Text S1. Supporting Information S1 shows the results with 5 parameters, Param1, 4, 5, 16 and 17 in Table 1 in the main manuscript. Figure 1 shows the example of parameter uncertainty quantification with 1%-error SST observation. Figure 2 shows the same results as Figure 4 in the main manuscript. In all cases, the observation helps reduce MSE compared to the prior uniform distribution. Figure 2 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. In Figure 3, the difference of the performance of parameter estimation between SSS and SST observations is further investigated. Figure 3(b) shows that introducing SSS as an observation variable instead of SST improves the marginal parameter estimation especially in parameter 16 and 17.

Figure 4 shows that compared to prior samples, the posterior samples achieved better accuracy to reproduce observed climate. 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 variable, which is the same result with the results in the main manuscript. 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.

We also tested the longer-term prediction uncertainty quantification with the posterior parameter samples. Figure 5 also shows 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. Figures 6 and 7 show that the error in AMOC strength is also reduced by parameter estimation. 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.

Table S1. Replica Exchange Monte Carlo Setting

Parameters	**Values**
Chain of Parallel Tempering	7
Total steps (/chain)	100000
Burn in steps (/chain)	10000

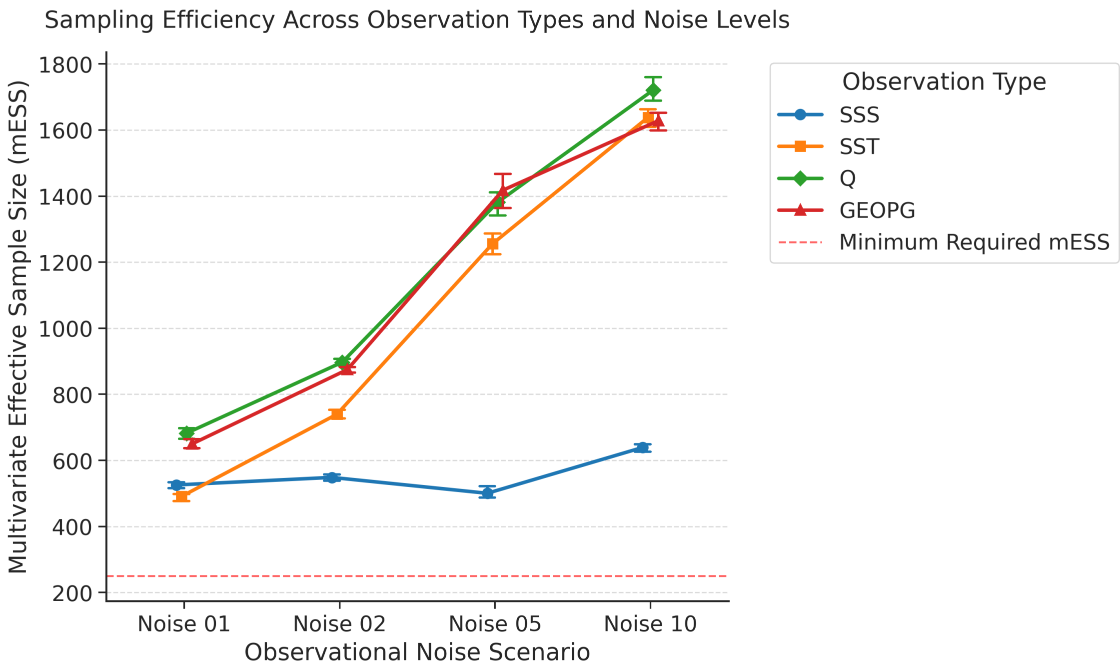


Figure S1. Comparison of Multivariate Effective Sample Size (mESS) across different observation types and observational noise levels. The blue, red, green, and orange lines show the mESS evaluated using Sea Surface Salinity (SSS), Sea Surface Temperature (SST), Specific Humidity (Q), and Geopotential Height (GEOPG) as the target observational data, respectively. Each point represents the mean mESS derived from the 25-dimensional joint posterior distribution, while the error bars indicate the standard deviation across five independent Replica Exchange Monte Carlo (REMC) runs. The red dashed line represents the heuristic minimum threshold (mESS = 250) required for a statistically reliable evaluation of the joint posterior distribution.

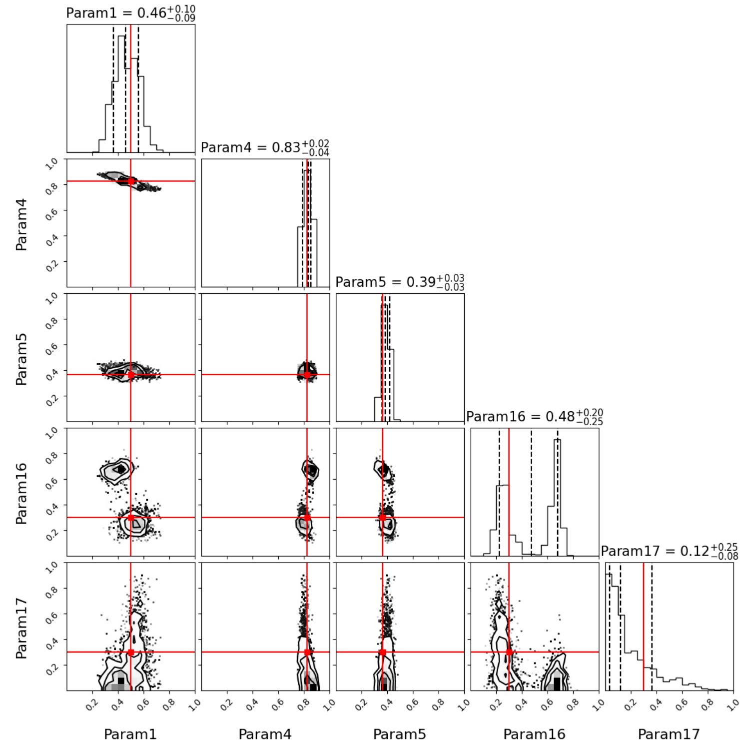


Figure S2. Corner plot of the estimated posterior parameter distributions using SST observations with a 1% error. Diagonal panels display the one-dimensional marginal posterior distribution for each parameter, while off-diagonal panels show the two-dimensional marginal posterior distributions for each 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.

