

Response to Reviewer #1

Comments from Reviewer #1:

In this manuscript the authors describe a series of twin experiments in which the Ensemble Adjustment Kalman Filter (EAKF) is used for state and parameter estimation in models of the Atlantic Meridional Overturning Circulation (AMOC). The models in question are box models, driven by atmospheric input taken from the output of chaotic, rather than stochastic systems. The authors demonstrate that their methods allow reliable estimation of unknown parameters in the model AMOCs, which exhibit multiple stable equilibria. The results described here will be of interest to a broad segment of NPG readers, but the manuscript as it stands needs a great deal of work to make it acceptable.

RE: A few conferences have been held for all co-authors to discuss the comments from reviewer #1. All authors converge to the point that the constructive comments and suggestions are extremely thoughtful and important for improving the manuscript and enhancing our understanding on the topic. Several extra experiments for addressing the concerns of the reviewer are performed. Thanks for your encouragement. All issues are replied point-by-point as below. We hope the whole manuscript has been essentially improved.

Application of data assimilation methods to use observations to track the evolution of solutions of systems that exhibit multiple stable modes has been described by many authors since the 1990s; see, e.g., Weir et al., Nonlin Proc Geophys 2013 and references therein. A look through the literature since the 1990s will also turn up examples of simultaneous state and parameter estimation in simple systems. The novelty of the present work lies in the specific application to regime transitions in the AMOC. I don't know of other examples of application of optimized methods such as the EAKF to joint state and parameter estimation of box models of the AMOC, which have been around since Stommel (1961).

RE: Excellent suggestion! We have added more detailed statements about the application of data assimilation methods (L112-118) and the examples of simultaneous state and parameter estimation in nonlinear systems having multiple stable modes (L122-128).

L112-118:

Tardif et al. (2014) implement data assimilation with EnKF to recover the AMOC with observations in a low-order coupled atmosphere-ocean climate model. They mainly explore the value of data assimilation for the initialization of the AMOC, while the effect of parameter errors in AMOC simulations needs further discussion. As another class of ensemble-based assimilation methods, particle filters, unlike the EnKF, are applicable to non-Gaussian probability distributions (e.g., Gordon et al., 1993; van Leeuwen, 2009). A mixture-based implicit particle method is presented and could detect transitions in an example with multiple attracting states (Weir et al., 2013a). However, the particle filter is plagued by the curse of dimensionality as the system dimension increases (Snyder et al., 2008; Carrassi et al., 2018).

L122-128:

Based on EAKF, a data assimilation scheme for enhance parameter correction is designed to improve parameter estimation using observations (Zhang et al., 2012). Zhao et al. (2019) perform this scheme in a simple AMOC box model, and the model parameters are successfully optimized when the model errors are caused by only erroneously set parameters. Although the AMOC regime transition is not addressed in their study, their exploration of model sensitivities regarding parameters serves as a guideline for our research. Many efforts have been made to advance the application of data assimilation and parameter estimation in nonlinear systems having multiple equilibrium states (e.g., Miller et al., 1994, 1999; Khalil et al., 2009; Weir et al., 2013b; Bisaillon et al., 2015).

The authors need to be more specific about exactly how their results differ from existing results, and why they are interesting. The closest thing that I can find to a statement of purpose appears at the end of the introduction, p4, beginning on line 111: “Here we present a method for improving the modeling of AMOC multi-equilibria. The new method is shown to simulate the AMOC transition between different equilibrium states accurately in two simple coupled models ...” As noted above, others have shown the ability to simulate transition between different equilibrium states in other systems. If the authors’ methods are novel, they should point out their differences from other methods that have been applied to similar systems.

RE: Thank you for this suggestion! In this revision, we have reorganized the existing literature on AMOC transitions (L81-90), as well as more specifically illustrated how this study differs from previous studies and pointed out the novelty of this paper (L96-103, L128-133).

L81-90:

AMOC transitions can occur due to external forcing or internal feedback (Klockmann et al., 2020). The external forcing applied in systems may include freshwater forcing (e.g., Cessi, 1994; Castellana et al., 2019), wind forcing (e.g., Ashkenazy and Tziperman, 2007; Kleppin et al., 2015), ice sheet forcing (e.g., Zhang et al., 2014; Mitsui and Crucifix, 2017), CO2 forcing (e.g., Zhang et al., 2017). The physical processes in the model are changed by external forcing, resulting in the transition between different states of the AMOC. For the AMOC model without external forcing, the transition is triggered by complex internal interactions within the model, such as salt oscillations (Peltier and Vettoretti, 2014), internal oceanic processes (Sévellec and Fedorov, 2014), thermohaline oscillations (Brown and Galbraith, 2016), intermittencies in the sea-ice cover (Gottwald, 2021). Regardless of whether it is due to external forcing or internal feedback, AMOC transitions could be influenced by complex physical processes in models, and the parameters involved in these physical processes are usually fixed.

L96-103:

Observation-constrained model parameters no longer keep fixed values but are constantly varying over time. The purpose of this paper is to explore whether the variations of observation-constrained parameters that allow the physical processes of model to evolve over time can make the simulation results closer to the “observed” feature of regime transitions. The models in this paper are obtained by coupling AMOC box model with Lorenz’s model, similar to the work by Roebber (1995) or Gottwald

(2021), where the variation of AMOC is driven by the chaotic dynamical system. The thermal mode and the reverse haline mode correspond to different equilibrium states of the AMOC. For simplicity, we will refer to these different states as the stronger AMOC (on-state) and the weaker AMOC (off-state) in simple conceptual models (e.g., Weijer et al., 2019).

L128-133:

Although the AMOC model eventually exhibits multiple equilibria, the AMOC is not a direct model state but is indirectly derived from model states such as atmospheric wind, ocean temperature and salinity. Instead of adjusting AMOC directly, the model states are adjusted through data assimilation. When constraining model parameters by observational information, the parameters that constantly vary with observations may provide more diversity in the physical processes involved with AMOC regime transition, so that the model can simulate more AMOC transition paths.

Their presentation of their three box model, equations (5)-(10) is confusing. Why different systems of equations for the thermal and haline modes? For details of the model they refer the reader to Shen et al. (2011) and the model described there is a single system that exhibits both haline and thermal modes. The model they finally use, defined in equation (12), is a complex system with many switches. I don't understand why this is necessary. Why not just use some form of (5) from Shen et al. (2011)?

RE: Sorry we haven't explained this clearly. Shen et al. (2011) described two different systems of equations for the thermal and haline modes, which are exactly the same as the equations (5)-(10) in this paper. They do not show the equations for the haline mode, but describe the equations for the thermal mode in detail. The haline mode is described briefly in a sentence between equation (2d) and equation (3) in Shen et al. (2011). Following this suggestion, we have added new lines to clarify this issue in L301-303, L318-320, and L401-402.

L301-303:

The governing equations for the haline mode also follow the study of Shen et al. (2011). They didn't show those equations, but only described them briefly. In this paper, to describe the construction of MOC3B-5V more clearly later, those equations are shown here.

L318-320:

Similar equations for the thermal and haline modes could be found in Guan and Huang (2008) for Eq. (1) (thermal mode) and Eq. (2) (haline mode), and in Shen and Guan (2015) for Eqs. (1)-(6) (thermal mode) and Eqs. (7)-(9) (haline mode).

L401-402:

A similar AMOC box model with many switches could be found in Castellana et al., (2019).

