Response to Reviewer #2

Comments from Reviewer #2:

General Comments:

This paper addresses regime transition in AMOC though 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

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

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

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

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

and mean

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

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 $\mathbf{A}$:

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

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

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: $p_y(y^u) = p(y^o|y^p)p_y(y^p)/(\mathrm{norm})$, 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:

$$\Delta y_i = y_i^u - y_i^p \ , \tag{S5}$$

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.

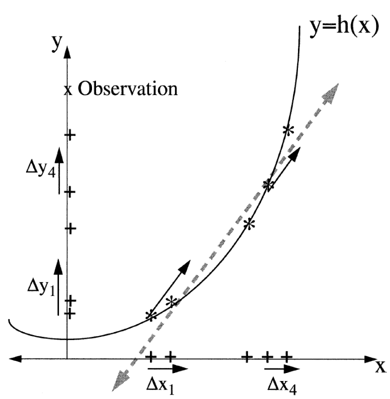

Figure S1. An idealized representation showing the relation between update increments for a state variable x and an observation variable y for a five member ensemble represented by asterisks. The projection of the ensemble on the x and y axes is represented by a plus sign and the observation y^o is represented 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

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

where $cov(x_i, y_i)$ 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

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

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

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

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

$$y_i^u = \alpha\left(y_i^p - \bar{y}^p\right) + \bar{y}^u \ , \tag{S9}$$

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

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

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 $\boldsymbol{y^o}$ (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 $\boldsymbol{y_k^p}$ (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

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

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

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

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

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 $\boldsymbol{x}$ (a matrix of size $N_{ens} \times N_{state}$) and this assimilation process can be expressed as

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

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 $x_{j,i}$ ($j = 1 \sim N_{state}$). $cov(\boldsymbol{x_j}, \boldsymbol{y_k^p})$ is the error covariance between the prior ensemble of the j-th model variable $\boldsymbol{x_j}$ (in N_{ens} dimensions) and the prior (model-estimated) ensemble of the k-th observation $\boldsymbol{y_k^p}$ (in N_{ens} dimensions), and is calculated as

$cov(\boldsymbol{x_j}, \boldsymbol{y_k^p}) = \frac{\sum_{i=1}^{N_{ens}} (x_{j,i} - \bar{x}_j)(y_{k,i}^p - \bar{y}_k^p)}{N_{ens}}$, where $\bar{x}_j$ is the ensemble mean of j-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

$$\Delta\beta_{j,i} = \frac{cov(\boldsymbol{\beta_j}, \boldsymbol{y_k^p})}{(\sigma_k^p)^2} \Delta y_{k,i}^o \ , \tag{4}$$

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 $\beta_{j,i}$ $(j = 1 \sim N_{para})$. $cov(\boldsymbol{\beta_j}, \boldsymbol{y_k^p})$ is the error covariance between the prior ensemble of the *j*-th model parameter $\boldsymbol{\beta_j}$ (in N_{ens} dimensions) and the prior (model-estimated) ensemble of the *k*-th observation $\boldsymbol{y_k^p}$ (in N_{ens} dimensions), and is calculated as $cov(\boldsymbol{\beta_j}, \boldsymbol{y_k^p}) = \frac{\sum_{i=1}^{N_{ens}}(\beta_{j,i} - \bar{\beta}_j)(y_{k,i}^p - \bar{y}_k^p)}{N_{ens}}$, 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 $\tilde{\beta}_{j,i} = \bar{\beta}_j + max\left(1, \frac{\alpha_0\sigma_0}{\sigma_j\sigma_t}\right)\left(\beta_{j,i} - \bar{\beta}_j\right)$, 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)

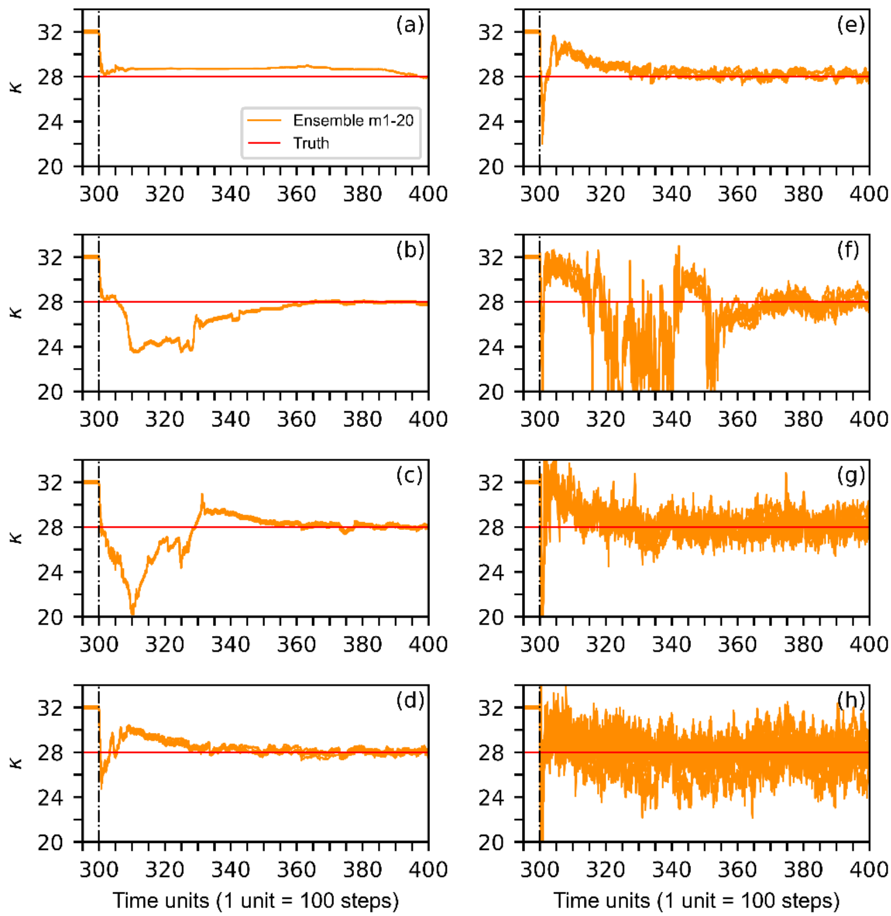

Figure S2. Time series of the estimated κ values with a Gaussian perturbation that has a mean value of 32 and a standard deviation of 0.1 in the individual ensemble members (orange) in the free model control ensemble simulations with data assimilation and parameter estimation. The solid red line denoting κ = 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, 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, 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.