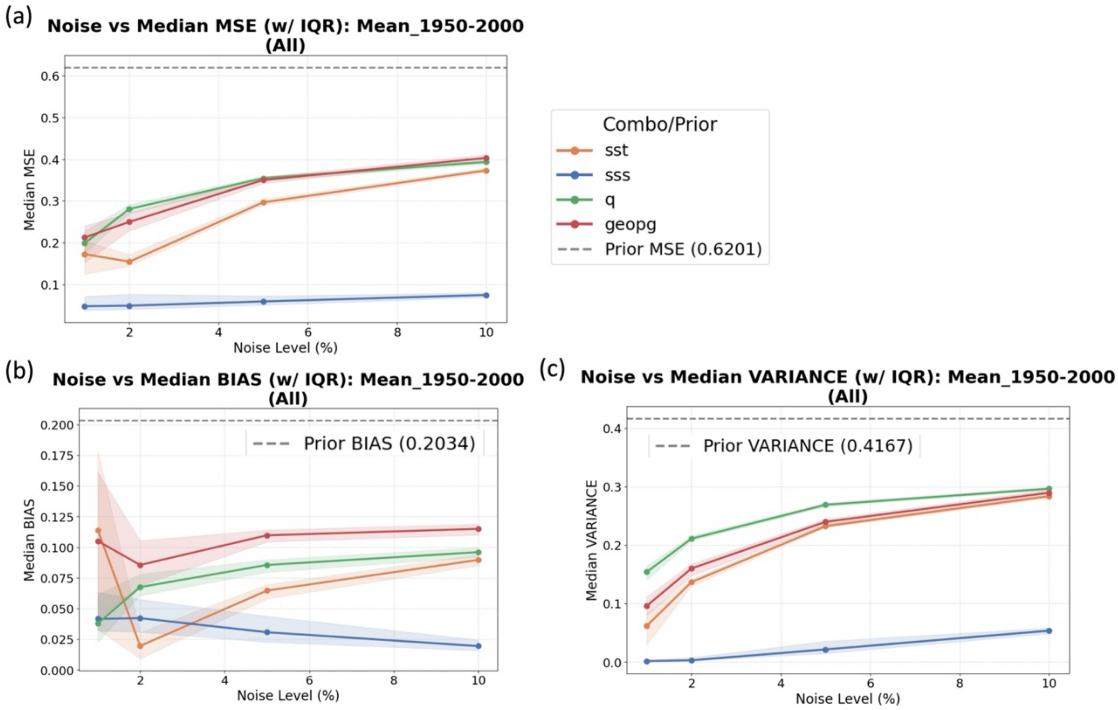


Figure S3. Sensitivity of parameter estimation accuracy and uncertainty to observation noise levels across different observation variable settings. (a), (b) and (c) show the mean squared error, squared bias and variances in its vertical axis, respectively. The orange, blue, green and red lines show the results with SST, SSS Q and GEOPG as observation variable, respectively. The horizonal axes are observation noise amplitude in (a), (b) and (c).

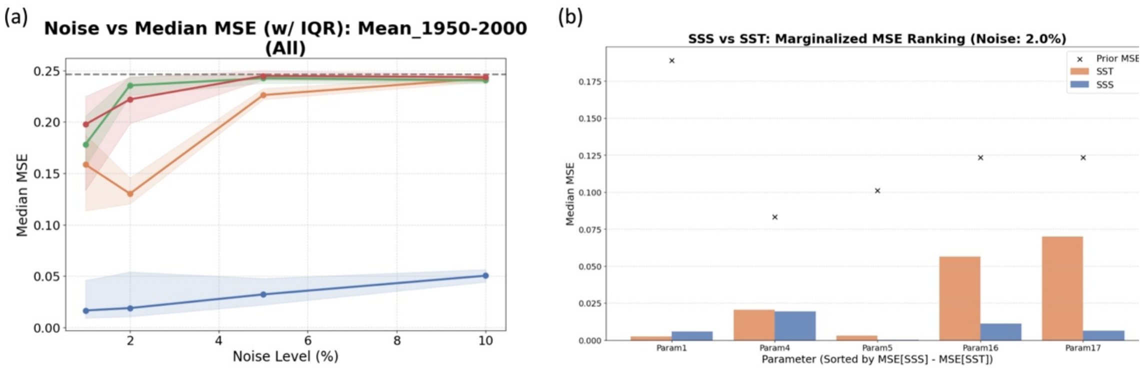


Figure S4. Comparison of the estimated parameter distributions of SSS with those with the other observation variables. (a) The MSE of parameter in the three-dimensional subspace of parameter spaces which consist of parameter 16 and 17. (b) The marginal MSE of posterior parameter distribution. The orange 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.

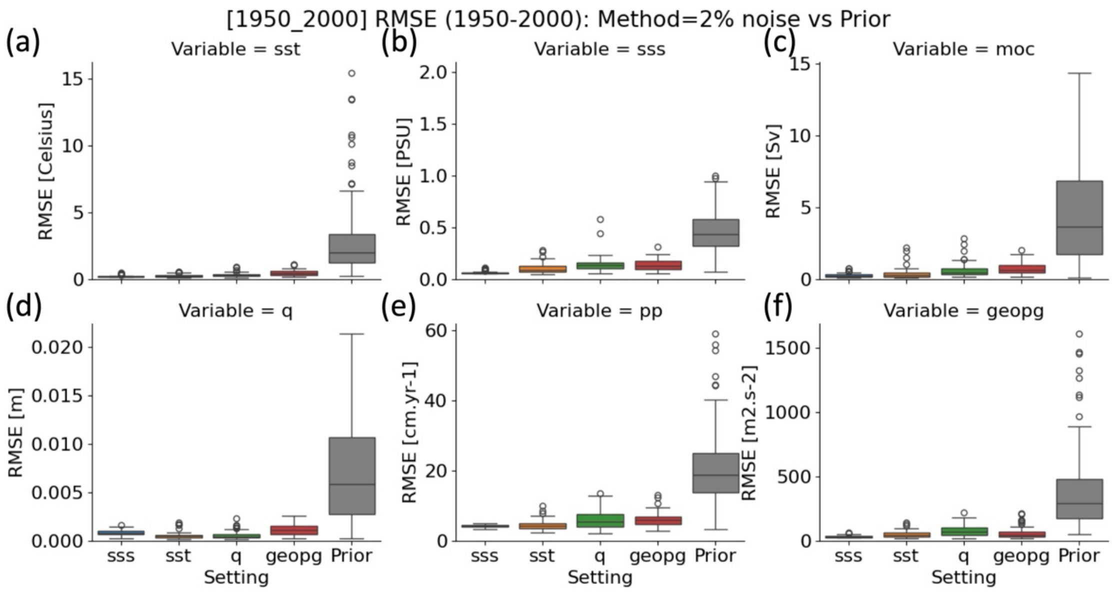


Figure S5. Simulation Accuracy of posterior simulation data in the observation period (50-year averaged climatology in 1950-2000 model year). (a-f) RMSEs with 2%-error observation. The blue, orange, green and red boxplots show the RMSE distribution of posterior simulation in the observation period with SSS, SST, Relative humidity and geopotential height as observation variable, respectively. The horizontal and vertical axes are observation setting and RMSE 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.

