https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

Parametric Uncertainty in Prediction of AMOC weakening by an Earth System Model

Amane Kubo[1], Yohei Sawada[1]

[1]Department of Civil Engineering, The University of Tokyo, Tokyo, Japan

Correspondence to: Amane Kubo (amanetf0323@g.ecc.u-tokyo.ac.jp)

Abstract. The Atlantic Meridional Overturning Circulation (AMOC) is a critical tipping element in the Earth system. While parametric uncertainty in climate models represents a dominant source of uncertainty in future projections, particularly regarding non-linear tipping phenomena, it has been largely ignored or overlooked in previous studies based on the CMIP6 model ensemble which have predominantly focused on structural uncertainties. To address this fundamental gap, this study investigates AMOC predictability by performing an uncertainty quantification (UQ) that explicitly accounts for parametric uncertainty. Crucially, this work represents the first application of an Observing System Simulation Experiment (OSSE) framework to an Earth System Model (ESM)-based AMOC predictability study under parametric uncertainty. Utilizing the Earth system model of intermediate complexity, LOVECLIM, we conduct an OSSE by varying 25 key uncertain parameters under a freshwater hosing scenario. To mitigate the prohibitive computational cost of coupled model simulations, we construct a high-fidelity surrogate model combining Principal Component Analysis (PCA) and Gaussian Process Regression (GPR) and realize efficient Surrogate-model Uncertainty Quantification, which effectively utilizes Perturbed Parameter Ensembles (PPEs). Our pseudo-observation experiments demonstrate that assimilating contemporary Sea Surface Salinity (SSS) data effectively constrains critical parameters governing freshwater transport, showing the best efficacy in reducing future AMOC projection uncertainty. Even though Sea Surface Temperature (SST) or other atmospheric observations reduce parametric and simulation accuracy in the observation period, the current AMOC and future AMOC prediction uncertainty is not reduced very well.

Short Summary

The Atlantic Ocean's current system stabilizes global climate, but scientists fear it could collapse. Climate assessments often overlook how model uncertainties, especially uncertainty in the model parameters, affect collapse predictions. This study addresses this gap and provides a framework for assessing the effects from the uncertainty in the model parameters. Findings show that sea surface salinity data reduces prediction uncertainty regarding how fast the current will slow down, whereas sea surface temperature data does not. This research guides future ocean monitoring priorities to better foresee and prepare for major climate risks.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

1. Introduction

Climate tipping is defined as a sudden, significant, and irreversible change in the Earth system. The Atlantic Meridional Overturning Circulation (AMOC), a deep ocean circulation that transports heat to higher latitudes, is recognized as a key tipping element (Lenton et al., 2008; Armstrong McKay et al., 2022). The AMOC plays a crucial role in transporting heat from the lower latitudes and the Southern Hemisphere to the high latitudes of the Northern Hemisphere (Buckley & Marshall, 2016; Jackson et al., 2015), thereby moderating the European climate. Therefore, preventing AMOC tipping is vital for avoiding radical changes in the current social system (Srokosz & Bryden, 2015).

Although we have evidence of past AMOC transitions from paleoclimatic reconstructions, which suggest an oscillation between on and off states during the Dansgaard-Oeschger events (Ganopolski & Rahmstorf, 2001; Vettoretti et al., 2023), our direct observational records, including RAPID transect (26N) since 2004 (Johns et al., 2023), the SAMBA transect (34.5S) since 2009 (Chidichimo et al., 2023), and the OSNAP transect (53N to 60N) since 2014 (Lozier, 2023), still cover only recent several decades and does not include direct record of its tipping. This raises a question whether climate tipping or its risk is foreseeable without any directs records of climate tipping in the past.

Earth System Models (ESMs) are powerful tools to predict future climate conditions including AMOC tipping. Community Earth System Model (CESM), which is one of the full-scale ESMs, showed climate tipping in AMOC (van Westen et al., 2024). However, ESM-based climate projections are yet to be uncertain especially about extreme events (Bevacqua et al., 2026) and longer-term AMOC stability (Bellomo et al., 2021). It is of paramount importance to reduce the uncertainty in ESMs' climate predictions, including those of AMOC tipping, leveraging observations to be integrated into ESMs.

Emergent Constraint (EC) has been widely adopted to constrain the multi-model climate change projections such as the CMIP6 ensemble using observation. Examples of applications cover wide range of Earth systems, such as carbon budget (Cox et al., 2024), hydrological cycles (Simpson et al., 2021, Shiogama et al., 2025) and biochemical cycles (Terhaar et al., 2020). EC also contributed to the reduction of uncertainty in AMOC weakening simulated by the CMIP6 model ensemble. Bonan et al., 2025 found an emergent relationship among CMIP6 model ensemble simulations that the AMOC strength approximately follows thermal wind relationship and reduces CMIP6 model output uncertainty to obtain more constrained future AMOC prediction.

It should be noted that EC focuses on model structural uncertainty by analysing the existing output from multi-model ensemble, while we focus on parametric uncertainty in a single model. Since many small-scale phenomena are parameterized and approximated in ESMs, parametric uncertainty is one of the major sources of uncertainty in ESM-based projections. Kubo and Sawada., 2025 performed parametric uncertainty quantification using conceptual and low-dimensional climate tipping models

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

and found that observed internal variability is helpful to constrain unknown parameters and the simulated long-term behavior of AMOC. They demonstrated the effectiveness and importance of Observing System Simulation Experiments (OSSEs) to assess the predictability of future climate tipping (Dijkstra et al., 2026). In this paper, we would extend Kubo and Sawada, 2025 to the AMOC prediction with a coupled ESM.

Several methods for ESM parametric uncertainty quantification have been developed. A surrogate-model Uncertainty Quantification method is one of the major methods for parameter uncertainty quantification of ESMs. In this method, a machine learning-based surrogate model is adopted to enable posterior sampling with computationally demanding ESMs (Dunbar et al., 2021, Yarger et al., 2024). This method does not require the derivative of the model and applicable to the ESMs. In addition, this method does not set any assumptions in the probabilistic distribution of model parameters, so that it is applicable even when the model output nonlinearly depends on model parameters. It has been applied not only for ESMs but also many other numerical models such as land surface models (Sawada et al., 2020) and hydrological models (Zhang et al., 2024). Shin et al., 2026 combined this method with sensitivity analysis to reduce the number of targeted parameters. A history matching iteratively performs this surrogate model training and sampling process (Raoult et al., 2024) and gathers samples only in the high-posterior parameter space.

Ensemble Kalman Sampler (EKS) is an alternative to realize parameter uncertainty quantification in ESMs. EKS is an iterative algorithm that repeatedly updates parameter ensembles without complex computational model derivatives remaining computational efficiency under gaussian assumption on the parameter distribution. Chen and Oliver., 2012 and Emerick and Reynolds., 2013 advanced Kalman Filter by approximating Kalman gain using ensembles and Kovachki et al., 2018 suggested those methods are applicable for the point estimate in high-dimensional parameter spaces, called Ensemble Kalman Inversion (EKI). Lopez-Gomez et al., 2022 applies ensemble Kalman inversion to learn 31 parameters of a turbulence parameterization scheme called the eddy-diffusivity mass-flux scheme in the general circulation model. Garbuno-Inigo et al., 2020 extended this framework to Ensemble Kalman Sampler (EKS) to facilitate uncertainty quantification (UQ) by approximating Bayesian posterior distributions. However, because the underlying update mechanism of these ensemble-based methods fundamentally relies on Gaussian approximations, they may fail to accurately capture the target distribution in systems characterized by strong non-Gaussianity (Cleary et al., 2021, Oliver et al., 2015). In addition, EKS requires the recursive update of parameters and must iterate long-term simulations of dynamic models, which is often infeasible in the case of ESMs.

Despite a lot of methodological efforts on parametric uncertainty quantification of ESMs, few studies discussed predictive uncertainty of long-term and non-linear future changes of the Earth system, such as AMOC climate tipping. In this study, we adopted surrogate model uncertainty quantification (Yarger et al. 2024, Keetz et al., 2025, Sawada 2020) for an ESM-based AMOC prediction, since it can handle non-Gaussianity in the probability distribution of model parameters, which is an important feature considering strong non-linear dynamics of AMOC tipping, contrary to EKS. Surrogate-model uncertainty quantification has been used for a sub-system parameter uncertainty quantification in such as Yarger et al., 2024, but few studies apply the method to coupled models. We performed OSSE using a fully coupled intermediate complexity of ESM, LOVECLIM, to assess the predictability of the AMOC long-term behaviour. We set 25 parameters uncertain and constrained parametric uncertainty using synthetic observation to assess how impactful the observations are to reduce the uncertainties in model parameters, simulations of observed climate, and prediction of future AMOC weakening. In this study, we utilize a single intermediate resolution ESM, LOVECLIM, as a pragmatic and essential first step, given a computational resource to systematically establish our framework. Carslaw et al., 2026 provides an opinion about the importance of Perturbed Parameter Ensembles (PPEs). Considering this emerging topic of PPE in climate models, we could extend our framework into the experiment with CMIP ensembles, which paves the way to further reduce uncertainty.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

2. Method

2.1 LOVECLIM

We chose LOVECLIM (Goosse et al., 2010), which is an intermediate complexity of Earth system model and can simulate AMOC tipping-like behavior. LOVECLIM is one of the most computationally efficient models which enables iterative evaluation of posterior parameter distribution and projection uncertainty. Kim et al. (2022) used it to analyze how freshwater forcing rate affects the rapid weakening of the AMOC. LOVECLIM was also used to analyze Dansgaard-Oeschger (D-O) events taking the advantage of its low computational burden (Liu et al., 2022).

Although the full LOVECLIM framework encompasses a wide range of Earth system processes, we restricted our experimental setup to three active core components, which are the atmosphere, the ocean, and the terrestrial vegetation, to isolate specific interactions relevant to our study. We used the quasi-geostrophic atmospheric module ECBilt (Opsteegh et al., 1998), the ocean general circulation and sea-ice module CLIO (Goosse and Fichefet, 1999), and the dynamic global vegetation model VECODE (Brovkin et al., 1997). Its atmospheric component is a T21, 3-level quasi-geostrophic model and its oceanic component has horizontal resolution of 3° by 3° and vertically 20 levels. By deactivating the continental ice sheet (AGISM) and marine carbon cycle (LOCH) modules, we constructed three-component coupled configuration that efficiently captures the essential physical climate and biosphere dynamics required for our analysis. The scenario of freshwater hosing and carbon emission we used in our experiments is shown in the section 3.1.

2.2 Parametric Uncertainty Quantification Method

2.2.1 Surrogate-model Uncertainty Quantification

Surrogate model is a machine learning model to estimate some representatives of ESM outputs from model parameters. Let $x \in \mathbb{R}^N$ a model state variable which has a large spatial dimension. Note that x is already time-averaged (in this paper 50-year averages) and has only spatial dimensions. We generate the model variable x by running the ESM with a fixed parameter set $\theta \in \mathbb{R}^n$. An ideal surrogate model F_{ideal} is a map from parameter θ to probability distribution on x. Using the finite size of training set D consisting of pairs of a parameter θ and model realization x, we obtain the surrogate model F trained on D as an approximation of the F_{ideal}.

