Detection of climate system bifurcations by degenerate fingerprinting

H. Held and T. Kleinen

Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany

Received 9 July 2004; revised 3 September 2004; accepted 9 November 2004; published 10 December 2004.

[1] A method is introduced to estimate the proximity of climate sub-systems to non-linear thresholds. We suggest to measure the smallest decay rate of the system under investigation and to consider its trend. We argue that this is the diagnostic variable most directly linked to the distance from a bifurcation threshold. With the climate model of intermediate complexity CLIMBER2 we demonstrate our method for the North Atlantic thermohaline circulation. It is shown that proper analysis of paleo information could significantly reduce the uncertainty which plagues current estimates of the distance from the shutdown of the thermohaline circulation. *INDEX TERMS:* 3220 Mathematical Geophysics: Nonlinear dynamics; 1620 Global Change: Climate dynamics (3309); 1635 Global Change: Oceans (4203); 4267 Oceanography: General: Paleoceanography; 6309 Policy Sciences: Decision making under uncertainty. **Citation:** Held, H., and T. Kleinen (2004), Detection of climate system bifurcations by degenerate fingerprinting, *Geophys. Res. Lett.*, *31*, L23207, doi:10.1029/2004GL020972.

1. Introduction

[2] The climate modeling community increasingly recognizes multi-stability as an important concept that allows for abrupt climate changes in response to gradual changes in forcing [e.g., *Rial et al.*, 2004]. In that respect, the North Atlantic thermohaline circulation (THC) represents the most prominent example, as it may switch equilibria [e.g., *Stommel*, 1961; *Bryan*, 1986] under global warming [e.g., *Rahmstorf and Ganopolski*, 1999]. The fresh water flux μ into the North Atlantic is regarded as one of the key parameters controlling the THC overturning strength q. The possibility that anthropogenic global warming may shift this control parameter μ over a critical value μ^*, called bifurcation point, which marks the end of the THC-switched-on equilibrium, has intensified efforts to determine trends in overturning q (RAPID-MOC, 2004, available at www.soc.soton.ac.uk/rapidmoc) and to estimate what additional change in freshwater input $\Delta\mu$ we could still afford before a collapse of the THC is induced. At present, any attempt to estimate $|\Delta\mu|$ rests on tuning a climate model to available climatological data and determining $|\Delta\mu|$ from that model. Unfortunately, such estimates of $|\Delta\mu|$ are heavily model-dependent: *Schmittner and Weaver* [2001] identified the parameterization of sub-scale processes in terms of diffusivity as a main source of this uncertainty.

[3] *Kleinen et al.* [2003] injected a fresh aspect into the debate by suggesting a change in spectral properties of q as a fundamental indicator for $|\Delta\mu| \rightarrow 0$. However, this idea was demonstrated only for low-dimensional THC box models and very long time-series. Here, we investigate whether similar indicators can also be constructed for complex natural or numerical systems, and how long the time-series need to be in a more parsimonious approach. We base our approach on the most fundamental, model-independent property of any bifurcation: the fact that the smallest system-immanent decay rate κ for perturbations of the equilibrium vanishes and the variability of the related mode diverges.

2. Model and Forcing

[4] We use the oceanic output of CLIMBER2 [*Petoukhov et al.*, 2000; *Ganopolski et al.*, 2001] which is a coupled climate model of intermediate complexity. The ocean component is similar to the ocean module employed by *Schmittner and Weaver* [2001] as both originate from the module by *Wright and Stocker* [1991]. We demonstrate our method for the following model time-series: (i) 5,000 yrs equilibrium runs for 6 different average freshwater forcings at 44° northern latitude with increasing proximity to the bifurcation threshold, i.e. decreasing $|\Delta\mu|$; each of the average forcings is superimposed with the identical Gaussian white noise time series, (ii) the same for two further non-standard sets of vertical and horizontal ocean diffusivities, (iii) a 50,000 yrs transient run with a linear increase in atmospheric CO_2 from 280 ppm to 800 ppm, implying an increased average fresh water forcing which we again perturb stochastically, resulting in a THC collapse at the end of the run.

