Bayesian analysis of early warning signals using a time-dependent model

Eirik Myrvoll-Nilsen[1], Luc Hallali[1], and Martin Wibe Rypdal[1]

[1]Department of Mathematics and Statistics, UiT The Arctic University of Norway, N-9037 Tromsø, Norway

Correspondence: Eirik Myrvoll-Nilsen (eirik.myrvoll-nilsen@uit.no)

Abstract.

A tipping point is defined by the IPCC as a critical threshold beyond which a system reorganizes, often abruptly and/or irreversibly. Tipping points can be crossed solely by internal variation in the system or by approaching a bifurcation point where the current state loses stability~~and~~ , which forces the system to move to another stable state. It ~~is currently debated whether or not Dansgaard-Oeschger (DO) events, abrupt warmings occurring during the last glacial period, are noise-induced or caused by the system reaching a bifurcation point. It~~ can be shown that before a bifurcation point is reached there are observable changes in the statistical properties of the state variable. These are known as early warning signals and include increased fluctuation and correlation time. ~~To express this~~ It is currently debated whether or not Dansgaard-Oeschger (DO) events, abrupt warmings of the North Atlantic region which occurred during the last glacial period, are preceded by early warning signals. To express the changes in statistical behaviour we propose a ~~new~~ model based on the well-known first order autoregressive process (AR), with modifications to the correlation parameter such that it depends linearly on time. In order to estimate the time evolution of the correlation parameter we adopt a hierarchical Bayesian modeling framework, from which Bayesian analysis can be performed using the methodology of integrated nested Laplace approximations. We then apply the model to segments of the oxygen isotope ratios from the Northern Greenland Ice Core Project record corresponding to 17 DO events. ~~Early warning signals were detected and found statistically significant~~ Statistically significant early warning signals are detected for a number of DO events, ~~suggesting~~ which suggests that such events could indeed ~~be~~ exhibit signs of ongoing destabilization and may have been caused by approaching a bifurcation point. The methodology developed to perform the given early warning analyses can be applied more generally ~~,~~ and is publicly available as the R-package `INLA.ews`.

1 Introduction

~~An equilibrium state is said to be stable if the system returns to the same state following a small perturbation in any direction.~~ If the state of a component of the climate system ~~,~~ changes from one stable equilibrium to another, either by crossing some threshold in the form of ~~an unstable barrier~~ a boundary of unstable fixed points separating two basins of attraction ~~, changes from one stable equilibrium to another~~ or by having the initial equilibrium destabilize, it is said to have ~~reached~~ crossed a tipping point. Components of the Earth system ~~has~~ have experienced tipping points numerous times in the past, leading to abrupt transitions in the climate system. These transitions are well documented in paleoclimatic proxy records. Notably, in

Greenland ice core records of oxygen isotope ratios (δ^{18}O) and dust concentrations there is evidence that large and abrupt climatic transitions from Greenland stadial (GS) to Greenland interstadial (GI) conditions took place in the last glacial interval (110,000–12,000 years before 2000 AD, hereafter denoted yr b2k). These transitions are known as Dansgaard-Oeschger (DO) events (Dansgaard et al., 1984, 1993) and are characterized by initialize climatic cycles where the temperature increased substantially, up to 16.5°C for single events, over the course of a few decades. This is followed by a more gradual cooling, over centuries to millenia, back returning to the GS state. A total of 17 DO events (Svensson et al., 2008) (Rasmussen et al., 2014) have been found for the past 60 kyr before present (BP) and they represent some of the most pronounced examples of abrupt transitions in past climate observed in paleoclimatic records.

It is widely accepted that such transitions are associated with related to a change in the meriodional overturning circulation (MOC) (Bond et al., 1999; Li et al., 2010) causing a (Lynch-Stieglitz, 2017; Henry et al., 2016; Menviel et al., 2014, 2020; Bond et al., 19 , possibly caused by loss of sea ice in the North Atlantic. However, the physical mechanisms that caused these changes in the MOC and how they triggered DO events are less understood. Some studies have found that DO events exhibit a periodicity of 1470 years (Schulz, 2002), which have has made some scientists suggest that the events have been triggered by changes in the earth system caused by changing quasi-periodic changes in the solar forcing (Braun et al., 2005). Others suggest that the transitions have been triggered by random fluctuations in the Earth system, without any significant changes to the underlying system caused by external forcing (Ditlevsen et al., 2007). Treating the GS and GI states as stable equilibria in a dynamical system representing the Greenland climate, and studying the statistical behaviour related to the stability of the system in the period preceeding DO events, can help determine whether or not they are forced or random and thus possibly constrain the number of plausible physical causes that trigger the events.

The behaviour around a tipping point can be analyzed by expressing the changes of the state-variable using a potential, wherein valleys represent the basins of attraction that are separated by an unstable fixed point. If the tipping point is reached solely from perturbations caused by internal variation of the system, then it is said to be noise-induced. However, if the dynamics of the system depend on some slowly varying control parameter the equilibrium points may shift, vanish or spawn as a function of the control parameter. This means that the stability of a fixed point can change over time and eventually be lost, making the system move to another equilibrium. Points in the control parameter space for which the qualitative behaviour of a system changes, e.g. change in stability or the number of fixed points, are called bifurcation points, and tipping points caused by the control parameter crossing a bifurcation point are said to be bifurcation-induced.

By assuming that Let a time-dependent state-variable $x(t)$, representing for example the δ^{18}O ratio, vary over some potential $V(x)$ with stochastic forcing corresponding to a white noise process $dB(t)$, expressed as the derivative of a Brownian motion, then the stability of the system can be modeled using the stochastic differential equation

$$dx(t) = F(x(t))dt + \sigma dB(t). \tag{1}$$

One could interpret this equation as describing the motion of some particle in the presence of a potential $V(x)$, with drift expressed by $F(x) = -V'(x)$ and a diffusion term $\sigma dB(t)$ describing the noise that acts on the particle.

Take for example the cusp catastrophe model where the potential is given by

$$V(x,\mu,\xi) = \frac{x^4}{4} - \xi\frac{x^2}{2} - \mu x,$$

where $\mu(t)$ is a slowly changing control parameter and ξ is a shape parameter that we in this example set equal to $\xi = 1$. The change in position $dx(t)$ at some time t is then given by

$$dx(t) = -x^3 + \xi x + \mu + \sigma dB(t).$$

It can be shown that the bifurcation pointsare

$$\mu_1 = -\frac{2}{3}\sqrt{\xi^3/3} \quad \text{and} \quad \mu_2 = \frac{2}{3}\sqrt{\xi^3/3}.$$

Crossing the bifurcation points changes the number of fixed points. For $\mu_1 < \mu < \mu_2$ there are two Points where $dx(t)/dt = 0$ are fixed points. These are stable if a small perturbation of the state variable near the fixed point decays in time and unstable otherwise. Fig. 1a illustrates an example of a potential with two valleys corresponding to stable fixed points and one unstable , and for $\mu < \mu_1$ or $\mu > \mu_2$ there is only one (stable) fixed point . The change of stability can be depicted by plotting the potential before and after the bifurcation points , see x_1 and x_3 that are separated by an unstable fixed point x_2. If a state variable near x_1 crosses x_2 into the basin of attraction of x_3 solely from perturbations caused by internal variation of the system, then the associated tipping point is said to be noise-induced. However, if the dynamics of the system depend on some slowly varying control parameter $\mu(t)$, then the equilibrium points may shift, vanish or spawn as a function of $\mu(t)$. This means that an equilibrium state of a system can change over time and eventually be lost, making the system move to another equilibrium, as illustrated in Fig. 1for an illustration where the control variable varies around μ_2.The change in values and stability of the fixed points as we increase the control parameter is illustrated in the bifurcation diagram a–c. Points in the control parameter space for which the qualitative behaviour of a system changes, e.g. changes in stability or the number of fixed points, are called bifurcation points. Fig. **??**, which include 1d illustrates these changes using a bifurcation diagram, where the stable fixed points x_1 (lower solid curve) and x_3 (upper solid curve) and are separated by the unstable fixed points x_2 (middle dashed curve), representing the separating barrier. The diagram also includes a simulated process generated by the same potential which demonstrates how abruptly the state variable changes when the system crosses the tipping threshold x_2, which happens before the control parameter reaches the bifurcation point μ_2 due to the diffusion term $\sigma dB(t)$. . Critical transitions caused by the control parameter crossing a bifurcation point are said to be bifurcation-induced tipping points.

The bifurcation diagram of the cusp catastrophe model. The black curve represent the fixed points of the state variable x given the changing control parameter $\mu \in (-1,1)$. The solid curves represent stable fixed points x_1 and x_3, and the dashed curve represent unstable fixed points x_2. The red line represent a simulation of Eq. (**??**) with $\sigma = 0.2$. As the control parameter μ approaches the bifurcation point μ_2 the stability of x_1 decreases which is expressed by increased variance and correlation in the simulated process, causing the system to cross the tipping point x_2 prematurely.

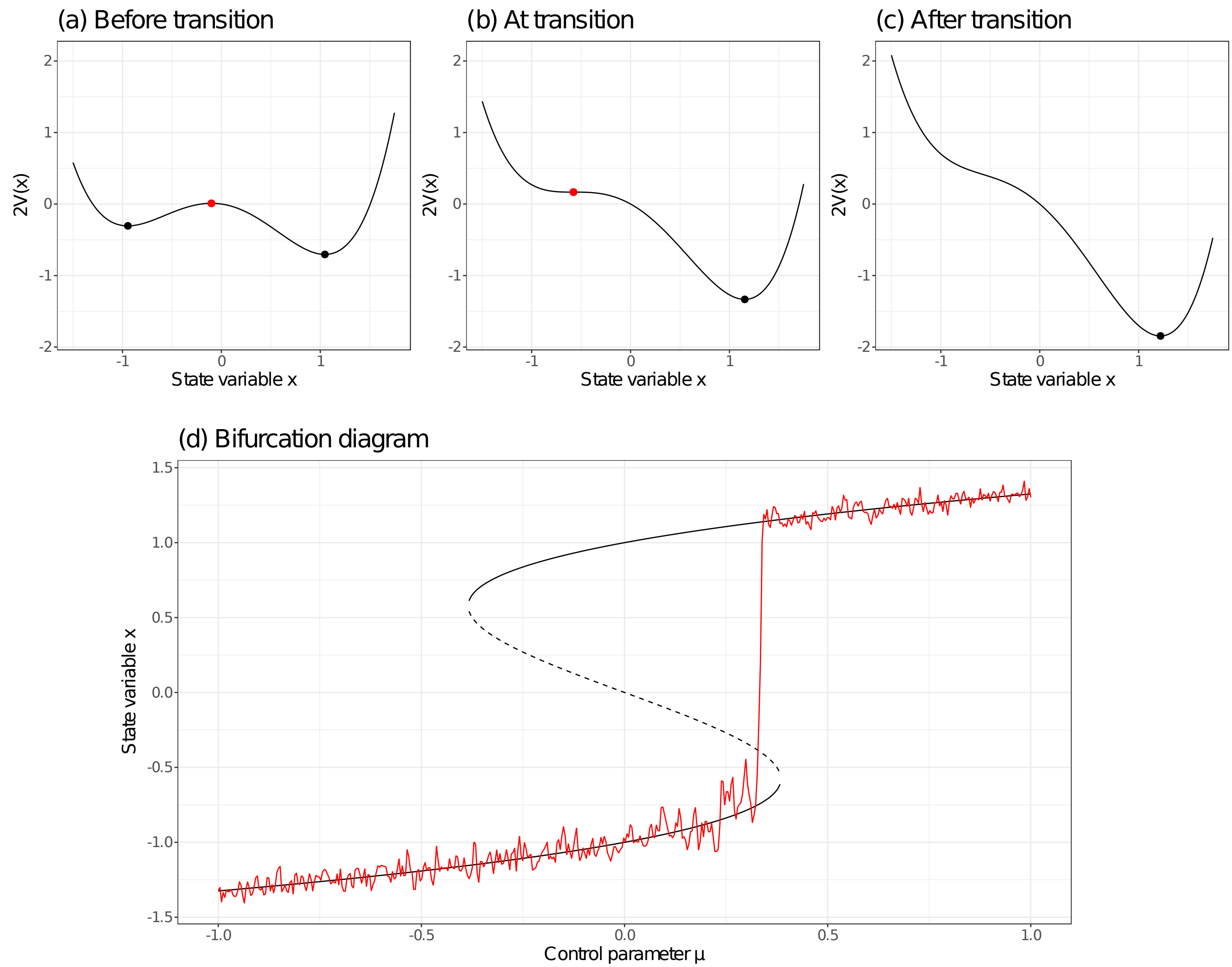

Figure 1. ~~The~~ Panels (a)–(c) show the potential over the ~~set of~~ state variables before, at and after the control parameter has reached the bifurcation point ~~μ_2~~μ^*. Panel (a) shows the potential and fixed points x_1, x_2 and x_3 for some ~~$\mu < \mu_2$~~$\mu < \mu^*$, and panels (b)–(c) ~~shows~~ show the same for ~~$\mu = \mu_2$~~ $\mu = \mu^*$ and ~~$\mu > \mu_2$~~$\mu > \mu^*$, respectively. When the control parameter approaches the bifurcation point ~~μ_2~~μ^*, the stability of the stable fixed point x_1 decreases and eventually collapses ~~at $x_1 = x_2 = -\sqrt{\xi/3}$~~with x_2, leaving x_3 as the only (stable) fixed point. Panel (d): Bifurcation diagram describing a bifurcation induced tipping point. The black curve represent the fixed points of the state variable x given the linearly changing control parameter $\mu \in (-1, 1)$. The solid curves represent stable fixed points x_1 and x_3, and the dashed curve represent unstable fixed points x_2. The red line represent a simulated process. As the control parameter μ approaches the bifurcation point μ^* the stability of x_1 decreases which is expressed by increased variance and correlation in the simulated process, causing the system to cross the tipping point x_2 prematurely.

