

Supporting Information for

“Initialization and ensemble generation for decadal climate predictions: A comparison of different methods”

Iuliia Polkova¹, Sebastian Brune¹, Christopher Kadow², Vanya Romanova³,
Gereon Gollan⁴, Johanna Baehr¹, Rita Glowienka-Hense³,
Richard J. Greatbatch⁴, Andreas Hense³, Sebastian Illing², Armin Köhl¹,
Jürgen Kröger⁵, Wolfgang A. Müller⁵, Klaus Pankatz⁶ and Detlef Stammer¹

¹Institute of Oceanography, Universität Hamburg, CEN, Hamburg, Germany.

²Institute for Meteorology, Freie Universität Berlin, Berlin, Germany.

³Meteorological Institute, University Bonn, Bonn, Germany.

⁴GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany.

⁵Max Planck Institute for Meteorology, Hamburg, Germany.

⁶Deutscher Wetterdienst, Offenbach am Main, Germany.

Contents

1. Figure S1 to Section 3.1 on SAT correlation skill.
2. Figures S2 to Section 3.1 on HC700 correlation skill.
3. Figure S3 to Section 3.3 on HC700 anomalies of the North Atlantic subpolar gyre.
4. Figures S4 to Section 3.3 on the basin-wide AMOC mean for lead year 5.
5. Figures S5 to Section 3.3 on the basin-wide AMOC standard deviation for lead year 5.
6. Figures S6 to Section 3.4 on SST ESS, ensemble spread and root-mean-squared error for lead year 1.
7. Figures S7 to Section 3.4 on SST ESS, ensemble spread and root-mean-squared error for lead year 5.
8. Figure S8 to Section 3.1 on SAT skill for a “multi-model” ensemble.

Corresponding author: Iuliia Polkova, iuliia.polkova@uni-hamburg.de

Additional Figures to Section “3.1 Prediction Skill for Surface Air Temperature (SAT) and the upper-ocean heat content (HC700)”

The SAT and HC700 prediction skill from the test-suite experiments and the reference system Preop-LR in terms of correlation coefficients in Figs. S1-S2. Apart from the eastern and central Pacific Ocean and the frontal area of the western-boundary currents, Preop-LR correlates well with SAT from the HadCRUT4 verification data set (Fig. S1). In the Pacific Ocean, there is an area of reduced skill in lead years 2-5 resembling a characteristic oscillation pattern, i.e., the Pacific Decadal Oscillation. Though correlation skill in this area of the Pacific Ocean tends to come closer to positive values in all test-suite experiments, it is still not statistically significant.

For HC700, in comparison to Preop-LR, the EnKF and FAI reveal better agreement with the NOAA/NODC heat content in the eastern and central Pacific Ocean (Fig. S2). The EnKF shows negative correlation along the Canary Current and the Indian Ocean. The EDF and the BV, which build on the Preop-LR assimilation, mimic low skill areas of Preop-LR.

Additional Figures to Section “3.3 Skill for the North Atlantic Ocean”

As discussed in the main manuscript, Preop-LR and the EDF and to a smaller extent BV and FAI show a “spike” around 2005. A similar “spike” is also present for the previous MiKlip systems. It comes from hindcasts started in 2000 and 2001 shown in Figure S3. For these initialization cases, the BV and FAI hindcasts show cooler SPG HC700 than in the verification dataset, and Preop-LR and the EDF show a decrease of the SPG HC700.

In Preop-LR, the AMOC cell has an expected structure: with a clear separation between the deep-water and the bottom-water cells (Figure S4). The BV and FAI show comparable to Preop-LR overturning cell patterns. The EnKF and the EDF shows extended to the ocean bottom deep-water cell. In terms of the AMOC standard deviation (Figure S5), all the initialized hindcasts simulate reduced variability as compared to Preop-LR, most of them (Preop-LR, BV and FAI) even lower than what is shown by the historical simulation. In yr5, the EDF exhibits stronger variability than in Preop-LR, particularly in the tropical North Atlantic. The EnKF shows high variability in the tropics. In BV, the variability pattern has a similar shape as in Preop-LR though somewhat

weaker. The FAI AMOC variability, apart from the latitudinal band around 40N, is about half as strong as in Preop-LR.

Additional Figures to Section “3.4 Ensemble Spread Performance”

The spread-to-error diagnostic shown in Figure S6 suggests that Preop-LR and the test-suite largely underestimate the spread of surface temperature in the first lead year. The advanced ensemble generation methods such as BV and EnKF improve the spread at this lead time. Also MODINI increases the spread in the first lead year. The EnKF seems to generate too large spread, especially in the extratropics. FAI does not have impact on the spread but has smaller RMSE. EDF by design reduces spread, but the RMSE is close to that of Preop-LR. For later leadtimes, the spread approaches the level of the errors; this feature is common to all the initialized hindcasts in the test-suite (lead year 5 in Figure S7). Note that ESS shown in the manuscript is equivalent to comparing ensemble variance and MSE, while Figures S6 and 7 compare ensemble spread and RMSE. Also the ESS in the manuscript is calculated for the standardized variables, while the spread-to-error ratio not.

