and the bootstrap test to the uncertain sample of Ca^{2+}–Na^{+} lags $\mathbf{\Delta t}$ to eventually rule out a generating population with mean greater than or equal to zero which would correspond to a Na^{+} transition onset lead or a simultaneous transition, respectively. Tab. 3 summarizes the results obtained by application of the tests under full consideration of the uncertainty, together with the p-values associated with the expected sample of time lags $\langle\mathbf{\Delta t}\rangle = (\langle\Delta t_1\rangle,...,\langle\Delta t_{16}\rangle)$ with $\langle\Delta t_i\rangle = \int \Delta t_i \rho_i(\Delta t_i) d\Delta t_i$. The corresponding expected lags for each DO-event are listed in Tab. 1. Strikingly, all tests yield significant results at a significance

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions

Table 3. Results from the application of the t-test, the WSR test and a bootstrap (BS) test to the uncertain sample of DO transition onset lags $\mathbf{\Delta t}$. $\langle p \rangle$ denotes the expected p-value, derived from the uncertainty-propagated p-value distribution. The probability for a significant test results associated with the same distribution is indicated by $P(p < 0.05)$. For comparison, the p-values from the application of the tests to the expected sample $\langle \mathbf{\Delta t} \rangle = \int \rho_{\mathbf{\Delta t}}(\mathbf{\Delta t}) \mathbf{\Delta t} d\mathbf{\Delta t}$ are shown in the bottom row.

	t-test	WSR	BS
$\langle p \rangle$	0.219	0.168	0.220
$P(p < 0.05)$	0.258	0.325	0.256
$p(\langle \mathbf{\Delta t} \rangle)$	0.044	0.009	0.018

level of $\alpha = 0.05$ if applied to the expectation of $\mathbf{\Delta t}$. In contrast, the same tests yield non-significant results, if uncertainties are rigorously propagated to the p-values. Neither of the two criteria formulated in Sec. 3.7 is fulfilled at a significance of $\alpha = 0.05$ and a required probability for significance of $\eta = 90\%$ The mean p-values range between 0.17 and 0.22 and the probability for a significant test result does not exceed $32\,\%$ across all tests. This said, the null hypothesis cannot be rejected. In other words, in view of the data we cannot rule out that the observed tendency for Ca^{2+} to lead the transition is a purely stochastic feature and that the population of lags is actually centered around zero or even a value greater than zero. The histograms associated with the empirical p-value distributions are displayed in Fig. 5. We find broad ranges of p-values spanning from 0 to 1 for all three tests with elevated frequencies below the significance threshold of $\alpha = 0.05$ that decay towards the higher p-values. The probability for a significant result is highest for the WSR test ($32\,\%$) and equally low for the t-test and the bootstrap test ($26\,\%$). The discrepancy between the p-values of the expected samples $p(\langle \mathbf{\Delta t} \rangle)$ and the expectation of the uncertainty propagated p-value $\int p \rho_p^{\text{emp}}(p) dp$ is due to the non-linearities in both, the mapping of $\mathbf{\Delta t}$ to the test statistic and the mapping of the test statistic to the p-value.

5 Discussion

Following Erhardt et al. (2019) we derived an uncertain sample $\mathbf{\Delta t}$ of the $Ca^{2+} - Na^{+}$ lags at the DO transition onset from 16 events. There is an apparent tendency for the transition onset in Ca^{2+} to lead the transition in Na^{+}. Application of the t-test, Wilcoxon-signed-rank test and a bootstrap test to the expected sample of lags $\langle \mathbf{\Delta t} \rangle$ confirms the impression of a significant lag between the two proxies. However, if the uncertainties from the determination of transition onset timings are rigorously propagated to the corresponding p-values, the picture becomes more ambiguous with higher probability for non-significant than for significant results across all employed tests. In fact, the proposed approach to project the p-value distributions to a binary decision fails to reject the hypothesis of a population with zero lag for all tests. Applied in combination, the tests compensate individual weaknesses and thus confer credibility to the conclusion that a population mean equal to zero – that is, a simultaneous transition of the two proxies – cannot be excluded. The derived shape of the distribution $\rho_{U_{\mathbf{\Delta t}}}(u_{\mathbf{\Delta t}})$ is in line with the results of the hypothesis tests. The tendency for a calcium lead at the transition onset remains, but with a probability

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions
Open Access
EGU

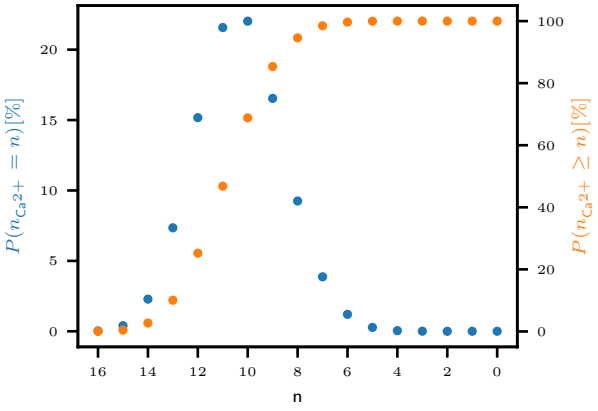

Figure 6. Probability distribution for $n_{Ca^{2+}}$ out of the 16 DO events under study to exhibit a calcium lead at the onset of the abrupt transition (blue) together with the cumulative probability of $n_{Ca^{2+}}$ or more DO transitions to be lead by calcium (orange) according to Eq. 35 and Eq. 36, respectively .