The results they get from this model, driven by a chaotic atmosphere, are encouraging. Figures 5-7 show that the data assimilation/parameter estimation system works well, reproducing the “true” trajectory quite accurately and producing a good estimate of the unknown parameter, while leaving out the parameter estimation steps, and using EAKF for state estimation without adjusting the parameter to its true value does not do nearly so well.

RE: Thanks for your encouragement.

At the end of this section the authors state: “The MOC3B-5V model is just a simple conceptual model, and the model states x_2 , w , and η simply conceptually simulate the variation characteristics of the atmosphere and the ocean. Although the transitions of AMOC are simulated by the MOC3B-5V model, the specific physical meaning of the model is not explicit enough. The method of capturing regime transitions in Sect. 2 is proved to be feasible in the simple model, and the next step is to apply the method to a physics-based MOC box model.” The next section describes experiments with a “physics-based MOC box model,” which is no more complicated than the MOC3B-5V model in the previous section, and the authors do not make clear what conceptual points are made with the MOC3B-5V model that are any less clear in the “MOCBM.” It seems to me that the sections dealing with the MOC3B-5V model, i.e., much of section 2 and all of section 3 could be left out entirely without any loss of understanding on the part of the reader, however convincing figures 5-7 may be. It shouldn’t be too hard for the authors to include additional figures corresponding to figures 5b, 6 and 7 to illustrate the details of the MOCBM experiment

RE: Thanks for your thoughtful suggestion! In this revision, we address this comment by two parts. In the first part, we like to show the details of the MOCBM experiment (additional Figures S1, S2, and S3 below corresponding to Figures 5, 6 and 7 in the manuscript). Based on your suggestion, we have added Fig. S1b that shows the assimilation result to the revised manuscript (Fig. 9b) and more relevant discussions have been added in L613-615. Given that Figs. S2 and S3 are very similar to the behavior of the MOC3B-5V model, they are not shown in the manuscript. However, the corresponding descriptions have been added in L618-619.

In the second part, we have added more discussions to justify the necessity of MOC3B-5V and MOCBM models as three aspects. First, new lines have been added to emphasize the benefits of the MOC3B-5V model in L520-524. Second, more discussions about the important differences between these two models have been added in L620-623. Third, we have added more description and discussion about the different behaviors of different models in data assimilation and parameter estimation to optimize the simulation results in L664-675.

L613-615:

Figure 9b shows the time series of the MOC value in the “truth” simulation (red lines) and in the free model control ensemble simulations with only data assimilation. Due to the existence of parameter error, inaccurate analyses are obtained when only data assimilation was performed without parameter estimation.

L618-619:

Since the behavior of the MOCBM and MOC3B-5V models are very similar, the figures corresponding to Fig. 6 and Fig. 7 are not shown here.

L520-524:

The 5-variable conceptual climate model could simulate the interactions between the atmosphere and ocean, and coupling it with the three-box MOC model could accurately address the ideas in this paper. The transferring of the uncertainty of the MOC3B-5V model is particularly simple and easily understood. With the help of this model, we found that the coupled model parameter estimation with observations can significantly mitigate the model deviations, thus capturing regime transitions of the AMOC. As such, this simple model has more visibility to demonstrate the essence of the problem.

L620-623:

The box model in the MOCBM is based on the classical approach of adopting a buoyancy constraint, and the circulation is regulated by the surface buoyancy difference, implying that surface thermohaline forcing drives AMOC (Birchfield, 1989). In contrast, for the box model in the MOC3B-5V model, the constraint is based on mechanical energy sustaining diapycnal mixing, and the circulation is maintained by mechanical energy from wind stress and tides (Shen et al., 2011).

L664-675:

Since the circulation is driven only by the meridional gradients of the upper ocean temperature and salinity in the buoyancy-constraint MOCBM model, AMOC regime transitions can be captured to some extent when the upper ocean temperature and salinity are directly adjusted by data assimilation only, but the simulation results are not accurate enough. In this simple model, since the data assimilation has worked well, the contribution of parameter estimation is relatively small but still indispensable. The AMOC regime transitions are captured more accurately by parameter estimation. The degree of contributions of data assimilation or parameter estimation to the optimization of simulation results is different in these two models. Compared with the MOCBM model, the energy-constraint MOC3B-5V model is more representative for the role of parameter estimation because the circulation is maintained by mechanical energy. When leaving out the parameter estimation steps and constraining the model states only by data assimilation, the accuracy of state estimation is not high due to the existence of parameter errors. Given the fact that the circulation is driven in a more complex way in the real world, this simple model study only provides a conceptual understanding and guideline for more complex real systems such as Coupled General Circulation Model (CGCM).

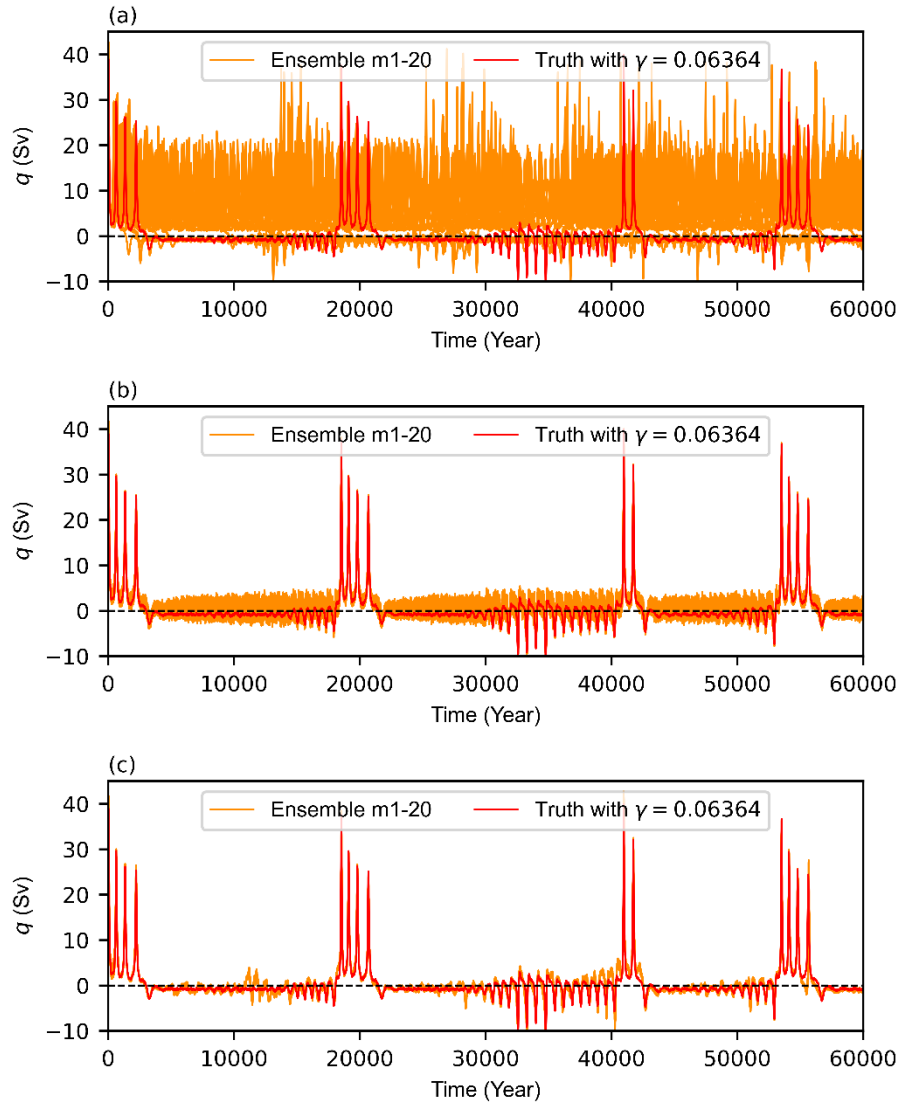


Figure S1. Time series of the meridional overturning circulation q in the “truth” simulation with $\gamma = 0.06364$ (red) and the individual ensemble members (orange) in the free model control ensemble simulations a) without data assimilation and parameter estimation, b) with data assimilation or c) with data assimilation and parameter estimation using erroneously-guessed γ values with a Gaussian perturbation that has a mean value of 0.070004 and a standard deviation of 10% times the standard value. The dashed black line denoting $q = 0$ is a division line between two equilibrium states.