3. Degenerate Fingerprinting

[5] In the following we motivate our bifurcation detection statistic by an idealized theoretical situation which provides valuable heuristics and also results in significant signals for our THC system. A bifurcation threshold is characterized by a critical value μ^* of a control parameter μ for which an equilibrium vanishes (or is born). When employing this concept in climate physics, one implicitly assumes that the quasi-stationary dynamics can be simplified as an equilibrium of a deterministic dynamics which is stochastically perturbed by weather-generated noise. In the small-noise limit, the response to noise can be approximated by the dynamics of linear modes. As a fundamental insight from dynamical systems theory, at any bifurcation, one mode becomes unstable, hence, the smallest decay rate κ of a perturbation vanishes [e.g., *Wiggins*, 1990]; we term the corresponding mode "critical mode." We further note that, still in the small-noise limit, referring to a linearized approximation, as $\Delta\mu \rightarrow 0$, the critical mode produces diverging variance $\propto 1/\kappa$. The latter implies that the leading

0094-8276/04/2004GL020972$05.00

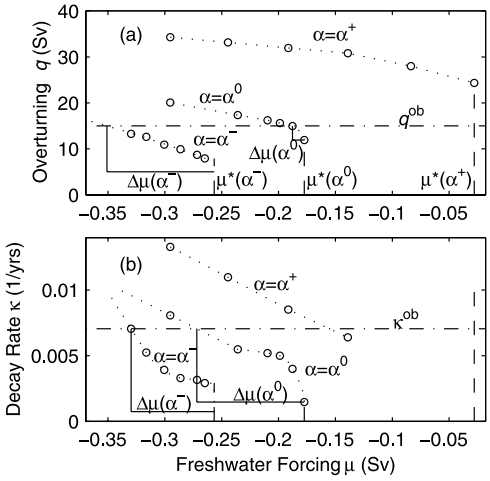

Figure 1. (a) North Atlantic overturning and (b) Atlantic salinity field decay rate as a function of freshwater forcing; CLIMBER2 results as circles. In Figure 1a, the best-guess value for THC overturning is indicated by a horizontal dashed-dotted line. Figure 1a clearly shows that from the measurement of overturning alone, the proximity $\Delta\mu$ to the bifurcation threshold μ^* cannot be determined. Quite the contrary, the decay rate (Figure 1b), if measured exactly, possesses a predetermined calibration $\kappa^* = 0$ at the bifurcation point. The standard CLIMBER2 value is indicated by a horizontal dash-dotted line. Note that here, the inferred $\Delta\mu$ depends much less on the unknown diffusion parameter α than in Figure 1a.

EOF [e.g., *von Storch and Zwiers*, 1999] approximates the critical mode. Hence, projecting onto the leading EOF provides the first step towards a parsimonious statistic. Thus we call the spatial aspect of our technique "degenerate fingerprinting." While in standard fingerprinting one strives to turn the detection and attribution pattern away from the noise patterns, here one has to accept that the main signal merges with the pattern of maximum noise. Our method simply rests on extracting κ from time-series. We determine the leading EOF of the Atlantic salinity field for standard diffusivities and for the freshwater forcing closest to the bifurcation. Time-series obtained from projection onto this EOF then serve as proxy for the critical mode dynamics. New paleo measurements would then strive at reconstructing salinity in the vicinity of the EOF. Note that it is not necessary to determine the THC overturning strength for our approach.

[6] The vicinity to a bifurcation allows for a simplification in the time-domain as well. If we assume that the other modes decay with much larger rates κ_i, their dynamics can be lumped into the weather noise, and we are allowed to pre-aggregate our EOF time-series into a time-discrete dynamics of fixed time-step Δt with $\Delta t \gg 1/\kappa_i$. If furthermore $1/\kappa \gg \Delta t$, the fluctuations of the critical mode can be modeled by a 1D-AR(1)-process $y_{n+1} = c\, y_n + \sigma\, \eta_n$, with $c = \exp(-\kappa\Delta t)$, and η_n Gaussian white noise. Hereby we assume that oscillatory modes do not play a major role in CLIMBER2. They could be easily included, however, by a 2D ansatz. We obtain c from lag-1 autocorrelation [*Priestley*, 2001], after a linear trend has been subtracted. A permanent choice of $\Delta t = 50$ yrs proves roughly consistent with the requirement $1/\kappa \gg \Delta t \gg 1/\kappa_i$. For each estimated c we then test whether the Δt-aggregated time series is compatible with AR(1).