Training an accurate surrogate model which directly maps from parameters to ESM high spatio-temporal dimension output is technically challenging because ESM output includes parameter-independent variabilities such as interannual variability coming from chaoticity. Therefore, we need to keep the reliability of the surrogate model by introducing some low-dimensional feature extraction as the representatives of ESM output to remove such noisy variations (Dunbar et al., 2021, Yarger et al., 2024). Before training the surrogate model, we applied Principal Component Analysis (PCA) to the high dimensional climate data used as outputs of surrogate models. Assuming that we have M training samples in D. Model variable x is a one-

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

dimensional vector with multiple variables on multiple grids. We construct a data matrix $X = [x_1, x_2, \dots, x_M]^T \in R^{N\times M}$, where M is the number of samples in the training set D. Then we standardized the data in the following way:

$$\tilde{X} = \frac{X - \bar{X}}{\sigma},$$

where $\bar{X} \in R^N$ is a mean of the training samples and σ is a standard deviation of all the elements of $X - \bar{X}$ calculated by $\sigma = \sqrt{\frac{1}{NM}\sum_i^N diag((X-\bar{X})(X-\bar{X})^T)_i}$, which means σ is a global mean of amplitude of annual fluctuations. We perform an eigenvalue decomposition of the sample covariance matrix C: $C = \frac{1}{M-1}\tilde{X}\tilde{X}^T = V\Omega V^T$. Here, $V = [v_1, v_2, \dots, v_M] \in R^{N\times M}$ is the matrix of eigenvectors (principal components) and Ω is a diagonal matrix of eigenvalues $\omega_1 \geq \omega_2 \geq \cdots \geq \omega_M \geq 0$. By selecting the first d principal components ($d \ll N$) that correspond to the largest eigenvalues, the high-dimensional vector x is projected into a low-dimensional subspace: $z = V_d^T(x - \bar{x})$ where V_d consists of the first d columns of V. This dimensionality reduction effectively filters out the high frequency noise and chaotic internal variability that are independent of the parameter θ. Consequently, the surrogate model F focus on capturing the fundamental relationships between the parameters θ and the dominant modes of variation in the ESM. This ensures that the surrogate model learns the structural response of the system to parameter changes rather than being misled by stochastic fluctuations.

2.2.2 Surrogate model

As a surrogate model, we employed Gaussian Process Regression (GPR), a non-parametric method that combines ridge regularization with the kernel trick to capture a non-linear relationship. Specifically, we constructed a separate GPR emulator for each dimension j of the reduced subspace. The GPR model learns the mapping function F_j by minimizing the regularized squared error in the Reproducing Kernel Hilbert Space (RKHS). To evaluate the efficacy of the emulation, we computed the coefficient of determination (R^2) for each dimension. We optimized kernels and their hyperparameter on each latent dimension emulator (see section 3.3 for its details).

2.2.3 Sampling

We used the GPR surrogate model to estimate posterior distribution. GPR maps parameter to gaussian probability distributions on $\mathbb{R}^d$ and prediction mean and prediction variance can be obtained from the GPR surrogate model. Likelihood function is calculated on the maharanobis distance. In addition to observation error and estimator uncertainty, we also considered the prediction uncertainty of the surrogate model and defined the prediction uncertainty of the surrogate model as Σ_{model}. Local observation error $\Sigma_{v,s}$ is the observation error at variable v and at grid s. Then, the observation error in the latent space is approximated by the following:

$$\epsilon_{obs} \sim \mathcal{N}\left(0, \Sigma_{latent}^{obs}\right), \quad \Sigma_{latent}^{obs} = V\left(\bigoplus_{s=1}^{N}\Sigma_{v,s}\right)V^T \tag{1}$$

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

where $V \in \mathbb{R}^d \times \mathbb{R}^N$ is a dimensionality reduction matrix which is used to transform the vector in the original N dimensional space into d dimensional latent space. The likelihood function is calculated in the latent space with d dimensions. We assumed that the observation data y_{obs} is mapped into u_{obs} in the latent space by $u_{obs} = Vy_{obs}$. Based on Bayes theory, the posterior distribution of parameter θ is derived from $p(\theta|u_{obs}) \propto p(u_{obs}|\theta)p(\theta)$. In our experiment, the prior distribution was modelled as uniform distribution, while the likelihood function $p(u_{obs}|x)$ was modelled as the following equation:

$$\ln p(u_{obs}|\theta) = -\frac{1}{2}\left(u_{obs} - m(\theta)\right)^T \left(\alpha\Sigma_{total}(\theta)\right)^{-1}\left(u_{obs} - m(\theta)\right) + \text{const} \quad (2)$$

where Σ_{total} is defined as $\Sigma_{total}(\theta) = \Sigma_{latent}^{obs} + \Sigma_{model}(\theta)$ and α is an inflation coefficient of the covariance matrix and set as 100. The prediction from the surrogate model after training follows gaussian distribution $F_{GPR}(\theta) \sim N\left(m(\theta), \Sigma_{model}(\theta)\right)$ where $m: \mathbb{R}^n \to \mathbb{R}^d$, $\Sigma_{model}: \mathbb{R}^n \to \mathbb{R}^d \times \mathbb{R}^d$ and the following equation is proved by Bayes theory:

$$\begin{aligned} p(u_{obs}|\theta) &= \int p(u_{obs}|f)p(f|\theta)df \\ &= \int df \frac{1}{\sqrt{2\pi\left|\Sigma_{latent}^{obs}\right|}} exp\left(-\frac{1}{2}(u_{obs} - f)^T (\Sigma_{latent}^{obs})^{-1}(u_{obs} - f)\right) \\ &\quad \cdot \frac{1}{\sqrt{2\pi|\Sigma_{model}(\theta)|}} exp\left(-\frac{1}{2}\left(f - m(\theta)\right)^T \left(\Sigma_{model}(\theta)\right)^{-1}\left(f - m(\theta)\right)\right) \end{aligned} \quad (3)$$

where $|\cdot|: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is determinant of the matrix and f is a realization from the distribution $F_{GPR}(\theta)$. Using Gaussian Integral, the equation (2) is derived. In numerical implementation, we apply Cholesky decomposition on the observation error covariance matrix $\Sigma_{total} = LL^T$ and calculate the likelihood function

$$\ln p(u_{obs}|x) = -\frac{1}{2}\left[L\left(u_{obs} - m(x)\right)\right]^T\left[L\left(u_{obs} - m(x)\right)\right] + \text{const.} \quad (4)$$

In this study, we employed the Replica Exchange Monte Carlo (REMC) method (Hukushima and Nemoto, 1996) to overcome the multiple-minima problem and efficiently sample the parameter space with high dimensionality. While the generalized replica exchange framework incorporates multi-dimensional parameter exchanges, such as Hamiltonian or pressure variations, we restricted our experimental setup to a straightforward one-dimensional temperature exchange scheme. Within each non-interacting replica, the system evolved using solely standard Metropolis-Hastings update algorithms (Metropolis et al., 1953; Hastings, 1970). By limiting our configuration to temperature-only exchanges across a discrete set of replicas, we maintained a streamlined and computationally efficient framework that successfully captures the essential thermodynamic properties and structural transitions required for our analysis, without the overhead of more complex extended ensemble techniques.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

The algorithmic workflow of our temperature-exchange REMC implementation consists of simulating M non-interacting replicas parallelly, each assigned to a distinct inverse temperature $\beta_m = 1/(k_B T_m)(m = 1, \ldots, M)$. During the local sampling phase, each replica m independently updates its state variable x_m using a standard Metropolis-Hastings transition kernel that targets the invariant distribution $p(\theta_m; \beta_m) \propto \exp(-\beta_m logp(p(\theta_m|u_{obs})))$. At predetermined step intervals, global moves are introduced by proposing configuration swaps between adjacent temperature pairs, $(m, m+1)$. To strictly preserve the detailed balance condition of the extended ensemble, the proposed exchange of configurations θ_m and θ_{m+1} is accepted with the probability

$$\min\left(1, \exp\left[(\beta_m - \beta_{m+1})\left(logp(p(\theta_m|u_{obs}) - logp(p(\theta_{m+1}|u_{obs}))\right)\right]\right). \tag{5}$$

By periodically alternating between these local exploration steps and global random walks in temperature space, the algorithm allows states trapped in local minima at lower temperatures to easily overcome thermodynamic barriers by temporarily migrating to higher-temperature regimes. We used REMC for the posterior sampling of the parameter and used the representative samples from them to run the posterior simulations with LOVECLIM. Refer to Table S1 and Figure S1 in the Supplements to see the detailed settings of REMC.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

3. Experiment Settings

3.1 Freshwater Hosing and Carbon Emission scenario

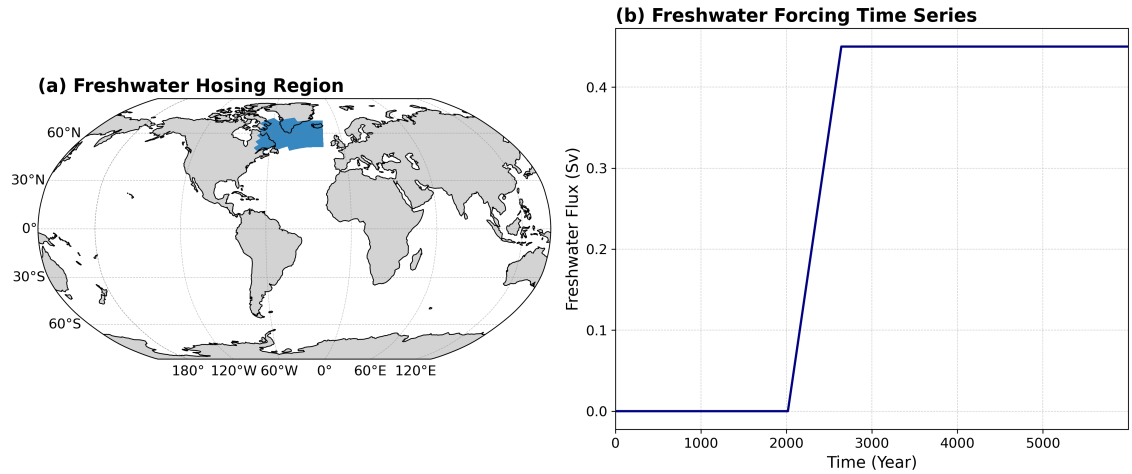

Figure 1. The area where freshwater is poured and the scenario of freshwater increase used in our experiment. (a) The region of freshwater hosing is shown as a blue shaded region. (b) Time profile of freshwater hosing scenario. The horizontal and vertical axes are model year and Freshwater Flux, respectively. The background map was generated by the authors using Python (Cartopy package).