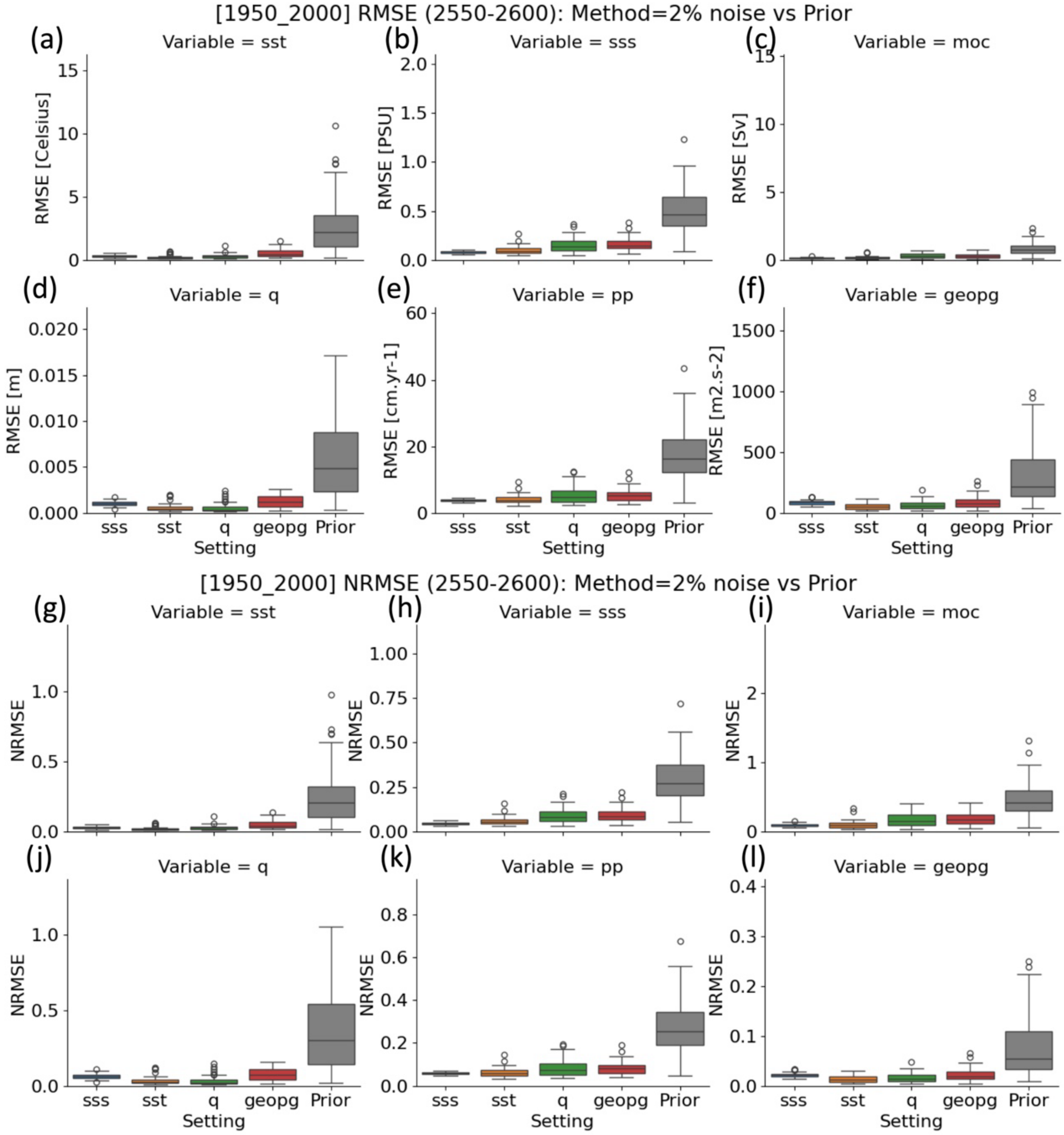


Figure S6. Prediction Accuracy of posterior simulation data in 50-year mean climatology in 2550-2600 model year. (a-f) RMSE with the multiple observation variables. (g-l) NRMSE with the multiple observation variables. The blue, orange, green, red and grey box plots show the RMSEs distribution with 2%-error 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 SST, SSS, stream function of meridional overturning circulation, surface air temperature, precipitation and geopotential height, respectively.

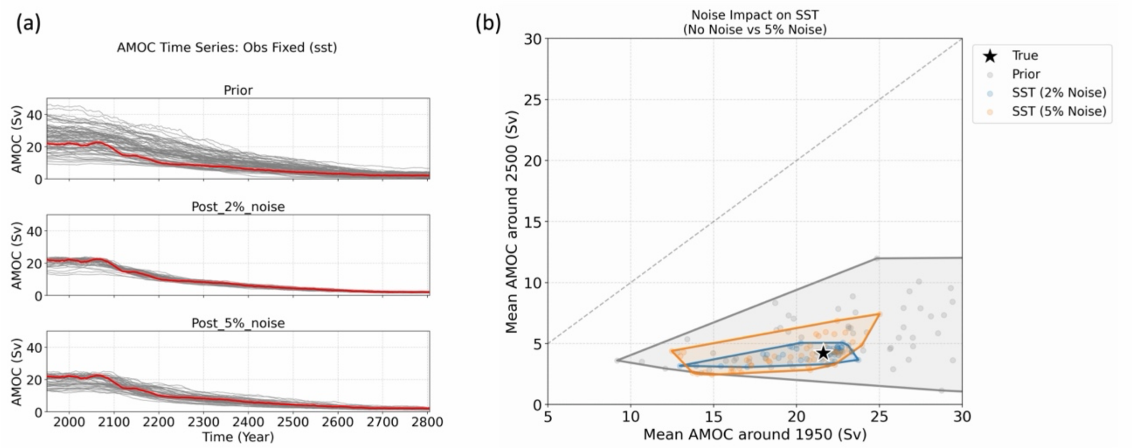


Figure S7. (a-c) AMOC timeseries of the posterior simulations (a) Without any observation, (b) with 2%-error SST observation, (c) with 5%-error SST observation. Grey and red lines in the panels are the posterior and synthetic true simulations, respectively. (d) The scatter plots of reproduced and predicted 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.