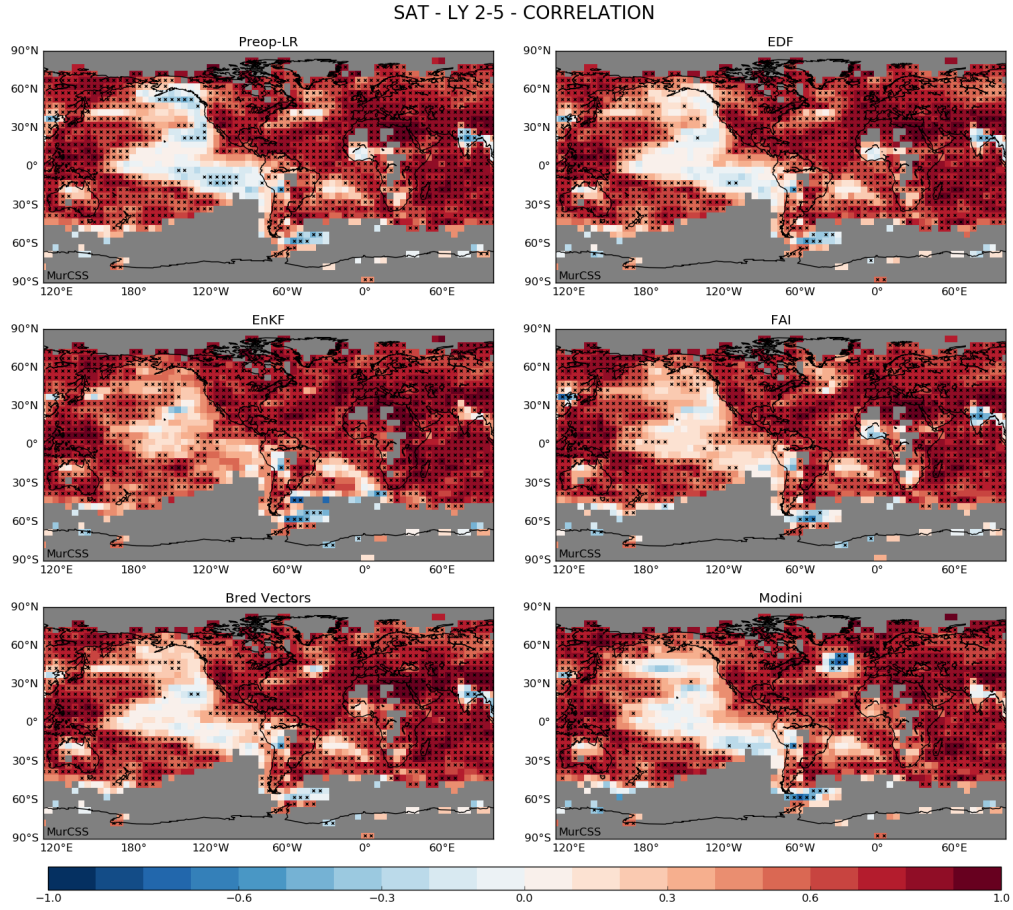


Figure S 1. Correlation skill for SAT w.r.t. HadCRUT4 for lead years 2-5 from Preop-LR (top left panel) and the test-suite experiments (first, second and third rows). Hindcasts initialized from 1960 to 2011 are evaluated. Stippling indicates significant correlation values as estimated with the bootstrap method at the 95% confidence level.