We applied linearly increasing freshwater flux at a fixed rate, which is uniformly distributed in the North Atlantic Region (50°–70°N, 70°–15°W) as shown in Figure 1. We referred to Kim et al., 2022 to determine the input freshwater forcing scenario to see AMOC response to the freshwater forcing. Kim et al., 2022 mentions the freshwater forcing with the increasing rate of 7.2×10^{-4}Sv/year, which is the rate of Greenland Icesheet melting obtained from CMIP6 projection from 2017-2100. We kept the increasing rate until the freshwater forcing reaches 0.45 Sv. We ran the posterior simulation until 3000 model year with the default carbon emission in the LOVECLIM version 1.4 which is PMIP 3 carbon emission scenario from 1 to 2100 model year and then be fixed to the value at 2100 model year.

3.2 Uncertain Parameters

We selected 25 unknown parameters. The prior is assumed to be uniformly distributed within specific ranges, as adopted in Shi et al. (2019). The detailed parameter settings, including the synthetic true values, are summarized in Table 1. All the parameters are standardized into [0,1] in our experiments. We also did the same experiments with 5 uncertain parameters selected from Table 1, specifically parameter 1, 4, 5, 16 and 17. Refer to Figure S2, S3, S4, S5, S6, S7 and S8 in the Supplements to see the results.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

Table 1: List of Uncertain Parameters

Index	Parameter	Definition	Synthetic Truth	Range
P1	ampwir	Scaling coefficient for the longwave radiation scheme (amplw)	1	[0.5, 1.5]
P2	ampeqir	Scaling coefficient for the longwave radiation scheme (amplw)—for the equator area between 15S and 15N.	1.8	[1.0, 2.5]
P3	expir	Exponent for the longwave radiation scheme	0.4	[0.2, 0.6]
P4	relhmax	The relevant threshold pfthe total precipitable water below 500 hPa	0.83	[0.50, 0.90]
P5	cwdrag	Drag coefficient to compute wind stress	2.1E-3	[1.0E-3, 4.0E-3]
P6	cdrag	Drag coefficient to compute sensible and latent heat fluxes	1.4E-3	[1.0E-3, 2.0E-3]
P7	uv10rfx	Reduction in wind speed between 800 hPa and 10 m	0.8	[0.7, 0.9]
P8	dragan	Rotation of the wind vector in the boundary layer (unit: degree)	15	[10, 20]
P9	alphd	Albedo of snow	0.72	[0.60, 0.90]
P10	alphdi	Albedo of bare ice	0.62	[0.50, 0.80]
P11	alphs	Albedo of melting snow	0.53	[0.30, 0.60]
P12	albice	Albedo of melting ice (general)	0.44	[0.30, 0.60]
P13	albin	Albedo of melting ice (Arctic)	0.44	[0.30, 0.60]
P14	albis	Albedo of melting ice (Antarctic)	0.44	[0.30, 0.60]
P15	cgren	Increase in snow/ice albedo under cloudy conditions	0.04	[0.01, 0.10]
P16	corAN	Adjestment parameter of precipitation in Atlantic Ocean (North)	-0.085	[−0.10, −0.05]
P17	corAS	Adjestment parameter of precipitation in Atlantic Ocean (South)	-0.085	[−0.10, −0.05]
P18	corAC	Reduction in precipitation over the Arctic	-0.25	[−0.30, − 0.20]
P19	evfac	Maximum evaporation factor over land	1	[0.5, 1]
P20	bmoismfx	Maximum bucket depth (unit: m)	0.15	[0.01, 0.50]
P21	bering	Scaling factor for computing the Bering Strait throughflow	0.3	[0.2, 0.5]
P22	ai	Coefficient of isopycnal diffusion (unit: m2 s−1)	300	[200, 400]
P23	aitd	Gent-McWilliams thickness diffusion coefficient (unit: m2 s−1)	300	[200, 400]
P24	ahs	Horizontal diffusivity for scalars (unit: m2 s−1)	100	[50, 150]
P25	ahu	Horizontal viscosity (unit:m2 s−1)	1E5	[0.5E5, 1.5E5]

We generated 1000 training samples with the perturbed 25 uncertain parameters using Quasi Monte Carlo method. We refer to Shi et al., 2019 and use the good lattice points methods (GLP) in this study. The GLP design is generated by the following equations:

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

$$\begin{cases} q_{ki} = kh_i (mod\ n) \\ x_{ki} = (2q_{ki} - 1)/n \end{cases}, k = 1,2,\dots,n; i = 1,2\dots,s$$

where n represents the number of samples, s represents the number of dimensions, x_{ki} represents the coordinate of the kth sample point in the ith dimension, q_{ki} represents an ancillary variable, and h_i represents an element in the generating vector. The range of coordinate x_{ki} is restricted to [0,1]. The h_i and n must be selected so that the greatest common divisor of the h_i and n is 1.

3.3 Observing System Simulation Experiments

Observing System Simulation Experiments (OSSEs) framework was adopted in this study. We first fixed the synthetic true parameter and generated nature run, which is the synthetic true simulation with the synthetic true parameter. In this framework, observation variable and error settings were assumed to obtain different sets of observation used in the subsequent posterior parameter sampling process. We obtained the posterior parameter distribution and then got representative parameter samples from the posterior distribution. Then, we ran the posterior simulations with the representative parameter samples and compared them with the synthetic true simulation to check whether the observation and the uncertainty quantification method reduce the uncertainty.

We performed observation sensitivity test to discuss which observation is informative to reduce the uncertainty of the parameter and the AMOC projection. We also checked whether the results are robust to the observation error. The observation resolution is the same with the LOVECLIM grids. We used annual data as observation data and generated the observation data from the synthetic true simulation data by selecting variables and adding observation noises.

The parameter uncertainty quantification was performed with 4 variables, sea surface temperature (SST), sea surface salinity (SSS), relative humidity (Q) and geopotential height (GEOPG), as synthetic observation variables. We set the period from 1950 to 2000 model year as a stationary climate observation period and the ratio of random observation error to the mean annual variability of all the grids was tested with four patterns, 1%, 2%, 5% and 10%. The same observation error across all the grid points, including on the grid point where the variability is small, was used. Note that we ran 1950-year spin up simulations for every training sample (Figure S9 in the Supplements to see the spin up is sufficient).

The GPR surrogate models on each observation variable was trained. In the process of GPR training, the selection of kernels and hyper parameters are optimized. The kernel function is selected from the following kernels, Radial Basis Function (RBF), Matérn kernels and Quadradic kernels. We selected a kernel and optimized hyperparameters on each kernel by maximizing the loglikelihood with prior regularization using 5-folded cross validation framework. The prior regularization on the length scale is introduced to avoid overfitting of the GPR to the training data samples.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

3.4 Uncertainty Quantification Setting for Parameter, Simulation in the observation period, and Prediction

The sampling process of REMC provides us with a large number of samples, and we cannot directly run ESM simulations with all the sampled parameters. It is necessary to extract the limited number of samples from the posterior parameter distribution. In our experiments, we randomly sampled from the posterior distribution and got 200 samples every observation setting. Note that in those settings, we ran the posterior simulations with 2%-error and 5%-error noise settings to see the effect of observation noise experimentally on the parameter, and prediction uncertainty.

Even after limiting the number of posterior parameter samples, it is necessary to reduce computational costs to get posterior simulations by avoiding the spin up processes. We used 1000 spin-up simulations already available, to get the samples used for training the surrogate-models (i.e., prior samples) and utilized those simulation data to reduce the spin up period. the closest prior samples for each posterior parameter sample in the parameter space were chosen and the corresponding simulation data as the initial condition of the posterior simulations was used. The simulation data of the closest samples were assumed to be close to the stable states under the new posterior samples and could be used for spun up initial conditions.

3.5 Evaluation

To evaluate the effectiveness of different observation settings in constraining climate model uncertainty, we utilized the bias, RMSE and MSE across the posterior/prior member ensemble as our primary metric. The bias is the difference in mean climatology between the posterior parameter ensembles and the synthetic true simulations, while the MSE and RMSE consist of the squared bias and the variance of the ensemble members. We quantified how much selecting observation variable and error settings affects the projection uncertainty. We also calculated the difference of the squared bias and RMSE between two observation settings to see how the projection uncertainty depends on the estimated uncertainty.

$$MSE = \frac{\sum_i^N \left|\left|x_i - x_{true}\right|\right|_2^2}{N},$$

$$RMSE = \sqrt{\frac{\sum_i^N \left|\left|x_i - x_{true}\right|\right|_2^2}{N}},$$

$$BIAS = \sqrt{\frac{\sum_i^N \left|\left|x_i - \bar{x}\right|\right|_2^2}{N}},$$

where x_i is a vector in parameter space or state space of the ensemble member i.

The longer-term climatology to see the stability in AMOC was evaluated by looking at the histogram of AMOC strength at different timesteps and its trend. Next, we analyzed how the parameter posterior uncertainty affects other climatology after AMOC weakening such as precipitation. The squared bias and MSE were used to check whether the reduction in the parameter

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

uncertainty also leads to the improvement in the projection of longer-term climatology. We also drew a scatter plot of AMOC strength or other climatology before and after freshwater hosing to see how correlated they and their uncertainty are, in other words, we would like to check whether reduction in the current climate uncertainty leads to decrease in the future projection uncertainty. The climate condition before and after freshwater hosing increases should have a different scale of interannual variability, so when we compared the representation/projection uncertainty before and after the hosing increases, the square root of MSE normalized by the interannual variability, called as Normalized Root Mean Squared Error (NRMSE), was adopted.

$$NRMSE = \sqrt{\frac{\sum_j^N \sum_i^k \left\| x_{i,j} - x_{true,j} \right\|_2^2 / \sigma_j^2}{N}},$$

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

4. Results

4.1 Performance of Surrogate model

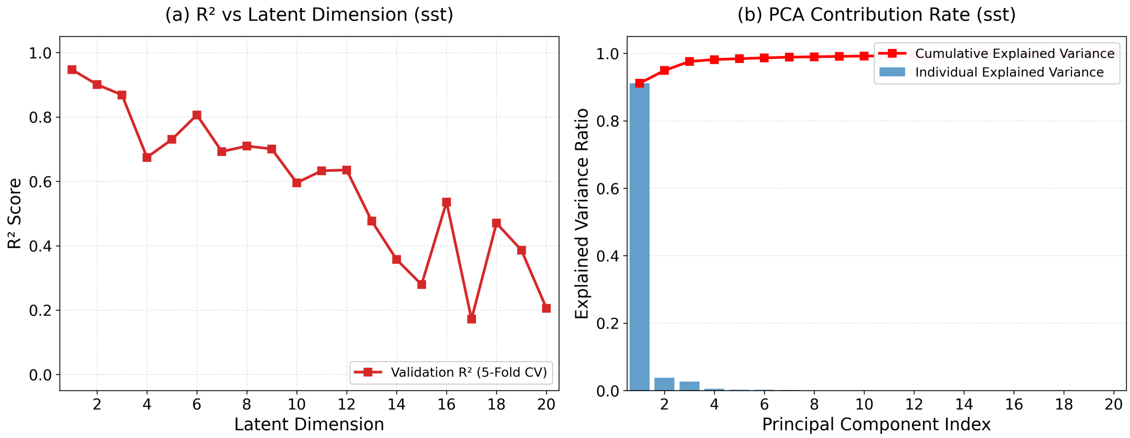

Figure 2. The performance of the surrogate model for estimating SST distribution from the input parameters. (a) Validation (red) determinant coefficients of the surrogate model. 5-folded Cross validation framework is adopted to evaluate the validation accuracy on each principal component. (b) The Explained variance ratio and its cumulative value of the principal component analysis. The horizontal axes on both panels are the index of principal components. The components are indexed in order that the explained variance ratio decreases with the index.