of 13 % the sample features a Na^+ lead on average. The probability of the sample to stem from a population $\mathcal{P}_{\Delta t}$ with mean $\mu_{\Delta t}$ greater than zero was calculated to be 22 %. This said, the analyzed data cannot serve as an unambiguous evidence for a significant Ca^{2+} lead at the DO transition onset with respect to Na^+. The rigorous propagation of uncertainties qualitatively changes the interpretation of the data and the observed tendency towards a Ca^{2+} lead might simply be a stochastic feature. As already metioned, we cannot infer an absence of causality from the fact that we cannot identify a systematic lag in the data. Therefore, in the following we present an important argument that contradicts the idea that either of the two – atmospheric change and sea ice change – is an exclusive trigger of the other.

While the discussion has focused on the population mean $\mu_{\Delta t}$ so far, for the search of a DO trigger mechanism it is important to note that the order of causes and consequences should be preserved across all studied DO events, provided they were driven by the same physics. Hence, if atmospheric circulation changes (indicated by a Ca^{2+} change) were to trigger the North Atlantic sea ice retreat (corresponding to a Na^+ decrease), the changes in Na^+ should lag the changes in Ca^{2+} across all analyzed DO events. If instead both, DO events with negative and positive lags between the two proxy variables can be evidenced, this would suggest that neither of the two exclusively triggers the other, or that the physical mechanism is not the same across all DO events. Following this thought, we calculated the probability for $n_{Ca^{2+}}$ out of the 16 events to feature a Ca^{2+} lead at the transition onset given by

$$P(N_{\Delta t_i<0} = n_{Ca^{2+}}) = \sum_{\substack{i=1 \\ j>i \\ k>j \\ \dots}}^{n} \underbrace{\prod_{l\neq i,j,k,\dots} P(\Delta t_k < 0)}_{n_{Ca^{2+}}\text{ terms}} \underbrace{P(\Delta t_i > 0)P(\Delta t_j > 0)P(\Delta t_k > 0)\dots}_{16-n_{Ca^{2+}}\text{ terms}}, \tag{35}$$

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020


Climate of the Past Discussions

where $N_{\Delta t_i<0}$ denotes the number of events with negative time lag. Since this probability is indifferent to the choice of the events with a Ca^{2+} lead, we have to sum over all constellations with a Ca^{2+} lead in $n_{Ca^{2+}}$ events. Fig. 6 shows the results for $P(N_{\Delta t_i<0} = n_{Ca^{2+}})$ together with the probabilities for the number of events with a Ca^{2+} lead being greater or equal to $n_{Ca^{2+}}$, given by the cumulative probability:

$$P(N_{\Delta t_i<0} \geq n_{Ca^{2+}}) = \sum_{n=n_{Ca^{2+}}}^{16} P(N_{\Delta t_i<0} = n). \tag{36}$$

The probability for all (but one) events to be led by calcium is as small as $0.03\,\%(0.39\,\%)$. Only configurations with 8 to 13 transitions being led by Ca^{2+} are associated with probabilities higher than $5\,\%$, with a maximum of $22\,\%$ likelihood for the configuration with 10 or 11 events being led by Ca^{2+} and 6 or 7 events being led by Na^{+}, respectively. Also the cumulative probability surpasses $5\,\%$ only at $n_{Ca^{2+}} = 13$, implying that the probability that at least 14 events are led by Ca^{2+} is still below $5\,\%$. For $n_{Ca^{2+}} = 10$ the cumulative probability exceeds $50\,\%$. Hence, it is more likely than not that at least 10 events exhibit a calcium transition lead. Finally, the probability for a minimum of 7 transitions with a Ca^{2+} lead is already above $95\,\%$. These results corroborate the conclusion that there is a tendency yet no guarantee for Ca^{2+} to transition first across the DO events under study. The fact that there is only a $10\,\%$ chance that 13 or more transitions are led by Ca^{2+} can be regarded as evidence for the presence of transitions with a Na^{+} lead. Therefore, in terms of a physical interpretation, we can state that this data contradicts the hypothesis that atmospheric changes exclusively trigger the North Atlantic sea ice retreat.

The question arises, which DO events are the ones that exhibit a Na^{+} lead. However, since we are dealing with stochastic lags it cannot be answered ultimately based on the empirical data alone. In Tab. 1 the probability for Na^{+} to lead the transition is listed for all events under study. A first review of the events with high probability does not reveal any systematic pattern, yet this topic merits further investigation.