[7] In summary, to estimate the proximity to a bifurcation, one assumes that in an equilibrium, the dynamics of the system can be approximated by its linearized version subject to stochastic forcing. One then strives at isolating the dynamics of the most unstable mode, the latter being characterized by the smallest decay rate. To achieve this, a spatial and a temporal filter are utilized. Any of the two would be sufficient if either the first EOF resembled closely enough the critical mode or the time-scale separation between the critical and the other modes was large enough. The closer the bifurcation is approached, the more justified either of the two assumptions is. In the next section we test for CLIMBER2 whether our ansatz delivers meaningful results.

4. Model Parameter Uncertainty Reduction

[8] In the following, we focus on the analysis of CLIMBER2-output. This allows us to not only determine the detection statistic κ, but also, more specifically, to demonstrate how the determination of κ decisively reduces the uncertainty in $|\Delta\mu|$, caused by uncertainty in diffusivity parameters. If a perfect ocean model was available, we could determine the proximity to the THC threshold from the standard values for overturning strength, $\bar{q} = 15 \pm 2$ Sv [*Ganachaud and Wunsch*, 2000]. Unfortunately, different ocean models would relate the same mean overturning $\bar{q}$ to different values of $|\Delta\mu|$. To a certain degree we emulate this effect following *Schmittner and Weaver* [2001] and change vertical and horizontal diffusivity parameters, k_v and k_h, respectively, of the CLIMBER2 ocean module. To keep the discussion simple, we construct a most influential (on q) diffusivity parameter α from this 2D-parameter manifold by setting $(k_v, k_h)(\alpha^0) = (0.8 \times 10^{-4}, 2 \times 10^3)$ m^2/s (standard values), $(k_v, k_h)(\alpha^-) = (0.605 \times 10^{-4}, 5.16 \times 10^3)$ m^2/s and $(k_v, k_h)(\alpha^+) = (1.4 \times 10^{-4}, 0.3 \times 10^3)$ m^2/s. This most influential parameter combination has been read from an ensemble of equilibrium runs, scanning the 2D parameter space. In particular the regime $\alpha^- \ldots \alpha^0$ within the α-manifold cannot be restricted well from observational data according to *Schmittner and Weaver* [2001]. Figure 1a displays the known equilibrium curves $\bar{q}(\mu)$ for the three choices of diffusivities α^-, α^0, α^+. From Figure 1a, a well-known dilemma in determining a potential THC threshold becomes obvious: in spite of a known overturning strength q^{ob}, virtually any value for $\Delta\mu$ could be concluded due to large uncertainty in α. In particular, we find $|\Delta\mu(q^{ob},\alpha^0)| \ll |\Delta\mu(q^{ob},\alpha^-)|$. We now show that κ is a much better proxy for $\Delta\mu$ than q.

[9] Figure 1b displays $\kappa(\mu)$ for the same equilibrium runs. We only plot results for those runs which are compatible with our AR(1)-ansatz, according to the following requirements: (i) $c > 0$, (ii) the residuals are white noise, tested by auto-covariance of lag 1, (iii) the residuals are Gaussian according to a χ^2 test of goodness of fit. The Gaussian to be tested against is estimated from the residuals' variance (i.e. $r = 1$ parameter is estimated) and then compared to a histogram of the residuals with an expected bin-occupation of approximately 5, leading to a k-bin histogram with $k = 19$.

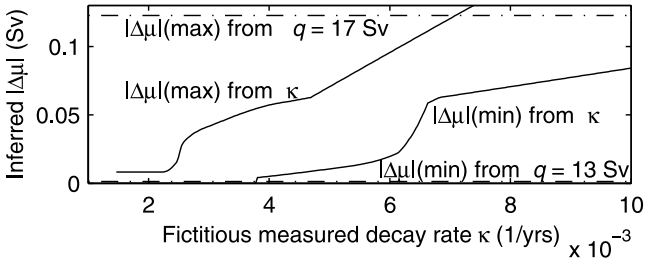

Figure 2. Learning from fluctuations. Maximum and minimum estimates for $|\Delta\mu|$ from an overturning of 13 to 17 Sv (dash-dotted lines), extracted from Figure 1a for the ocean diffusivities α^-, α^0. The difference is huge. The same in solid lines for potential measurements of the decay rate κ (abscissa), extracted from Figure 1b with error bars according to $\delta\kappa \approx \sqrt{2\kappa/T}$, $T = 10$ kyrs. The combined knowledge of q and κ implies a strong reduction of uncertainty in $|\Delta\mu|$.