Figure 2 shows the performance of the surrogate model used in these experiments. Figure 2(a) shows that there is a tendency that the determination coefficients of the surrogate model are small when the explained variance of the component is small. The determinant coefficients of the top three components, whose cumulative explained variance ratio is larger than 95 % (Figure 2b), are more than 0.8 even with validation data. This shows that the performance of surrogate model is high and can estimate SST accurately from the parameters. We showed the other surrogate model performance with SSS, GEOPG and Q in Figure S10 in the Supplements. We regarded these surrogate models as those satisfying required performance and used them in the subsequent processes.

4.2 Estimated posterior parameter distribution

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

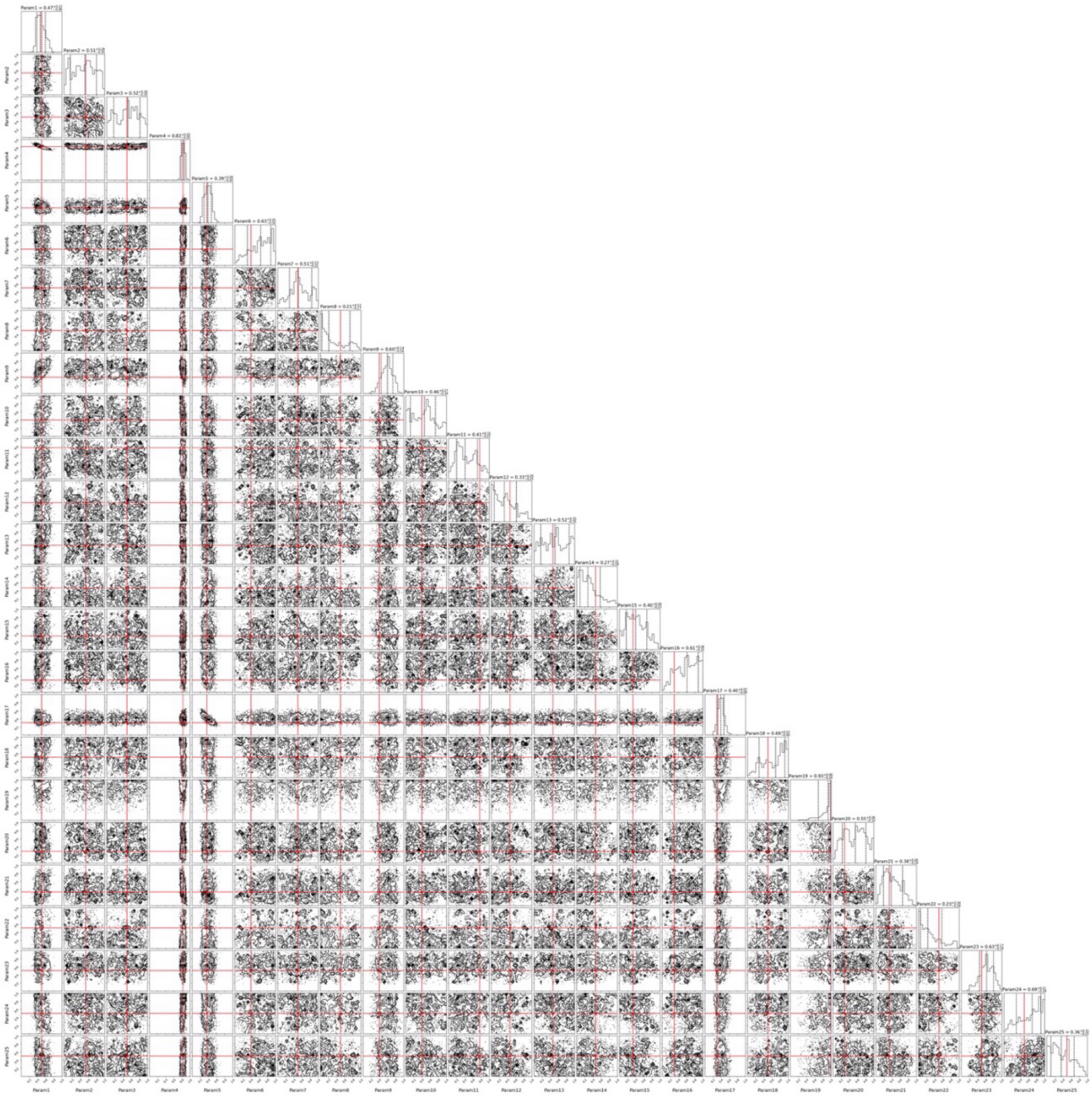

Figure 3. Corner plot of the estimated posterior parameter distributions using SST observations with a 1% error. Diagonal panels display the one-dimensional marginal posterior distribution for each parameter, while off-diagonal panels show the two-dimensional marginal posterior distributions for each pair of parameters. The red lines in the diagonal panels and the red crosses in the off-diagonal panels indicate the synthetic true parameter values.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

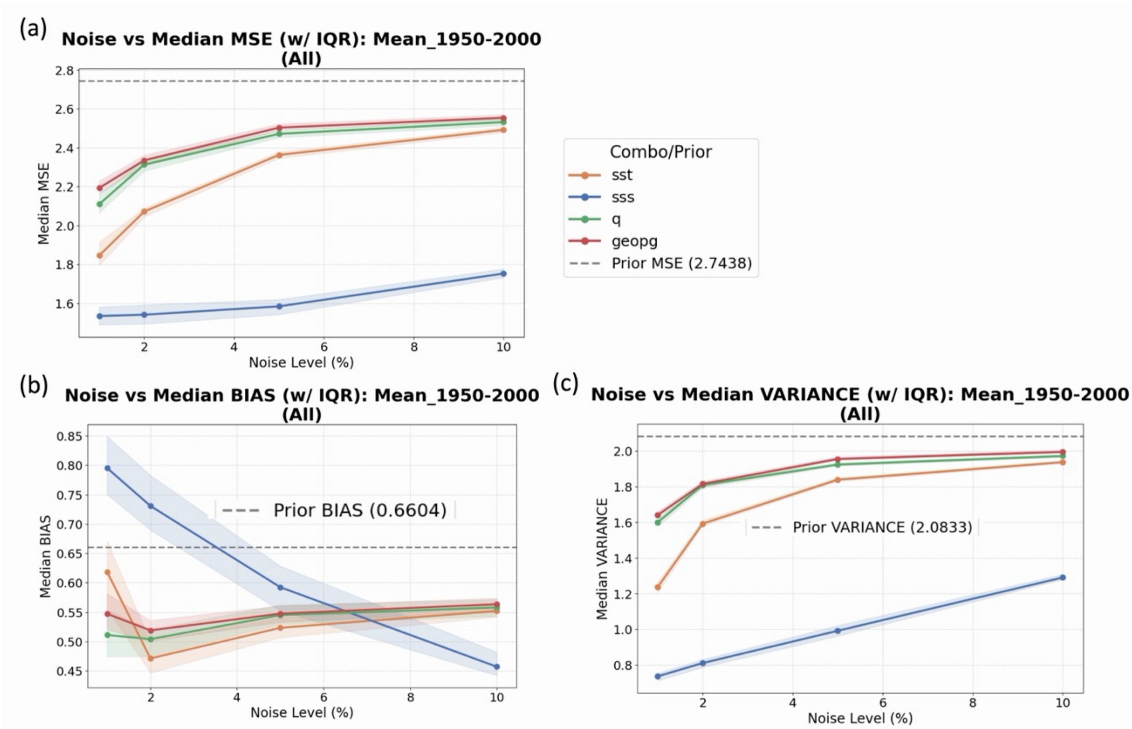

Figure 4. Sensitivity of parameter estimation accuracy and uncertainty to observation noise levels across different observation variable settings. (a), (b) and (c) show the mean squared error, squared bias and variances in its vertical axis, respectively. The orange, blue, green and red lines show the results with SST, SSS Q and GEOPG as observation variable, respectively. The horizonal axes are observation noise amplitude in (a), (b) and (c).

Figure 3 demonstrated the posterior distribution of the 25 model parameters obtained by our surrogate model uncertainty quantification method. We found that the posterior distribution of some of the parameters are well constrained around the synthetic truth. The 1st, 4th, 5th, 17th and 19th parameters are especially well estimated with small variance, while the other parameters remain relatively uncertain. See Figure S4 for the posterior distribution with 1%-error SSS, Relative humidity (Q) and Geopotential height (GEOPG). We calculated the estimated posterior parameter distribution error metrics, mean squared error, squared bias and variances across all the observation variables and error settings. Figure 4 shows that, in all cases, the observation reduces MSE compared to the prior uniform distribution. Figure 4 also shows that the performance of parameter estimation depends strongly on the observed variables. The SSS observation contributes to constraining model parameters differently from the other observations. When we use SSS as observation variable, the MSE is significantly small compared to the other settings, while whether the bias with SSS as observation variable is smaller or not depends on the observation noise settings.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

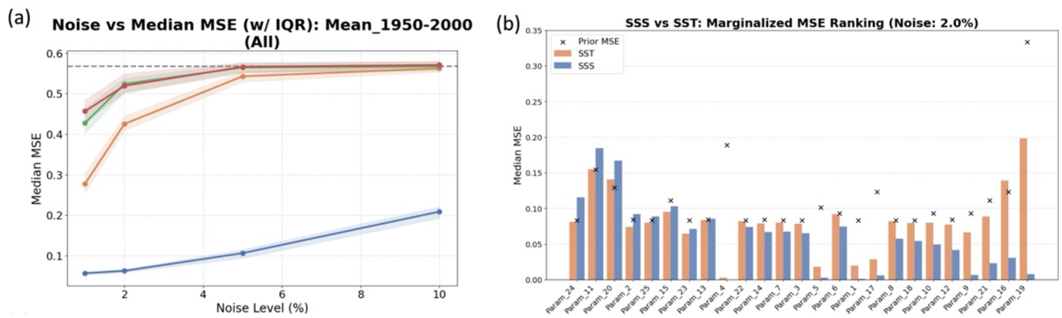

Figure 5. Comparison of the estimated parameter distributions of SSS with those with the other observation variables. (a) The MSE of parameter in the three-dimensional subspace of parameter spaces which consist of parameter 16, 19 and 21. (b) The marginal MSE of posterior parameter distribution. The orange and blue bar show the MSE with SST and SSS as an observation variable, respectively. The parameters on the righter side have smaller the marginal MSE with SSS than the marginal MSE with SST. The cross shows the prior marginal MSE.