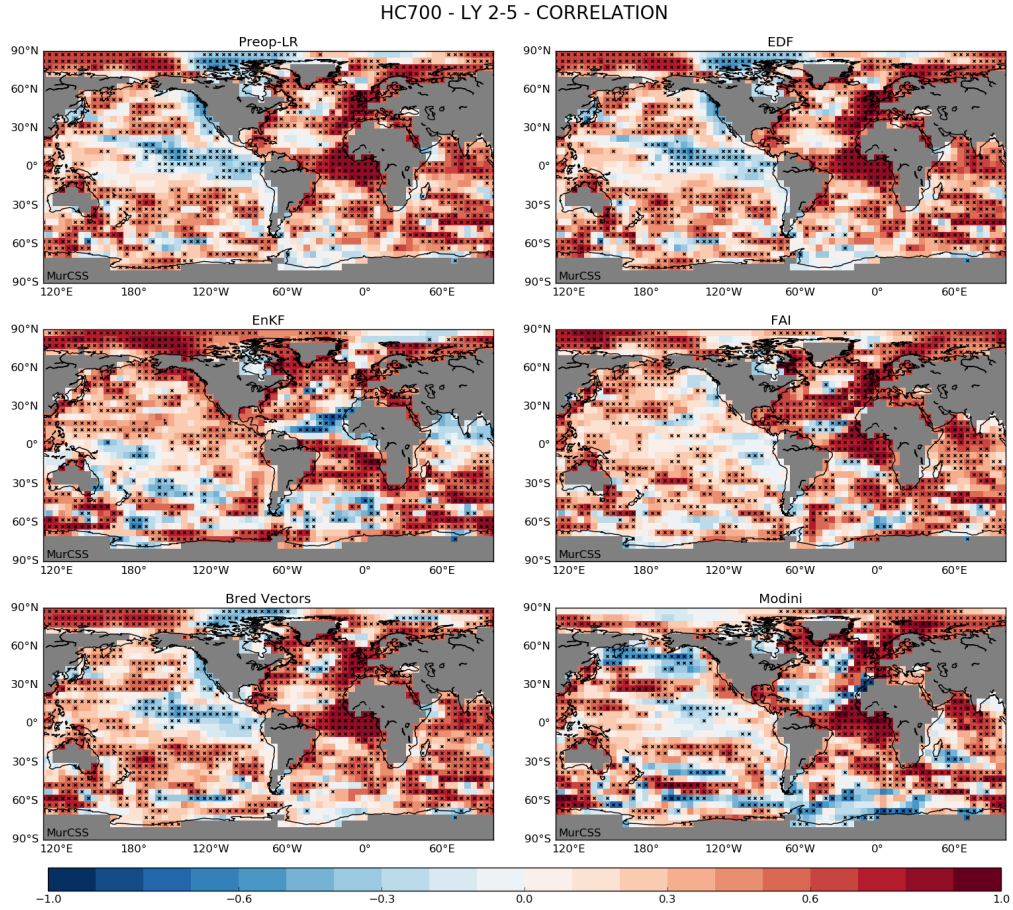


Figure S 2. MESS for SAT w.r.t. the HadCRUT4 climatology for lead years 2-5 from Preop-LR (top left panel). MESS for SAT w.r.t. Preop-LR for the test-suite experiments (first, second and third rows). Hindcasts initialized from 1960 to 2011 are evaluated. The range of MESS is from $-\infty$ to $+1$. Stippling indicates significant MESS values as estimated with the bootstrap method at the 95% confidence level.