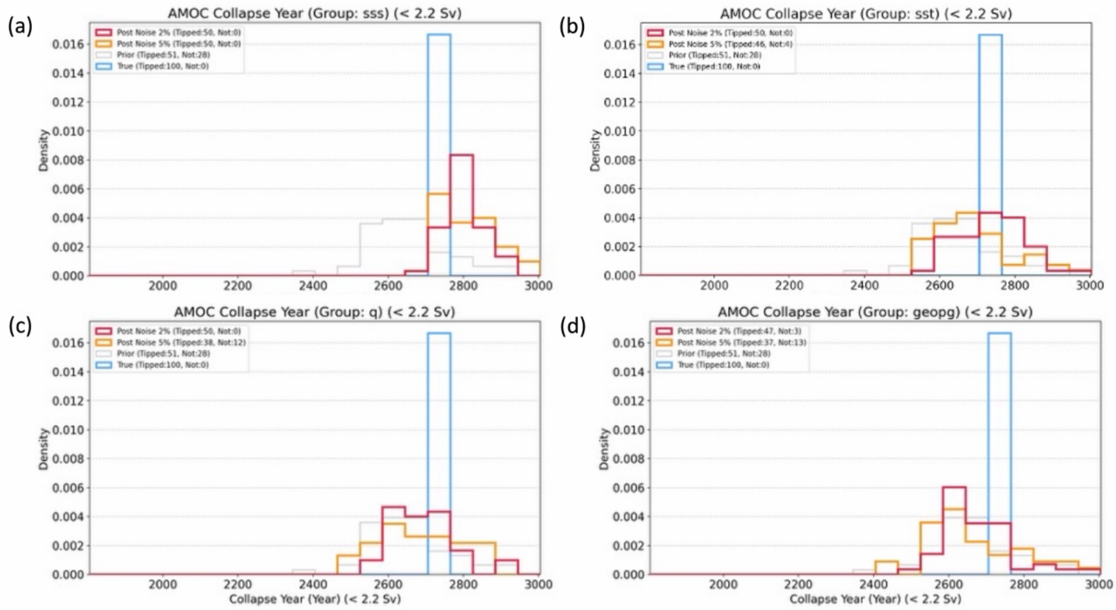


Figure S8. Tipping timing distribution across all the observation (a) SSS, (b) SST, (c) Relative humidity (Q), (d) Geopotential height (Geopg). The blue, grey, red and orange histograms show the tipping timing distributions of synthetic true simulations, prior simulations, posterior simulations with 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.

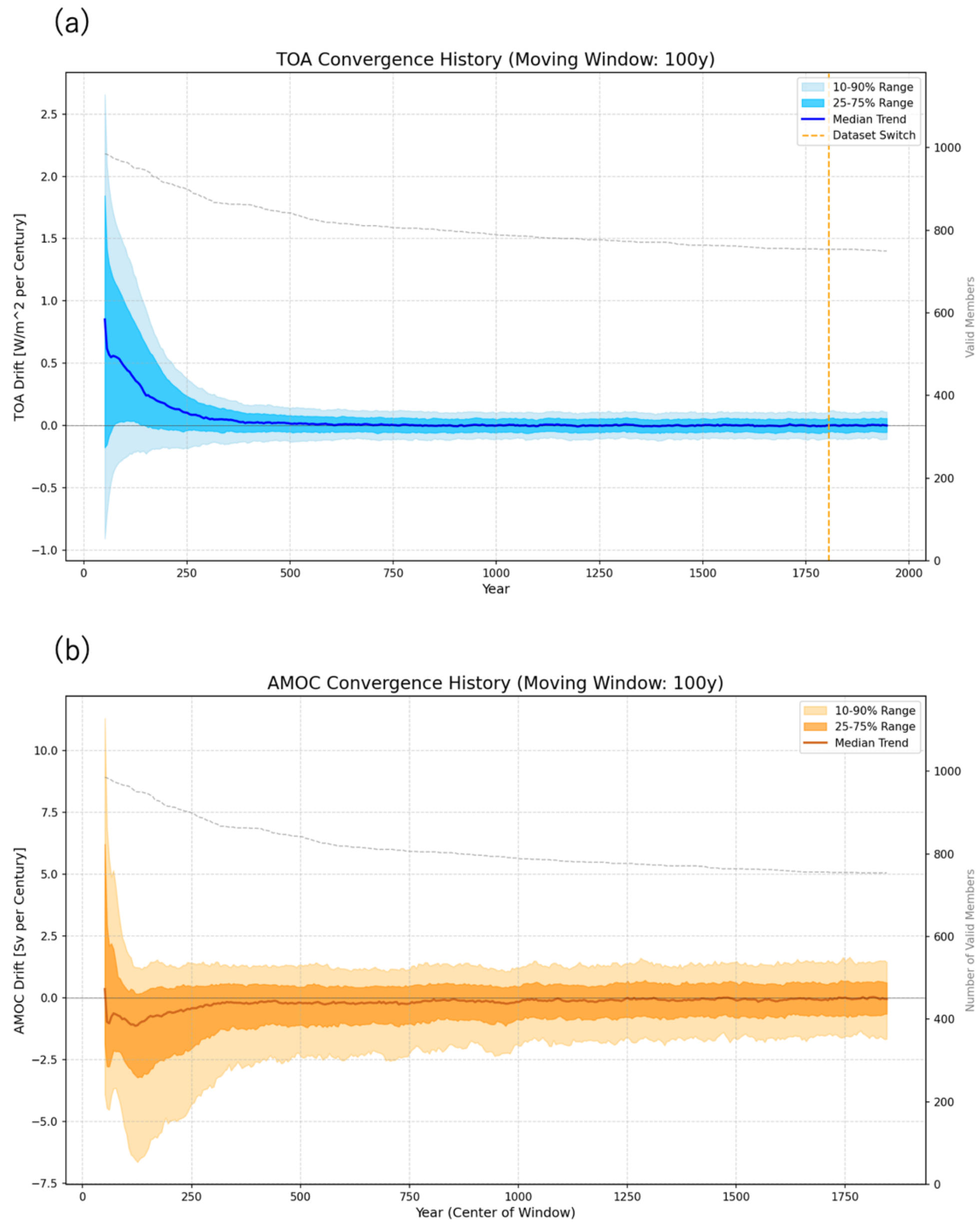


Figure S9. Convergence diagnostics for the coupled climate model ensemble using a 100-year moving window analysis. The panels show the temporal evolution of drift in (a) global mean top-of-atmosphere (TOA) net radiation and (b) the Atlantic Meridional Overturning Circulation (AMOC) maximum strength. The drift value at time t is a mean in the window from t-50 to t+50. In both panels, the solid lines represent the ensemble median trend, while the dark and light shaded regions correspond to the interquartile range (25th–75th percentiles) and the 10th–90th percentile range, respectively.