In Figure 5, the difference of the performance of parameter estimation between SSS and SST observations is further investigated. Figure 5(b) shows that introducing SSS as an observation variable instead of SST improves the marginal parameter estimation especially in parameter 16, 19 and 21. Figure 5(a) shows that most of the MSE difference between the results with SSS and SST observations can be explained by the three-dimensional parameter subspace. Figure 4(a) shows that the differences in parameter MSE between SSS observation settings and SST observation settings are approximately 0.5, which is almost the same with the MSE difference in the subspace consisting of the dominant parameters. Table 1 shows that the top three improved parameters are related to freshwater transport into North Atlantic Ocean. The precipitation adjustment parameter in North Atlantic (P16) controls freshwater input to the North Atlantic Ocean in a form of precipitation. The scaling factor for the Bering Strait throughflow (P21) also directly adjusts freshwater input from the region around north pole to the North Atlantic Ocean. The other parameter (P19) is also related to freshwater transport through evapotranspiration process over the land.

4.3 Simulation Accuracy in the observation period

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

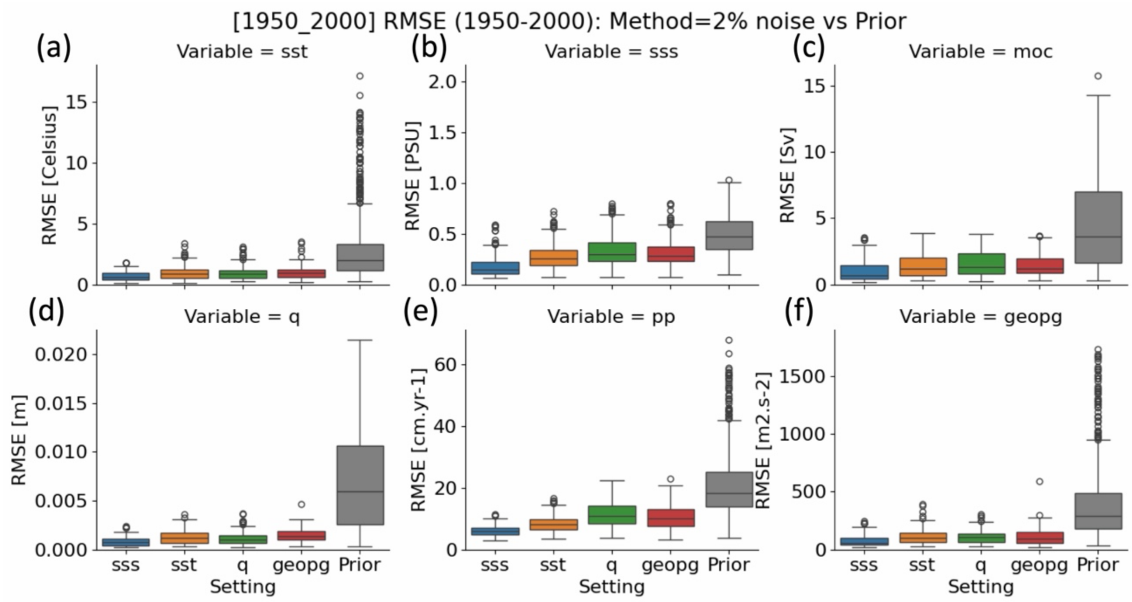

Figure 6. Simulation Accuracy of posterior simulation data in the observation period (50-year averaged climatology in 1950-2000 model year). (a-f) RMSEs with 2%-error observation. The blue, orange, green and red show the RMSE distribution of posterior simulation in the observation period with SSS, SST, Relative humidity and geopotential height as observation variable, respectively. The grey boxplots show the RMSE distribution of the prior simulations. The horizontal and vertical axes are observation setting and RMSEs of predicted variables. Different panel shows RMSE of different predicted variable. (a), (b), (c), (d), (e) and (f) show the results with SST, SSS, stream function of meridional overturning circulation, relative humidity (q), precipitation (pp) and geopotential height (geopg), respectively.

Figure 6 shows that compared to prior samples, the posterior samples achieved better accuracy to reproduce observed climate. Observation noise worsens the posterior simulation accuracy in the observation period within most cases. Refer to Figure S12 in the Supplements to see the corresponding results with difference observation noise settings. Figure 6 also shows that all the observation settings, with different observation variable and noise amplitude, RMSE in the state estimation is reduced. Consistent with the results of parameter uncertainty reduction, for instance, SSS, which is the most impactful observation to reduce the uncertainty in the model parameters, achieves the best performances in reducing the error in state variables.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

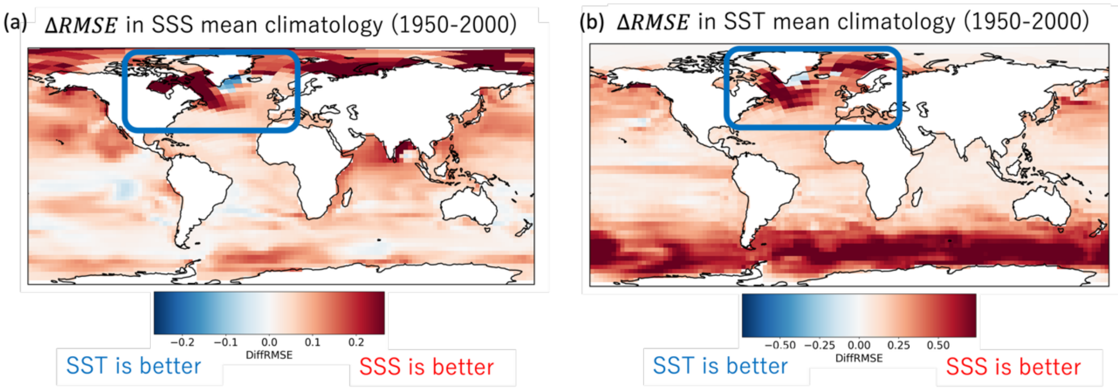

Figure 7. The difference in Simulation error in 50-year mean (a) SSS and (b) SST distribution with SSS and SST observation (2% error) in the observation period. When the RMSEs with SSS observation is smaller than those with SST observation, the map shows darker red color, while the RMSEs with SSS is larger than those with SSS, the map shows darker blue color. The area surrounded by blue rectangle is North Atlantic region where large difference in RMSEs is observed in both panel (a) and (b). he background map was generated by the authors using Python (Cartopy package).

Figure 7 shows that SSS observation reduces the RMSE of both SSS and SST in North Atlantic more effectively than SST observation. As described in the previous section, the parameters related to freshwater transportation process remain more uncertain when we used the SST as observation variable than when with SSS as observation variable. This difference of the impact on parameter estimation between SSS and SST brings the RMSE differences in North Atlantic SSS and SST, which leads to the difference in AMOC historical simulation and prediction uncertainty as we will show in the subsequent sections. It is not intuitive that SST observation is worse to constrain SST distribution than SSS observation. Figure S13 in the Supplements explains the reason of the superior performance of SSS observation to mitigate the error of SST simulation. Figure S14 in the Supplements shows that SST observation has larger variations compared to the observation noise when the parameters are perturbed. This high signal-to-noise ratio in the high-latitude area leads to better performance of SSS observation to reduce parameter uncertainty related to the freshwater flux in North Atlantic Ocean than SST observation. When SSS is used as an observation variable, the observation constrains the parameter distribution for accurate estimation of high-latitude area in the North hemisphere and this also leads to the improvement in the tropical and mid-latitude area. On the other hand, when SST is used as an observation, the tropical and mid-latitude areas mark high signal-to-noise ratio and the reduction in parameter uncertainty for accurate estimation of the tropical and mid-latitude areas, which does not help improving simulation in high-latitude area. Note that the large RMSE in SST in high latitude of Southern Ocean is explained by the fact that the thermal condition in North Atlantic Ocean and Southern Antarctic circulation current is strongly linked via thermohaline balance in the Atlantic Ocean (Menviel et al., 2013).