6 Conclusions

We regarded the Ca^{2+}– Na^{+} transition onset lag distributions $\rho_i(\Delta t_i)$ for 16 DO events, derived from the data set provided by Erhardt et al. (2019), as an uncertain sample generated from a population $\mathcal{P}_{\Delta t}$. Within this framework, we have shown that the data does not significantly contradict the hypothesis that the mean $\mu_{\Delta t}$ of a generating population equals or even exceeds zero when taking into account the uncertainties involved in the transition detection. Three different hypothesis tests, all targeting the population mean, consistently yield non-significant results under the proposed decision criterion for application to uncertain samples. In agreement with this result, the derived distribution $\rho_{\mu_{\Delta t}}(\mu_{\Delta t})$ features a probability of $22\,\%$ for the population mean to exceed zero under the assumption of normality. Even for the stochastic sample mean $U_{\mathbf{\Delta t}}$ we find a probability of $13\,\%$ to be larger than zero, as indicated by the distribution $\langle\rho\rangle_{U_{\mathbf{\Delta t}}}(u_{\mathbf{\Delta t}})$. Despite the non-significance, our results confirm the tendency observed by Erhardt et al. (2019) of calcium to lead the transition. However, we must state that this tendency cannot be discriminated from a purely stochastic feature and therefore – provided the physical proxy interpretation holds true – cannot serve as evidence for atmospheric changes to trigger sea ice retreat during DO events. In fact, by calculating the likelihood of lead-lag configurations across the 16 events under study, we find evidence for the data to comprise both, transitions led

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020


Climate of the Past Discussions

by Ca^{2+} and transitions led by Na^{+}. This finding implies that atmospheric changes cannot be viewed as an exclusive trigger mechanism for the North Atlantic sea ice retreat during DO events, because such a causal relationship would require the temporal sequence to be preserved across all events. In the one-mechanism framework, our results could still be explained by an external trigger not represented by Ca^{2+} nor Na^{+}, which drives both atmospheric changes and sea ice retreat, with the atmosphere simply responding faster. Other than that, our findings question the assumption that all DO events followed the same physical pattern of causes and consequences. In order to gain more insight in the temporal order of events we suggest the following: First, studies similar to this one should be carried out targeting the phasing of other proxy variables. The stochastic ramp fit introduced by Erhardt et al. (2019) already constitutes a sound approach for the derivation of transition leads and lags in proxy time series. Second, the interpretation of proxy variables requires further refinement. Especially the interpretation of Na^{+} as a sea ice proxy remains debated and new results are on their way. Third, effort in conducting model studies should be sustained. Especially proxy enabled modeling bears the potential to improve comparability between model results and paleoclimate records. Together, these lines of research are promising to further constrain the sequence of events that have caused the abrupt climate changes during the last glacial.

Code and data availability. 10-year resolution time series of Na^{+} and Ca^{2+} concentrations and $\delta^{18}O$ values from the NGRIP ice core shown in Fig. 1 are retrieved from PANGAEA (Erhardt et al., 2018, https://doi.org/10.1594/PANGAEA.896743). So are the high-resolution Na^{+} and Ca^{2+} concentration time series centered around DO transitions which were used to derive the time lags between the transition onsets of the two proxies. The code used to generate the empirical densities of transition onsets is available at https://github.com/terhardt/DO-progression (last access: 21 October 2020). The code used to carry out the statistical analysis of the sample of empirical transition onset distributions is available from the author upon request and will be published once the manuscript is accepted.

Acknowledgements. We thank Norbert Marwan for very helpful comments and discussions. This is TiPES contribution #X; the TiPES (Tipping Points in the Earth System) project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement 178 No. 820970. NB acknowledges funding by the Volkswagen foundation.

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions

References

Andersen, K. K., Azuma, N., Barnola, J. M., Bigler, M., Biscaye, P., Caillon, N., Chappellaz, J., Clausen, H. B., Dahl-Jensen, D., Fischer, H., Flückiger, J., Fritzsche, D., Fujii, Y., Goto-Azuma, K., Grønvold, K., Gundestrup, N. S., Hansson, M., Huber, C., Hvidberg, C. S., Johnsen, S. J., Jonsell, U., Jouzel, J., Kipfstuhl, S., Landais, A., Leuenberger, M., Lorrain, R., Masson-Delmotte, V., Miller, H., Motoyama, H., Narita, H., Popp, T., Rasmussen, S. O., Raynaud, D., Rothlisberger, R., Ruth, U., Samyn, D., Schwander, J., Shoji, H., Siggard-Andersen, M. L., Steffensen, J. P., Stocker, T., Sveinbjörnsdóttir, A. E., Svensson, A., Takata, M., Tison, J. L., Thorsteinsson, T., Watanabe, O., Wilhelms, F., and White, J. W.: High-resolution record of Northern Hemisphere climate extending into the last interglacial period, Nature, 431, 147–151, https://doi.org/10.1038/nature02805, 2004.

Boers, N., Goswami, B., and Ghil, M.: A complete representation of uncertainties in layer-counted paleoclimatic archives, Climate of the Past, 13, 1169–1190, https://doi.org/10.5194/cp-13-1169-2017, 2017.

Boers, N., Ghil, M., and Rousseau, D. D.: Ocean circulation, ice shelf, and sea ice interactions explain Dansgaard–Oeschger cycles, Proceedings of the National Academy of Sciences of the United States of America, 115, E11 005–E11 014, https://doi.org/10.1073/pnas.1802573115, 2018.

Bond, G., Broecker, W., Johnsen, S., McManus, J., Labeyrie, L., Jouzel, J., and Bonanl, G.: Correlations between climate records from North Atlantic sediments and Greenland ice, Nature, 365, 143–147, 1993.

Breitenbach, S. F., Rehfeld, K., Goswami, B., Baldini, J. U., 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, https://doi.org/10.5194/cp-8-1765-2012, 2012.

Broecker, W. S., Peteet, D. M., and Rind, D.: Does the ocean-atmosphere system have more than one stable mode of operation?, Nature, 315, 21–26, https://doi.org/10.1038/315021a0, 1985.