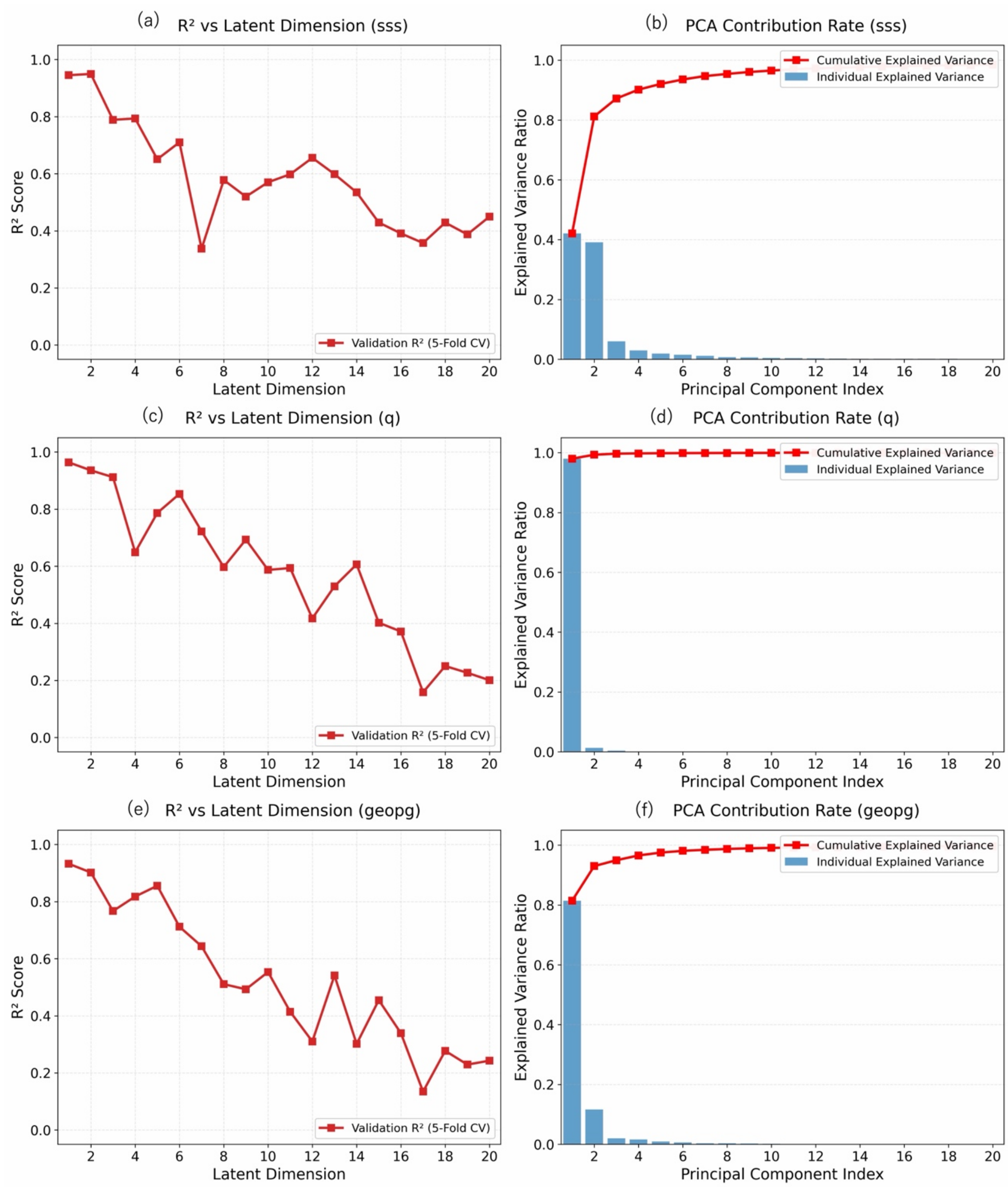
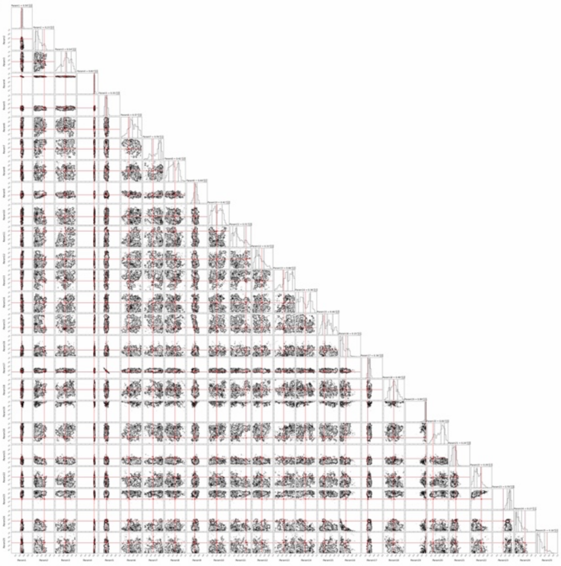
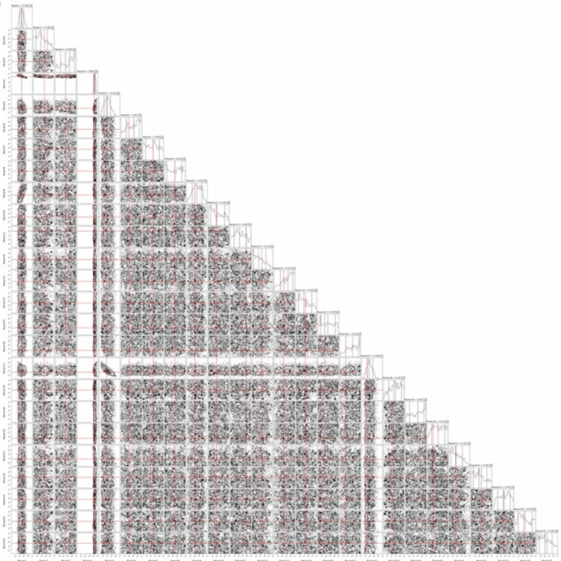


Figure S10. The performance of the surrogate model for estimating SSS, Q, GEOPG distribution from the input parameters. (a, c, e) Training (blue) and Validation (red) determinant coefficients of the surrogate model. (b, d, f) The Explained variance ratio and its cumulative value of the principal component 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. (a,b), (c,d) and (e,f) are the results with SSS, Q and GEOPG, respectively.