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

4.4 Long-term prediction

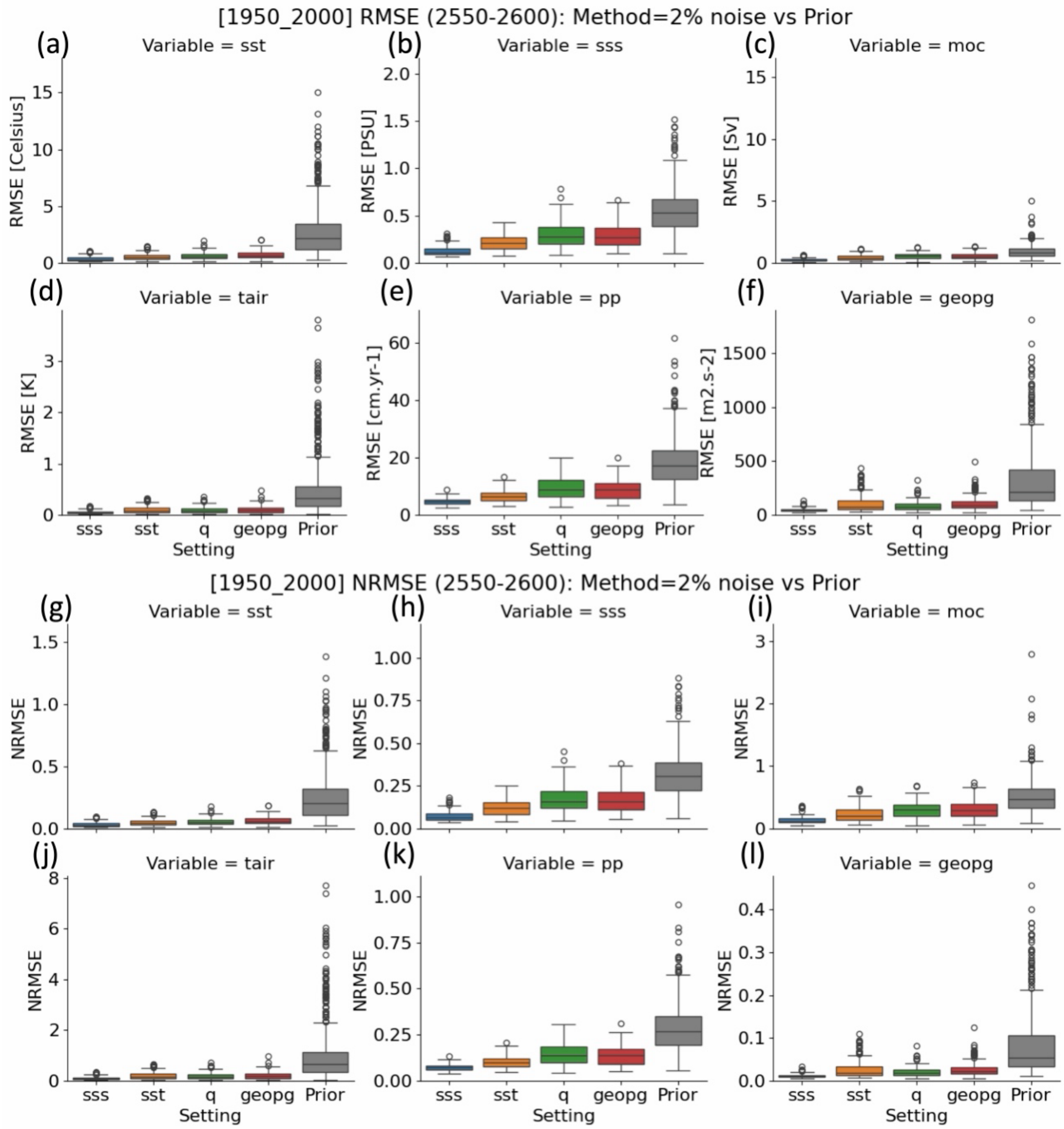

Figure 8. Prediction Accuracy of posterior simulation data in 50-year mean climatology in 2550-2600 model year. (a-f) RMSE with the multiple observation variables. (g-l) NRMSE with the multiple observation variables. The blue, orange, green, red and grey box plots show the RMSEs distribution with 2%-error SSS, SST, Relative humidity (Q) and Geopotential Height (GEOPG) observation, respectively. The horizontal and vertical axes are observation setting and RMSE of predicted variables. Different panel shows RMSE of different predicted variable. (a,g), (b, h), (c, i), (d, j), (e, k) and (f, l) show the results with

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

SST, SSS, stream function of meridional overturning circulation, surface air temperature, precipitation and geopotential height, respectively.

As shown in Figure 8, we found that the prediction accuracy in future unobserved climate was improved across all the observation settings. This means that the parameter uncertainty reduction based on observed climate contributes to the reduction of prediction uncertainty across all the predicted variables. Figure 8 also showed the Normalized RMSE because the MOC gets significantly weaker when the freshwater forcing is strong, and RMSEs of MOC is so small even with prior simulations. Even when we focus on the Normalized RMSEs, the results do not change, and the observation reduces future prediction errors across all the variables. This is also the case when we evaluated 2950-3000 model year climatology (Figure S14 in the Supplements.)

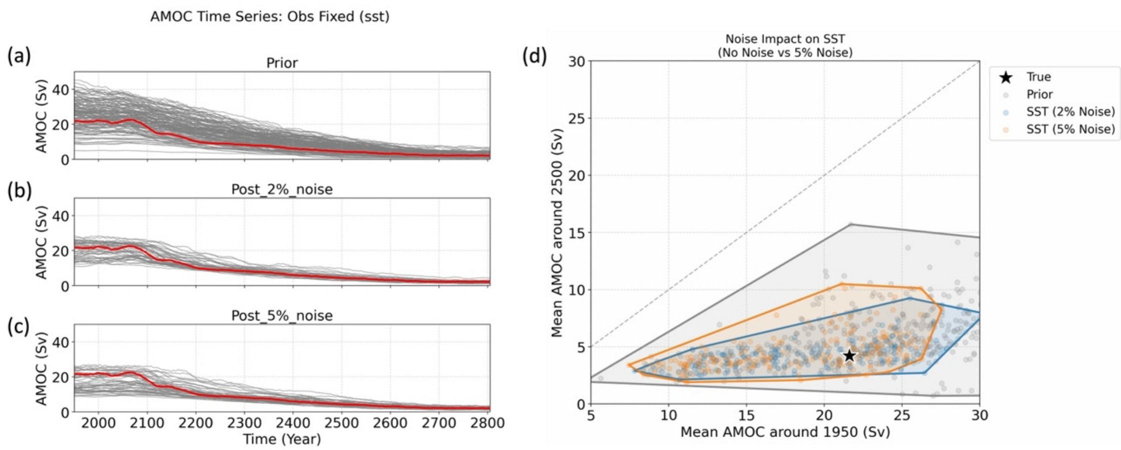

Figure 9. (a-c) AMOC timeseries of the posterior simulations (a) Without any observation, (b) with 2%-error SST observation, (c) with 5%-error SST observation. Grey and red lines in the panels are the posterior and synthetic true simulations, respectively. (d) The scatter plots of reproduced and predicted AMOC strength from the posterior simulations. The horizontal and vertical axis is historical and predicted AMOC strength, respectively. The grey, blue and orange dots are the parameter ensemble simulations with no observations, the 2%-error SST observations and the 5%-error SST observations, respectively. The star shows the AMOC strength of synthetic true data. We drew the convex hull to cover all the scatter plots every observation setting.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

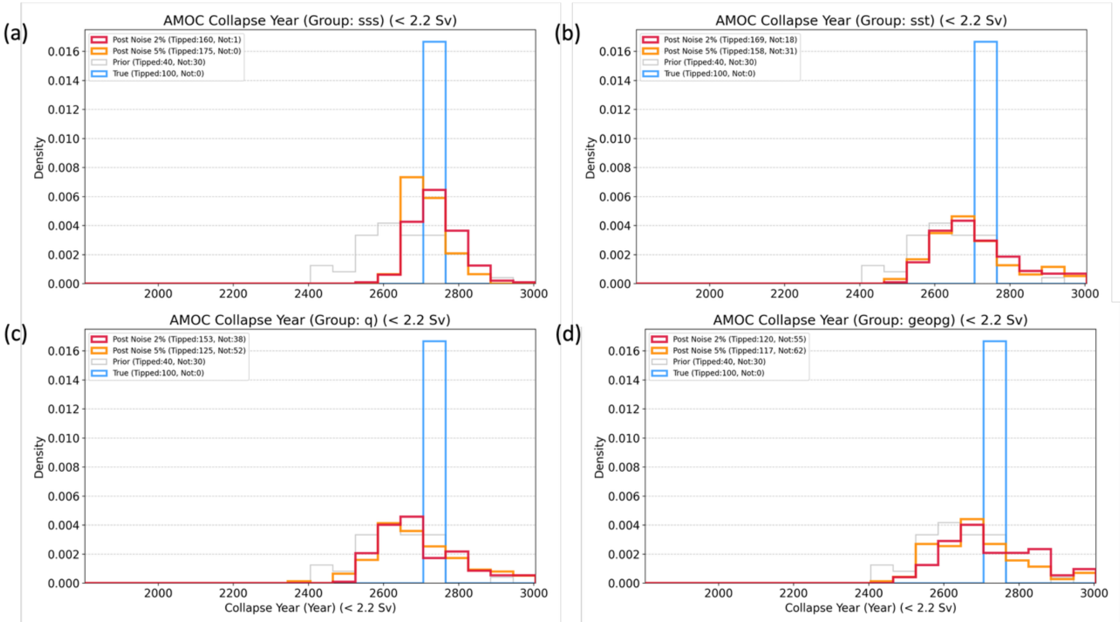

Figure 10. Tipping timing distribution across all the observation (a) SSS, (b) SST, (c) Relative humidity (Q), (d) Geopotential height (Geopg). The blue, grey, red and orange histograms show the tipping timing distributions of synthetic true simulations, prior simulations, posterior simulations with 2%-error observation and posterior simulations with 5%-error observation, respectively. The number of samples which do not go below the tipping threshold until 3000 year is shown in the legend of each panel.

Table 2: AMOC strength intervals in the posterior simulations with different observation settings. Each cell has an interval in AMOC strength and its width.

		2%-Error	**5%-Error**
SST	25-75%	14.18~22.94Sv (8.76Sv)	13.46~22.79Sv (9.33Sv)
	10-90%	12.08~24.88Sv (12.80Sv)	11.10~24.30Sv (13.20Sv)
SSS	25-75%	15.79~21.36Sv (5.57Sv)	13.91~22.32Sv (8.41Sv)
	10-90%	12.71~23.55Sv (10.84Sv)	12.71~24.15Sv (11.44Sv)
Q	25-75%	12.96~21.91Sv (8.76Sv)	13.44~22.37Sv (8.93Sv)
	10-90%	10.57~23.83Sv (13.26Sv)	10.35~24.20Sv (13.85Sv)
GEO PG	25-75%	14.29~22.21Sv (7.92Sv)	13.25~21.97Sv (8.72Sv)
	10-90%	11.20~24.27Sv (13.07Sv)	10.12~23.76Sv (13.64Sv)

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

Figures 9, 10 and Table 2 show that the error in AMOC strength is also reduced by parameter estimation. The tendency of uncertainty reduction is the same with all the observation settings (as in Figure S15 in the Supplements). However, when it comes to the timing of tipping, defined as the time when the AMOC strength becomes lower than the 2.2 Sv for the first time, only SSS observation reduces its uncertainty. Even though observations can reduce parameter uncertainty and improve simulation accuracy in the observation period, it does not so much contribute to reducing the uncertainty in estimating tipping timing.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

5. Conclusion

5.1 How much AMOC simulation uncertainty in observation period and prediction uncertainty is reduced

We established a framework of predictability study of climate tipping combining Earth system model and observation, focusing on the long-term behaviors of AMOC. We ran OSSEs to see how AMOC simulations in the observation period and long-term prediction can be constrained by a specific observation using parametric uncertainty quantification method. We controlled observation variable and noise setting to test their sensitivity on the predictability in AMOC tipping. We found that SSS observation was the most impactful to reduce uncertainty in model parameters, especially the parameters related to freshwater transport in the North Atlantic Ocean. We tested how parametric uncertainty affects simulation uncertainty in the observation period and future projection uncertainty. The parameters which are effectively constrained around the synthetic true parameters lead to improvement in prediction of many variables. However, the AMOC strength and its tipping timing remained more uncertain when we used SST, relative humidity and geopotential height than when we used SSS observation.

The key factor in AMOC strength is freshwater dynamics in the North Atlantic Ocean (Ma et al., 2024), which implies that our results of the effectiveness of SSS observation are physically consistent. Our uncertainty quantification is designed to fit the model to the global distribution of observed state variables, while the performance of the model to reproduce regional climate in North Atlantic greatly affects the AMOC behaviour. Therefore, we still have relatively large uncertainty in the AMOC tipping estimates. These results are consistent with Park et al., 2023 that uses the present-day sea surface salinity condition in the North Atlantic to constrain future warming of the Northern Hemisphere. Terhaar et al., 2021 also uses present-day regional sea surface salinity in the subtropical-polar frontal zone to reduce uncertainty in the future Southern Ocean carbon uptake.