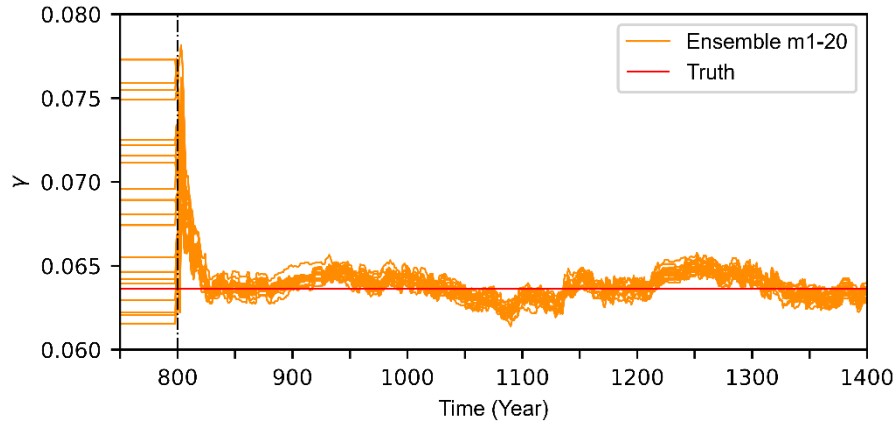


Figure S2. Time series of the estimated γ values with a Gaussian perturbation that has a mean value of 0.070004 and a standard deviation of 10% times the standard value in the individual ensemble members (orange) in the free model control ensemble simulations with data assimilation and parameter estimation. The solid red line denoting $\gamma = 0.06364$ marks the true value of γ being estimated. The dotted-dashed black line denoting 800th-year marks the start of parameter estimation using the observations of the atmosphere states and the temperature and salinity of the surface ocean.

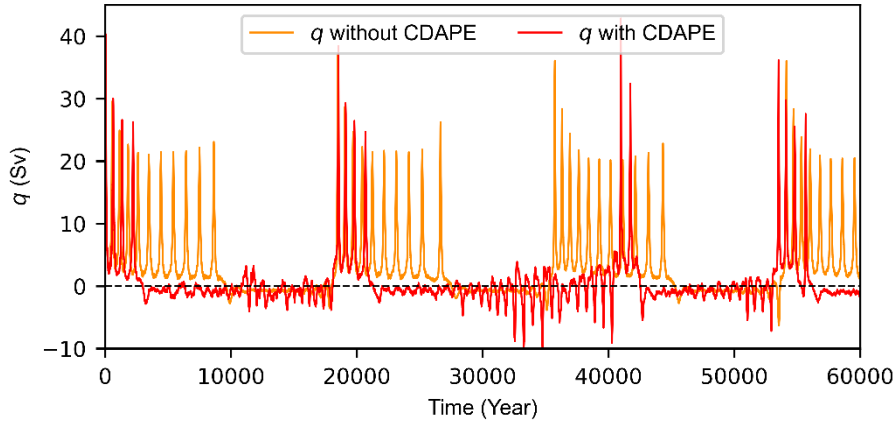


Figure S3. Time series of the meridional overturning circulation q in one of the ensemble members in the free model control ensemble simulations with (red) or without (orange) data assimilation and parameter estimation, where γ is erroneously guessed as 0.067419. The dashed black line denoting $q = 0$ is a division line between the thermal mode equilibrium state and the haline mode equilibrium state.

The two systems dealt with here are highly parameterized, but the parameter the authors chose to estimate, in both cases, was a parameter in the atmospheric model, equations (11) and (16). Why didn't they choose a parameter in the box models?

RE: Thanks for your suggestion. We have re-chosen the parameter to be estimated in the second system. The parameter b in the atmospheric model is replaced by the parameter γ in the box model. Following this suggestion, the first paragraph in Section 4.2 has been rewritten (L595-598), and Fig. 9 has been replaced by a new figure (L603-608). The new results are similar to the original results for parameter b . Both can demonstrate that the AMOC regime transitions can be captured by constraining the model parameters with observations.

Besides, the discussion about parameter selection has been added in L598-602. Considering the length limitation, the process of investigating the model sensitivity is not shown in the manuscript. However, the relevant literature has been added in L601-602. The process of investigating the model sensitivity will be described later, and we first show the results in Figure S4. The most sensitive parameter T_{A2} directly affects the temperature in the box model, so it is more appropriate to choose the parameter γ in comparison.

L595-598:

In the twin experimental framework, the assimilation model is similar to the truth model except that parameter γ in the box model is assumed to be incorrectly estimated, with error that is 10% greater than the standard value 0.06364. Thus, the mean value of all parameters from the twenty assimilation models is 0.070004 and their standard deviation is 10% of the standard value.

L603-608: (The same as Fig. S1).

L598-602:

The parameters in the atmosphere model or in the ocean model could be selected for parameter estimation to address the points in this paper. Given that we have experimented with the parameter in the atmosphere model before, here we show the experiment with the parameter of the ocean model. Again, the parameter being estimated is based on the model sensitivities regarding all parameters in the box model (Zhao et al., 2019).

L601-602:

Again, the parameter being estimated is based on the model sensitivities regarding all parameters in the box model (Zhao et al., 2019).

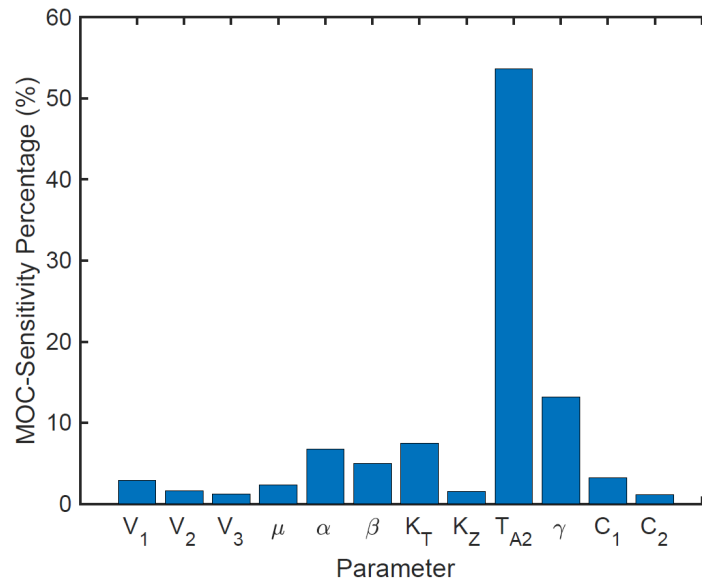


Figure S4. The sensitivity percentage of MOC in the space of the parameters. The sensitivity percentage refers to the ratio of the specified parameter sensitivity to the total sum of the sensitivities for the 12 parameters in the box model. The time-averaged value of the MOC spread is calculated by data over the last 200 years.