Hence for (iii), we involve a χ^2-distribution of $k - 1 - r = 17$ degrees of freedom. For both (ii) and (iii) we require at least 0.4% error of first kind (i.e., false rejection of the null hypothesis).

[10] The decisive difference between the graphs (a) and (b) lies in the fact that the critical overturning value $\bar{q}^*$ strongly depends on α while κ^* displays the a priori *absolute calibration* $\kappa^* = 0 \iff c^* = 1$. The curve for α^0 resembles this quite well, while for α^- the theoretical limit is not reached as closely. This can be attributed to the fact that the fingerprinting was optimized for α^0. More sophisticated statistics can be thought of which filter out contaminating effects from the non-critical modes. However, Figure 1b still reveals a drastic reduction in uncertainty on $|\Delta\mu|$.

[11] The error bars we find on c [*Priestley*, 2001] are in rough agreement with the expected variance $\delta c^2 = (1 - c^2)\Delta t/T \Rightarrow \delta\kappa \approx \sqrt{2\kappa/T}$, T denoting the length of the time series, adding further trust to the justification of our approximations. We now continue our analysis for the plausible scenario that the main uncertainty is spanned by the regime $\alpha^- \ldots \alpha^0$ which is still compatible with $q = 15$ Sv. For fictitious measurements of κ we read from Figure 1b which error bar on $|\Delta\mu|$ would follow then, and add the uncertainty $\delta\kappa \approx \sqrt{2\kappa/T}$ stemming from the finite length of a 10 kyr time series (see Figure 2, solid lines). Analogously, within the framework of the traditional method, a combined error bar is read from Figure 1a assuming a current best-guess value for the uncertainty in overturning. Figure 2 reveals a strong information gain if κ is obtained from a potential Holocene time series: the knowledge of κ would imply a reduction of uncertainty about $\Delta\mu$ by a factor 2 to 6.

5. Determination of Transient Effects

[12] In Figure 3 we display a 50,000 yrs CLIMBER2-run for α^0 and a transient yet quasi-static increase in μ until the THC collapses, and compare the transient behavior of overturning strength and c. The overturning shows a quasi-linear time-dependence; given just the overturning time-series, an observer would receive no early warning for lurking abrupt changes as the square-root shape visible in the equilibrium curve of Figure 1a is masked by transient effects. On the contrary, the time-series of c, although noisy, allows to conclude where that threshold is roughly to be expected. This advantage again roots in the fundamental systems-theoretical property $c^* = 1 \iff \kappa^* = 0$. A linear fit has been obtained by cutting $c(t)$ into quasi-uncorrelated segments of $1/\kappa(\bar{c} \approx 0.9) \approx 500$ yrs. We consider linear fits compatible with the standard 95% confidence ellipsoid, according to an F-distribution [*Priestley*, 2001].

[13] Finally we would like to know what the typical length of a time series needs to be such that our method becomes applicable. Inversion of the formula on $\delta\kappa$ reveals for a desired maximum error $\delta\kappa/\kappa$

$$T_{\min} > \frac{2}{\kappa}\left(\frac{\delta\kappa}{\kappa}\right)^{-2}. \quad (1)$$

[14] We finally address a further property of our statistic κ being beneficial independent of whether the investigated system possesses a threshold or not, and again choose the THC for illustration. Suppose that under RAPID-MOC and follow-up projects a trend in q is observed, resulting in a net shift of $\delta\bar{q}_{\text{obs}}$ after an observational period of $T_{20} < 20$ yrs. In order to conclude the long-term development of q from $\delta\bar{q}_{\text{obs}}$, knowledge of the internal THC response times, to be obtained from Holocene data, is crucial. If, for example, for $t > T_{20}$ an aggressive climate-protection policy was implemented stabilizing the THC forcing μ, then in the long run, THC would obtain the new equilibrium $\bar{q}_0 + \delta\bar{q}_{\text{equ}}$. If we assume for illustration that also for a short period like T_{20} the internal dynamics are governed by κ only and that a