Buizert, C., Adrian, B., Ahn, J., Albert, M., Alley, R. B., Baggenstos, D., Bauska, T. K., Bay, R. C., Bencivengo, B. B., Bentley, C. R., Brook, E. J., Chellman, N. J., Clow, G. D., Cole-Dai, J., Conway, H., Cravens, E., Cuffey, K. M., Dunbar, N. W., Edwards, J. S., Fegyveresi, J. M., Ferris, D. G., Fitzpatrick, J. J., Fudge, T. J., Gibson, C. J., Gkinis, V., Goetz, J. J., Gregory, S., Hargreaves, G. M., Iverson, N., Johnson, J. A., Jones, T. R., Kalk, M. L., Kippenhan, M. J., Koffman, B. G., Kreutz, K., Kuhl, T. W., Lebar, D. A., Lee, J. E., Marcott, S. A., Markle, B. R., Maselli, O. J., McConnell, J. R., McGwire, K. C., Mitchell, L. E., Mortensen, N. B., Neff, P. D., Nishiizumi, K., Nunn, R. M., Orsi, A. J., Pasteris, D. R., Pedro, J. B., Pettit, E. C., Price, P. B., Priscu, J. C., Rhodes, R. H., Rosen, J. L., Schauer, A. J., Schoenemann, S. W., Sendelbach, P. J., Severinghaus, J. P., Shturmakov, A. J., Sigl, M., Slawny, K. R., Souney, J. M., Sowers, T. A., Spencer, M. K., Steig, E. J., Taylor, K. C., Twickler, M. S., Vaughn, B. H., Voigt, D. E., Waddington, E. D., Welten, K. C., Wendricks, A. W., White, J. W., Winstrup, M., Wong, G. J., and Woodruff, T. E.: Precise interpolar phasing of abrupt climate change during the last ice age, Nature, 520, 661–665, https://doi.org/10.1038/nature14401, 2015.

Cheng, H., Sinha, A., Cruz, F. W., Wang, X., Edwards, R. L., D'Horta, F. M., Ribas, C. C., Vuille, M., Stott, L. D., and Auler, A. S.: Climate change patterns in Amazonia and biodiversity, Nature Communications, 4, https://doi.org/10.1038/ncomms2415, 2013.

Clark, P. U., Pisias, N. G., Stocker, T. F., and Weaver, A. J.: The role of the thermohaline circulation in abrupt climate change, Nature, 415, 863–869, https://doi.org/https://doi.org/10.1038/415863a, 2002.

Dansgaard, W., Johnsen, S. J., Clausen, H. B., Dahl-Jensen, D., Gundestrup, N. S., Hammer, C. U., Hvidberg, C. S., Steffensen, J. P., Sveinbjörnsdottir, A. E., Jouzel, J., and Bond, G.: Evidence for general instability of past climate from a 250-kyr ice-core record, Nature, 364, 218–220, https://doi.org/10.1038/364218a0, 1993.

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Erhardt, T., Capron, E., Rasmussen, S. O., Schüpbach, S., Bigler, M., Adolphi, F., and Fischer, H.: High resolution aerosol, layer thickness and δ^{18}O data around Greenland warming events (10-60ka) from NGRIP and NEEM ice cores., https://doi.org/10.1594/PANGAEA.896743, https://doi.org/10.1594/PANGAEA.896743, 2018.

Erhardt, T., Capron, E., Olander Rasmussen, S., Schüpbach, S., Bigler, M., Adolphi, F., and Fischer, H.: Decadal-scale progression of the onset of Dansgaard-Oeschger warming events, Climate of the Past, 15, 811–825, https://doi.org/10.5194/cp-15-811-2019, 2019.

Fleitmann, D., Cheng, H., Badertscher, S., Edwards, R. L., Mudelsee, M., Göktürk, O. M., Fankhauser, A., Pickering, R., Raible, C. C., Matter, A., Kramers, J., and Tüysüz, O.: Timing and climatic impact of Greenland interstadials recorded in stalagmites from northern Turkey, Geophysical Research Letters, 36, 1–5, https://doi.org/10.1029/2009GL040050, 2009.

Fuhrer, K., Wolff, E. W., and Johnsen, S. J.: Timescales for dust variability in the Greenland Ice Core Project (GRIP) ice core in the last 100,000 years, Journal of Geophysical Research Atmospheres, 104, 31 043–31 052, https://doi.org/10.1029/1999JD900929, 1999.

Ganopolski, A. and Rahmstorf, S.: Rapid changes of glacial climate simulated in a coupled climate model, Nature, 409, 153–158, www.nature.com, 2001.

Gkinis, V., Simonsen, S. B., Buchardt, S. L., White, J. W., and Vinther, B. M.: Water isotope diffusion rates from the NorthGRIP ice core for the last 16,000 years - Glaciological and paleoclimatic implications, Earth and Planetary Science Letters, 405, 132–141, https://doi.org/10.1016/j.epsl.2014.08.022, http://dx.doi.org/10.1016/j.epsl.2014.08.022, 2014.

Goswami, B., Boers, N., Rheinwalt, A., Marwan, N., Heitzig, J., Breitenbach, S. F., and Kurths, J.: Abrupt transitions in time series with uncertainties, Nature Communications, 9, 1–10, https://doi.org/10.1038/s41467-017-02456-6, http://dx.doi.org/10.1038/s41467-017-02456-6, 2018.