Referring to Section 3.2.3 in Zhao et al. (2019), the process of investigating the model sensitivity regarding each parameter is as follows. Each parameter is individually perturbed into 20 members from the ensemble as white noise, with a standard deviation of 10% of its standard value, while the remaining 11 parameters remain fixed at their standard values. Then, from the same initial states, all 20 ensemble members are freely run for 250 years. Only the model outputs of the last 200 years are used to quantitatively calculate the relative sensitivities (Owing to the MOC value is a long-time-scale variable). We use the standard deviation in the MOC strength to evaluate the model sensitivity regarding each parameter. The time-averaged sensitivity percentages of the MOC strength with respect to all parameters are shown in Fig. S4.

References

- Ashkenazy, Y. and Tziperman, E.: A wind-induced thermohaline circulation hysteresis and millennial variability regimes, *J. Phys. Oceanogr.*, 37, 2446–2457, <https://doi.org/10.1175/JPO3124.1>, 2007.
- Birchfield, G. E.: A coupled ocean-atmosphere climate model: temperature versus salinity effects on the thermohaline circulation, *Clim. Dynam.*, 4, 57–71, <https://doi.org/10.1007/BF00207400>, 1989.
- Bisaillon, P., Sandhu, R., Khalil, M., Pettit, C., Poirel, D., and Sarkar, A.: Bayesian parameter estimation and model selection for strongly nonlinear dynamical systems, *Nonlinear Dynam.*, 82, 1061–1080, <https://doi.org/10.1007/s11071-015-2217-8>, 2015.
- Brown, N. and Galbraith, E. D.: Hosed vs. unhosed: interruptions of the Atlantic Meridional Overturning Circulation in a global coupled model, with and without freshwater forcing, *Clim. Past*, 12, 1663–1679, <https://doi.org/10.5194/cp-12-1663-2016>, 2016.

- Carrassi, A., Bocquet, M., Bertino, L., and Evensen, G.: Data assimilation in the geosciences: An overview of methods, issues, and perspectives, *Wires. Clim. Change*, 9, e535, <https://doi.org/10.1002/wcc.535>, 2018.
- Castellana, D., Baars, S., Wubs, F. W., and Dijkstra, H. A.: Transition probabilities of noise-induced transitions of the Atlantic Ocean circulation, *Sci. Rep.*, 9, 20284, <https://doi.org/10.1038/s41598-019-56435-6>, 2019.
- Cessi, P.: A simple box model of stochastically forced thermohaline flow, *J. Phys. Oceanogr.*, 24, 1911–1920, [https://doi.org/10.1175/1520-0485\(1994\)024<1911:ASBMOS>2.0.CO;2](https://doi.org/10.1175/1520-0485(1994)024<1911:ASBMOS>2.0.CO;2), 1994.
- Gordon, N. J., Salmond, D. J., and Smith, A. F. M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *Iee Proc. F*, 140, 107–113, <https://doi.org/10.1049/ip-f-2.1993.0015>, 1993.
- Gottwald, G. A.: A model for Dansgaard–Oeschger events and millennial-scale abrupt climate change without external forcing, *Clim. Dynam.*, 56, 227–243, <https://doi.org/10.1007/s00382-020-05476-z>, 2021.
- Guan, Y. P. and Huang, R. X.: Stommel’s box model of thermohaline circulation revisited—The role of mechanical energy supporting mixing and the wind-driven gyration, *J. Phys. Oceanogr.*, 38, 909–917, <https://doi.org/10.1175/2007jpo3535.1>, 2008.
- Khalil, M., Sarkar, A., and Adhikari, S.: Nonlinear filters for chaotic oscillatory systems, *Nonlinear Dynam.*, 55, 113–137, <https://doi.org/10.1007/s11071-008-9349-z>, 2009.
- Kleppin, H., Jochum, M., Otto-Bliesner, B., Shields, C. A., and Yeager, S.: Stochastic atmospheric forcing as a cause of Greenland climate transitions, *J. Climate*, 28, 7741–7763, <https://doi.org/10.1175/JCLI-D-14-00728.1>, 2015.
- van Leeuwen, P. J.: Particle filtering in geophysical systems, *Mon. Weather Rev.*, 137, 4089–4114, <https://doi.org/10.1175/2009mwr2835.1>, 2009.
- Miller, R. N., Ghil, M., and Gauthiez, F.: Advanced data assimilation in strongly nonlinear dynamical systems, *J. Atmos. Sci.*, 51, 1037–1056, [https://doi.org/10.1175/1520-0469\(1994\)051<1037:ADAISN>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<1037:ADAISN>2.0.CO;2), 1994.
- Miller, R. N., Carter, E. F., and Blue, S. T.: Data assimilation into nonlinear stochastic models, *Tellus A*, 51, 167–194, <https://doi.org/10.3402/tellusa.v51i2.12315>, 1999.
- Mitsui, T. and Crucifix, M.: Influence of external forcings on abrupt millennial-scale climate changes: a statistical modelling study, *Clim. Dynam.*, 48, 2729–2749, <https://doi.org/10.1007/s00382-016-3235-z>, 2017.
- Peltier, W. R. and Vettoretti, G.: Dansgaard-Oeschger oscillations predicted in a comprehensive model of glacial climate: A “kicked” salt oscillator in the Atlantic, *Geophys. Res. Lett.*, 41, 7306–7313, <https://doi.org/10.1002/2014GL061413>, 2014.
- Roebber, P. J.: Climate variability in a low-order coupled atmosphere-ocean model, *Tellus A*, 47, 473–494, <https://doi.org/10.3402/tellusa.v47i4.11534>, 1995.
- Sévellec, F. and Fedorov, A. V.: Millennial variability in an idealized ocean model: predicting the AMOC regime shifts, *J. Climate*, 27, 3551–3564, <https://doi.org/10.1175/JCLI-D-13-00450.1>, 2014.
- Shen, Y. and Guan, Y. P.: Feature of thermohaline circulation in two-layer conceptual model based on energy constraint, *Sci. China: Earth Sci.*, 58, 1397–1403, <https://doi.org/10.1007/s11430-015-5092-8>, 2015.
- Shen, Y., Guan, Y. P., Liang, C. J., and Chen, D. K.: A three-box model of thermohaline circulation under the energy constraint, *Chinese Phys. Lett.*, 28, 059201, <https://doi.org/10.1088/0256-307x/28/5/059201>, 2011.
- Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J.: Obstacles to high-dimensional particle filtering, *Mon. Weather Rev.*, 136, 4629–4640, <https://doi.org/10.1175/2008mwr2529.1>, 2008.