Although the dynamics of SSS and SST are known to show early warning signals from the Earth system model-based experiments, their effectiveness on the reduction of AMOC predictive uncertainty is different. Boulton et al., 2014 showed that the dynamics of SST was affected by global-scale atmospheric interannual variability, which makes it difficult to construct the early warning signals from SST. On the other hand, they found that the dynamics of SSS was more likely to preserve its early warning signals than SST (see also Smolders et al. 2025). van Westen et al., 2024 showed that the statistical early warning signals are not robust and suggested freshwater salinity flux at 34 S which has a strength connection with global ocean salinity distribution. These works suggest that SSS is better than SST as an early warning signal. Our results are consistent with those results in the sense that future AMOC strength can be constrained when SSS observation is integrated into the ESM.

We compared our estimated uncertainty in AMOC weakening with previous works. Table 2 shows our accuracy and uncertainty in simulated AMOC strength by LOVECLIM constrained by the observations of SSS or SST. Bonan et al., 2025 described uncertainty in the prediction of AMOC strength in the CMIP6 model ensemble. They showed that the range of AMOC strength between the 10th and 90th percentiles in the original CMIP6 ensemble was approximately 14 Sv. They also

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

showed that this interquartile range is reduced to approximately 3 Sv by their emergent constraint method. Our AMOC simulation uncertainty in the observation period is equivalent to or slightly smaller than the existing CMIP6 uncertainty, while after EC, the uncertainty of current CMIP model is far smaller than ours. The large parametric uncertainty in our work implies that existing uncertainty estimates underestimated whole uncertainty in AMOC simulation because parametric uncertainty existing in the Earth system models is ignored. We describe the potential source of this uncertainty in AMOC prediction uncertainty in the next subsection. Our results also imply that there is a possibility that current CMIP6 and Emergent Constraint-based projection uncertainty quantification underestimates the structural uncertainty in the future climate projection.

The conclusion that parametric uncertainty significantly contributes to AMOC projection uncertainty is consistent to recent PPE studies. For instance, Yamazaki et al. (2024) conducted PPE using the HadGEM3-GC3.05 GCM and demonstrated that even when parameters are constrained by observational datasets such as the RAPID array (McCarthy et al., 2015), substantial uncertainty remains in future AMOC strength projections (approximately 7 Sv). Their findings highlight that parametric uncertainty cannot be neglected, even after constraining models with present-day climate observations. Combining our framework with methods designed to accelerate the spin-up process offers a promising avenue to comprehensively evaluate the impact of parametric uncertainty on AMOC projections.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

5.2 Parametric uncertainty quantification

We directly applied a surrogate-model uncertainty quantification to quantify 25 parameters in LOVECLIM using OSSE as a framework. Evaluating parametric uncertainty in Earth System Models (ESMs) typically involves a trade-off between parameter dimensionality and methodological rigor. For instance, while Yarger et al. (2024) rigorously quantified uncertainty for 4 parameters using E3SMv2, Elsaesser et al. (2025) successfully tackled a massive 45-parameter space in the NASA GISS ModelE (version E3). Jiang et al. (2026) performed uncertainty quantification for 26 parameters in the E3SM land model but relied on prior sensitivity analysis (e.g., Sobol's method) to reduce dimensionality by screening out parameters insensitive to observables. However, these studies use actual observation data and cannot verify whether the estimated results are actually correct in the sense of parameter estimation. Our experiments are the first trial to verify how well the uncertainty quantification of the $O(10)$ dimensional parameters works to identify the accurate parameter and reduces the parametric uncertainty using OSSE framework.

We deliberately avoided the sensitivity test for pre-screening our 25-parameter configuration for three critical reasons. First, variance-based sensitivity tests require a massive number of ensemble evaluations. Second, screening based solely on observed variables ignores the hidden contributions of seemingly insensitive parameters to the predictive uncertainty of variables that cannot be fully directly observed, such as our targeted AMOC strength. Instead of fixing marginally insensitive parameters to their default values, re-introducing prior distribution to the insensitive parameters is a possible way to obtain posterior parameter samples. However, this approach suffers from severe computational inefficiency due to equifinality and parameter correlations; when high-dimensional parameters are practically collapsed onto a lower-dimensional manifold, independent sampling from the priors blindly explores redundant parameter spaces. Furthermore, if we attempt to lower the screening threshold to rigorously retain all parameters with potential or hidden contributions to the target physics, the dimensionality reduction effect of the sensitivity analysis is almost lost, rendering the pre-screening step obsolete.

By directly applying the surrogate model to the full 25-dimensional space, we aimed to rigorously preserve these physical uncertainties without the theoretical compromises of prior dimension reduction. However, this approach resulted in a relatively large AMOC predictive uncertainty. This large uncertainty is primarily attributed to observation covariance inflation and an insufficient number of training samples. We generated 1,000 training samples for the 25-dimensional input space. While standard practice suggests training on $O(d)$ samples to capture a coarse global landscape (Loeppky et al., 2009, Elsaesser et al. (2025)), mapping a precise Bayesian posterior requires resolving much more detailed localized structures. Specifically, independent follow-up tests using an Ensemble Kalman Smoother (EKS) confirmed the presence of strong parameter equifinality (Figure S16 in the Supplements). Yet, the current global surrogate model lacked the sampling density to delineate these narrow, highly correlated posterior distributions within the vast unconstrained parameter space, ultimately smearing out the constraints.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

Given the curse of dimensionality, it is computationally intractable to generate enough samples to uniformly cover the entire 25-dimensional prior parameter space with sufficient density. To overcome this limitation and properly capture parameter equifinality without erroneously dropping dimensions, a hybrid approach bridging calibration and uncertainty quantification is required. The Calibrate-Emulate-Sample (CES) framework (Cleary et al., 2021) represents a promising solution. Instead of artificially reducing dimensions based on marginal sensitivities (as in Jiang et al., 2026), CES restricts the parameter region by first applying a calibration step (conceptually like the goals of Elsaesser et al., 2025) to identify the high-probability domain. Future work will focus on integrating our method with the CES framework, enabling the surrogate to be trained densely within a physically relevant, restricted parameter space. This will allow us to capture complex equifinality and substantially reduce AMOC predictive uncertainty while retaining the full dimensionality of the model.

Our results are only limited with LOVECLIM, which is not complex as ESMs in CMIP ensembles. Therefore, it is necessary to extend this experiment into CMIP ensembles in the future. The Surrogate-model Uncertainty Quantification method matches the current trend to generate multiple PPE simulations. Carslaw et al., 2026 is a comprehensive review paper about existing experiments about the PPE generation with multiple Earth System Models. For instance, Yamazaki et al., 2024 assessed how the AMOC strength depends on the parameters in HadGEM3-GC3.05 using PPEs. Although computational costs to generate PPEs are still infeasible especially for the spin up, the advance in the computational resources and feasible algorithms, such as Anderson Acceleration (Khatiwala., 2024), enables the PPEs and extension of our framework with more complex ESMs.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

Acknowledgement

This work was supported by the Recommendation Program for Young Researchers and Woman Researchers, and the Self-directed Research and Activity Funding for Fiscal Year 2025. This research was also partially supported by JSPS KAKENHI (Grant Number 26K00012) and JST Moonshot R&D project (JMPJMS2281). Furthermore, financial support was provided by the International Graduate Program of Innovation for Intelligent World, and the JST Support for Pioneering Research Initiated by the Next Generation (SPRING), Project for Fostering Advanced Human Resources to Lead Green Transformation (SPRING GX). We have updated the figure captions for Figures 1, 7, and S6 to state that the maps were created by the authors using Python.

Code and Data Availability

The data and scripts used to produce the results in this study, including the surrogate modeling and parametric uncertainty quantification of the AMOC weakening, are publicly available on Zenodo at https://doi.org/10.5281/zenodo.20805429 (Kubo, 2026). The simulations were conducted using the Earth system model of intermediate complexity, LOVECLIM version1.4. The source code and official documentation for LOVECLIM are available freely (Goosse et al., 2010).

Author Contribution

AK and YS designed the research. AK developed the methodology, conducted analyses, acquired funding and drafted and edited the manuscript. YS supervised the project, reviewed the manuscript, and acquired funding.

Competing Interests

The authors declare that they have no conflict of interest.

References

Armstrong McKay, D. I., Staal, A., Abrams, J. F., Winkelmann, R., Sakschewski, B., Loriani, S., Fetzer, I., Cornell, S. E., Rockström, J., and Lenton, T. M.: Exceeding 1.5°C global warming could trigger multiple climate tipping points, Science, 377, eabn7950, https://doi.org/10.1126/science.abn7950, 2022.

Ashton-Key, H., Mecking, J., Marsh, R., Drijfhout, S., Oltmanns, M., and Sanchez-Franks, A.: AMOC weakening across latitude and time in CMIP6 future scenarios, EGUsphere, 1–41, https://doi.org/10.5194/egusphere-2026-575, 2026.

Baker, J. A., Bell, M. J., Jackson, L. C., Renshaw, R., Vallis, G. K., Watson, A. J., and Wood, R. A.: Overturning Pathways Control AMOC Weakening in CMIP6 Models, Geophysical Research Letters, 50, e2023GL103381, https://doi.org/10.1029/2023GL103381, 2023.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

Boulton, C. A., Allison, L. C., and Lenton, T. M.: Early warning signals of Atlantic Meridional Overturning Circulation collapse in a fully coupled climate model, Nat Commun, 5, 5752, https://doi.org/10.1038/ncomms6752, 2014.

Brovkin, V., Ganopolski, A., and Svirezhev, Y.: A continuous climate-vegetation classification for use in climate-biosphere studies, Ecological Modelling, 101, 251–261, https://doi.org/10.1016/S0304-3800(97)00049-5, 1997.

Buckley, M. W. and J. Marshall.: Observations, inferences, and mechanisms of the Atlantic Meridional Overturning Circulation: A review, Rev. Geophys., 54, doi:10.1002/2015RG000493, 2016.

Carslaw, K. S., Regayre, L. A., Proske, U., Gettelman, A., Sexton, D. M. H., Qian, Y., Marshall, L. R., Wild, O., van Lier-Walqui, M., Oertel, A., Peatier, S., Yang, B., Johnson, J. S., Li, S., McCoy, D. T., Sanderson, B. M., Williamson, C. J., Elsaesser, G. S., Yamazaki, K., and Booth, B. B. B.: Opinion: The importance and future development of perturbed parameter ensembles in climate and atmospheric science, Atmos. Chem. Phys., 26, 4651–4667, https://doi.org/10.5194/acp-26-4651-2026, 2026.

Chidichimo, M. P., Perez, R. C., Speich, S., Kersalé, M., Sprintall, J., Dong, S., Lamont, T., Sato, O. T., Chereskin, T. K., Hummels, R., and Schmid, C.: Energetic overturning flows, dynamic interocean exchanges, and ocean warming observed in the South Atlantic, Commun Earth Environ, 4, 10, https://doi.org/10.1038/s43247-022-00644-x, 2023.

Cox, P.M., Williamson, M.S., Friedlingstein, P. et al. Emergent constraints on carbon budgets as a function of global warming. Nat Commun 15, 1885, https://doi.org/10.1038/s41467-024-46137-7, 2024.