Hall, P. and Wilson, S. R.: Two Guidelines for Bootstrap Hypothesis Testing, International Biometric Society, 47, 757–762, https://doi.org/10.2307/2532163, 1991.

Henry, L. G., McManus, J. F., Curry, W. B., Roberts, N. L., Piotrowski, A. M., and Keigwin, L. D.: North Atlantic ocean circulation and abrupt climate change during the last glaciation, Science, 353, 470–474, https://doi.org/10.1126/science.aaf5529, 2016.

Huber, C., Leuenberger, M., Spahni, R., Flückiger, J., Schwander, J., Stocker, T. F., Johnsen, S., Landais, A., and Jouzel, J.: Isotope calibrated Greenland temperature record over Marine Isotope Stage 3 and its relation to CH4, Earth and Planetary Science Letters, 243, 504–519, https://doi.org/10.1016/j.epsl.2006.01.002, 2006.

Johnsen, S. J., Clausen, H. B., Dansgaard, W., Fuhrer, K., Gundestrup, N., Hammer, C. U., Iversen, P., Jouzel, J., Stauffer, B., and Steffensen, J.: Irregular glacial interstadials recorded in a new Greenland ice core, Nature, 359, 311–313, 1992.

Kim, J. H., Romero, O. E., Lohmann, G., Donner, B., Laepple, T., Haam, E., and Sinninghe Damsté, J. S.: Pronounced subsurface cooling of North Atlantic waters off Northwest Africa during Dansgaard-Oeschger interstadials, Earth and Planetary Science Letters, 339-340, 95–102, https://doi.org/10.1016/j.epsl.2012.05.018, 2012.

Kindler, P., Guillevic, M., Baumgartner, M., Schwander, J., Landais, A., and Leuenberger, M.: Temperature reconstruction from 10 to 120 kyr b2k from the NGRIP ice core, Climate of the Past, 10, 887–902, https://doi.org/10.5194/cp-10-887-2014, 2014.

Kleppin, H., Jochum, M., Otto-Bliesner, B., Shields, C. A., and Yeager, S.: Stochastic atmospheric forcing as a cause of Greenland climate transitions, Journal of Climate, 28, 7741–7763, https://doi.org/10.1175/JCLI-D-14-00728.1, 2015.

Landais, A., Jouzel, J., Masson-Delmotte, V., and Caillon, N.: Large temperature variations over rapid climatic events in Greenland: a method based on air isotopic measurements, Comptes Rendus - Geoscience, 337, 947–956, https://doi.org/10.1016/j.crte.2005.04.003, 2005.

Lehmann, E. L. and Romano, J. P.: Testing Statistical Hypothesis, vol. 102, Springer US, New York, third edit edn., https://doi.org/10.1016/j.peva.2007.06.006, 2006.

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Lenton, T., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., and Schellnhuber, H. J.: Tipping elements in the Earth's climate system, Proceedings of the National Academy of Sciences of the United States of America, 105, 1786–1793, https://doi.org/https://doi.org/10.1073/pnas.0705414105, 2008.

Lenton, T. M. and Schellnhuber, H. J.: Tipping the Scales, Nature Climate Change, 1, 97–98, 2007.

Lenton, T. M., Rockström, J., Gaffney, O., Rahmstorf, S., Richardson, K., Steffen, W., and Schellnhuber, H. J.: Climate tipping points — too risky to bet against, Nature, 575, 592–595, https://doi.org/doi: 10.1038/d41586-019-03595-0, https://www.nature.com/articles/d41586-019-03595-0, 2019.

Li, C. and Born, A.: Coupled atmosphere-ice-ocean dynamics in Dansgaard-Oeschger events, Quaternary Science Reviews, 203, 1–20, https://doi.org/10.1016/j.quascirev.2018.10.031, https://doi.org/10.1016/j.quascirev.2018.10.031, 2019.

Lynch-Stieglitz, J.: The Atlantic Meridional Overturning Circulation and Abrupt Climate Change, Annual Review of Marine Science, 9, 83–104, https://doi.org/10.1146/annurev-marine-010816-060415, 2017.

Moseley, G. E., Spötl, C., Brandstätter, S., Erhardt, T., Luetscher, M., and Edwards, R. L.: NALPS19: Sub-orbital scale climate variability recorded in Northern Alpine speleothems during the last glacial period, Climate of the Past Discussions, 16, 1–25, https://doi.org/10.5194/cp-16-29-2020, https://doi.org/10.5194/cp-16-29-2020, 2020.

Rasmussen, S. O., Bigler, M., Blockley, S. P., Blunier, T., Buchardt, S. L., Clausen, H. B., Cvijanovic, I., Dahl-Jensen, D., Johnsen, S. J., Fischer, H., Gkinis, V., Guillevic, M., Hoek, W. Z., Lowe, J. J., Pedro, J. B., Popp, T., Seierstad, I. K., Steffensen, J. P., Svensson, A. M., Vallelonga, P., Vinther, B. M., Walker, M. J., Wheatley, J. J., and Winstrup, M.: A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: Refining and extending the INTIMATE event stratigraphy, Quaternary Science Reviews, 106, 14–28, https://doi.org/10.1016/j.quascirev.2014.09.007, http://dx.doi.org/10.1016/j.quascirev.2014.09.007, 2014.