- Tardif, R., Hakim, G. J., and Snyder, C.: Coupled atmosphere–ocean data assimilation experiments with a low-order climate model, *Clim. Dynam.*, 43, 1631–1643, <https://doi.org/10.1007/s00382-013-1989-0>, 2014.
- Weijer, W., Cheng, W., Drijfhout, S. S., Fedorov, A. V., Hu, A., Jackson, L. C., Liu, W., McDonagh, E. L., Mecking, J. V., and Zhang, J.: Stability of the Atlantic Meridional Overturning Circulation: A review and synthesis, *J. Geophys. Res.: Oceans*, 124, 5336–5375, <https://doi.org/10.1029/2019JC015083>, 2019.
- Weir, B., Miller, R. N., and Spitz, Y. H.: A potential implicit particle method for high-dimensional systems, *Nonlinear Proc. Geoph.*, 20, 1047–1060, <https://doi.org/10.5194/npg-20-1047-2013>, 2013a.
- Weir, B., Miller, R. N., and Spitz, Y. H.: Implicit estimation of ecological model parameters, *B. Math. Biol.*, 75, 223–257, <https://doi.org/10.1007/s11538-012-9801-6>, 2013b.
- Zhang, S., Liu, Z., Rosati, A., and Delworth, T.: A study of enhance parameter correction with coupled data assimilation for climate estimation and prediction using a simple coupled model, *Tellus A*, 64, 10963, <https://doi.org/10.3402/tellusa.v64i0.10963>, 2012.
- Zhang, X., Lohmann, G., Knorr, G., and Purcell, C.: Abrupt glacial climate shifts controlled by ice sheet changes, *Nature*, 512, 290–294, <https://doi.org/10.1038/nature13592>, 2014.
- Zhao, Y., Deng, X., Zhang, S., Liu, Z., Liu, C., Vecchi, G., Han, G., and Wu, X.: Impact of an observational time window on coupled data assimilation: simulation with a simple climate model, *Nonlinear Proc. Geoph.*, 24, 681–694, <https://doi.org/10.5194/npg-24-681-2017>, 2017.
- Zhao, Y., Deng, X., Zhang, S., Liu, Z., and Liu, C.: Sensitivity determined simultaneous estimation of multiple parameters in coupled models: part I—based on single model component sensitivities, *Clim. Dynam.*, 53, 5349–5373, <https://doi.org/10.1007/s00382-019-04865-3>, 2019.

Response to Reviewer #2

Comments from Reviewer #2:

General Comments:

This paper addresses regime transition in AMOC through state and parameter estimation, via application of the EAKF. While the idea of parameter estimation is not new in ensemble-based data assimilation and there are many published papers addressing it, I believe that the novelty of the paper is the application of EAKF to regime transition in AMOC. Still, the paper needs to explain what the main findings are and how those findings could improve our knowledge about AMOC. In addition, I have some specific comments requiring clarification of the parameter estimation approach and the EAKF equations.

RE: A few conferences of all co-authors have been held to discuss the comments of reviewer #2. All authors appreciate greatly for the encouragements and comments. All the comments are very important and useful for authors to improve the quality of this manuscript. The paper is renewed as the reviewer's suggestions. Thanks for your encouragement. In this revision, a more detailed explanation has been added to the Introduction section. We have expressed better what the main findings of this paper are (L96-103, L664-675) and how these findings could improve our knowledge about AMOC (L128-133). All specific comments are replied point-by-point as below. We hope the whole manuscript has been essentially improved.

L96-103:

Observation-constrained model parameters no longer keep fixed values but are constantly varying over time. The purpose of this paper is to explore whether the variations of observation-constrained parameters that allow the physical processes of model to evolve over time can make the simulation results closer to the “observed” feature of regime transitions. The models in this paper are obtained by coupling AMOC box model with Lorenz’s model, similar to the work by Roebber (1995) or Gottwald (2021), where the variation of AMOC is driven by the chaotic dynamical system. The thermal mode and the reverse haline mode correspond to different equilibrium states of the AMOC. For simplicity, we will refer to these different states as the stronger AMOC (on-state) and the weaker AMOC (off-state) in simple conceptual models (e.g., Weijer et al., 2019).

L664-675:

Since the circulation is driven only by the meridional gradients of the upper ocean temperature and salinity in the buoyancy-constraint MOCBM model, AMOC regime transitions can be captured to some extent when the upper ocean temperature and salinity are directly adjusted by data assimilation only, but the simulation results are not accurate enough. In this simple model, since the data assimilation has worked well, the contribution of parameter estimation is relatively small but still indispensable. The AMOC regime transitions are captured more accurately by parameter estimation. The degree of contributions of data assimilation or parameter estimation to the optimization of simulation results is different in these two models. Compared with

the MOCBM model, the energy-constraint MOC3B-5V model is more representative for the role of parameter estimation because the circulation is maintained by mechanical energy. When leaving out the parameter estimation steps and constraining the model states only by data assimilation, the accuracy of state estimation is not high due to the existence of parameter errors. Given the fact that the circulation is driven in a more complex way in the real world, this simple model study only provides a conceptual understanding and guideline for more complex real systems such as Coupled General Circulation Model (CGCM).

L128-133:

Although the AMOC model eventually exhibits multiple equilibria, the AMOC is not a direct model state but is indirectly derived from model states such as atmospheric wind, ocean temperature and salinity. Instead of adjusting AMOC directly, the model states are adjusted through data assimilation. When constraining model parameters by observational information, the parameters that constantly vary with observations may provide more diversity in the physical processes involved with AMOC regime transition, so that the model can simulate more AMOC transition paths.

Specific Comments:

(1) Please explain how equations (1)-(4) were derived from the EAKF.

RE: Good suggestion! In this revision, a detailed derivation process is shown below. Given the length limitation, these derivations are not added to the manuscript. However, a concise description and the corresponding references have been added in L192-197.

L192-197:

The ensemble adjustment Kalman filter (Anderson, 2001) is used for data assimilation and parameter estimation in this study. The basic process of the two-step EAKF (Anderson, 2003; Zhang and Anderson, 2003; Zhang et al., 2007) is to project the observational increment onto model states (relevant parameters) by calculating the error covariance between the prior ensemble of the model variable (parameter) and the model-estimated ensemble. The core of the two-step EAKF is to calculate the increment of each state variable by a global least squares fit (linear regression), and the calculation of the observational increment is related to the scalar application of the equations of EAKF (Anderson, 2003).

Based on the EAKF (Anderson, 2001), Anderson (2003) described a two-step data assimilation procedure for ensemble filtering under a local least squares framework.

The joint state-observation space is defined by the joint space state vector: $\mathbf{z} = [\mathbf{x}, \mathbf{y}]$, where \mathbf{x} is the model state vector; $\mathbf{y} = h(\mathbf{x})$, where h is the forward observation operator. Using Bayesian statistics, the distribution of the posterior (or updated) distribution can be computed from the prior distribution, as

$$\mathbf{p}(\mathbf{z}^u) = \mathbf{p}(\mathbf{y}^o | \mathbf{z}^p) \mathbf{p}(\mathbf{z}^p) / (\text{norm}) \quad . \quad (\text{S1})$$

At the heart of the ensemble Kalman filter is the fact that the product of the joint prior Gaussian with mean $\bar{\mathbf{z}}^p$, covariance

Σ^p , and the Gaussian observation distribution with mean \mathbf{y}^o and error variance \mathbf{R} has covariance

$$\Sigma^u = [(\Sigma^p)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1} , \quad (\text{S2})$$

and mean

$$\bar{\mathbf{z}}^u = \Sigma^u [(\Sigma^p)^{-1} \bar{\mathbf{z}}^p + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y}^o] . \quad (\text{S3})$$

The EAKF constructs an updated ensemble with a mean and sample variance that satisfy Eq. (S2) and Eq. (S3). In Anderson (2001), this is done by shifting the mean of the ensemble and then adjusting the spread of the ensemble around the updated mean using a linear operator \mathbf{A} :

$$\mathbf{z}_i^u = \mathbf{A}(\mathbf{z}_i^p - \bar{\mathbf{z}}^p) + \bar{\mathbf{z}}^u \quad (\text{S4})$$

where \mathbf{A} satisfies $\Sigma^u = \mathbf{A} \Sigma^p \mathbf{A}^T$.