The nature of an equilibrium can be investigated by examining the linear approximation in its nearby domain. Linearizing (1) around some stable fixed point x_s yields

$$dx(t) = -\lambda(x(t) - x_s)dt + \sigma dB(t), \tag{2}$$

where $\lambda = -F'(x_s)$. This is known as the Langevin stochastic differential equation and has the solution

$$x(t) = x_0 + \int_{-\infty}^{t} g(t-s)dB(s), \tag{3}$$

with Green's function

$$g(t) = \begin{cases} \exp(-\lambda t), & x \geq 0 \\ 0, & x < 0 \end{cases}. \tag{4}$$

This solution forms an Ornstein-Uhlenbeck (OU) process ~~, which under discretization~~ with variance $\mathrm{Var}(x_t) = \sigma^2/(2\lambda)$. Under discretization this is a first order autoregressive (AR) process ~~with variance $\mathrm{Var}(x_t) = \sigma^2/(2\lambda)$ and~~

$$x_t = \phi x_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}\left(0, \frac{1-\phi^2}{2\lambda}\sigma^2\right) \tag{5}$$

with lag-one autocorrelation parameter ~~$\phi(t) = \exp(-\lambda)$.~~

~~$$x_t = \phi x_{t-1} + \varepsilon_t, \qquad \varepsilon \sim \mathcal{N}\left(0, \frac{1-\phi^2}{2\lambda}\sigma^2\right)$$~~

$\phi = \exp(-\lambda \Delta t)$.

When the control parameter approaches a bifurcation point ~~we expect increased~~ the restoring rate λ goes to zero and consequently the variance and correlation of the state variable will increase, as could be observed in Fig. ~~??.~~ 1d. This phenomenon was first demonstrated by inspecting the power spectra of a simple physical model by Wiesenfeld (1985). The idea was later extended to a complex earth system model by Held and Kleinen (2004) and first applied to real data by Dakos et al. (2008) . These changes in statistical behaviour are called early-warning signals (EWS) of the bifurcation point, or critical slowing down (Lenton et al., 2012; Dakos et al., 2008), and can be used as precursors to help determine whether or not a tipping point is imminent. ~~In fact, recent~~ EWS are derived from the linearization of the system around its fixed points, however even in cases where the system dynamic is far from its equilibrium the same changes in statistical behaviour can be found with a delay (Ritchie and Sieber, 2016).

Recent studies have discovered that ~~more~~ several components in the earth system exhibit EWS and are at risk of approaching or have already reached a tipping point. This include the western Greenland ice sheets (Boers and Rypdal, 2021), the Atlantic meridional overturning circulation (Boers, 2021) and the Amazon rainforest (Boulton et al., 2022).

Analysis of EWS for DO events in the ~~high-dimensional~~ Greenland ice core record has been conducted by others, e.g. Ditlevsen and Johnsen (2010) ~~whom applied a Monte Carlo approach to detect increased~~ who estimated the variance and autocorrelation ~~in a system~~ over a sliding window where the system was assumed to be driven by white noise. Under these assumptions they were unable to detect a statistical significant increase in EWS suggesting that DO events are noise-induced. However, using different model assumptions, Rypdal (2016) was able to detect statistically significant EWS in an ensemble of DO events. This was achieved by analyzing individual frequency bands separately, using a fractional Gaussian noise (fGn) (Mandelbrot and Van Ness, 1968) model to describe the noise. ~~Fractional Gaussian noise is a long-range dependent model for which the Green's function in (3) is scale-invariant~~

$$g(t) = \begin{cases} t^{H-3/2}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

~~$H \in (0.5, 1)$ is the memory coefficient known as the Hurst exponent. Fractional Gaussian noise have been shown to be more realistic for describing components in the Earth system where the power spectrum does not follow an exponential decay, such as monthly to centennial global and local mean surface temperature data (Lovejoy and Schertzer, 2013; Huybers and Curry, 2006; Rybski et a . Rypdal (2016) was able to detect an increase of variance of the high-frequency fluctuations for the ensemble average of the 17 DO events at a 5% significance level, and individually for five separate events.~~ These results were corroborated by Boers (2018) ~~whom~~ who applied a similar strategy to ~~the higher resolution~~ a higher resolved version of the NGRIP δ^{18}O data set (Andersen et al., 2004; Gkinis et al., 2014)~~on which he applied interpolation to obtain time series with regular 5-year sampling steps.~~ .

Most approaches for detecting EWS in the current literature require estimation of statistical properties such as variance and correlation in a sliding window~~, e. g. by producing Fourier surrogates and estimating the Kendall's τ statistic for each iteration.~~ . Consequently, this ~~presents~~ requires a choice on the length of the window. Using a small window will allow for the momentary state to be better depicted, but there will be fewer points used in the estimation hence accuracy will suffer. On the other hand, if a larger window is used the estimated statistics will be more accurate, but less representative of the momentary state as it represents an average over a larger time scale. The optimal choice of window length should ideally represent a good trade-off between accuracy and ability to represent momentary evolution, but this can be hard to determine in practice. In this paper we circumvent this issue and present a model-based approach where such a compromise is not required. By assuming that the correlation parameter is time-dependent, following a specific linear structure, it is possible to formulate this into a hierarchical Bayesian model for which well-known computational frameworks can be applied. A Bayesian approach has the additional benefit of providing uncertainty estimates in the form of posterior distributions.

The paper is structured as follows. A description of the data used in this paper is included in section 2. Section 3 details our methodology, including how we treat time-dependence, how to formulate our model as a hierarchical Bayesian model and how to perform statistical inference efficiently. Results are presented in section 4 where our framework is applied first to simulated data, then to Dansgaard-Oeschger events observed in the δ^{18}O data from the NGRIP record. Our results are compared with those obtained by Ditlevsen and Johnsen (2010), Rypdal (2016) and Boers (2018). Further discussion and conclusions are provided in section 5.

2 NGRIP ice core data

The δ^{18}O ratios are frequently used in paleoscience as proxies for temperature at the time of precipitation (Johnsen et al., 1992, 2001; Dansgaard et al., 1993; Andersen et al., 2004), where higher ratios signals colder climates and, conversely, warmer climates tend to result in lower ratios. We employ the δ^{18}O proxy record from the Northern Greenland Ice core Project (NGRIP) (North Greenland Ice Core Project members, 2004; Gkinis et al., 2014; Ruth et al., 2003). There are currently two different versions of the NGRIP/GICC05 data, at different resolutions. We will apply our methodology to the higher resolution record, which is sampled every 5cm in depth. The NGRIP δ^{18}O proxy record is defined on a temporal axis given by the Greenland Ice Core Chronology 2005 (GICC05) (Vinther et al., 2006; Rasmussen et al., 2006; Andersen et al., 2006; Svensson et al., 2008) which thus pairs the δ^{18}O measurements with a corresponding age, stretching back to 60 kyrs b2k. We use segments of the δ^{18}O record corresponding to Greenland stadial phases preceding DO onsets, as given by Table 2 of Rasmussen et al. (2014). The data used in this paper can be downloaded from https://www.iceandclimate.nbi.ku.dk/data/ (last accessed: **day month year**July 30, 2024)

3 Methodology

During critical slowing down stationarity can no longer be assumed as we expect both the correlation and variance to increase. For an AR(1) process $\boldsymbol{x} = (x_1, ..., x_n)^\top$ sampled at times $t_1, ..., t_n$, we assume that the increase in correlation can be expressed by representing the lag-one autocorrelation parameter as a linear function of time

$$\phi(t) = a + bt, \qquad 0 \leq t \leq 1, \tag{6}$$

where a and b are two unknown parameters. The time-dependent AR(1) process is expressed by the difference equation given in (5) and the joint vector of variables $\boldsymbol{x} = (x_1, ..., x_n)^\top$ forms a multivariate Gaussian process

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}), \tag{7}$$

where the covariance matrix is given by

$$\Sigma_{ij} = \text{Cov}(x_i, x_j)., \tag{8}$$

The time-dependent AR(1) process is expressed by the difference equation

$$x_t = \phi(t)x_{t-1} + \varepsilon_t, \qquad \varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2), \qquad t = t_1, ..., t_n,$$

and we assume that

$$\text{Var}(x_1) = \frac{1-\phi^2}{2\lambda}\sigma^2. \tag{9}$$

for which the covariance between two variables x_i and x_j is given by $\text{Cov}(x_i, x_j)$.

Since the covariance matrix is almost always dense always dense for $\phi \in (0,1)$ it is computationally beneficial to instead work with the inverse-covariance matrix, also known as the precision matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$. For consistency we hereafter use precision

$$\kappa = \frac{2\lambda}{(1-\phi^2)\sigma^2} \tag{10}$$

instead of the variance as the unknown parameter of interest, and denote $\kappa(t_k) = 2\kappa\lambda(t_k)$ for $t_k = t_2, \ldots, t_n$.

It can be shown that for a time-dependent AR(1) process the precision matrix is sparse and equal to

$$\boldsymbol{Q} = \frac{1}{\sigma^2} \begin{pmatrix} \left(\kappa + \kappa(t_2)\phi(t_2)^2\right) & -\kappa(t_2)\phi(t_2) & & & \\ -\kappa(t_2)\phi(t_2) & \left(\kappa(t_2) + \kappa(t_3)\phi(t_3)^2\right) & -\kappa(t_3)\phi(t_3) & & \\ & \ddots & \ddots & \ddots & \\ & & -\kappa(t_{n-1})\phi(t_{n-1}) & \left(\kappa(t_{n-1}) + \kappa(t_n)\phi(t_n)^2\right) & -\kappa(t_n)\phi(t_n) \\ & & & -\kappa(t_n)\phi(t_n) & \kappa(t_n) \end{pmatrix}. \tag{11}$$

To allow for non-constant time steps $\Delta t_k = t_k - t_{k-1}$ we define

$$\phi(t_k) = e^{-\lambda(t_k)\Delta t_k / c}, \tag{12}$$

where

$$\lambda(t_k) = -\log(a + bt_k), \tag{13}$$

$c = \sum_{k=2}^{n} \Delta t_k/(n-1)$ and t_k has been rescaled such that $t_k \in (0,1)$. This modification guarantees that $\phi(t) \longrightarrow 1$ as $\Delta t_k \longrightarrow 0$ and $\phi(t) \longrightarrow 0$ as $\Delta t_k \longrightarrow \infty$. It also ensures that $\phi(t_1) = a$ and $\phi(t_n) = a + b$ which makes the interpretability of the parameters easier.

Gaussian processes with sparse precision matrices are known as Gaussian Markov random fields, and there is a wealth of efficient algorithms for fast Bayesian inference, see e.g. Rue and Held (2005) for a comprehensive discussion on this topic. These computationally efficient properties are not shared by the fractional Gaussian noise for which both the covariance matrix and the precision matrix are dense. This means that essential matrix operations such as computing the Cholesky decomposition will have a computational cost of $\mathcal{O}(n^3)$ floating point operations (flops), as opposed to $\mathcal{O}(n)$ flops for the AR(1) process. Inference might still be possible to achieve in a reasonable amount of time if the size of the data set remains sufficiently small. For larger data sets, however, both time and memory consumption may become an issue.

In fitting the model to data it is beneficial that the model parameters are defined on an unconstrained parameter space. We therefore introduce a suitable parameterization for the model parameters. For the precision κ we take the logarithm, $\theta_\kappa = \log\kappa$, and for a and b using we use variations of the logistic transformation. Our reasoning is as follows. Assuming the lag-one autocorrelation parameter is defined on the interval $(0,1)$, and since $t \in [0,1]$, then the slope must be constrained by

$$|b| < 1, \tag{14}$$

An unconstrained parameterization for b thus reads

$$\theta_b = \log\left(\frac{1+b}{1-b}\right) \iff b = -1 + \frac{2}{1+\exp(-\theta_b)}, \qquad \theta_b \in (-\infty, \infty). \tag{15}$$

The parameter space for a depend on the current state of b

$$0 < a + bt < 1 \iff -bt < a < 1 - bt. \tag{16}$$

Let

$$a_{\text{lower}} = -\min(b, 0) \quad \text{and} \quad a_{\text{upper}} = 1 - \max(b, 0), \tag{17}$$

then an unconstrained parameterization for a is given by

$$\theta_a = \log\left(\frac{a - a_{\text{lower}}}{a_{\text{upper}} - a}\right) \iff a = a_{\text{lower}} + \frac{a_{\text{upper}} - a_{\text{lower}}}{1+\exp(-\theta_a)}, \qquad \theta_a \in (-\infty, \infty). \tag{18}$$

3.1 Latent Gaussian model formulationBayesian inference

In this paper we adopt a Bayesian framework to estimate the model parameters. This means that parameters

Bayesian analysis presents a powerful framework for estimating model parameters that provides uncertainty quantification and allows us to incorporate prior knowledge about the parameters. These benefits are both very valuable in making informed decisions regarding climate action. In the Bayesian paradigm the model parameters $\boldsymbol{\theta} = (\kappa, b, a)$ are treated as stochastic variables for which prior knowledge , expressed using prior distributions, is incorporated and updated by the likelihood is expressed using a predefined distribution $\pi(\boldsymbol{\theta})$. These are updated by new information expressed by the likelihood function of the observations using Bayes' theorem. Bayesian inference can be obtained by expressing our model as a hierarchical Bayesian model, wherein the observed state variables are modeled in terms of a random predictor $\boldsymbol{x} = (x_1, \ldots, x_n)$, which is here a Gaussian distribution

$$= \beta_0 + \sum_{i=1}^{n_\beta} \beta_i i + \varepsilon(\boldsymbol{\theta}) \pi(\boldsymbol{x} \mid \boldsymbol{\theta}) = \boldsymbol{\mu}(\boldsymbol{\beta}) + \boldsymbol{\varepsilon}(\boldsymbol{\theta}) \mathcal{N}\left(\mathbf{0}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1}\right), \tag{19}$$

Here, β_0 represent an intercept, β_i are fixed effects corresponding to covariates z_i and ε are random effects representing some time-dependent noise that depend on some parameters with precision matrix $\boldsymbol{Q}(\boldsymbol{\theta})$ as given by (11). The updated belief is expressed by the posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{x})$, which is obtained using Bayes' rule

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{\pi(\boldsymbol{x} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{x})}, \tag{20}$$

where the model evidence $\pi(\boldsymbol{x})$ is a normalizing constant with respect to $\boldsymbol{\theta}$. Notably, in the Bayesian framework fixed effects are treated as stochastic variables and must be assigned prior distributions. If the data is already detrended then $\boldsymbol{\mu} = \mathbf{0}$. The covariance structure of the different components in the model are expressed by a latent field of random variables containing the predictor and all stochastic terms therein , i.e. $\boldsymbol{x} = (\boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\varepsilon})$. Assigning a Gaussian prior on $\boldsymbol{x}$ the model becomes a latent Gaussian model, a subset of Bayesian hierarchical modelsfor which there exists additional computational frameworks. The latent Gaussian model is specified in three stages as follows.