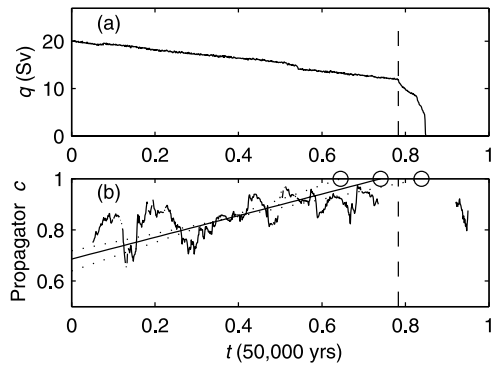

Figure 3. (a) THC overturning and (b) Atlantic AR(1) propagator c over time, for a transient CO_2-forcing run of CLIMBER2 ($\alpha = \alpha^0$). Here c is obtained from a moving 10% time-window. Although Figure 3a shows artificially low variability, no early warning sign for a threshold behavior is observable. Figure 3b is noisy but has the benefit of the absolute calibration $c^* = 1$, and hence is much more informative on $\Delta\mu$ than the overturning. The linear fit is accompanied by 95% error bars maximizing the intersection with the critical line $c \equiv 1$. The critical time is within the error bar. The latter is $0.2 \ll 1$, and hence the method is informative. The kink which is visible on the right side of the critical time is a fluctuation property of the new THC equilibrium, the shut-down mode.

linear trend in forcing has already persisted another T_{20} before starting the observations, then

$$\left|\delta\bar{q}_{\mathrm{equ}}\right| \approx \frac{4}{3}\left|\delta\bar{q}_{\mathrm{obs}}\right| \left(\kappa\ T_{20}\right)^{-1} \gg \left|\delta\bar{q}_{\mathrm{obs}}\right|. \quad (2)$$

[15] Hence, κ determines the long-term impacts (in terms of q) of the observed shift in overturning. The observed shift will most probably underestimate the long-term shift by orders of magnitude. Present estimates of $1/\kappa$ range from the order of 10 yrs from an adiabatic Stommel model [*Monahan*, 2002] to 1,000 yrs [*Weijer and Dijkstra*, 2003]. Therefore even an order-of-magnitude, yet model-independent, paleo-time series based knowledge of κ would be of tremendous value. In the most spectacular case, a threshold exists and $\delta\bar{q}_{\mathrm{obs}}$, $1/\kappa$ are found so large that the threshold could already have been passed.

6. Discussion

[16] The time-series method presented here consists of extracting the smallest decay rate κ in the time-series of the first principal component, which is beneficial for three purposes: for determining distances to bifurcation thresholds in a static situation, given a particular climate model with large parameter uncertainty, for model-independent distance calibration with slow climate trends present, and for the correct interpretation of fast climate trends detected by traditional methods. For the first two, explicit examples have been generated from the climate model CLIMBER2 of intermediate complexity. Referring to the third, if a climate shift was observed over a short period, caused by some forcing in the past, the major response to that forcing may yet have to be expected for the future (see section 5).

[17] For the first of the three purposes, the validity of the approximations and the underlying physical picture we presented in section 3 is helpful but not necessary. In the worst case, our statistic does not reduce uncertainty beyond what can be achieved with traditional methods, but it does not lead to wrong conclusions: extraction of κ just implies a filter which allows to relate observational data and model through a further channel. The improvement on model parameters implies an enhanced reliability of model-based climate prediction. The remaining two purposes, however, are more demanding: if our assumptions were not justified, a wrong absolute calibration would follow. Our results suggest that such a calibration is possible within a certain error bar. More sophisticated methods mentioned below provide considerable potential for improvement in terms of accuracy. However, we find it promising that already our rather elementary method, chosen for the benefit of transparency, extracts the desired key effects.

[18] We demonstrate with CLIMBER2 that in case of large model uncertainty as for ocean diffusivity, the knowledge of the decay rate decisively reduces uncertainty about the proximity to the bifurcation threshold. Holocene duration proves sufficiently long to allow for an estimate of this rate. Furthermore, even in an almost-static yet transient CLIMBER2-run with final THC-collapse, no early-warning signs in terms of a square-root shaped behavior of the overturning can be obtained. On the contrary, the trend in decay rate allows for an approximate estimation of the distance to the threshold.