Rice, J. A.: Mathematical Statistics and Data Analysis, Thomson Brooks/Cole, third edit edn., 2007.

Ruth, U., Bigler, M., Röthlisberger, R., Siggaard-Andersen, M. L., Kipfstuhl, S., Goto-Azuma, K., Hansson, M. E., Johnsen, S. J., Lu, H., and Steffensen, J. P.: Ice core evidence for a very tight link between North Atlantic and east Asian glacial climate, Geophysical Research Letters, 34, 1–5, https://doi.org/10.1029/2006GL027876, 2007.

Schüpbach, S., Fischer, H., Bigler, M., Erhardt, T., Gfeller, G., Leuenberger, D., Mini, O., Mulvaney, R., Abram, N. J., Fleet, L., Frey, M. M., Thomas, E., Svensson, A., Dahl-Jensen, D., Kettner, E., Kjaer, H., Seierstad, I., Steffensen, J. P., Rasmussen, S. O., Vallelonga, P., Winstrup, M., Wegner, A., Twarloh, B., Wolff, K., Schmidt, K., Goto-Azuma, K., Kuramoto, T., Hirabayashi, M., Uetake, J., Zheng, J., Bourgeois, J., Fisher, D., Zhiheng, D., Xiao, C., Legrand, M., Spolaor, A., Gabrieli, J., Barbante, C., Kang, J. H., Hur, S. D., Hong, S. B., Hwang, H. J., Hong, S., Hansson, M., Iizuka, Y., Oyabu, I., Muscheler, R., Adolphi, F., Maselli, O., McConnell, J., and Wolff, E. W.: Greenland records of aerosol source and atmospheric lifetime changes from the Eemian to the Holocene, Nature Communications, 9, https://doi.org/10.1038/s41467-018-03924-3, http://dx.doi.org/10.1038/s41467-018-03924-3, 2018.

Steffensen, J. P., Andersen, K. K., Bigler, M., Clausen, H. B., Dahl-jensen, D., Fischer, H., Goto-azuma, K., Hansson, M., Johnsen, S. J., Jouzel, J., Masson-delmotte, V., Popp, T., Rasmussen, S. O., Röthlisberger, R., Ruth, U., Stauffer, B., Sveinbjörnsdóttir, Á. E., Svensson, A., and White, J. W. C.: High-Resolution Greenland Ice Core Data Show Abrupt Climate Change Happens in Few Years, Science, 321, 680–684, https://doi.org/10.1126/science.1157707, 2008.

Student: The probable error of a mean, Biometrika, 6, 1–25, https://doi.org/10.2307/2331554, https://www.jstor.org/stable/2331554, 1908.

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Thomas, E. R., Wolff, E. W., Mulvaney, R., Johnsen, S. J., Steffensen, J. P., and Arrowsmith, C.: Anatomy of a Dansgaard-Oeschger warming transition: High-resolution analysis of the North Greenland Ice Core Project ice core, Journal of Geophysical Research Atmospheres, 114, 1–9, https://doi.org/10.1029/2008JD011215, 2009.

Vettoretti, G. and Peltier, W. R.: Fast physics and slow physics in the nonlinear Dansgaard-Oeschger relaxation oscillation, Journal of Climate, 31, 3423–3449, https://doi.org/10.1175/JCLI-D-17-0559.1, 2018.

Voelker, A. H.: Global distribution of centennial-scale records for Marine Isotope Stage (MIS) 3: A database, Quaternary Science Reviews, 21, 1185–1212, https://doi.org/10.1016/S0277-3791(01)00139-1, 2002.

Wilcoxon, F.: Individual comparisons of grouped data by ranking methods., Biometrics Bulletin, 1, 80–83, https://doi.org/10.2307/3001968, http://links.jstor.org/sici?sici=0099-4987%28194512%291%3A6%3C80%3AICBRM%3E2.0.CO%3B2-P, 1945.

Zhang, X., Lohmann, G., Knorr, G., and Purcell, C.: Abrupt glacial climate shifts controlled by ice sheet changes, Nature, 512, 290–294, https://doi.org/10.1038/nature13592, http://dx.doi.org/10.1038/nature13592, 2014.

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

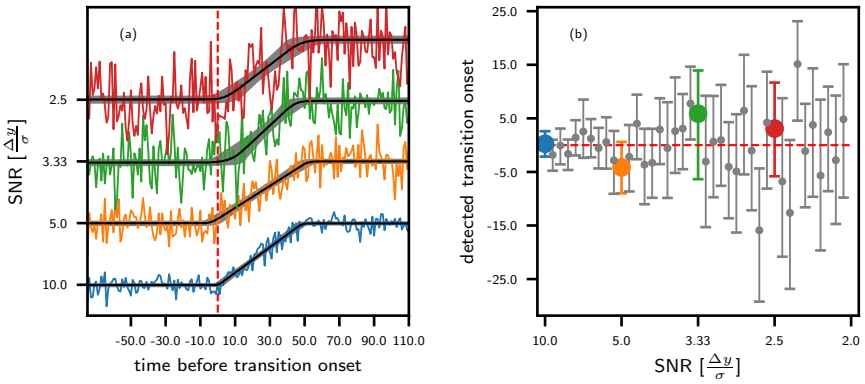

Figure A1. (a) Synthetically generated time series featuring a linear transition of fixed height Δy. The time series are perturbed by Gaussian white noise with different standard deviations σ and therefore exhibit different signal-to-noise ratios (SNR) $\frac{\Delta y}{\sigma}$. The black line is the mean over the ensemble of 6000 ramps sampled with the MCMC ramp-fit algorithm, while the gray shaded area indicates the 5th to 95th percentiles of the same ensemble. (b) Mean transition onsets $\langle t_0 \rangle = \int \rho^{\text{emp}}(t_0)\, t_0\, dt_0$ derived from the ensembles sampled by the MCMC ramp-fit algorithm for synthetically generated noisy ramps with different signal to noise ratios. The error bars indicate the 5th to 95th percentiles for t_0.