The first stage is to specify the likelihood of the model. We assume the likelihood to be conditionally independent given the latent field $\boldsymbol{x}$, and expressed by a Gaussian distribution with some small negligible fixed variance $\sigma_y^2 \approx 0$ and mean equal to the predictor

$$\boldsymbol{y} \mid \boldsymbol{x} \sim \prod_{i=1}^{n} \mathcal{N}(\eta_i, \sigma_y^2).$$

The second stage in specifying a latent Gaussian model is to specify a Gaussian prior distribution for the latent field $\boldsymbol{x}$, with mean vector $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{\theta})$ and precision matrix $\boldsymbol{Q}$. This may depend on some unknown hyperparameters $\boldsymbol{\theta}$ and expresses the covariance structure of the latent variables. $\boldsymbol{\beta}$ are assigned vague Gaussian priors and the noise term Specifically, for the linear predictor we assume

$$\boldsymbol{x} \mid \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1}),$$

Prior selection is an essential part of any Bayesian analysis, and presents a great strength of the Bayesian framework by allowing prior knowledge to be incorporated into models. Since we do not incorporate prior knowledge in this paper, and we wish to maintain objectivity, we will adopt vague prior distributions. These are distributions with large variances that express minimal information about the parameters and allows inference to be primarily driven by the data, as opposed to more informative priors which can guide the posterior to reflect prior knowledge or assumptions. Since the parameter a depend on the value of another parameter b we assign a conditional prior, $\pi(a \mid b)$, such that the latent variables corresponding to a potential $\boldsymbol{\beta}$ component represent vague Gaussian priors and those corresponding to $\boldsymbol{\varepsilon}$ represent the chosen model. The precision matrix is given by Eq. (??).

The final stage concerns the prior distributions of the model parameters, which we assign independently

$$\boldsymbol{\theta} \sim \pi(\kappa)\pi(\theta_a)\pi(\theta_b).$$

joint prior is expressed by

$$\pi(\kappa, b, a) = \pi(\kappa)\pi(b)\pi(a \mid b). \tag{21}$$

For the analysis

Specifically, all analyses performed in this study we have assigned a penalised complexity prior (Simpson et al., 2017) for the scaling parameter $\kappa = 1/\sigma^2$ and Gaussian priorsfor the parameterized memory parameters θ_a and θ_b.

3.2 Inference

In the Bayesian paradigm inference is expressed by the posterior distribution which provides a complete description of the probabilistic nature of the model parameters and latent variables. The joint posterior distribution can be found relatively easily by

$$\pi(\boldsymbol{x},\boldsymbol{\theta} \mid \boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x} \mid \boldsymbol{\theta}) \prod_{i=1}^{n} \pi(y_i \mid \boldsymbol{x}).$$

We want to estimate paper, assume the same set of priors, unless otherwise specified. κ is assigned a gamma distribution with shape 1 and rate 0.1. b is assigned a uniform prior on $(-1,1)$ and $a \mid b$ is assigned a uniform prior on $(a_{\text{lower}}, a_{\text{upper}})$.

The main goal of this paper is to detect whether or not an early warning signal can be observed in the data. We are therefore primarily interested in the marginal posterior distribution for all hyperparameters and latent variables. These are computed by evaluating the integrals

$$\pi(x_i \mid \boldsymbol{y}) = \int \pi(x_i \mid \boldsymbol{\theta}, \boldsymbol{y}) \pi(\boldsymbol{\theta} \mid \boldsymbol{y}) d\boldsymbol{\theta}$$

$$\pi(\theta_j \mid \boldsymbol{y}) = \int \pi(\boldsymbol{\theta} \mid \boldsymbol{y}) d\boldsymbol{\theta}_{-j}.$$

of the slope parameter b,

$$\pi(b \mid \boldsymbol{x}) = \int \pi(\boldsymbol{\theta} \mid \boldsymbol{x}) \mathrm{d}\kappa \mathrm{d}a. \tag{22}$$

These integrals are often impossible to evaluate analytically and are typically computed numerically Typically, marginal posterior distributions can be evaluated using Markov chain Monte Carlo approaches (Robert et al., 1999). However, these can sometimes be very time consuming for hierarchical models. For latent Gaussian models , but these are very often time-consuming and could potentially be sensitive to convergence issues. However, since our model is Gaussian with a sparse precision matrix there exists a we instead use the computationally superior alternative in using of integrated nested Laplace approximations approximation (INLA) (Rue et al., 2009, 2017). Instead of using simulations, INLA use various numerical optimization techniques to compute an accurate approximation of the posterior marginal distributions. Most importantly is the Laplace approximation (Tierney and Kadane, 1986), which is used to approximate the joint posterior distribution

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y}) \approx \left. \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi_G(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})} \right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})},$$

where $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the mode of the latent field $\boldsymbol{x}(\boldsymbol{\theta})$ and $\pi_G(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation of

$$\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y}) \propto \pi(\boldsymbol{x} \mid \boldsymbol{\theta}) \pi(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}).$$

The methodology is available as the open source R package `R-INLA`, which can be downloaded (Rue et al., 2009), which is available as an `R` package at www.r-inla.org(last access: **day month year**).

As there are currently no model components already implemented for `R-INLA` that meet our specifications we are required to implement the model components ourselves using the custom modeling framework of R-INLA called `rgeneric`. This adds more work and complexity in implementing our model, and adds an additional barrier to further adoptation of our methodology . To increase accessibility we have implemented the code and made it available as a user-friendly R-package To make the methodology more accessible we have released the code associated with this model as an R package titled INLA.ews, available which can be downloaded at www.github.com/eirikmn/INLA.ews (last access: **day month year**). Inference can then be produced by executing `inla.ews(y, formula=formula)`, where `y` is a numeric vector containing the data and `formula` describes the trends included in the model. July 30, 2024). A demonstration of the `INLA.ews` package applied to simulated data this package can be found in Appendix A, and a detailed description of its features can be found in its accompanying documentation.

3.2 Non-constant time steps

To allow for non-constant time steps $\Delta t_k = t_k - t_{k-1}$ we assume

$$\phi(t_k) = e^{-\lambda(t_k)\Delta t_k / c},$$

where

$$\lambda(t_k) = -\log(a + bt_k),$$

$c = \sum_{k=2}^{n} \Delta t_k / (n-1)$ and t_k has been normalized such that $t_k \in (0,1)$. This modification guarantees that $\phi(t) \longrightarrow 1$ as $\Delta t_k \longrightarrow 0$ and $\phi(t) \longrightarrow 0$ as $\Delta t_k \longrightarrow \infty$. It also ensures that $\phi(t_1) = a$ and $\phi(t_n) = a + b$ which makes the interpretability of the parameters easier.

If we denote $\sigma(t_k)^2 = \sigma^2/(2\lambda(t_k))$, and assume

$$x_1 \sim \mathcal{N}\left(0, \sigma(t_1)^2\right),$$

then the precision matrix for non-constant time steps yields

$$\boldsymbol{Q} = \begin{pmatrix} \left(\frac{1}{\sigma(t_1)^2} + \frac{\phi(t_2)^2}{\sigma(t_2)^2}\right) & -\frac{\phi(t_2)}{\sigma(t_2)^2} & & \\ -\frac{\phi(t_2)}{\sigma(t_2)^2} & \left(\frac{1}{\sigma(t_2)^2} + \frac{\phi(t_3)^2}{\sigma_3^2}\right) & -\frac{\phi(t_3)}{\sigma(t_3)^2} & \\ & \ddots & \ddots & \ddots \\ & -\frac{\phi(t_{n-1})}{\sigma(t_{n-1})^2} & \left(\frac{1}{\sigma(t_{n-1})^2} + \frac{\phi(t_n)^2}{\sigma(t_n)^2}\right) & -\frac{\phi(t_n)}{\sigma(t_n)^2} \\ & & -\frac{\phi(t_n)}{\sigma(t_n)^2} & \frac{1}{\sigma(t_n)^2} \end{pmatrix}.$$

Non-constant time steps can be specified in the `inla.ews` function by using the `timesteps` input argument.

3.2 Incorporating forcing

Climate components may also be affected by forcing, which can be measured alongside the climate variable of interest. How the observed component responds to such forcing will be influenced by time-dependence. In this subsection we adopt a similar strategy to ~~**myrvoll-nilsen2020**~~ Myrvoll-Nilsen et al. (2020) with changes to the Green's function to allow for time-dependence and non-constant time steps.

Let $F(t)$ denote the known forcing component such that

$$dx(t) = -\lambda x(t) + F(t)dt + dB(t). \tag{23}$$

~~The~~ As shown in C.W.Gardiner (1985), the model can then be expressed as the sum of two components

$$x(t) = \mu\nu(t) + \varepsilon\xi(t), \tag{24}$$

where ~~$\varepsilon(t)$~~ $\xi(t)$ is a time-dependent OU process and the forcing response is given by

$$\mu\nu(t) = \sigma_f(t)\frac{1}{2\lambda(t)\kappa_f(t)}\int_0^t e^{-\lambda(t)(t-s)}\left(F(s)+F_0\right)ds. \tag{25}$$

~~$\sigma_f^2(t) = \sigma_f^2/(2\lambda(t))$ is~~ Here, $\lambda(t) = -\log(\phi(t))$ is the restoring rate, $\kappa_f(t)$ is an unknown scaling parameter and F_0 is an unknown shift parameter.

~~Forcing can be incorporated into the model~~ These parameters can be estimated using the same Bayesian framework as before, which can be computed with `INLA.ews` by specifying the `forcing` argument in the `inla.ews` function.

4 Results

4.1 Accuracy test on simulated data

To test the accuracy and robustness of the time-dependent AR(1) model we fit the model to a number of simulations. Specifically, we perform accuracy tests using a grid of $b \in [-0.8, 0.8]$ with increments of 0.1, and choose the parameter a corresponding to $\theta_a = 0$. For each b we draw $n_r = 1000$ time series of length $n = 500$ and $n = 1000$ from the time-dependent AR(1) model. The model is fitted using `R-INLA` ~~with the same specifications as used in the `INLA.ews` package~~using priors specified in Section 3.1. To quantify the accuracy of the model we compare the posterior marginal mean of the slope $\hat{b} = \mathrm{E}(\pi(b \mid \boldsymbol{y}))$ to the true values b. We also compute the posterior probability of the slope being positive $P(b > 0)$. Ideally, we want $\hat{b}$ to be as close to b as possible, and $P(b > 0) > 0.5$ if $b > 0$ and, conversely, $P(b > 0) < 0.5$ if $b < 0$. We also count the number of simulations where EWS is detected, using threshold $P(b > 0) \geq 0.95$. Since σ only scales the amplitude of the data without affecting the correlation structure we expect similar estimations for a and b regardless of the value of σ. This was confirmed by testing the model on simulations using both $\sigma = 1$ and $\sigma = 10$.

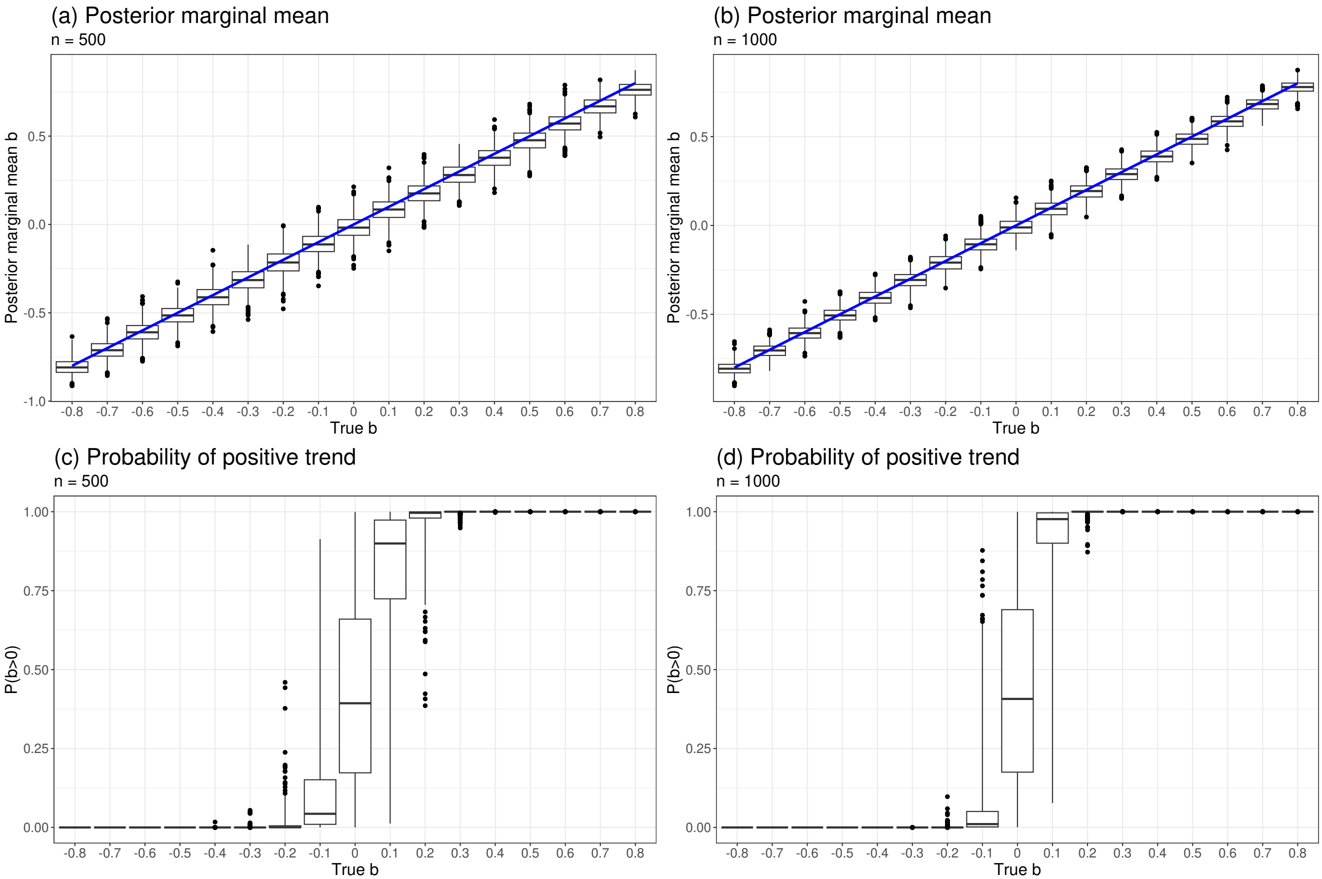

Figure 2. Box plots representing the results of the accuracy test for $n_r = 1000$ simulated time series of length $n = 500$ for each $b \in [-0.8, 0.8]$. The boxes cover the interquartile range (IQR) between the 25 and 75% quantiles, and the whiskers represent an adjustment to the more common boundaries of 1.5 times the IQR, to better describe skewed distributions. Points that fall outside the whiskers are classified as outliers. Panels (a) and (b) show box plots of the posterior marginal mean estimated by INLA for simulations of lengths $n = 500$ and $n = 1000$, respectively. The blue line shows the true b used in the simulation. ~~Panel~~ Panels (c) and (d) show box plots of the estimated posterior probability of the slope being positive ~~given the~~ against different true ~~value~~ values used for simulations of length $n = 500$ and $n = 1000$, respectively.