Dijkstra, H. A., Krauskopf, B., Börner, R., and van Westen, R. M.: Transitions of the Atlantic Ocean circulation, Nat Rev Phys, 1–15, https://doi.org/10.1038/s42254-026-00945-6, 2026.

Dunbar, O. R. A., Garbuno-Inigo, A., Schneider, T., and Stuart, A. M.: Calibration and Uncertainty Quantification of Convective Parameters in an Idealized GCM, Journal of Advances in Modeling Earth Systems, 13, e2020MS002454, https://doi.org/10.1029/2020MS002454, 2021.

Elsaesser, G. S., van Lier-Walqui, M., Yang, Q., Kelley, M., Ackerman, A. S., Fridlind, A. M., Cesana, G. V., Schmidt, G. A., Wu, J., Behrangi, A., Camargo, S. J., De, B., Inoue, K., Leitmann-Niimi, N. M., and Strong, J. D. O.: Using Machine Learning to Generate a GISS ModelE Calibrated Physics Ensemble (CPE), Journal of Advances in Modeling Earth Systems, 17, e2024MS004713, https://doi.org/10.1029/2024MS004713, 2025.

Ernst, O. G., Sprungk, B., and Starkloff, H.-J.: Analysis of the Ensemble and Polynomial Chaos Kalman Filters in Bayesian Inverse Problems, SIAM/ASA J. Uncertainty Quantification, 3, 823–851, https://doi.org/10.1137/140981319, 2015.

Fu, Y., Lozier, M. S., Biló, T. C., Bower, A. S., Cunningham, S. A., Cyr, F., de Jong, M. F., deYoung, B., Drysdale, L., Fraser, N., Fried, N., Furey, H. H., Han, G., Handmann, P., Holliday, N. P., Holte, J., Inall, M. E., Johns, W. E., Jones, S., Karstensen, J., Li, F., Pacini, A., Pickart, R. S., Rayner, D., Straneo, F., and Yashayaev, I.: Seasonality of the Meridional Overturning Circulation in the subpolar North Atlantic, Commun Earth Environ, 4, 181, https://doi.org/10.1038/s43247-023-00848-9, 2023.

Khatiwala, S.: Efficient spin-up of Earth System Models using sequence acceleration, Science Advances, 10, eadn2839, https://doi.org/10.1126/sciadv.adn2839, 2024.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

Ganopolski, A. and Rahmstorf, S.: Rapid changes of glacial climate simulated in a coupled climate model, Nature, 409, 153–158, https://doi.org/10.1038/35051500, 2001.

Goosse, H. and Fichefet, T.: Importance of ice-ocean interactions for the global ocean circulation: A model study, Journal of Geophysical Research: Oceans, 104, 23337–23355, https://doi.org/10.1029/1999JC900215, 1999.

Goosse, H., Brovkin, V., Fichefet, T., Haarsma, R., Huybrechts, P., Jongma, J., Mouchet, A., Selten, F., Barriat, P.-Y., Campin, J.-M., Deleersnijder, E., Driesschaert, E., Goelzer, H., Janssens, I., Loutre, M.-F., Morales Maqueda, M. A., Opsteegh, T., Mathieu, P.-P., Munhoven, G., Pettersson, E. J., Renssen, H., Roche, D. M., Schaeffer, M., Tartinville, B., Timmermann, A., and Weber, S. L.: Description of the Earth system model of intermediate complexity LOVECLIM version 1.2, Geoscientific Model Development, 3, 603–633, https://doi.org/10.5194/gmd-3-603-2010, 2010.

Hastings, W. K.: Monte Carlo Sampling Methods Using Markov Chains and Their Applications, Biometrika, 57, 97–109, https://doi.org/10.2307/2334940, 1970.

Hukushima, K. and Nemoto, K.: Exchange Monte Carlo Method and Application to Spin Glass Simulations, J. Phys. Soc. Jpn., 65, 1604–1608, https://doi.org/10.1143/JPSJ.65.1604, 1996.

Jackson, L.C., Kahana, R., Graham, T. et al. Global and European climate impacts of a slowdown of the AMOC in a high resolution GCM. Clim Dyn 45, 3299–3316, https://doi.org/10.1007/s00382-015-2540-2, 2015.

Jiang, Z., Isenberg, N. M., Subba, T., Woo, H.-M., Serbin, S. P., Urban, N. M., and Kuang, C.: A Framework for Parametric and Predictive Uncertainty Quantification in the E3SM Land Model: Assessing Site and Observable Generalizability, Journal of Advances in Modeling Earth Systems, 18, e2025MS005562, https://doi.org/10.1029/2025MS005562, 2026.

Johns, W. E., Elipot, S., Smeed, D. A., Moat, B., King, B., Volkov, D. L., and Smith, R. H.: Towards two decades of Atlantic Ocean mass and heat transports at 26.5° N, Philos Trans A Math Phys Eng Sci, 381, 20220188, https://doi.org/10.1098/rsta.2022.0188, 2023.

Kim, S.-K., Kim, H.-J., Dijkstra, H. A., and An, S.-I.: Slow and soft passage through tipping point of the Atlantic Meridional Overturning Circulation in a changing climate, npj Clim Atmos Sci, 5, 13, https://doi.org/10.1038/s41612-022-00236-8, 2022.

Kim, S.-K., Kim, H.-J., Dijkstra, H. A. and An, S.-I.: Slow and soft passage through tipping point of the Atlantic Meridional Overturning Circulation in a changing climate. npj Clim Atmos Sci 5, 13, 2022.

Kubo, A. and Sawada, Y.: Predictability of Climate Tipping Focusing on Internal Variability in the Earth System, Geophysical Research Letters, 52, e2024GL113146, https://doi.org/10.1029/2024GL113146, 2025.

Kumar, R., Carroll, C., Hartikainen, A., and Martin, O.: ArviZ a unified library for exploratory analysis of Bayesian models in Python, Journal of Open Source Software, 4, 1143, https://doi.org/10.21105/joss.01143, 2019.

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

Liu, M., Prentice, I. C., Menviel, L., and Harrison, S. P.: Past rapid warmings as a constraint on greenhouse-gas climate feedbacks, Commun Earth Environ, 3, 196, https://doi.org/10.1038/s43247-022-00536-0, 2022.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

Loeppky, J. L., Sacks, J., and Welch, W. J.: Choosing the Sample Size of a Computer Experiment: A Practical Guide, Technometrics, 51, 366–376, https://doi.org/10.1198/TECH.2009.08040, 2009.

McCarthy, G. D., Smeed, D. A., Johns, W. E., Frajka-Williams, E., Moat, B. I., Rayner, D., et al.: Measuring the Atlantic Meridional Overturning Circulation at 26°N, Prog. Oceanogr., 130, 91–111, https://doi.org/10.1016/j.pocean.2014.10.006, 2015.

Menviel, L., England, M. H., Meissner, K. J., Mouchet, A., and Yu, J.: Atlantic-Pacific seesaw and its role in outgassing CO2 during Heinrich events, Paleoceanography, 29, 58–70, https://doi.org/10.1002/2013PA002542, 2014.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of State Calculations by Fast Computing Machines, Journal of Chemical Physics, 21, 1087–1092, https://doi.org/10.1063/1.1699114, 1953.

Opsteegh, J. D., Haarsma, R. J., Selten, F. M., and Kattenberg, A.: ECBILT: a dynamic alternative to mixed boundary conditions in ocean models, Tellus, 50, https://doi.org/10.3402/tellusa.v50i3.14524, 1998.

Park, IH., Yeh, SW., Cai, W. et al.: Present-day North Atlantic salinity constrains future warming of the Northern Hemisphere. Nat. Clim. Chang. 13, 816–822, https://doi.org/10.1038/s41558-023-01728-y, 2023.

Raoult, N., Beylat, S., Salter, J. M., Hourdin, F., Bastrikov, V., Ottlé, C., and Peylin, P.: Exploring the potential of history matching for land surface model calibration, Geosci. Model Dev., 17, 5779–5801, https://doi.org/10.5194/gmd-17-5779-2024, 2024.

Sawada, Y.: Machine Learning Accelerates Parameter Optimization and Uncertainty Assessment of a Land Surface Model, Journal of Geophysical Research: Atmospheres, 125, e2020JD032688, https://doi.org/10.1029/2020JD032688, 2020.

Shi, Y., Gong, W., Duan, Q., Charles, J., Xiao, C., and Wang, H.: How parameter specification of an Earth system model of intermediate complexity influences its climate simulations, Prog Earth Planet Sci, 6, 46, https://doi.org/10.1186/s40645-019-0294-x, 2019.

Shiogama, H., Hayashi, M., Hirota, N.: Combined emergent constraints on future extreme precipitation changes. Nat Commun 16, 5293, https://doi.org/10.1038/s41467-025-60385-1, 2025.

Smolders, E. J. V., van Westen, R. M., and Dijkstra, H. A.: Optimal Observation Locations for Early Warning of the Onset of an AMOC Collapse, Geophysical Research Letters, 52, e2025GL116242, https://doi.org/10.1029/2025GL116242, 2025.

Srokosz, M. A. and Bryden, H. L.: Observing the Atlantic Meridional Overturning Circulation yields a decade of inevitable surprises, Science, 348, 1255575, https://doi.org/10.1126/science.1255575, 2015.

Terhaar, J., Kwiatkowski, L. Bopp, L.: Emergent constraint on Arctic Ocean acidification in the twenty-first century. Nature 582, 379–383. https://doi.org/10.1038/s41586-020-2360-3, 2020.

van Westen, R. M., Kliphuis, M., and Dijkstra, H. A.: Physics-based early warning signal shows that AMOC is on tipping course, Science Advances, 10, eadk1189, https://doi.org/10.1126/sciadv.adk1189, 2024.

Yamazaki, K., Jackson, L. C., and Sexton, D. M. H.: Prediction of slowdown of the Atlantic Meridional Overturning Circulation in coupled model simulations, Clim Dyn, 62, 5197–5217, https://doi.org/10.1007/s00382-024-07159-5, 2024.

https://doi.org/10.5194/egusphere-2026-3915
Preprint. Discussion started: 17 August 2026

Yarger, D., Wagman, B. M., Chowdhary, K., and Shand, L.: Autocalibration of the E3SM Version 2 Atmosphere Model Using a PCA-Based Surrogate for Spatial Fields, Journal of Advances in Modeling Earth Systems, 16, e2023MS003961, https://doi.org/10.1029/2023MS003961, 2024.

Zhang, Y., Blöschl, G., Wei, H., Kong, D., Ma, N., Wagener, T., Tian, J., Xia, J., Li, C., Wang, L., Chiew, F. H. S., Leung, L. R., Liu, X., Zheng, H., Zhang, X., and Liu, C.: Overestimation of past and future increases in global river flow by Earth system models, Nat. Geosci., 19, 311–316, https://doi.org/10.1038/s41561-025-01897-9, 2026.