Appendix A: Sensitivity of the transition detection

Na^+ transitions exhibit a lower signal to noise ratio than corresponding Ca^{2+} transitions for many of the events under study. Therefore, we carried out a performance test for our stochastic transition detection by applying the algorithm to synthetically generated linear ramps disturbed by Gaussian white noise. To investigate the influence of the noise, we kept the ramp amplitude Δy constant and step-wise increased the standard deviation σ of the perturbing noise. As can be seen in Fig. A1, the detection of the transition onset becomes less precise with decreasing signal to noise ratio $\frac{\Delta y}{\sigma}$. That is, the mean of the posterior distribution $\rho^{\text{emp}}_{t_0}(t_0)$ may differ stronger from the true transition onset for lower signal to noise ratios while simultaneously the standard deviation of $\rho^{\text{emp}}_{t_0}(t_0)$ increases. Importantly, there is no systematic bias of the detected transition onset in the one or the other direction. Also, we find that for most of the synthetic time series, the true transition onset lies inside the uncertainty interval of the derived distributions $\rho^{\text{emp}}_{t_0}(t_0)$. Thus, the algorithm is suited for application to data with different signal-to-noise ratios. The lower signal to noise ratio is reflected in the broader distribution for t_0 computed by the algorithm.

Appendix B: Distribution of the uncertain statistics

In Sec. 3.6 we discuss the propagation of the sample uncertainty to the p-value of any given hypothesis test. Certainly, this relies on an intermediate propagation of the uncertainty to the corresponding test statistic. Fig. B1 shows the density histograms

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

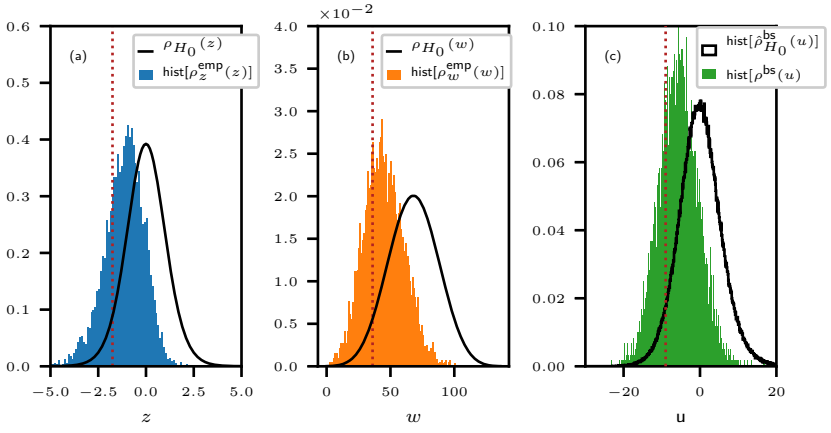

Figure B1. Density histograms of the uncertain test statistics Z (a), W (b) and U (c) associated with their empirical density distributions, which were in turn derived from the empirical density $\rho^{\mathrm{emp}}_{\mathbf{\Delta t}}(\mathbf{\Delta t})$ of the uncertain sample of transition lags. The statistics correspond to the t-test, the WSR-test and the bootstrap test presented in Sec. 3.6. The distributions of the statistics under the null hypothesis are displayed for comparison (black). In the case of the t-test, under the null hypothesis z follows the t-distribution with 15 degrees of freedom. The distribution of the w-statistic under the null hypothesis is given by a lengthy analytical expression and is defined only for integer values. For the bootstrap test, the null-hypothesis distribution is derived empirically. The red dotted lines separate the significant from the non-significant values in all three panels.

associated with the empirical distributions for Z, W and U that result from the distribution $\rho^{\mathrm{emp}}_{\mathbf{\Delta t}}(\Delta t)$ according to

$$\rho^{\mathrm{emp}}_{\phi}(\phi) = \int \rho^{\mathrm{emp}}_{\mathbf{\Delta t}}(\mathbf{\Delta t})\delta(\phi(\mathbf{\Delta t}) - \phi)\, d\Delta t. \tag{B1}$$

Here, $\phi(\mathbf{\Delta t})$ represents a test statistic as a function of the uncertain sample $\mathbf{\Delta t}$, while ϕ independent of any arguments denotes possible values that $\phi(\mathbf{\Delta t})$ may assume. For comparison, the distribution of each test statistic under the null-hypothesis is displayed as well. The histograms are not supposed to coincide with the null distributions. Clearly, for all three tests, the histograms of the uncertain test statistics empirical probability densities comprise both, significant and non-significant values of the statistic.