The results of the analysis ~~is presented in~~ are presented in Table 1 and displayed graphically as box plots in Fig. 2. Since the posterior distribution of b is skewed, especially when its absolute value approaches 1, ordinary box plots would classify a larger number of points as outliers. We use instead an adjusted box plot proposed by (Hubert and Vandervieren, 2008) which is better suited for skewed distributions. We obtain decent accuracy of the posterior marginal means $\hat{b}$, with a small~~underestimation when $b \longrightarrow -1$ and a small overestimation when $b \longrightarrow 1$. The posterior probabilities suggests that when $|b| \geq 0.2$ there is both a low chance of false negatives (high sensitivity) and false positives (high specificity).~~, but consistent, underestimation which decreases as n increases. In panels (c) and (d) of Fig. 2 we observe some variation in $P(b > 0)$ for small values of

$|b|$, and less so for $|b| \geq 0.2$. ~~For smaller absolute values however, especially those generated under $b = 0$, more variation in posterior probabilities were observed.~~ This behaviour also improves when n increases from ~~500 to 1000. For~~ 500 to 1000. By counting the number of simulations with detected EWS for $b \leq 0$ using threshold $P(b > 0) \geq 0.95$ we find 54 false positives for $n = 500$~~we find that out of $n_r = 1000$ simulations there were zero false positives~~, and 56 for ~~$b \leq -0.1$,~~ $n = 1000$. Out of 8000 simulations for $b \geq 0.1$ we find 669 false negatives for $n = 500$ and ~~a single false negative at $b \geq 0.1$. For $n = 1000$, no false positives or negativeswere found.~~ 321 for $n = 1000$, as reported in Table 1.

In order to assess how sensitive the model is to the choice of prior distribution we repeat the same simulation procedure with $n = 500$ using different Gaussian priors on the parameters in their internal scaling

$$(\log(\kappa), \theta_a, \theta_b) \sim \mathcal{N}(\mathbf{0}, \sigma_\theta \boldsymbol{I}), \tag{26}$$

where $\sigma_\theta \in \{0.1, 1, 10\}$ and $\boldsymbol{I}$ is the identity matrix. The posterior marginal means and posterior probabilities $P(b > 0)$ for the different priors are compared to the default prior in Fig. 3. The results show that using the most informative prior, corresponding to $\sigma_\theta = 0.1$, will pull b too much towards the central value of $b = 0$, resulting in worse estimates for b. This pull is also reflected in the posterior probability estimates, $P(b > 0)$, where this prior performs less well compared to the others, although the effect is less strong here. Counting the number of misclassifications we find, for σ_θ equal to 0.1, 1 and 10, 4, 48 and 67 false positives our of 9000 simulations and 1255, 657 and 631 false negatives out of 8000 simulations, respectively. This is overall quite comparable to using the default priors, which found 54 false positives and 669 false negatives. Overall, we find the model to be quite robust to the choice of prior distributions as most of the priors perform very similar in terms of the posterior marginal mean and posterior probabilities. However, we also find that using too informative priors could cause the model to be overly cautious, making it less able to detect EWS.

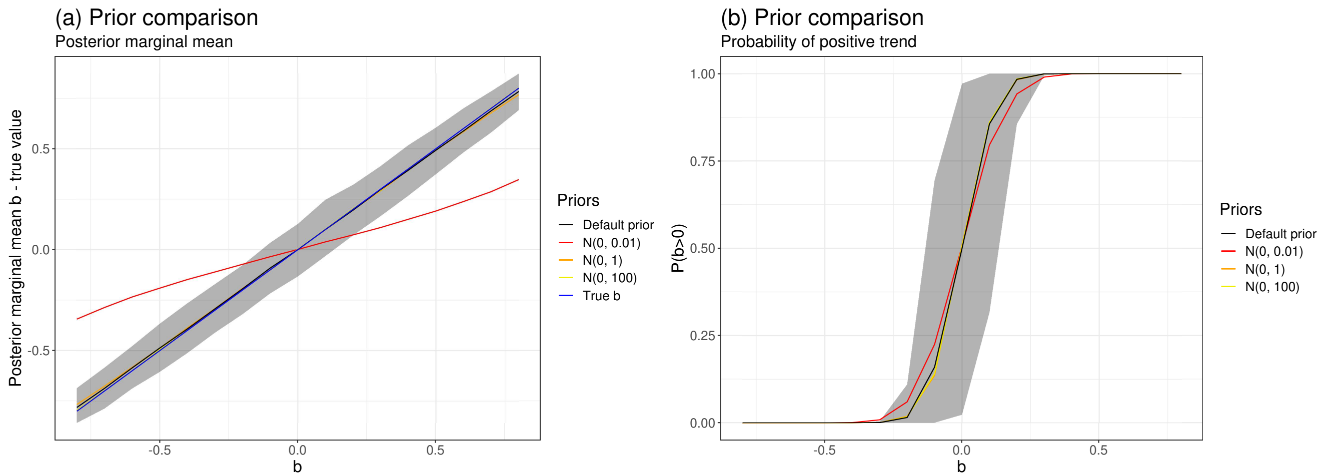

Figure 3. Prior sensitivity analysis on simulated data using the default priors (black) and Gaussian priors $\mathcal{N}(0, 0.1^2)$ (red), $\mathcal{N}(0, 1)$ (orange) and $\mathcal{N}(0, 10^2)$ (yellow). Panel (a) shows the ensemble average of the posterior marginal mean estimates for b compared to the true value (blue), and panel (b) shows the ensemble average for the posterior probabilities $P(b > 0)$. The shaded grey area in both panels represent the ensemble spread of the estimates from the default prior using the 2.5 and 97.5% quantiles.

	$n = 500$	$n = 1000$	$n = 500$	$n = 1000$	$n = 500$	$n = 1000$
True b	~~$\hat{b}$~~ $\langle\hat{b}\rangle$	~~$\hat{b}$~~ $\langle\hat{b}\rangle$	~~$P(b > 0)$~~ $\langle P(b > 0)\rangle$	~~$P(b > 0)$~~ $\langle P(b > 0)\rangle$	# $P(b > 0) > 0.95$	# $P(b > 0) > 0.95$
-0.8	~~-0.766~~ -0.781	~~-0.78~~ -0.792	0	0	0	0
-0.7	~~-0.67~~ -0.687	-0.694	0	0	0	0
-0.6	~~-0.578~~ -0.586	~~-0.591~~ -0.595	0	0	0	0
-0.5	~~-0.483~~ -0.487	~~-0.493~~ -0.496	0	0	0	0
-0.4	~~-0.385~~ -0.393	-0.398	0	0	0	0
-0.3	~~-0.291~~ -0.293	~~-0.295~~ -0.299	0.001	0	0	0
-0.2	~~-0.192~~ -0.194	~~-0.197~~ -0.195	0.015	0.001	0	0
-0.1	~~-0.095~~ -0.091	~~-0.1~~ -0.096	~~0.151~~ 0.16	~~0.065~~ 0.074	1	1
0	~~-0.004~~ 0	~~0.002~~ -0.001	~~0.478~~ 0.5	~~0.508~~ 0.498	53	55
0.1	~~0.097~~ 0.1	~~0.854~~ 0.096	~~0.937~~ 0.856	0.926	425	685
0.2	~~0.199~~ 0.196	~~0.197~~ 0.2	~~0.985~~ 0.984	~~0.999~~ 0.998	907	994
0.3	~~0.292~~ 0.298	~~0.294~~ 0.299	0.999	1	999	1000
0.4	~~0.392~~ 0.393	~~0.394~~ 0.396	1	1	1000	1000
0.5	~~0.485~~ 0.492	0.494	1	1	1000	1000
0.6	~~0.583~~ 0.588	~~0.592~~ 0.594	1	1	1000	1000
0.7	~~0.675~~ 0.688	0.695	1	1	1000	1000
0.8	~~0.768~~ 0.783	~~0.781~~ 0.793	1	1	1000	1000

Table 1. Results from accuracy tests on $n_r = 1000$ simulated time-dependent AR(1) series of length n for each b ranging from -0.8 to 0.8. The table includes the ensemble ~~average~~ averages of the posterior marginal means $\hat{b}$ and posterior probabilities of positive slope $P(b > 0)$ for each value of b, and for time series' lengths of $n = 500$ and $n = 1000$. We also show the number of detected early warning signals using threshold $P(b > 0) \geq 0.95$.

4.2 DO-events

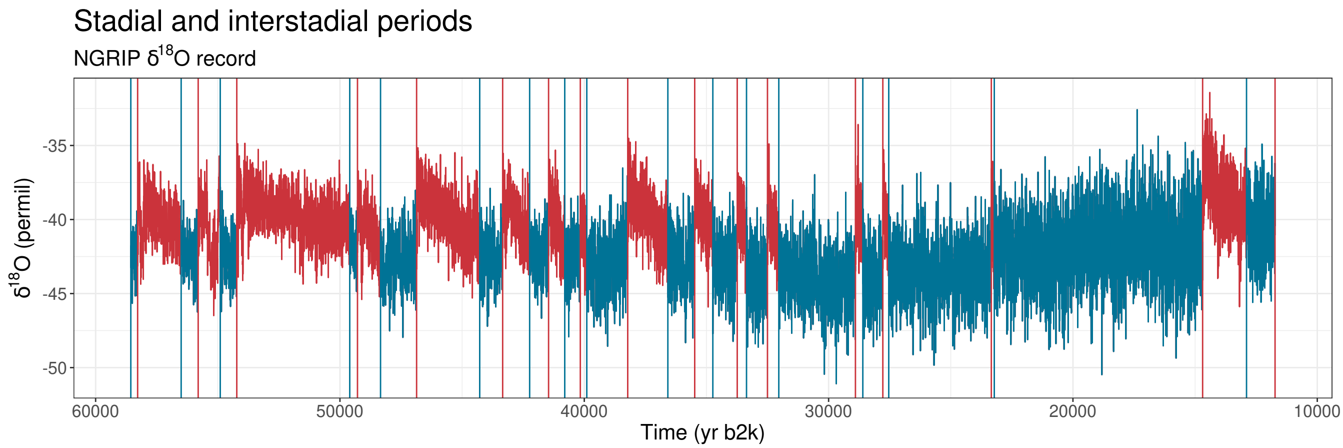

Figure 4. NGRIP δ^{18}O proxy record. The time-series used in our study are the parts of the curves drawn in ~~blue~~ turquoise which are the cold stadial periods ~~preceeding~~ preceding the onsets of interstadial periods drawn in red. The red and ~~blue~~ turquoise vertical bars represent ~~respectively~~ the start and ~~the~~ end points of ~~inter-stadial (warm)~~ the interstadial periods, respectively.

We apply our time-dependent AR(1) model on the high resolution NGRIP δ^{18}O record, which is partitioned into stadial and interstadial periods as shown in Fig. 4. This version of the NGRIP record is sampled regularly every 5cm steps in depth, but is non-constant in time. Having modified our model to allow for irregular time points we are able to use the raw NGRIP record without having to perform interpolation or other types of pre-processing, such as that of Boers (2018). This grants us a larger dataset for each event which could significantly improve parameter estimation. Having implemented the model using INLA we are able to take advantage of this extra resolution while keeping computational time low.

Some of these datasets appear non-stationary and thus require trend estimation. Since there is no obvious choice of forcing we consider different alternatives for trend components which are then compared. The `R-INLA` framework allows us to very easily incorporate these trends into our model and estimates all model components simultaneously. First, we fit our model to the data without any additional trend, then we assume a linear trend, followed by a 2nd order polynomial trend. Finally, we model the trend using a continuous 2nd order random walk (RW2) spline. More details on the comparison between the different trends are included in appendix C, which also includes a plot of how well each trend ~~fit~~ fits the data.

~~Having looked~~ Looking at the fits for each event we observe that most events can be fitted easily with linear or even ~~constant~~ no trend, but a few events require non-linearity. We choose the 2nd order polynomial trend as this gives a nice trade-off between flexibility and simplicity and appears to provide a decent fit for all events. The $\phi(t) = a + bt$ evolutions for all events using 2nd order polynomial detrending ~~is~~ are included in Fig. 5~~.~~, and the posterior marginal distribution of the trend, $\pi(b \mid \boldsymbol{x})$, is included in Fig. 6.

The models are fitted to the stadial ~~period~~ periods preceding each of the 17 DO events and the posterior probability of $\phi(t) = a + bt$ being increasing, $P(b > 0)$, is compared for all events and trend assumptions. ~~These~~ The results are included in table Table 2. Using the conventional threshold of $P(b > 0) \geq 0.95$ we are able to detect early warning signals in 4 events using no detrending , and 5 events 5 events both without detrending and while using linear or 2nd order polynomial trends, and 6 events using the continuous RW2 model as a trend. Averaging over the estimated $P(b > 0)$ for all events we are not able to conclude that early warning signals has have been found over the ensemble of events for any detrending model.

Having found EWS in multiple stadial periods preceding DO events therefore indicates that DO events are not solely noise-induced unlike the hypothesis can exhibit evidence of ongoing destabilization unlike the conclusion formulated in Ditlevsen and Johnsen (2010). These differences in results can be explained by both the use of a higher-resolution dataset and a methodology not involving time windows. However,

However, given the absence of EWS in the ensemble of events does not support the hypothesis that all DO events are bifurcation-induced and hence exhibiting signs of ongoing destabilization and hence one cannot exclude the possibility for some events to be purely noise-induced and not approaching a bifurcation point. Our results do, however, suggest that some specific transitions may be bifurcation induced, have undergone destabilization which is in line with the results of Rypdal (2016) and Boers (2018), in which wherein significant EWS have also been found only for some specific events. These studies use different versions of the NGRIP record from our study and their methodologies differ from oursas they use a scale-invariant fGn model to describe the noise, as opposed to an AR(1) process. .