(a)



(b)



(c)

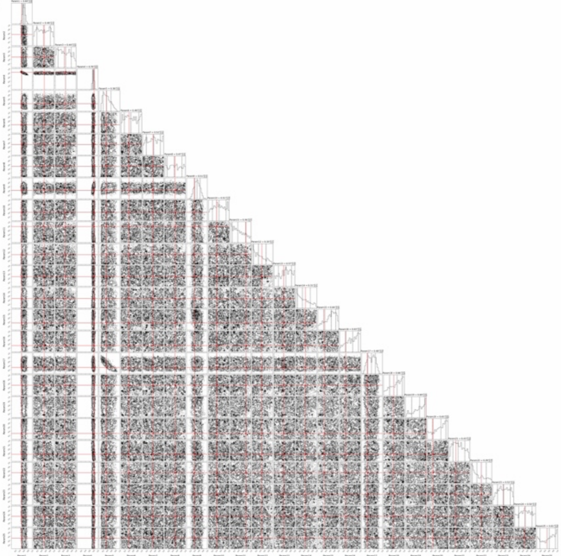


Figure S11. Corner plot of the estimated posterior parameter distributions using (a) SST, (b) Relative Humidity (Q) and (c) Geopotential height (GEOPG) observations with a 1% error. Diagonal panels display the one-dimensional marginal posterior distribution for each parameter, while off-diagonal panels 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.

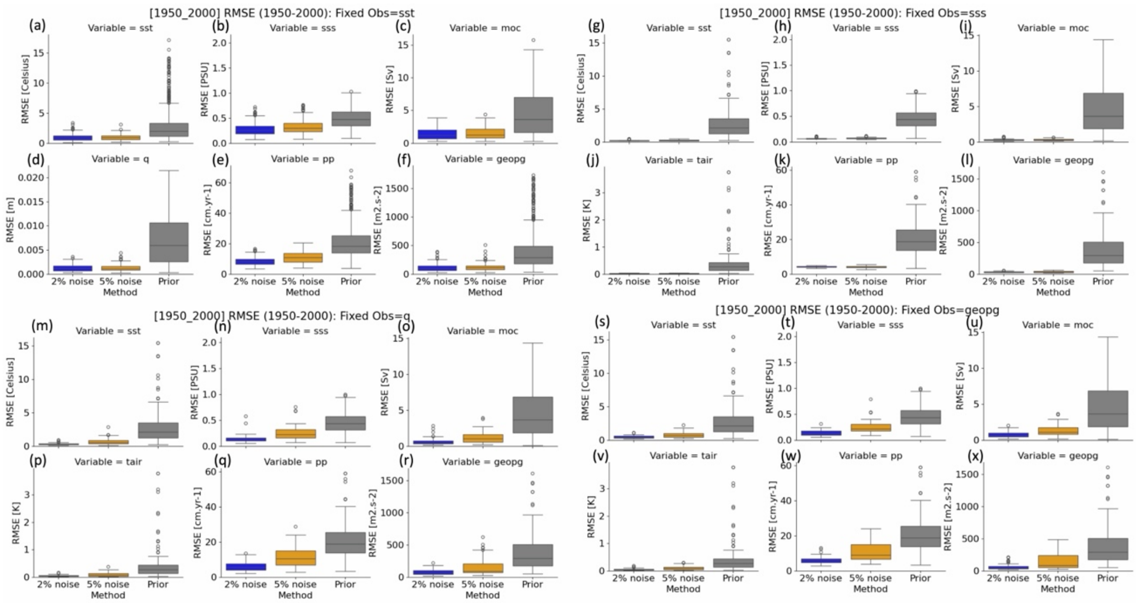


Figure S12. The dependency of simulation accuracy in the observation period on the observation noise setting. Blue, orange and grey boxplots show the results with 2%-error observation, 5%-error observation and without any observation, respectively. (a,b,c,d,e,f), (g,h,ij,k,l), (m,n,o,p,q,r) and (s,t,u,v,w,x) show the results with SST, SSS, Relative humidity (Q) and geopotential height (GEOPG), respectively. (a,g,m,s), (b,h,n,t), (c,i,o,u), (d,j,p,v), (e,k,q,w) and (f,l,r,x) show the simulation accuracy in the observation period of sst, sss, meridional overturning circulation stream function (moc), relative humidity (q), orecipitation (pp) and geopotential height (geopg), respectively.

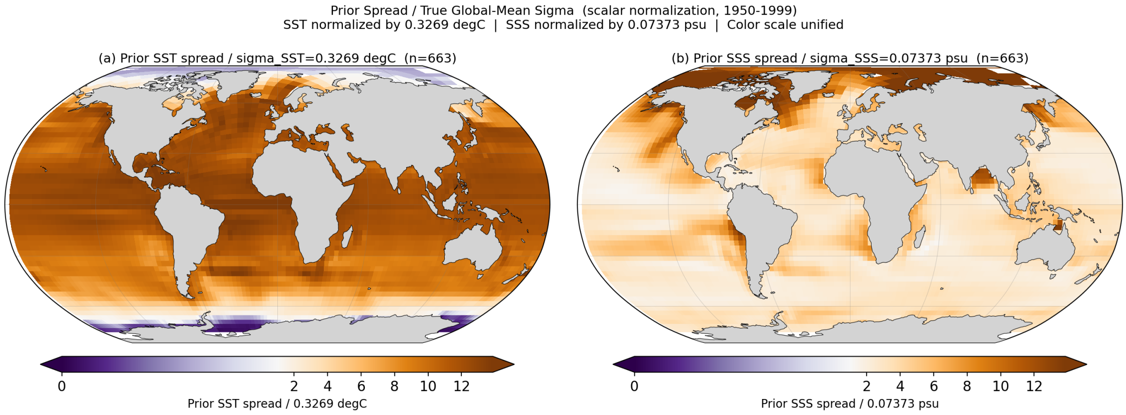


Figure S13. The distribution of signal-to-noise ratio of (a) SST and (b) SSS. The color is determined based on the signal to noise ratio which is calculated by prior ensemble spread divided by standard deviation of global interannual variability. The darker brown color shows the signal is larger at the grid while the darker purple color shows the noise is larger at the grid. While SST shows high signal-to-noise ratio in the low-latitude and mid-latitude region and low signal-to-noise ratio in the high-latitude region, SSS records high signal-to-noise ratios in the high-latitude area in the north hemisphere, which implies that we can resolve high-latitude ocean surface situations more precisely using SSS than SST. The maps are generated by the authors using Python's Cartopy package.