[19] Finally, when addressing observational data from a real-world system, one faces three major obstacles: First, the system may not be as close to a bifurcation such that our unimodal approximation was justified. Second, when using paleo data, these typically imply a reconstruction of only part of the 3D field, for example, of temperature or salinity for the ocean. Third, the data will contain observational noise.

[20] However, our method to extract the critical decay rate κ could be refined by choosing modern methods of time-series analysis [*Timmer et al.*, 2000] which simultaneously address all three of those obstacles. This opens the fascinating possibility that κ could be inferred from paleo time series even if the forcing of the system could never be reconstructed.

[21] **Acknowledgments.** We are grateful to T. Schneider von Deimling for a hint on the construction of α, and to W. Ihra, E. Kriegler, T. Kuhlbrodt, H. Rust, and H. U. Voss for helpful discussions. H.H. and T.K. were financed by BMBF- and VW-research grants.

References

Bryan, F. (1986), High-latitude salinity effects and interhemispheric thermohaline circulations, *Nature*, *323*, 301–304.

Ganachaud, A., and C. Wunsch (2000), Improved estimates of global ocean circulation, heat transport and mixing, *Nature*, *408*, 453–457.

Ganopolski, A., V. Petoukhov, S. Rahmstorf, V. Brovkin, M. Claussen, A. Eliseev, and C. Kubatzki (2001), CLIMBER-2: A climate system model of intermediate complexity: II. Model sensitivity, *Clim. Dyn.*, *17*, 735–751.

Kleinen, T., H. Held, and G. Petschel-Held (2003), The potential role of spectral properties in detecting thresholds in the Earth System: Application to the thermohaline circulation, *Ocean Dyn.*, *53*, 53–63.

Monahan, A. (2002), Stabilization of climate regimes by noise in a simple model of the thermohaline circulation, *J. Phys. Oceanogr.*, *32*, 2072–2085.

Petoukhov, V., A. Ganopolski, V. Brovkin, M. Claussen, A. Eliseev, C. Kubatzki, and S. Rahmstorf (2000), CLIMBER-2: A climate system model of intermediate complexity: I. Model description and performance for present climate, *Clim. Dyn.*, *16*, 1–17.

Priestley, M. B. (2001), *Spectral Analysis and Time Series*, pp. 352, 353, 366, Academic, San Diego, Calif.

Rahmstorf, S., and A. Ganopolski (1999), Long-term global warming scenarios computed with an efficient coupled climate model, *Clim. Change*, *43*(2), 353–367.

Rial, J. A., R. A. Pielke Sr., M. Beniston, M. Claussen, P. Cox, H. Held, N. de Noblet-Ducoudr, R. Prinn, and J. Salas (2004), Nonlinear responses to global environmental change: Critical thresholds and feedbacks, *Clim. Change*, *65*, 11–38.

Schmittner, A., and A. J. Weaver (2001), Dependence of multiple climate states on ocean mixing, *Geophys. Res. Lett.*, *28*(6), 1027–1030.

Stommel, H. (1961), Thermohaline convection with two stable regimes of flow, *Tellus*, *13*, 224–241.

Timmer, J., U. Schwarz, H. U. Voss, I. Wardinski, T. Belloni, G. Hasinger, M. van der Klis, and J. Kurths (2000), Linear and nonlinear time series analysis of the black hole candidate Cygnus *X*-1, *Phys. Rev. E*, *61*, 1342–1352.

von Storch, H., and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, Cambridge Univ. Press, New York.

Weijer, W., and H. A. Dijkstra (2003), Multiple oscillatory modes of the global ocean circulation, *J. Phys. Oceanogr.*, *33*, 11.

Wiggins, S. (1990), *Introduction to applied nonlinear dynamical systems and chaos*, Springer, New York.

Wright, D. G., and T. F. Stocker (1991), A zonally averaged ocean model for the thermohaline circulation. Part I: Model development and flow dynamics, *J. Phys. Oceanogr.*, *21*, 1713–1724.

H. Held and T. Kleinen, Potsdam Institute for Climate Impact Research (PIK), PO Box 60 12 03, 14412 Potsdam, Germany. (held@pik-potsdam.de)