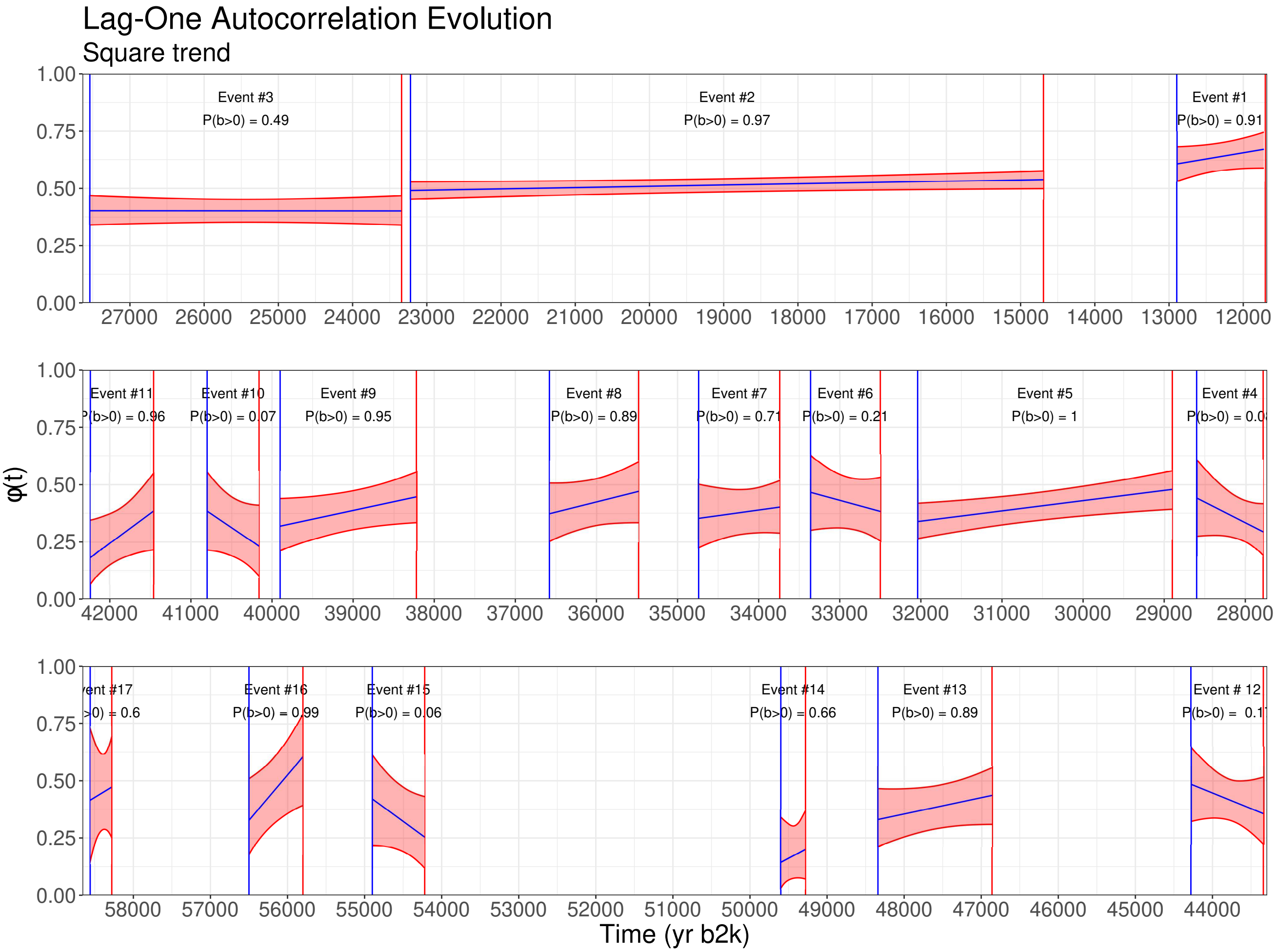

Figure 5. The evolution of the lag-one autocorrelation parameter $a + bt$ for each of the 17 transitions analyzed in this paper. The blue lines represents the posterior marginal means of each Greenland stadial phase, and the red shaded areas represent the 95% credible intervals (corresponding to region between the 2.5 and 97.5% quantiles of the posterior distribution). The δ^{18}O proxy measurements have been detrended using a second order polynomial. The probability of an increasing slope, $P(b > 0)$, given the posterior distribution, is also included.

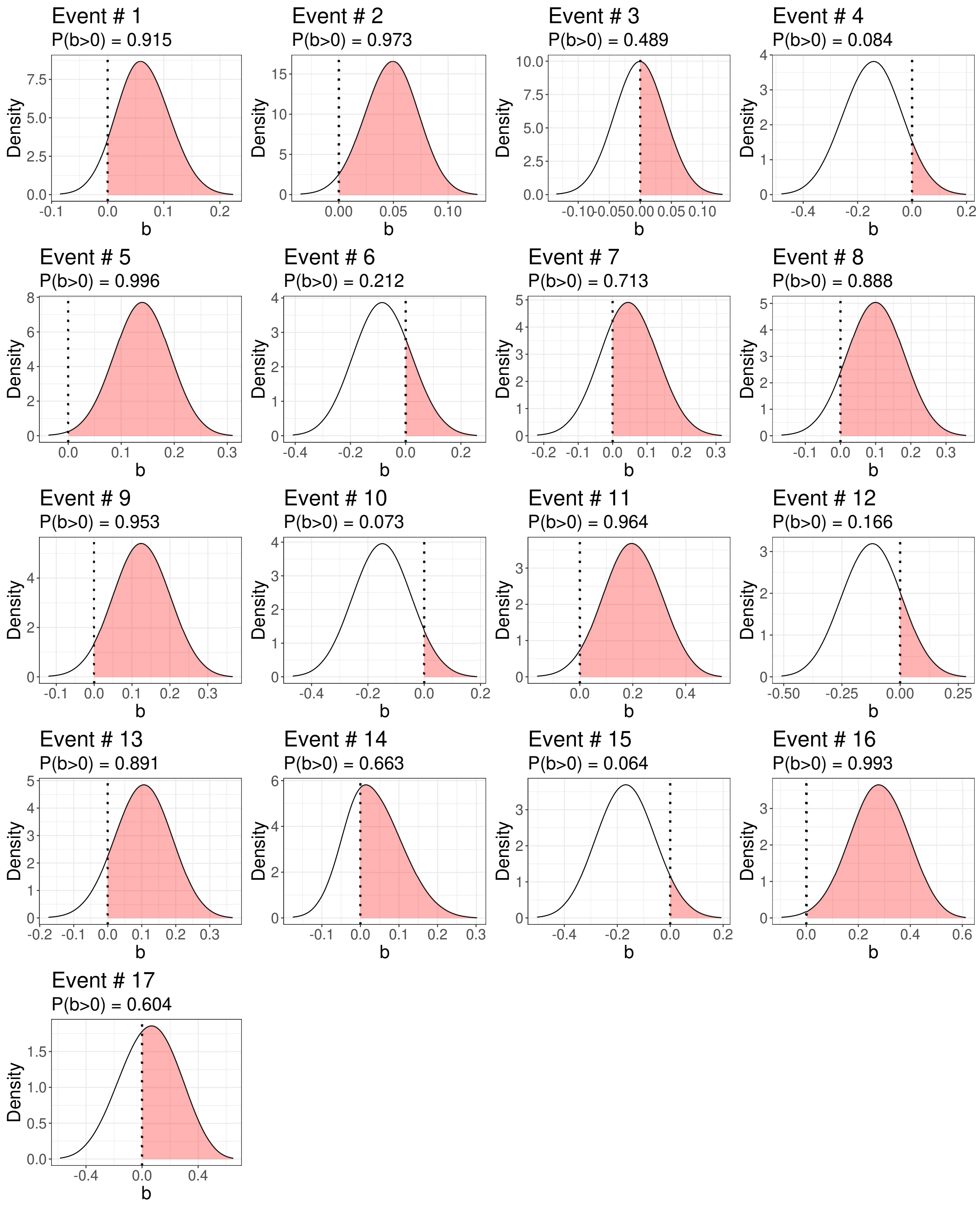

Figure 6. The marginal posterior distribution of the trend parameter b. The red dotted line represent $b = 0$ and the red shaded area illustrate the density $P(b > 0)$. If the shaded area is larger than 0.95 e conclude that EWS is detected.

Event	No trend	Linear	Square	RW2	Rypdal	Boers
1	~~0.8824~~ 0.8765	~~0.8558~~ 0.854	~~0.9022~~ 0.9146	~~**0.9512**~~ 0.9552	$p = 0.02$	–
2	~~**0.9660**~~**0.9658**	~~**0.9886**~~**0.9902**	~~**0.973**~~**0.9728**	~~**0.9655**~~**0.9672**	$p = 0.008$	$p < 0.05$
3	~~0.4949~~ 0.5252	~~0.4983~~ 0.5174	~~0.5005~~ 0.4893	~~0.6137~~ 0.647	–	–
4	~~0.0714~~ 0.0684	~~0.0878~~ 0.078	~~0.0821~~ 0.084	~~0.0928~~ 0.1007	–	$p < 0.05$
5	**0.9958**	~~**0.9956**~~**0.9959**	~~**0.9952**~~**0.9959**	~~**0.9918**~~**0.9893**	$p = 0.13$	–
6	~~0.2924~~ 0.28	~~0.3052~~ 0.3174	~~0.2159~~ 0.2123	~~0.2249~~ 0.2068	–	~~p<0.05~~ $p < 0.05$
7	~~0.7569~~ 0.7703	~~0.7182~~ 0.7517	~~0.6695~~ 0.7132	~~0.9819~~ 0.79	–	–
8	~~0.9117~~ 0.8976	~~0.9141~~ 0.9189	~~0.8747~~ 0.8878	~~0.8394~~ 0.8524	–	–
9	~~**0.9862**~~**0.9857**	~~**0.9669**~~**0.9628**	~~**0.9563**~~**0.953**	~~**0.9557**~~**0.9818**	$p = 0.16$	–
10	~~0.0415~~ 0.019	~~0.1549~~ 0.1413	~~0.0942~~ 0.0732	~~0.1311~~ 0.1001	–	–
11	~~0.9325~~ **0.967**	~~**0.9516**~~**0.967**	~~**0.9614**~~**0.9643**	~~0.9488~~ **0.9635**	–	$p < 0.05$
12	~~0.1393~~ 0.1483	~~0.1441~~ 0.1319	~~0.1268~~ 0.1662	~~0.0287~~ 0	–	–
13	~~0.8898~~ 0.8872	~~0.8864~~ 0.8953	~~0.8923~~ 0.8912	~~0.9059~~ 0.9043	$p = 0.39$	$p < 0.05$
14	~~0.7261~~ 0.7781	~~0.866~~ 0.914	~~0.684~~ 0.6629	~~0.7511~~ 0.7808	–	$p < 0.05$
15	~~0.0304~~ 0.0227	~~0.0599~~ 0.0546	~~0.0695~~ 0.0637	~~0.0754~~ 0.0446	–	$p < 0.05$
16	~~**0.9903**~~**0.9885**	~~**0.9918**~~**0.9915**	**0.9935**	~~**0.9953**~~**0.9939**	–	–
17	~~0.5889~~ 0.6748	~~0.5581~~ 0.6366	~~0.5766~~ 0.6043	~~0.5285~~ 0.5855	–	–
Ensemble	~~0.6292~~ 0.6383	~~0.6437~~ 0.654	~~0.6216~~ 0.626	~~0.646~~ 0.639	–	–

Table 2. Table comparing the probability of positive slope $P(b > 0)$ for each event given posterior distributions obtained using the time-dependent AR(1) model. We ran the model using different trends including no trend (except for the intercept), a linear effect, a second order polynomial and a 2nd order random walk spline. Our results are also compared with the p values obtained from Rypdal (2016) and Boers (2018).

5 Conclusions

~~This paper presents~~

In this paper we introduce a Bayesian framework to analyze early warning signals ~~, using an~~ in time series data. Specifically, we define a time-dependent AR(1) process where the lag-one correlation parameter is assumed to increase linearly over time. ~~Bayesian inference is obtained using a latent Gaussian model formulation and implemented using the R−INLA framework. In addition to computing the posterior marginal distribution for all variables and parameters in the model, implementation in the R−INLA framework grants a number of benefits. First, it provides a great reduction in computational cost, both in terms of speed and memory . Second, the framework is very versatile and other model components such as trends can be easily added to the predictor. Third, R−INLA uses posterior prediction to impute missing data automatically. The model has been applied to simulated data and shows decent accuracy.~~ The slope of this parameter indicates whether or not memory is increasing and thus if early warning signals are detected. Using a Bayesian approach we automatically obtain uncertainty quantification expressed by posterior distributions and allow for prior knowledge to be utilized.

To detect early warning signals of DO events we have applied our model to interstadial periods of the raw 5cm NGRIP water isotope record. This record is sampled evenly in depth, but not in time, requiring us to make some necessary modifications to allow for ~~non-equidistant~~ non-constant time steps.

Using the time-dependent AR(1) model we were unable to detect statistically significant EWS for the ensemble of 17 DO events, and only detected EWS individually for six ~~5~~ events using a second-order polynomial detrending. Unlike Ditlevsen and Johnsen (2010), we find evidence of EWS in some events, corroborating Rypdal (2016) and Boers (2018). We were, however, unable to conclude that DO events are individually or generally bifurcation-induced. To better compare with ~~Rypdal (2016) and Boers (2018)~~other studies, we would have liked to employ a long-range dependent process such as the fGn. However, this task is more difficult than for the AR(1) process, as necessary modifications have to be made to the model. Moreover, this would also require working with non-sparse precision matrices which are far more computationally demanding. We did attempt to implement the time-dependent fGn model presented by Ryvkina (2015), but we were unable to ensure sufficient stability. This is, however, a very interesting topic for future work.

Currently, our model can only fit an AR(1) process where the lag-one correlation parameter is expressed as a linear function of time, which is not realistic. Although this is sufficient for detecting whether or not there ~~has been a statistically significant increase in EWS~~ are EWS expressed by a linear trend, our model is unable to perform predictions or give an indication of when the tipping point could be reached. More advanced functions for the evolution of the lag-one correlation parameter should be possible, but would have to be implemented. One possible extension would be to include a break point such that the memory is constant for all steps before this point, and starts increasing or decreasing afterwards. A simple implementation and discussion of this is included in Appendix D. Another extension would be to formulate a model where the memory parameter follows a polynomial $\phi(t) = a + bt^c$, where the exponent term $c > 0$ is an additional hyperparameter. This would perhaps help give an indication of the rate of which the correlation has increased. However, when adding more parameters one needs to be careful to avoid overfitting.