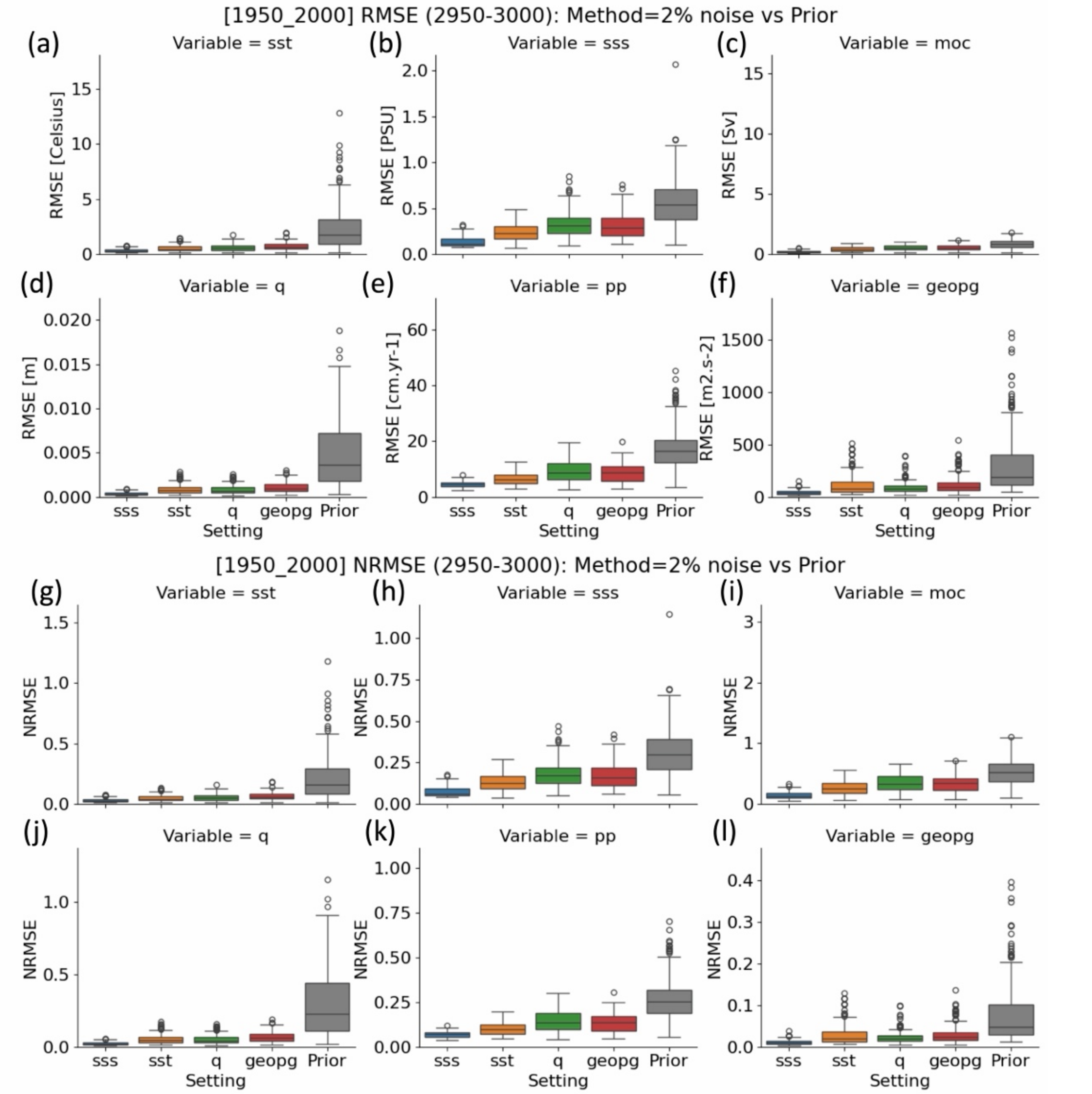


Figure S14. Prediction Accuracy of posterior simulation data in 2950-3000 model year. (a-f) RMSE with the multiple observation variables. (g-l) NRMSE with the multiple observation variables. The blue, orange, green, red and grey box plots show the RMSEs distribution with 2%-error SSS, SST, Relative humidity (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 SST, SSS, stream function of meridional overturning circulation, surface air temperature, precipitation and geopotential height, respectively.