Appendix C: Results of the analysis for the control group

As explained in Sec. 3.3, we drastically reduce the set of vectors considered in the representation of $\rho^{\mathrm{emp}}_{\mathbf{\Delta t}}(\mathbf{\Delta t}) = \frac{1}{6000}\sum_{i=1}^{6000}\delta(\mathbf{\Delta t} - \mathbf{\Delta t}_i^{\mathrm{emp}})$ from 6000^{16} theoretically available vectors to 6000 vectors $\mathbf{\Delta t}_i$. To cross-check that the results obtained within the limits of this approximation, we applied our analysis to a control group of 10 alternative realizations of the the probability density $\rho^{\mathrm{emp}}_{\mathbf{\Delta t},j}(\mathbf{\Delta t}) = \frac{1}{6000}\sum_{i=1}^{6000}\delta(\mathbf{\Delta t} - \mathbf{\Delta t}^{\mathrm{emp}}_{i,j})$, all drawing on different representative sets $\{\mathbf{\Delta t}_i^{\mathrm{emp}}\}_j$ with $i \in \{1, 2, ..., 6000\}$ and $j \in \{1, 2, ..., 10\}$. These sets have been generated by drawing 6000 vectors at random from the pool of the theoretically available vectors. We find that all original results that rely on the empirical density $\rho^{\mathrm{emp}}_{\mathbf{\Delta t}}(\mathbf{\Delta t})$ fall well into the uncertainty range

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions

Table C1. Results obtained from the application of hypothesis tests to the control group. Reported are the mean p-values $\langle p(\mathbf{\Delta t})\rangle$ together with the p-values of the expected samples $p(\langle \mathbf{\Delta t}\rangle)$ and the probability of the uncertain sample to be significant $P(p < 0.05)$ for all three tests. All results were derived from the corresponding empirical densities $\rho^{\text{emp}}_{\mathbf{\Delta t},j}(\mathbf{\Delta t})$ (labeled by r_j). Also listed are the mean and the standard deviation (std) for each quantity across the control group. For comparison, the results from the original analysis are given as well.

	t-test			WSR-test			bootstrap test		
no.	$\langle p(\mathbf{\Delta t})\rangle$	$p(\langle \mathbf{\Delta t}\rangle)$	$P(p < 0.05)$	$\langle p(\mathbf{\Delta t})\rangle$	$p(\langle \mathbf{\Delta t}\rangle)$	$P(p < 0.05)$	$\langle p(\mathbf{\Delta t})\rangle$	$p(\langle \mathbf{\Delta t}\rangle)$	$P(p < 0.05)$
r_1	0.220	0.044	0.258	0.166	0.011	0.324	0.220	0.018	0.250
r_2	0.219	0.041	0.252	0.167	0.011	0.317	0.218	0.026	0.244
r_3	0.218	0.042	0.260	0.167	0.009	0.321	0.219	0.016	0.251
r_4	0.216	0.042	0.249	0.164	0.009	0.319	0.214	0.019	0.244
r_5	0.221	0.048	0.243	0.167	0.011	0.316	0.222	0.018	0.238
r_6	0.217	0.043	0.256	0.164	0.011	0.318	0.216	0.027	0.252
r_7	0.219	0.044	0.251	0.165	0.011	0.317	0.219	0.016	0.244
r_8	0.216	0.042	0.257	0.165	0.011	0.322	0.216	0.017	0.251
r_9	0.221	0.046	0.249	0.169	0.009	0.320	0.221	0.031	0.249
r_{10}	0.221	0.043	0.254	0.169	0.009	0.314	0.220	0.019	0.250
mean	0.219	0.044	0.253	0.166	0.010	0.319	0.219	0.021	0.247
std	0.002	0.002	0.005	0.002	0.001	0.003	0.002	0.005	0.004
original	0.219	0.044	0.258	0.168	0.009	0.325	0.220	0.018	0.253

determined by the control group. Tab. C2 summarizes the results derived from the control group members for the population mean, while Tab. C1 shows the results obtained by application of the hypothesis tests to the control group.

https://doi.org/10.5194/cp-2020-136
Preprint. Discussion started: 10 November 2020

Climate of the Past Discussions

Table C2. Expected values for the population mean derived from the control group probability densities together with the corresponding probability for the population mean to be less than zero. The left part of the table refers population mean distributions obtained by inverting the t-distribution, while the right part refers to population mean distributions obtained from bootstrapping sample means from the empirical probability distribution of the uncertain sample. Also listed are the mean values and standard deviations across the control group and for comparison the results from the original probability density as reported in the main part of this article.

	$\langle\rho\rangle_\mu^{\mathrm{emp}}(\mu)$		$\langle\rho\rangle_\mu^{\mathrm{bs}}(\mu)$	
no.	$\langle\mu\rangle$	$p(\mu<0)$	$\langle\mu\rangle$	$p(\mu<0)$
r_1	-5.60	0.783	-5.61	0.788
r_2	-5.66	0.784	-5.66	0.788
r_3	-5.74	0.787	-5.75	0.793
r_4	-5.63	0.784	-5.62	0.788
r_5	-5.56	0.782	-5.56	0.787
r_6	-5.66	0.785	-5.65	0.789
r_7	-5.58	0.782	-5.56	0.786
r_8	-5.67	0.786	-5.68	0.791
r_9	-5.64	0.783	-5.65	0.788
r_{10}	-5.75	0.788	-5.75	0.793
mean	-5.65	0.784	-5.65	0.789
std	0.06	0.002	0.07	0.002
original	-5.66	0.784	-5.66	0.789