The ability to update prior beliefs in light of new evidence presents a great benefit of a Bayesian approach, and it presents an intuitive framework for iteratively updating the posterior distribution as new data becomes available by using the posterior distribution from previous analyses as the prior distribution in the the analysis. This is of great relevance for monitoring climatic systems suspected of approaching a tipping point.

To make the methodology more accessible we have released the code associated with this model as an `R` package titled `INLA.ews`. This package performs all analysis and includes functions to plot and print key results from the analysis very easily. Although this paper focuses on the detection of EWS in DO events observed in Greenland ice core records, our methodology is general and the `INLA.ews` package should be applicable to tipping points observed in other proxy records as well. We have also implemented the option of including forcing, for which the package will estimate the necessary parameters and compute the resulting forcing response. The package is demonstrated on simulated data in the appendix.

Code and data availability. The code and data sets used for this paper is available through the R-package, INLA.ews, which can be downloaded from: github.com/eirikmn/INLA.ews (last access July 30, 2024).

Appendix A: Demonstration of the INLA.ews package

We demonstrate the INLA.ews package on simulated forced data with non-equidistant time steps. The time steps t_k are obtained by adding Gaussian noise such that $\tilde{t}_k = k + \xi_k$ and normalized $t_k = (\tilde{t}_k - \tilde{t}_0)/(\tilde{t}_n - \tilde{t}_0)$ such that $t_0 = 0$ and $t_n = 1$. We assume a time dependent AR(1) process of length $n = 1000$ for the observations, sampled at times $\tilde{t}_1, \ldots, \tilde{t}_n$. The AR(1) process has scale parameter $\kappa = 0.04$ and time-dependent lag-one correlation $\phi(t) = a + bt_k$ given by $a = 0.3$ and $b = 0.2$. We also include a forcing $F(t)$, obtained by simulation from another AR(1) process with unit variance and lag-one correlation $\tilde{\phi} = 0.95$. The forcing response is approximated by

$$\nu(t_k) = \frac{\sigma_f}{\sqrt{2\lambda(t_k)}} \sum_{s=t_0}^{t_k} e^{-\lambda(t_k)(t_k - t_s)} \big(F_0 + F(s)\big), \tag{A1}$$

with parameters set to $\sigma_f = 0.1$ and $F_0 = 0$, and added to the simulated observations. We assign Gamma$(1, 0.01)$ priors for the precision parameters κ and κ_f, uniform priors on b and a and a Gaussian prior $\mathcal{N}(0, 10^2)$ on F_0. For INLA.ews, these priors must be transformed for the unconstrained parameterization $\boldsymbol{\theta} = (\log\kappa, \theta_b, \theta_a, \log\kappa_f, F_0)$, using the change-of-variable formula. The logarithm of the prior distributions are specified by creating a function

```
  my.log.prior <- function(theta) {
      lprior = dgamma(exp(theta[1]), shape=1, rate=0.1) + theta[1] + #kappa
          -theta[2] -2*log(1+exp(-theta[2])) + #b
          -theta[3] -2*log(1+exp(-theta[3])) + #a
          dgamma(exp(theta[4]), shape=1, rate=0.1) + theta[4] + #kappa_f
          dnorm(theta[5], sd=10, log=TRUE) #F0
  return(lprior)
 }
```

and passing it into inla.ews using the log.prior argument. The AR(1) model and forcing z sampled at time points time can be fitted to the data y with INLA using the inla.ews wrapper function:

```
results <- inla.ews(data=y, forcing=z, log.prior=my.log.prior, timesteps=time)
```

The inla.ews function computes all posterior marginal distributions, computes summary statistics, formats the results and returns all information as an inla.ews list object. Summary statistics and other important results can be extracted using the summary function:

```
> summary(results)
```

```

Call:
inla.ews(data = y, forcing = forcing, log.prior=my.log.prior, timesteps=time)

Time used:
    Running INLA Post processing            Total
        503.1737         151.8262         655.5909

Posterior marginal distributions for all parameters have been computed.

Summary statistics for using ar1 model (with forcing):
             mean     sd 0.025quant 0.5quant 0.975quant
a          0.3065 0.0546     0.1974   0.3072     0.4148
b          0.1929 0.0615     0.0524   0.2018     0.2878
sigma      7.0249 0.4522     6.3420   6.9522     8.0780
sigma_f    0.1000 0.0096     0.0862   0.0982     0.1229
F0        -0.0047 0.0223    -0.0514  -0.0034     0.0355

Memory evolution is sampled on an irregular grid.
Summary for first and last point in smoothed trajectory (a+b*time):
            mean     sd 0.025quant 0.5quant 0.975quant
phi0[1]  0.3065 0.0546     0.1974   0.3072     0.4148
phi0[n]  0.4980 0.0587     0.3667   0.5039     0.5945
Mean and 95% credible intervals for forced response have also been computed.

Probability of positive slope 'b' is 0.9954214
%DIF >
Marginal log-Likelihood: -3088.35
```

The results may be displayed graphically using the plot function:

```
> plot(results)
```

For this example the estimated memory evolution and forcing response are presented in Fig. A1. The estimated parameters are summarized in Table A1.

Parameter	True value	Posterior marginal mean	95% credible Interval
a	0.3	0.306	(0.197, 0.415)
b	0.2	0.193	(0.052, 0.288)
σ	5	7.025	(6.342, 8.078)
σ_f	0.1	0.1	(0.086, 0.123)
F_0	0	-0.005	(-0.051, 0.036)

Table A1. Underlying values used for simulating the data, along with estimated posterior marginal means and 95% credible intervals for all hyperparameters.

Combining forcing with irregular time steps requires more computationally intensive calculations within `rgeneric`, which increases the total computational time to around ten minutes, compared to 10 seconds using any other model configuration. To reduce this we have implemented the model in `cgeneric` which grants a substantial boost in speed. However, this requires pre-compiled `C` code using more simplistic priors for the parameters, which cannot be changed without recompiling the source code. Thus there could potentially be a small loss in accuracy of the fitted model at the cost of the improved speed. To use the `cgeneric` version of the model, set `do.cgeneric=TRUE` in the `inla.ews` function call.

Appendix B: Latent Gaussian model formulation

Section 3.1 defines our model within a Bayesian framework. However, in order for the model to be compatible with INLA we require some modifications such that it is expressed in terms of a latent Gaussian model. Latent Gaussian models represents a subset of hierarchical Bayesian models which are defined in three stages. First, the likelihood function is specified. The likelihood is then expressed using a latent field of unobserved Gaussian variables $\boldsymbol{w} = (w_1, ..., w_n)$ whose specification forms the second stage. These depend on a number of unknown hyperparameters. The final stage is to assign prior distributions to the hyperparameters.

Since our model is originally a two stage model, a Gaussian likelihood that depend on some parameters without an intermediate latent field $\boldsymbol{w}$, we reshape this into three stages by defining the former likelihood as the latent field $\boldsymbol{w}$ and the observations $\boldsymbol{x}$ to be the latent field with some additional negligible noise

$$\pi(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{\theta}) = \prod_{i=1}^{n} \mathcal{N}\left(w_i(\boldsymbol{\theta}), \sigma_x^2\right), \tag{B1}$$

where $\sigma_x^2 \approx 0$, essentially stating that $\boldsymbol{x} \approx \boldsymbol{w}$. This trick does not change our model, but creates a reformulation of the model into a latent Gaussian model where the latent field $\boldsymbol{w}$ is the prior of the mean of the likelihood.

The latent variables $\boldsymbol{w}$ follows a multivariate Gaussian process

$$\boldsymbol{w} \mid \boldsymbol{\theta} \sim \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1}\right), \tag{B2}$$

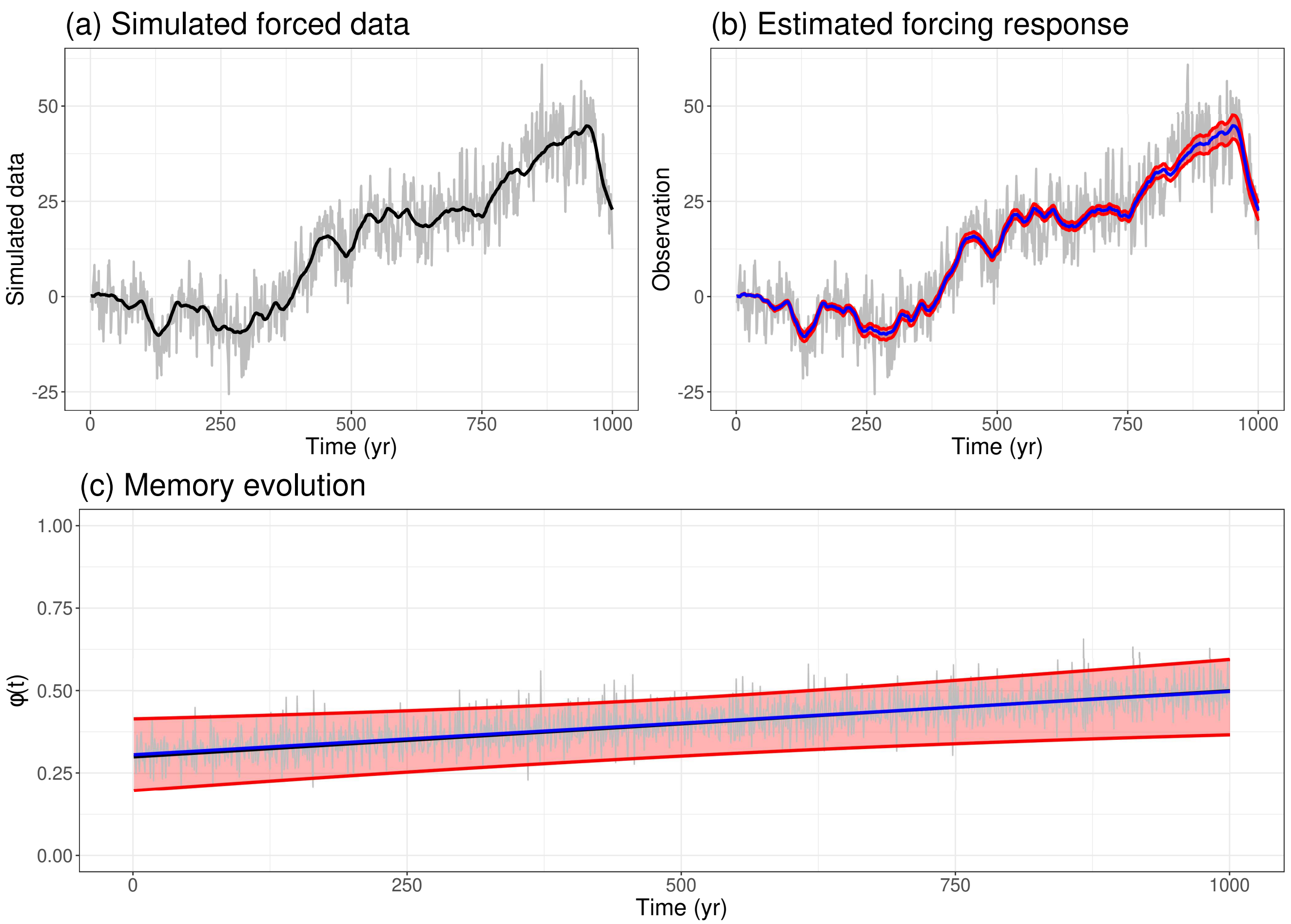

Figure A1. Panel(a) shows the simulated data (gray) where a simulated forcing response (black) has been added. Panel (b) shows the posterior marginal mean (blue) and 95% credible intervals (red) of the forcing response. Panel (c) shows the posterior marginal mean of the lag-one correlation parameter of the simulated data (gray). The fluctuations are caused by being sampled at non-constant time steps. The posterior marginal mean of the smoother evolution of $a+bt$ is included (blue), along with 95% credible intervals (red) and the true values (white).

where the precision matrix $\boldsymbol{Q}$ is given by (11) and $\boldsymbol{\mu}$ describe any potential trends as specified in the model. We will not discuss such trends here. When using INLA it is essential that the precision matrix is sparse in order to retain computational efficiency.

Since the Gaussian process now describe the latent variables instead of the likelihood, the parameters $\boldsymbol{\theta}$ which govern $\boldsymbol{w}$ will now be called hyperparameters, since they are the parameters of a prior distribution. The final step of defining the latent Gaussian model is to assign prior distributions to the hyperparameters, but since INLA prefers to work with unconstrained variables these are specified through the parameterizations derived in Section 3

$$\boldsymbol{\theta} \sim \pi(\log\kappa)\pi(\theta_b)\pi(\theta_a \mid \theta_b). \tag{B3}$$

Transforming priors chosen for (κ, b, a) to the corresponding priors chosen for $(\log\kappa, \theta_b, \theta_a)$ can be done using the change-of-variables formula.

We want to estimate the marginal posterior distribution for all hyperparameters and latent variables. These are computed by evaluating the integrals

$$\pi(w_i \mid \boldsymbol{x}) = \int \pi(w_i \mid \boldsymbol{\theta}, \boldsymbol{x})\pi(\boldsymbol{\theta} \mid \boldsymbol{x})d\boldsymbol{\theta} \tag{B4}$$

$$\pi(\theta_j \mid \boldsymbol{x}) = \int \pi(\boldsymbol{\theta} \mid \boldsymbol{x})d\boldsymbol{\theta}_{-j}. \tag{B5}$$

Of these we are primarily concerned with the latter, since the latent field will very similar to the observed values $\boldsymbol{x}$ since $\sigma_x \approx 0$. To compute these integrals INLA uses various numerical optimization techniques to obtain an appropriate approximation. Most importantly is the Laplace approximation (Tierney and Kadane, 1986), which is used to approximate the joint posterior distribution

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) \approx \left. \frac{\pi(\boldsymbol{w}, \boldsymbol{\theta}, \boldsymbol{x})}{\pi_G(\boldsymbol{w} \mid \boldsymbol{\theta}, \boldsymbol{x})} \right|_{\boldsymbol{w}=\boldsymbol{w}^*(\boldsymbol{\theta})}, \tag{B6}$$

where $\boldsymbol{w}^*(\boldsymbol{\theta})$ is the mode of the latent field $\boldsymbol{w}(\boldsymbol{\theta})$ and $\pi_G(\boldsymbol{w} \mid \boldsymbol{\theta}, \boldsymbol{x})$ is the Gaussian approximation of

$$\pi(\boldsymbol{w} \mid \boldsymbol{\theta}, \boldsymbol{y}) \propto \pi(\boldsymbol{w} \mid \boldsymbol{\theta})\pi(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{\theta}). \tag{B7}$$

The methodology is available as the open source R package `R-INLA`, which can be downloaded at www.r-inla.org (last access: July 30, 2024).