In everything that follows, results are presented only for assimilation of a single scalar observation. Define the joint state space forward observation operator for a single observation as the order $1 \times k$ linear operator $\mathbf{H} = [0, 0, \dots, 0, 1]$, where k is the joint state space size. The updated probability for the marginal distribution of the observation joint state variable y can be formed: $p_y(y^u) = p(y^o | y^p) p_y(y^p) / (\text{norm})$, where the subscript on the probability densities indicates a marginal probability on the observation variable. Note that this equation does not depend on any of the model state variables. This suggests a partitioning of the assimilation of an observation into two parts. The first determines updated ensemble members for the observation variable given the observation. To update the ensemble sample of y^p , an increment is computed for each ensemble member:

$$\Delta y_i = y_i^u - y_i^p , \quad (\text{S5})$$

where $i = 1, \dots, N$, and N is the ensemble size. The second step computes corresponding increments for i -th ensemble sample of j -th state variable $\Delta x_{i,j}$. This requires assumptions that the prior distribution is Gaussian. This equivalent to assuming that a least squares fit to the prior ensemble members summarizes the relationship between the joint state variables.

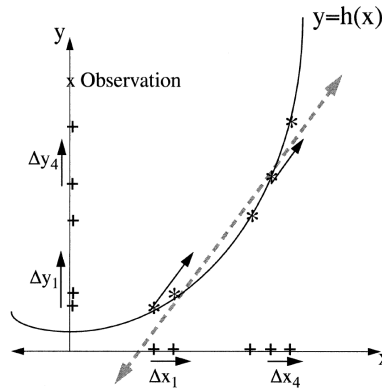


Figure S1. An idealized representation showing the relation between update increments for a state variable x and an observation variable y for a five member ensemble represented by asterisks. The projection of the ensemble on the x and y axes is represented by a plus sign and the observation y^o is represented by “*”. The gray dashed line shows a global least squares fit to the ensemble members. Update increments for ensemble members 1 and 4 for y are shown along with corresponding increments for the ensemble as a whole (thin vectors parallel to least squares fit) and for the x ensemble. From Anderson (2003)

Figure S1 depicts the simplest example in which there is only a single state variable x . The observation variable y is related to x by the operator h , which is nonlinear in the figure. Increments for each ensemble sample of y have been computed. The corresponding increments for x are then computed by a global least squares fit (linear regression) so that

$$\Delta x_i = \frac{\text{cov}(x_i, y_i)}{(\sigma^p)^2} \Delta y_i, \quad (\text{S6})$$

where $\text{cov}(x_i, y_i)$ is the prior covariance of x with y , $(\sigma^p)^2$ is the prior variance of y .

Equation (S1) implies that the observation variable can be updated independently of the other joint state variables. Using a scalar application of Eq. (S2), the updated variance for y can be written

$$(\sigma^u)^2 = [((\sigma^p)^2)^{-1} + ((\sigma^o)^2)^{-1}]^{-1} = \frac{(\sigma^o)^2(\sigma^p)^2}{(\sigma^o)^2 + (\sigma^p)^2}. \quad (\text{S7})$$

Applying a scalar version of Eq. (S3) to compute the updated mean

$$\begin{aligned} \bar{y}^u &= (\sigma^u)^2 [((\sigma^p)^2)^{-1} \bar{y}^p + ((\sigma^o)^2)^{-1} y^o] \\ &= \frac{(\sigma^o)^2}{(\sigma^o)^2 + (\sigma^p)^2} \bar{y}^p + \frac{(\sigma^p)^2}{(\sigma^o)^2 + (\sigma^p)^2} y^o. \end{aligned} \quad (\text{S8})$$

Using a scalar application of Eq. (S4), the updated value of y can be written

$$y_i^u = \alpha (y_i^p - \bar{y}^p) + \bar{y}^u, \quad (\text{S9})$$

$$\text{where } \alpha = \sqrt{(\sigma^u)^2 / (\sigma^p)^2} = \sqrt{\frac{(\sigma^o)^2}{(\sigma^o)^2 + (\sigma^p)^2}}.$$

Hence, the observational increment from Eq. (S5) can be written

$$\begin{aligned} \Delta y_i &= y_i^u - y_i^p = \alpha (y_i^p - \bar{y}^p) + \bar{y}^u - y_i^p \\ &= \sqrt{\frac{(\sigma^o)^2}{(\sigma^o)^2 + (\sigma^p)^2}} (y_i^p - \bar{y}^p) + \frac{(\sigma^o)^2}{(\sigma^o)^2 + (\sigma^p)^2} \bar{y}^p + \frac{(\sigma^p)^2}{(\sigma^o)^2 + (\sigma^p)^2} y^o - y_i^p. \end{aligned} \quad (\text{S10})$$

For more details, please see Sect. 2c and 3b in Anderson (2003).

Also, please explain exact meaning of each variable. For example, is $\Delta \beta^p$ a value of the parameter β , or prior ensemble perturbation of the parameter β ? Is Δy_i^p prior ensemble spread or prior ensemble perturbation? In addition, all vectors and matrices need to have clearly defined space (e.g., observation or state space) and dimensions (e.g., Nobs, Nstate, Nens, Nobs x Nens, ...).

RE: Excellent suggestion! To fully address this comment, we have substantially rewritten Section 2.2. The meaning of each variable has been explained more exactly, and the space and dimensions of all vectors and matrices have been clearly defined. We believe that the new Section 2.2 has been considerably improved (L197-220).

L197-220:

All observations at time t have the observation value \mathbf{y}^o (in N_{obs} dimensions). For a single observation y_k^o at the k -th observation location ($k = 1 \sim N_{obs}$), the standard deviation of observational error is σ^o (assumed to be Gaussian). The model states are mapped onto the observational space by applying a linear interpolation, and then the prior (model-estimated) ensemble of the k -th observation \mathbf{y}_k^p (in N_{ens} dimensions) can be obtained. $y_{k,i}^p$ is the i -th prior ensemble member of the k -th observation. The ensemble mean and standard deviation are \bar{y}_k^p and σ_k^p , respectively.

The first step is to compute the observational increment of the k -th observation ($k = 1 \sim N_{obs}$). The observational increment $\Delta y_{k,i}^o$ for the i -th ensemble member ($i = 1 \sim N_{ens}$) is formulated by

$$\Delta y_{k,i}^o = \bar{y}_k^u + \Delta y'_{k,i} - y_{k,i}^p, \quad (1)$$

where \bar{y}_k^u is the posterior ensemble mean of the k -th observation, representing the shift of the ensemble mean induced by this observation, $\Delta y'_{k,i}$ is the updated ensemble spread of the k -th observation, representing the reshaping of the model ensemble. They are respectively computed by

$$\bar{y}_k^u = \frac{(\sigma^o)^2}{(\sigma^o)^2 + (\sigma_k^p)^2} \bar{y}_k^p + \frac{(\sigma_k^p)^2}{(\sigma^o)^2 + (\sigma_k^p)^2} y_k^o, \text{ and}$$

$$\Delta y'_{k,i} = \sqrt{\frac{(\sigma^o)^2}{(\sigma^o)^2 + (\sigma_k^p)^2}} (y_{k,i}^p - \bar{y}_k^p), \quad (2)$$

where the first equation shows that the ensemble mean shifts toward the prior model ensemble mean \bar{y}_k^p or the observation value y_k^o , and whether it is \bar{y}_k^p or y_k^o depends on who has the smaller variance. The second equation denotes that the prior probability density function is squashed by a new observation.