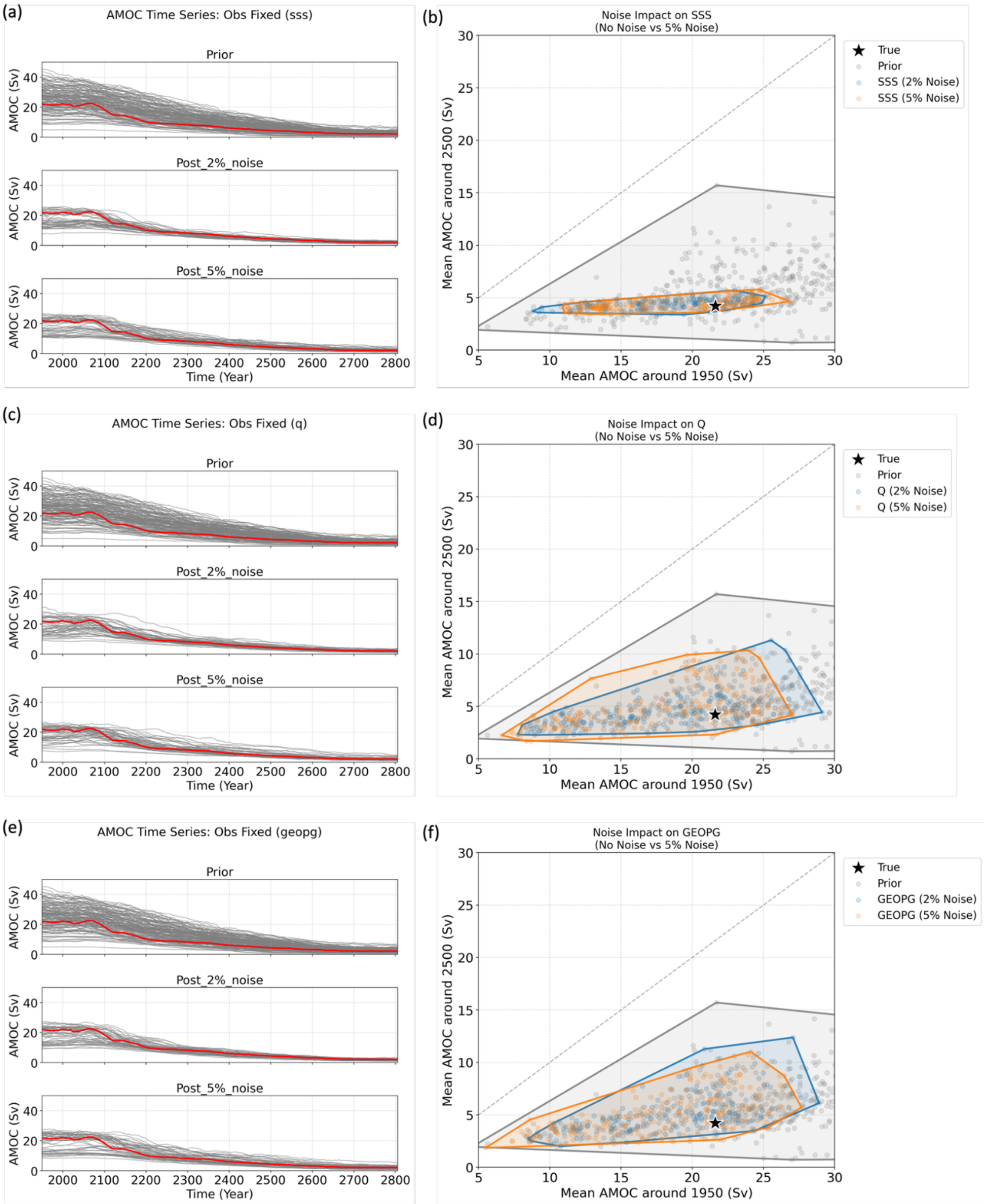


Figure S15. (a-c) AMOC timeseries of the posterior simulations (a) Without any observation, (b) with 2%-error SST observation, (c) with 5%-error SST observation. Grey and red lines in the panels are the posterior and synthetic true simulations, respectively. (d) The scatter plots of reproduced and predicted 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.

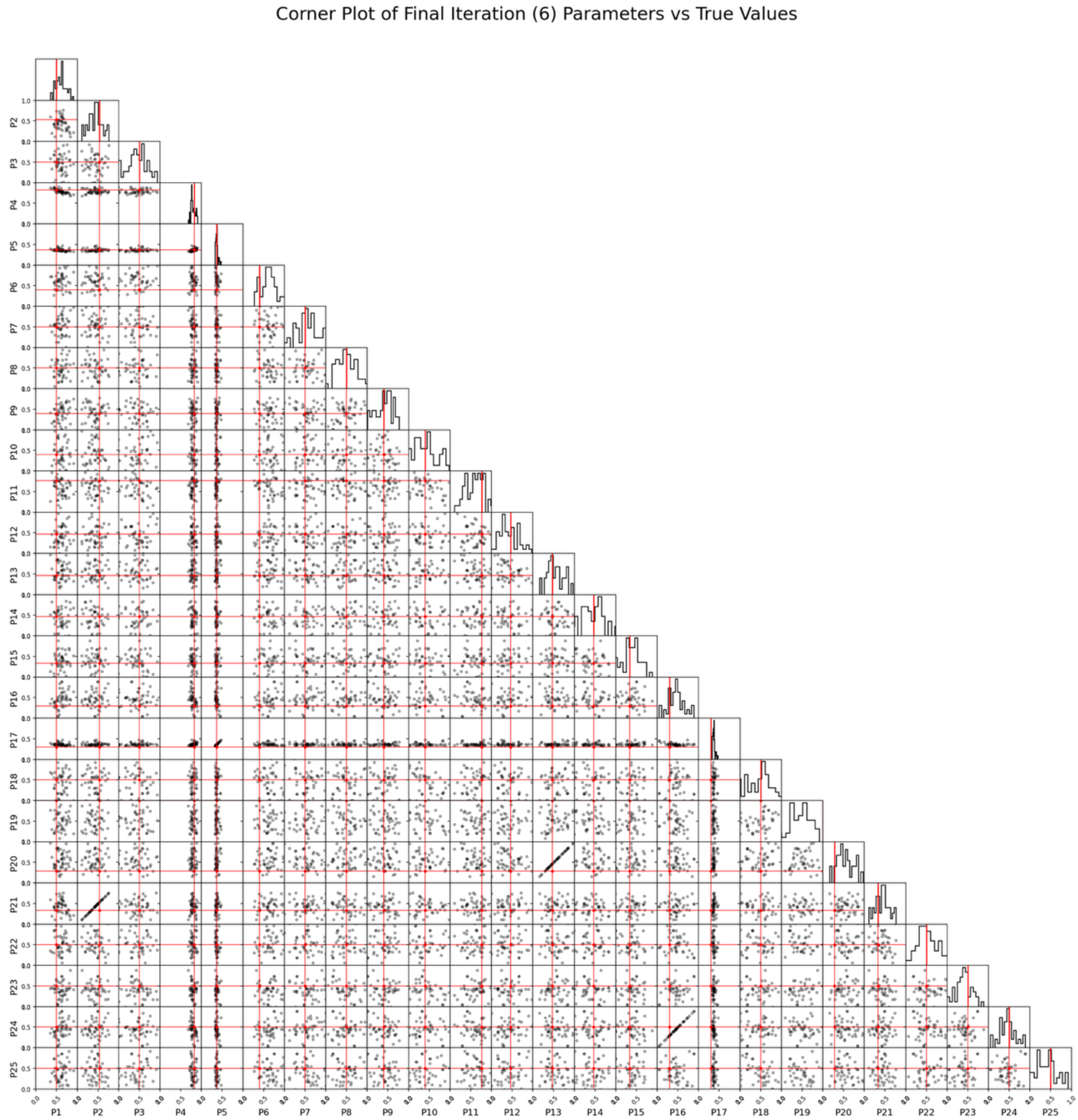


Figure S16. The scatter plot of the posterior 50 parameter obtained by Ensemble Kalman Sampler with 10%-error SST observation. We did five iterations to get obtain the posterior distribution. Diagonal panels display the one-dimensional marginal posterior distribution for each parameter, while off-diagonal 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.