Since there are no model components already implemented for `R-INLA` that meet our specifications we are required to implement the model components ourselves using the custom modeling framework of R-INLA called `rgeneric`. This adds more work and complexity in implementing our model, and adds an additional barrier to further adoption of our methodology, which motivated us to create a more user-friendly R-package titled `INLA.ews`, available at www.github.com/eirikmn/INLA.ews (last access: July 30, 2024).

Appendix C: Comparison of different detrending approaches

Since there is no clear choice of forcing for DO events, and not all data windows appear stationary, we assume that there is some unknown trend component reflected in the data. This trend needs to be managed or the estimates of other components will suffer. Often, this is done by first detrending the data, before the parameters of interest are estimated. This bears the risk that variation caused by the time-dependent noise component may be attributed to the trend, and it is therefore better to estimate both the trend and noise components simultaneously. This can be achieved using INLA, which supports many common model components. We perform the same analysis on the data windows preceding all 17 DO events using four different trend models.

– No trend: The data is explained using the time-dependent AR(1) noise component ε_t and an intercept β_0 only,

$$y_t \sim \beta_0 + \varepsilon_t. \tag{C1}$$

We only expect this to provide accurate results for stationary data windows. The results in this paper can be recreated using the INLA.ews package. Let `y` denote the δ^{18}O ratios and `time` denote the GICC05 chronology, then the model can be fitted by prompting

```
results = inla.ews(data=y, timesteps=time, formula = y ~ 1)
```

To omit the intercept term set the formula argument to `formula = y ~ -1` instead. The `rgeneric` model component corresponding to the time-dependent AR(1) noise is added automatically. To improve numerical convergence, we perform the analysis in iterations, restarting from the previous found optima with reduced step lengths. This can be specified using the `stepsize` argument in the `inla.ews` function. The length of this argument corresponds to the number of iterations. Here we used `stepsizes = c(0.01, 0.005, 0.001)`.

– Linear trend: We incorporate an additional linear effect β_1 in the model,

$$y_t \sim \beta_0 + \beta_1 t + \varepsilon_t. \tag{C2}$$

This can capture linear increases, but will not be able to model any non-linearity in the ~~model~~data. This model can be fitted using

```
results = inla.ews(data=data.frame(y=y, trend1=time_norm),
     timesteps=time, formula = y ~ 1 + trend1)
```

where `trend1 = time_norm` is the covariate corresponding to the normalized time steps,

```
time_norm = (time-time[1])/(time[n]-time[1])
```

– 2nd order polynomial: We add another effect β_2 which allows for non-linearity to be described using a second order polynomial trend,

$$y_t \sim \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t. \tag{C3}$$

This model can be fitted using

```
results = inla.ews(data=data.frame(y=y,trend1=time_norm,trend2=time_norm^2),
          timesteps=time, formula = y ~ 1 + trend1 + trend2)
```

where `trend2` specifies a linear response to the covariates defined as the square of the normalized GICC05 chronology `trend2=time_norm**2`.

– 2nd order random walk (RW2): We use a random effect $f(t)$ described by a continuous 2nd order random walk to describe the trend,

$$y_t \sim f(t) + \varepsilon_t. \tag{C4}$$

This is a continuous extension (Lindgren and Rue, 2008) of a stochastic spline model which assumes that the second-order increments are independent Gaussian processes

$$x_i - 2x_{i+1} + x_{i+2} \sim \mathcal{N}(0, \sigma^2_{\text{RW2}}). \tag{C5}$$

This model is able to capture more general non-linearities compared to the 2nd degree polynomial trend, but makes the model less interpretable. Similar as in `R-INLA`, the RW2 model is specified using the following call

```
results = inla.ews(data=data.frame(y=y, idx=time,
          timesteps=time, formula = y ~ 1 + f(idx, model="crw2"))
```

where `idx` specifies the time steps of the continuous RW2 trend.

In Table 2 we present the estimated posterior probability of a positive trend, $P(b > 0 \mid \boldsymbol{y})$, compared to the corresponding p-values by Rypdal (2016) and Boers (2018). We show the fitted trends for each data interval in Fig C1. We observe that the models tend to agree, with some exceptions where the assumed trend is unable to capture the variation of the data. Although the RW2 trend is the ~~more~~ most flexible model it ~~appear~~ appears to exhibit irregular fluctuation for several events. The second order polynomial trend ~~appear~~ appears to be sufficiently flexible for all events, and provides a much smoother and more interpretable fit.

Appendix D: Break-point in memory evolution

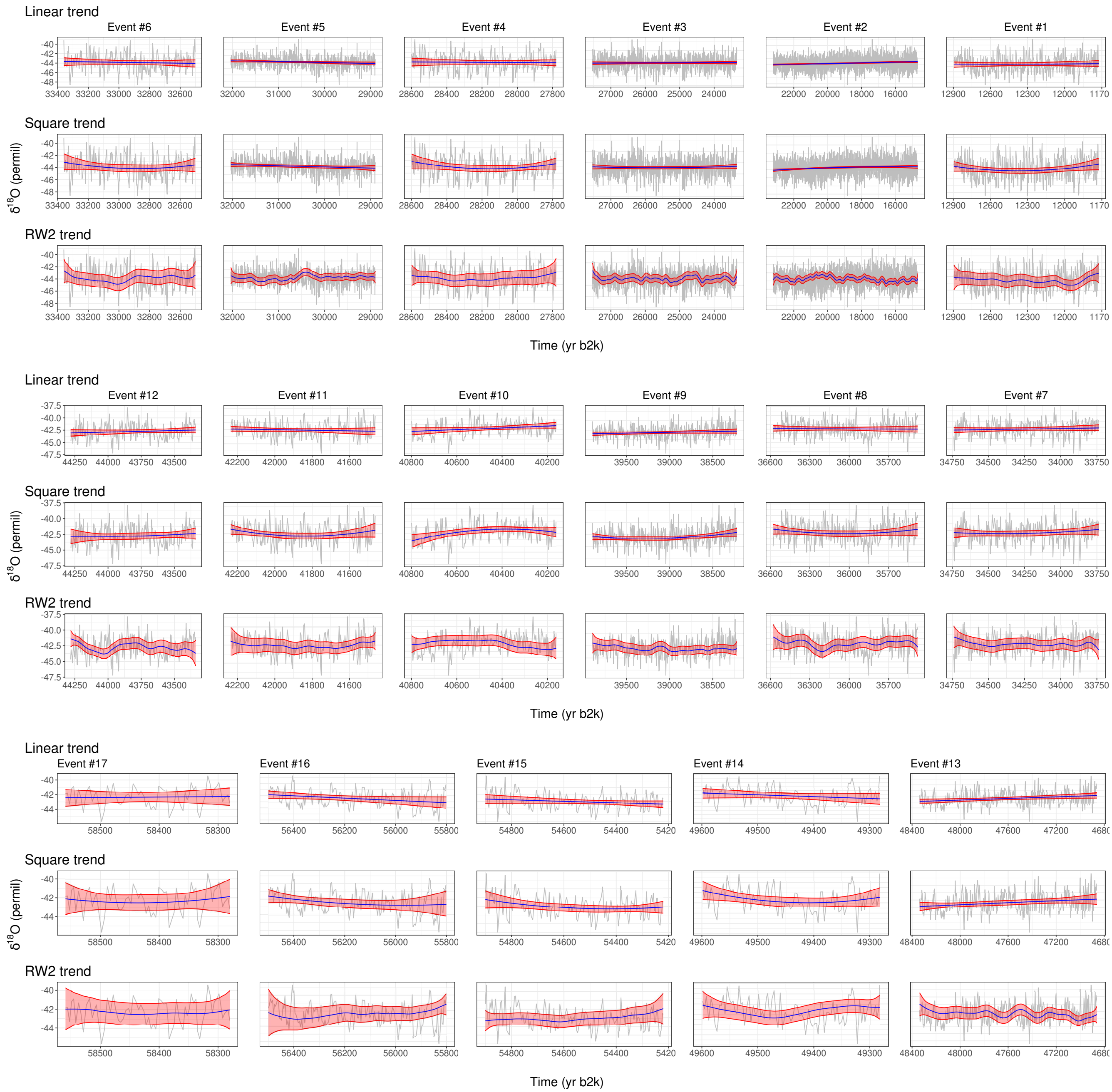

Figure C1. δ^{18}O proxy data from the NGRIP record (gray), with Greenland stadial phases highlighted. The posterior marginal mean (blue) and 95% credible intervals (red) of the fitted trends are included for each event.

Early warning signals are most easily detectable shortly before a bifurcation point. If the dataset covers a much larger period, for which most of it is stationary, it could be more difficult for the time-dependent AR(1) model to detect early warning signals if they are observable only for a much smaller subset of the data. To accommodate this one could add a break-point, a point in time where the lag-one correlation transitions from constant to linearly increasing.

Let t_{bp} denote a break-point, the lag-one correlation parameter is then defined by

$$\phi(t) = \begin{cases} a, & t \leq t_{\text{bp}} \\ a + \frac{b}{1-t_{\text{bp}}}(t - t_{\text{bp}}) & t > t_{\text{bp}} \end{cases}. \tag{D1}$$

For stability, we constrain the break-point parameter using parameterization $\theta_{\text{bp}} = 1/(1+\exp(-t_{\text{bp}}))$, such that $t_{\text{bp}} \in (0,1)$. This model is demonstrated by fitting it to simulated data where $t_{\text{bp}} = 0.5$. The results are presented visually in Fig. E1.

Appendix E: Demonstration of the `INLA.ews` package

We demonstrate the `INLA.ews` package on simulated forced data with non-equidistant time steps. The time steps t_k are obtained by adding Gaussian noise such that $\tilde{t}_k = k + \xi_k$ and normalized $t_k = (\tilde{t}_k - \tilde{t}_0)/(\tilde{t}_n - \tilde{t}_0)$ such that $t_0 = 0$ and $t_n = 1$. We assume a time dependent AR(1) process of length $n = 1000$ for the observations, sampled at times $t_1, ..., t_n$. The AR(1) process has standard deviation $\sigma = 5$ and time-dependent lag-one correlation $\phi(t) = a + bt_k$ given by $a = 0.3$ and $b = 0.2$. We also include a forcing $F(t)$, obtained by simulation from another AR(1) process with unit variance and $\tilde{\phi} = 0.95$. The forcing response is approximated by

$$\mu(t_k) = \frac{\sigma_f}{\sqrt{2\lambda(t_k)}} \sum_{s=t_0}^{t_k} e^{-\lambda(t_k)(t_k - t_s)} \left(F_0 + F(s)\right), \tag{E1}$$

with parameters set to $\sigma_f = 0.1$ and $F_0 = 0$, and added to the simulated observations. The AR(1) model and forcing `z` sampled at time points `time` can be fitted to the data `y`with INLA using the `inla.ews` wrapper function:

```
1: results <- inla.ews(data=y, forcing=z, formula=y ~ -1, timesteps=time)
```

The `inla.ews` function computes all posterior marginal distributions, computes summary statistics, formats the results 390 and returns all information as an `inla.ews` list object. Summary statistics and other important results can be extracted using the `summary` function:

```
1: > summary(results)
2: -
3: Call:
4: inla.ews(data = y, forcing = z, timesteps = time, formula = y ~ -1)
```

```
-
Time used:
   Running INLA Post processing            Total
        616.7390        142.4620        759.6259
-
Posterior marginal distributions for all parameters have been computed.
-
Summary statistics for using ar1 model (with forcing):
             mean      sd 0.025quant 0.5quant 0.975quant
a          0.2938 0.0353     0.2279   0.2927     0.3672
b          0.2127 0.0449     0.1350   0.2087     0.3091
sigma      7.1593 0.3223     6.5160   7.1651     7.7789
sigma_f    0.1019 0.0055     0.0910   0.1019     0.1127
F0        -0.0036 0.0202    -0.0453  -0.0028     0.0338
-
Memory evolution is sampled on an irregular grid.
Summary for first and last point in smoothed trajectory (a+b*time):
            mean      sd 0.025quant 0.5quant 0.975quant
phi0[1] 0.2938 0.0353     0.2279   0.2927     0.3672
phi0[n] 0.5060 0.0366     0.4370   0.5050     0.5798
Mean and 95% credible intervals for forced response have also been computed.
-
Probability of positive slope is 0.9999925
-
Marginal log-Likelihood: -3090.02
```

The results may be displayed graphically using the plot function:

```
> plot(results)
```

For this example the estimated memory evolution and forcing response is included in Fig. A1. The estimated parameters are summarized in Table A1.