The second step is to distribute the observational increments $\Delta y_{k,i}^o$ on to the related model states \mathbf{x} (a matrix of size $N_{ens} \times N_{state}$) and this assimilation process can be expressed as

$$\Delta x_{j,i} = \frac{cov(x_j, y_k^p)}{(\sigma_k^p)^2} \Delta y_{k,i}^o, \quad (3)$$

where $\Delta x_{i,j}$ is the contribution of the k -th observation to the i -th ensemble member of the j -th model variable $x_{j,i}$ ($j = 1 \sim N_{state}$). $cov(\mathbf{x}_j, \mathbf{y}_k^p)$ is the error covariance between the prior ensemble of the j -th model variable \mathbf{x}_j (in N_{ens} dimensions) and the prior (model-estimated) ensemble of the k -th observation \mathbf{y}_k^p (in N_{ens} dimensions), and is calculated as

$$cov(\mathbf{x}_j, \mathbf{y}_k^p) = \frac{\sum_{i=1}^{N_{ens}} (x_{j,i} - \bar{x}_j)(y_{k,i}^p - \bar{y}_k^p)}{N_{ens}}, \text{ where } \bar{x}_j \text{ is the ensemble mean of } j\text{-th model variable.}$$

(2) What dynamical model was used to propagate parameters in time? Was it identity model?

RE: Thanks for this comment. There is no dynamical model used to propagate parameters in this paper and it is identity model. Addressing this comment, we have added more description and discussion about the parameter estimation in L96-97 and L221-231.

L96-97:

Observation-constrained model parameters no longer keep fixed values but are constantly varying over time.

L221-231:

The model parameters are fixed when parameter estimation is not performed. The parameters vary with observational information by parameter estimation. The core of the parameter estimation is to obtain the increment of the estimated parameter by a linear regression that is based on the error covariance between the prior parameter ensemble and the state ensemble (Anderson, 2001, 2003). The error covariance used in regression is flow dependent and temporally varying (Zhang and Anderson, 2003). Therefore, for the model parameter estimation, the observational increments are distributed onto a relevant parameter and the equation is

$$\Delta\beta_{j,i} = \frac{cov(\beta_j, \mathbf{y}_k^p)}{(\sigma_k^p)^2} \Delta y_{k,i}^o, \quad (4)$$

where $\Delta\beta_{j,i}$ is the contribution of the k -th observation to the i -th ensemble member of the j -th parameter being estimated, called $\beta_{j,i}$ ($j = 1 \sim N_{para}$). $cov(\beta_j, \mathbf{y}_k^p)$ is the error covariance between the prior ensemble of the j -th model parameter β_j (in N_{ens} dimensions) and the prior (model-estimated) ensemble of the k -th observation \mathbf{y}_k^p (in N_{ens} dimensions), and is calculated as $cov(\beta_j, \mathbf{y}_k^p) = \frac{\sum_{i=1}^{N_{ens}} (\beta_{j,i} - \bar{\beta}_j)(y_{k,i}^p - \bar{y}_k^p)}{N_{ens}}$, where $\bar{\beta}_j$ is the ensemble mean of j -th model parameter being optimized.

Is there a more suitable model than identity? Please provide a brief reference review about different models used so far in literature and justify your choice.

RE: Good suggestion! We have added a brief review and discussion about different models in L715-722.

L715-722:

Aksoy et al. (2006b) proposed a spatial updating technique that recovers the globally uniform parameter value using a spatial average of the entire spatially varying parameter field. Wu et al. (2012, 2013) explored the impact of the geographic dependence of observing system on the parameters. The adjustment of the parameters is based on the spatial distribution of the model state sensitivity to parameters. Liu et al. (2014a, b) proposed the adaptive spatial average method that obtains the final global uniform posterior parameter based on spatially varying posterior estimated parameter values. In this study, considering that the simple box models are used as a first step to explore AMOC transitions, it is more appropriate to use the identity model. The impact of geographic-dependent parameter optimization on climate estimation and prediction can be considered in future studies for complex systems such as CGCMs.

(3) Have you applied any covariance inflation? Please explain.

RE: Thanks for this suggestion! In this revision, a more detailed description of the covariance inflation scheme has been added in L239-246.

L239-246:

To further improve the signal-to-noise ratio of parameter estimation, Zhang (2011a) introduced an inflation scheme based on model sensitivity with respect to the parameter being estimated. In this inflation scheme, the inflation amplitude of a parameter ensemble is inversely proportional to the sensitivity. It is formulated as $\tilde{\beta}_{j,i} = \bar{\beta}_j + \max\left(1, \frac{\alpha_0 \sigma_0}{\sigma_j \sigma_t}\right) (\beta_{j,i} - \bar{\beta}_j)$, where $\tilde{\beta}_{j,i}$ denotes the inflated version of the i -th ensemble member of the j -th parameter being estimated, σ_0 and σ_t are the prior ensemble spreads of this parameter at the initial time and time t , α_0 is a constant tuned by a trial-and-error procedure (e.g., Wu et al., 2016), and σ_j is the sensitivity of the model state with regard to j -th parameter. This indicates that if the prior ensemble spread of j -th parameter is smaller than $\frac{\alpha_0}{\sigma_j}$ times the initial spread, it will be enlarged to this amount (e.g., Wu et al., 2012; Han et al., 2014; Zhao et al., 2019).

(4) Please address the issue of ensemble spread vs. forecast skill. Are they correlated in your experiments? Is ensemble spread over-estimated or under-estimated? Are ensembles collapsing?

RE: Thanks for this comment. Ensemble spread and forecast skill are correlated in our experiments. There are two considerations regarding the ensemble spread of the parameter. As shown in the following figure, the ensemble spread is affected by the inflation level. On the one hand, overly small inflation factors and small ensemble spread will lead to too long fluctuation period of ensemble members, which affects the forecast skill and causes errors. Insufficient inflation may result in ensemble collapsing. On the other hand, overly large inflation factors lead the spread to jump out of the reasonable range. Considering the length limitation, the experiments on the ensemble spread and inflation factors are not shown in the manuscript. However, the relevant literature has been added in L244. Following this suggestion, Fig. 6 has been replaced by a new figure (Fig. S2d) in the revised manuscript. Accordingly, Figs. 5 and 7 have been updated due to the adjustment of the inflation scheme, and the new figures are almost identical to the original ones.