Combining forcing with irregular time steps requires more computationally intensive calculations within `rgeneric`, which increases the total computational time to around ten minutes, compared to 10 seconds using any other model configuration. To reduce this we have implemented the model in `cgeneric` which grants a substantial boost in speed. However, this requires pre-compiled `C` code using more simplistic priors for the parameters, which cannot be changed without recompiling the source

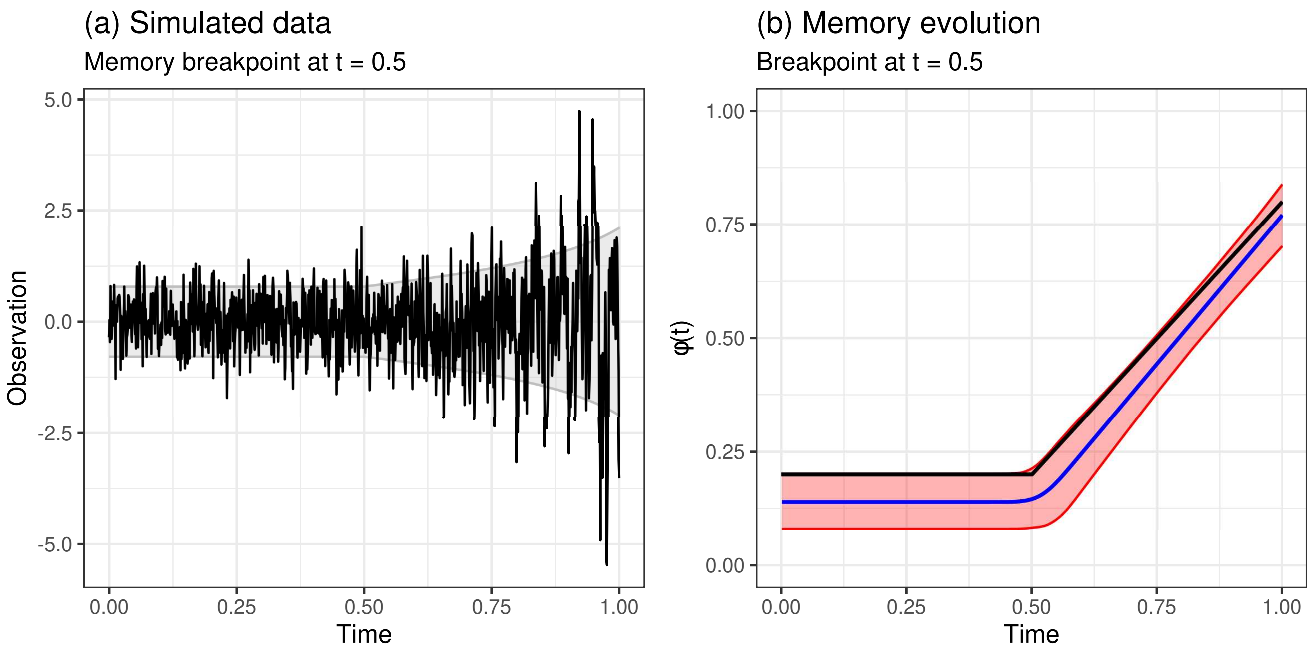

Figure E1. Panel (a) shows the posterior marginal mean of the lag-one correlation parameter of the simulated data (grayblack) . The fluctuations are caused by being sampled with a break-point located at non-constant time steps$t = 0.5$. The posterior marginal mean of shaded gray area describes the smoother evolution of $a + bt$ is included (blue), along with 95% credible intervals (red) and standard deviation derived from using the true values(white). Panel (b) shows the simulated observations (gray) along with the posterior marginal mean means (blue) and 95% credible intervals (red)of , with the forcing responsetrue memory evolution (black).

code. Thus there could potentially be a small loss in accuracy of the fitted model at the cost of the improved speed. To use the `cgeneric` version of the model, set `do.cgeneric=TRUE` in the `inla.ews` function call.

Parameter True value Posterior marginal mean 95% credible Interval a 0.3 0.294 (0.228, 0.367) b 0.20.213 (0.135, 0.309) σ 5 7.159 (6.516, 7.779) σ_f 0.1 0.102 (0.091, 0.113) F_0 0 -0.004 (-0.045, 0.034) Underlying values used for simulating the data, along with estimated posterior marginal means and 95% credible intervals for all hyperparameters.

Author contributions. All authors conceived and designed the study. EMN adopted the model for a Bayesian framework and wrote the code. LH and EMN carried out the examples and analysis. All authors discussed the results and drew conclusions. EMN and LH wrote the paper with input from MWR.

Competing interests. The authors declare that they have no conflict of interest

Acknowledgements. This project has received funding from the European Union's Horizon 2020 research and innovation programme (TiPES, grant no. 820970). EMN has also received funding from the Norwegian Research Council (IKTPLUSS-IKT og digital innovasjon, project no. 332901). We would like to thank Niklas Boers for helping us reproduce the results of Boers (2018), including providing code to obtain the interpolated 5-year sampled NGRIP/GICC05 data set.

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., White, J. W. C., and members, N. G. I. C. P.: High-resolution record of Northern Hemisphere climate extending into the last interglacial period, Nature, 431, 147–151, https://doi.org/10.1038/nature02805, 2004.

Andersen, K. K., Svensson, A., Johnsen, S. J., Rasmussen, S. O., Bigler, M., Röthlisberger, R., Ruth, U., Siggaard-Andersen, M.-L., Steffensen, J. P., Dahl-Jensen, D., et al.: The Greenland ice core chronology 2005, 15–42 ka. Part 1: constructing the time scale, Quaternary Science Reviews, 25, 3246–3257, 2006.

Boers, N.: Early-warning signals for Dansgaard-Oeschger events in a high-resolution ice core record, Nature communications, 9, 1–8, 2018.

Boers, N.: Observation-based early-warning signals for a collapse of the Atlantic Meridional Overturning Circulation, Nature Climate Change, 11, 680–688, 2021.

Boers, N. and Rypdal, M.: Critical slowing down suggests that the western Greenland Ice Sheet is close to a tipping point, Proceedings of the National Academy of Sciences, 118, e2024192 118, 2021.

Bond, G. C., Showers, W., Elliot, M., Evans, M., Lotti, R., Hajdas, I., Bonani, G., and Johnson, S.: The North Atlantic's 1-2 kyr climate rhythm: relation to Heinrich events, Dansgaard/Oeschger cycles and the Little Ice Age, Geophysical Monograph-American Geophysical Union, 112, 35–58, 1999.

Boulton, C. A., Lenton, T. M., and Boers, N.: Pronounced loss of Amazon rainforest resilience since the early 2000s, Nature Climate Change, 12, 271–278, 2022.

Braun, H., Christl, M., Rahmstorf, S., Ganopolski, A., Mangini, A., Kubatzki, C., Roth, K., and Kromer, B.: Possible solar origin of the 1,470-year glacial climate cycle demonstrated in a coupled model, Nature, 438, 208–211, 2005.

C.W.Gardiner: Handbook of Stochastic Methods, vol. 3, Springer Berlin, 1985.

Dakos, V., Scheffer, M., van Nes, E. H., Brovkin, V., Petoukhov, V., and Held, H.: Slowing down as an early warning signal for abrupt climate change, Proceedings of the National Academy of Sciences, 105, 14 308–14 312, 2008.

Dansgaard, W., Johnsen, S., Clausen, H., Dahl-Jensen, D., Gundestrup, N., Hammer, C., and Oeschger, H.: North Atlantic Climatic Oscillations Revealed by Deep Greenland Ice Cores, pp. 288–298, American Geophysical Union (AGU), https://doi.org/https://doi.org/10.1029/GM029p0288, 1984.

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., Jouzel, J., et al.: 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.

Ditlevsen, P. D. and Johnsen, S. J.: Tipping points: Early warning and wishful thinking, Geophys. Res. Lett., 37, https://doi.org/10.1029/2010GL044486, 2010.

Ditlevsen, P. D., Andersen, K. K., and Svensson, A.: The DO-climate events are probably noise induced: statistical investigation of the claimed 1470 years cycle, Climate of the Past, 3, 129–134, 2007.

Franzke, C. L. E., Osprey, S. M., Davini, P., and Watkins, N. W.: A Dynamical Systems Explanation of the Hurst Effect and Atmospheric Low-Frequency Variability, Scientific Reports, 5, 9068, https://doi.org/10.1038/srep09068, 2015.

Fredriksen, H.-B. and Rypdal, K.: Spectral Characteristics of Instrumental and Climate Model Surface Temperatures, Journal of Climate, 29, 1253–1268, https://doi.org/10.1175/JCLI-D-15-0457.1, publisher: American Meteorological Society Section: Journal of Climate, 2016.

Gkinis, V., Simonsen, S. B., Buchardt, S. L., White, J., 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, 2014.

Held, H. and Kleinen, T.: Detection of climate system bifurcations by degenerate fingerprinting, Geophysical Research Letters, 31, https://doi.org/https://doi.org/10.1029/2004GL020972, 2004.

Henry, L., McManus, J., Curry, W., Roberts, N., Piotrowski, A., and Keigwin, L.: North Atlantic ocean circulation and abrupt climate change during the last glaciation, Science, 353, 470–474, 2016.

Hubert, M. and Vandervieren, E.: An adjusted boxplot for skewed distributions, Computational Statistics & Data Analysis, 52, 5186–5201, https://doi.org/https://doi.org/10.1016/j.csda.2007.11.008, 2008.

Huybers, P. and Curry, W.: Links between annual, Milankovitch and continuum temperature variability, Nature, 441, 329–332, https://doi.org/10.1038/nature04745, 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. P.: Irregular glacial interstadials recorded in a new Greenland ice core (article) Author, Nature, 359, 311–313, https://doi.org/10.1038/359311a0, 1992.

Johnsen, S. J., Dahl-Jensen, D., Gundestrup, N., Steffensen, J. P., Clausen, H. B., Miller, H., Masson-Delmotte, V., Sveinbjörnsdottir, A. E., and White, J.: Oxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP, Journal of Quaternary Science: Published for the Quaternary Research Association, 16, 299–307, https://doi.org/10.1002/jqs.622, 2001.

Lenton, T., Livina, V., Dakos, V., Van Nes, E., and Scheffer, M.: Early warning of climate tipping points from critical slowing down: comparing methods to improve robustness, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 370, 1185–1204, 2012.

Li, C., Battisti, D. S., and Bitz, C. M.: Can North Atlantic sea ice anomalies account for Dansgaard–Oeschger climate signals?, Journal of Climate, 23, 5457–5475, https://doi.org/10.1175/2010JCLI3409.1, 2010.

Lindgren, F. and Rue, H.: On the Second-Order Random Walk Model for Irregular Locations, Scandinavian Journal of Statistics, 35, 691–700, https://doi.org/https://doi.org/10.1111/j.1467-9469.2008.00610.x, 2008.

Lovejoy, S. and Schertzer, D.: The Weather and Climate: Emergent Laws and Multifractal Cascades, Cambridge University Press, https://doi.org/10.1017/CBO9781139093811, 2013.

Løvsletten, O. and Rypdal, M.: Statistics of regional surface temperatures after 1900: Long-range versus short-range dependence and significance of warming trends, Journal of Climate, 29, 4057–4068, 2016.

Lynch-Stieglitz, J.: The Atlantic meridional overturning circulation and abrupt climate change, Annual review of marine science, 9, 83–104, 2017.

Mandelbrot, B. B. and Van Ness, J. W.: Fractional Brownian motions, fractional noises and applications, SIAM review, 10, 422–437, 1968.

Menviel, L., Timmermann, A., Friedrich, T., and England, M.: Hindcasting the continuum of Dansgaard–Oeschger variability: mechanisms, patterns and timing, Climate of the Past, 10, 63–77, 2014.

Menviel, L. C., Skinner, L. C., Tarasov, L., and Tzedakis, P. C.: An ice–climate oscillatory framework for Dansgaard–Oeschger cycles, Nature Reviews Earth & Environment, 1, 677–693, 2020.

Myrvoll-Nilsen, E., Fredriksen, H.-B., Sørbye, S. H., and Rypdal, M.: Warming trends and long-range dependent climate variability since year 1900: A Bayesian approach, Frontiers in Earth Science, 7, 214, 2019.

Myrvoll-Nilsen, E., Sørbye, S. H., Fredriksen, H.-B., Rue, H., and Rypdal, M.: Statistical estimation of global surface temperature response to forcing under the assumption of temporal scaling, Earth System Dynamics, 11, 329–345, 2020.

North Greenland Ice Core Project members: High-resolution record of Northern Hemisphere climate extending into the last interglacial period, Nature, 431, 147–151, https://doi.org/10.1038/nature02805, 2004.

Rasmussen, S. O., Andersen, K. K., Svensson, A., Steffensen, J. P., Vinther, B. M., Clausen, H. B., Siggaard-Andersen, M.-L., Johnsen, S. J., Larsen, L. B., Dahl-Jensen, D., et al.: A new Greenland ice core chronology for the last glacial termination, Journal of Geophysical Research: Atmospheres, 111, 2006.

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., et al.: 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, 2014.

Ritchie, P. and Sieber, J.: Early-warning indicators for rate-induced tipping, Chaos: An Interdisciplinary Journal of Nonlinear Science, 26, 2016.

Robert, C. P., Casella, G., and Casella, G.: Monte Carlo statistical methods, vol. 2, Springer, 1999.

Rue, H. and Held, L.: Gaussian Markov Random Fields: Theory And Applications (Monographs on Statistics and Applied Probability, London: Chapman and Hall-CRC Press, 2005.

Rue, H., Martino, S., and Chopin, N.: Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion)., J R Stat Soc Series B, 71, 319–392, 2009.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K.: Bayesian Computing with INLA: A Review, Annu. Rev. Stat. Appl., 4, 395–421, 2017.

Ruth, U., Wagenbach, D., Steffensen, J. P., and Bigler, M.: Continuous record of microparticle concentration and size distribution in the central Greenland NGRIP ice core during the last glacial period, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/https://doi.org/10.1029/2002JD002376, 2003.

Rybski, D., Bunde, A., Havlin, S., and von Storch, H.: Long-term persistence in climate and the detection problem, Geophysical Research Letters, 33, https://doi.org/https://doi.org/10.1029/2005GL025591, 2006.

Rypdal, M.: Early-warning signals for the onsets of Greenland interstadials and the younger Dryas–Preboreal transition, Journal of Climate, 29, 4047–4056, https://doi.org/10.1175/JCLI-D-15-0828.1, 2016.

Rypdal, M. and Rypdal, K.: Late Quaternary temperature variability described as abrupt transitions on a $1/f$ noise background, Earth System Dynamics, 7, 281–293, https://doi.org/10.5194/esd-7-281-2016, 2016.

Ryvkina, J.: Fractional Brownian Motion with Variable Hurst Parameter: Definition and Properties, Journal of Theoretical Probability, 28, 866–891, https://doi.org/10.1007/s10959-013-0502-3, 2015.

Schulz, M.: On the 1470-year pacing of Dansgaard-Oeschger warm events, Paleoceanography, 17, 4–1, 2002.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H.: Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors, Statistical Science, 32, 1 – 28, https://doi.org/10.1214/16-STS576, 2017.

Svensson, A., Andersen, K. K., Bigler, M., Clausen, H. B., Dahl-Jensen, D., Davies, S. M., Johnsen, S. J., Muscheler, R., Parrenin, F., Rasmussen, S. O., Röthlisberger, R., Seierstad, I., Steffensen, J. P., and Vinther, B. M.: A 60 000 year Greenland stratigraphic ice core chronology, Climate of the Past, 4, 47–57, https://doi.org/10.5194/cp-4-47-2008, 2008.

Tierney, L. and Kadane, J. B.: Accurate approximations for posterior moments and marginal densities, Journal of the american statistical association, 81, 82–86, 1986.

Vinther, B. M., Clausen, H. B., Johnsen, S. J., Rasmussen, S. O., Andersen, K. K., Buchardt, S. L., Dahl-Jensen, D., Seierstad, I. K., Siggaard-Andersen, M.-L., Steffensen, J. P., Svensson, A., Olsen, J., and Heinemeier, J.: A synchronized dating of three Greenland ice cores throughout the Holocene, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/https://doi.org/10.1029/2005JD006921, 2006.

Wiesenfeld, K.: Noisy precursors of nonlinear instabilities, Journal of Statistical Physics, 38, 1071–1097, 1985.