L244:

α_0 is a constant tuned by a trial-and-error procedure (e.g., Wu et al., 2016)

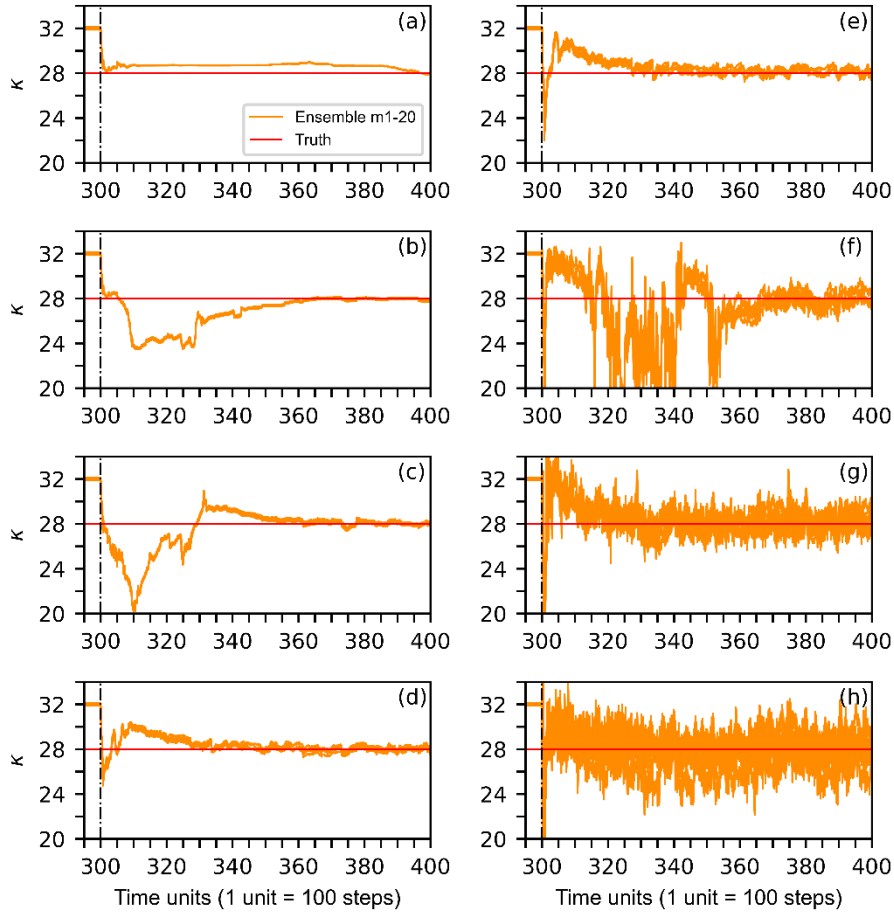


Figure S2. Time series of the estimated κ values with a Gaussian perturbation that has a mean value of 32 and a standard deviation of 0.1 in the individual ensemble members (orange) in the free model control ensemble simulations with data assimilation and parameter estimation. The solid red line denoting $\kappa = 28$ marks the true value of κ being estimated. The dotted-dashed black line denoting 300th-unit marks the start of parameter estimation using x_2 , w , and η observations. The limited inflation value is 0.01 (a), 0.05 (b), 0.10 (c), 0.15 (d), 0.20 (e), 0.40 (f), 0.60 (g), 1.0 (h) and the value used in the paper is 0.15.

References

- Aksoy, A., Zhang, F., and Nielsen-Gammon, J. W.: Ensemble-based simultaneous state and parameter estimation with MM5, *Geophys. Res. Lett.*, 33, L12801, <https://doi.org/10.1029/2006GL026186>, 2006.
- Anderson, J. L.: An ensemble adjustment Kalman filter for data assimilation, *Mon. Weather Rev.*, 129, 2884–2903, [https://doi.org/10.1175/1520-0493\(2001\)129<2884:Aeakff>2.0.Co;2](https://doi.org/10.1175/1520-0493(2001)129<2884:Aeakff>2.0.Co;2), 2001.
- Anderson, J. L.: A local least squares framework for ensemble filtering, *Mon. Weather Rev.*, 131, 634–642, [https://doi.org/10.1175/1520-0493\(2003\)131<0634:Allsff>2.0.Co;2](https://doi.org/10.1175/1520-0493(2003)131<0634:Allsff>2.0.Co;2), 2003.
- Gottwald, G. A.: A model for Dansgaard–Oeschger events and millennial-scale abrupt climate change without external forcing, *Clim. Dynam.*, 56, 227–243, <https://doi.org/10.1007/s00382-020-05476-z>, 2021.
- Han, G. J., Zhang, X. F., Zhang, S., Wu, X. R., and Liu, Z.: Mitigation of coupled model biases induced by dynamical core misfitting through parameter optimization: simulation with a simple pycnocline prediction model, *Nonlinear Proc. Geoph.*, 21, 357–366, <https://doi.org/10.5194/npg-21-357-2014>, 2014.
- Liu, Y., Liu, Z., Zhang, S., Rong, X., Jacob, R., Wu, S., and Lu, F.: Ensemble-based parameter estimation in a coupled GCM using the adaptive spatial average method, *J. Climate*, 27, 4002–4014, <https://doi.org/10.1175/JCLI-D-13-00091.1>, 2014a.
- Liu, Y., Liu, Z., Zhang, S., Jacob, R., Lu, F., Rong, X., and Wu, S.: Ensemble-based parameter estimation in a coupled general circulation model, *J. Climate*, 27, 7151–7162, <https://doi.org/10.1175/jcli-d-13-00406.1>, 2014b.
- Roebber, P. J.: Climate variability in a low-order coupled atmosphere-ocean model, *Tellus A*, 47, 473–494, <https://doi.org/10.3402/tellusa.v47i4.11534>, 1995.
- Weijer, W., Cheng, W., Drijfhout, S. S., Fedorov, A. V., Hu, A., Jackson, L. C., Liu, W., McDonagh, E. L., Mecking, J. V., and Zhang, J.: Stability of the Atlantic Meridional Overturning Circulation: A review and synthesis, *J. Geophys. Res.: Oceans*, 124, 5336–5375, <https://doi.org/10.1029/2019JC015083>, 2019.
- Wu, X., Zhang, S., Liu, Z., Rosati, A., Delworth, T. L., and Liu, Y.: Impact of geographic-dependent parameter optimization on climate estimation and prediction: simulation with an intermediate coupled model, *Mon. Weather Rev.*, 140, 3956–3971, <https://doi.org/10.1175/MWR-D-11-00298.1>, 2012.
- Wu, X., Zhang, S., Liu, Z., Rosati, A., and Delworth, T. L.: A study of impact of the geographic dependence of observing system on parameter estimation with an intermediate coupled model, *Clim. Dynam.*, 40, 1789–1798, <https://doi.org/10.1007/s00382-012-1385-1>, 2013.
- Wu, X., Han, G., Zhang, S., and Liu, Z.: A study of the impact of parameter optimization on ENSO predictability with an intermediate coupled model, *Clim. Dynam.*, 46, 711–727, <https://doi.org/10.1007/s00382-015-2608-z>, 2016.
- Zhang, S.: A study of impacts of coupled model initial shocks and state-parameter optimization on climate predictions using a simple pycnocline prediction model, *J. Climate*, 24, 6210–6226, <https://doi.org/10.1175/jcli-d-10-05003.1>, 2011.
- Zhang, S., Harrison, M. J., Rosati, A., and Wittenberg, A. T.: System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies, *Mon. Weather Rev.*, 135, 3541–3564, <https://doi.org/10.1175/mwr3466.1>, 2007.
- Zhao, Y., Deng, X., Zhang, S., Liu, Z., and Liu, C.: Sensitivity determined simultaneous estimation of multiple parameters in coupled models: part I—based on single model component sensitivities, *Clim. Dynam.*, 53, 5349–5373, <https://doi.org/10.1007/s00382-019-04865-3>, 2019.