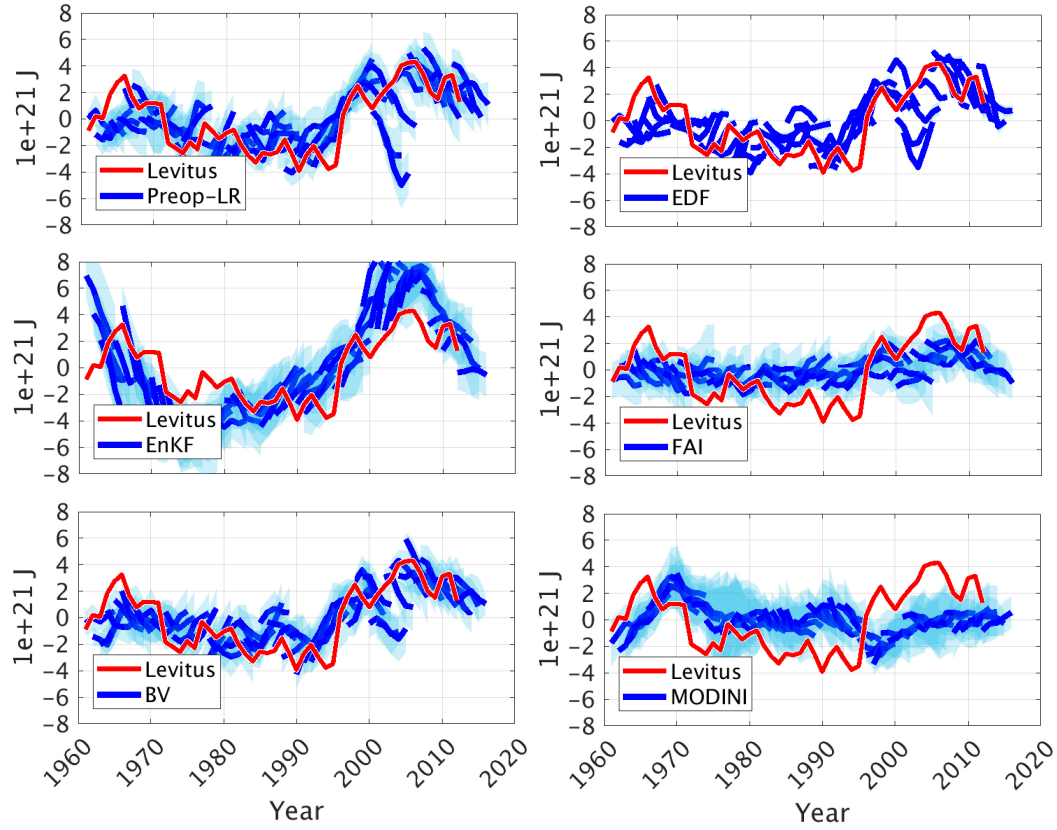


Figure S 3. Time series of HC700 anomalies (10^{21} J) for the North Atlantic subpolar gyre from 5 lead years of the initialized hindcasts (blue) and the NOAA/NODC data (red). Shading represents the ensemble spread.

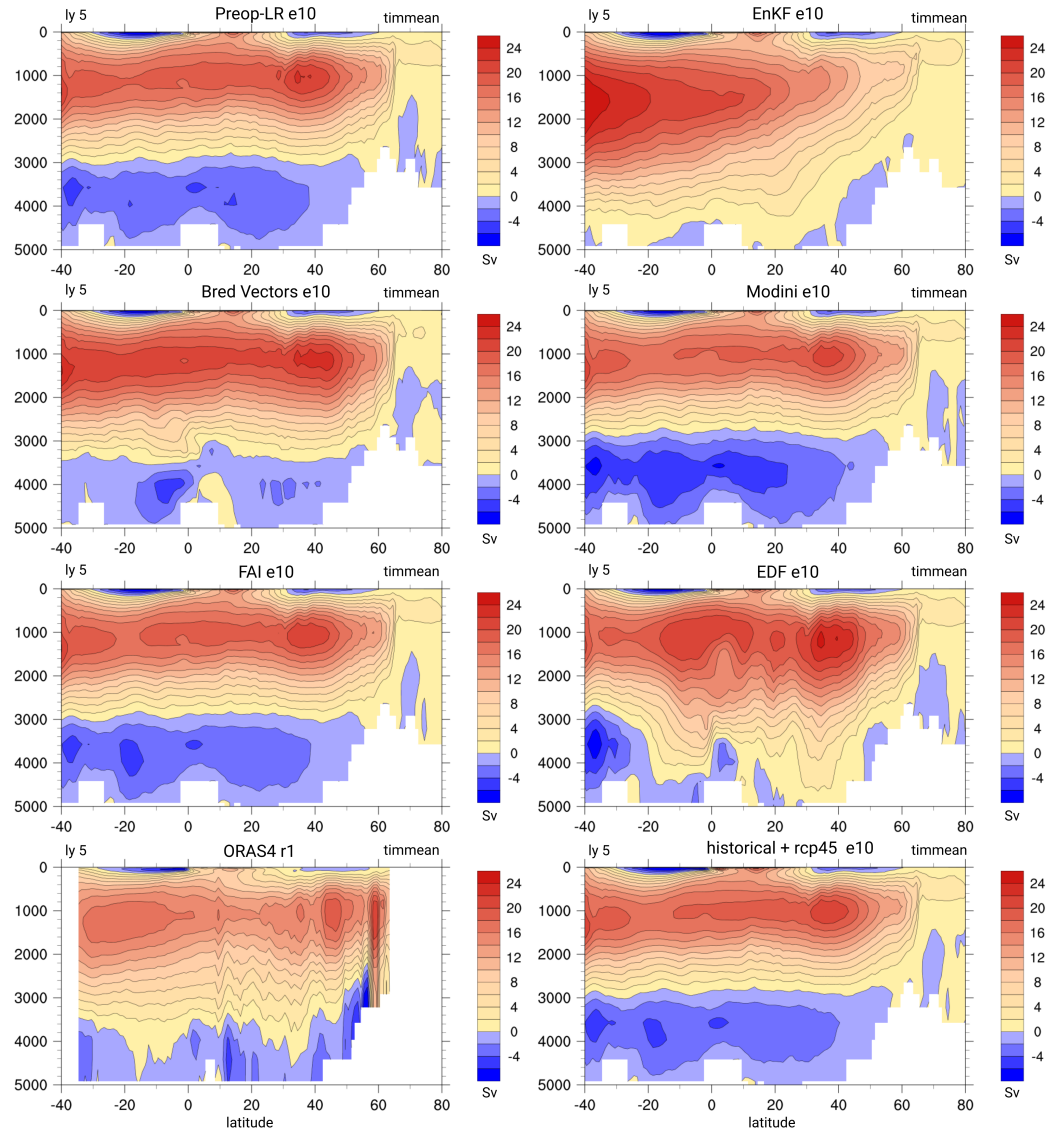


Figure S 4. Figure 14. The time mean of the Atlantic Meridional Overturning Circulation (in Sv) from the initialized hindcasts in the 5th lead year.

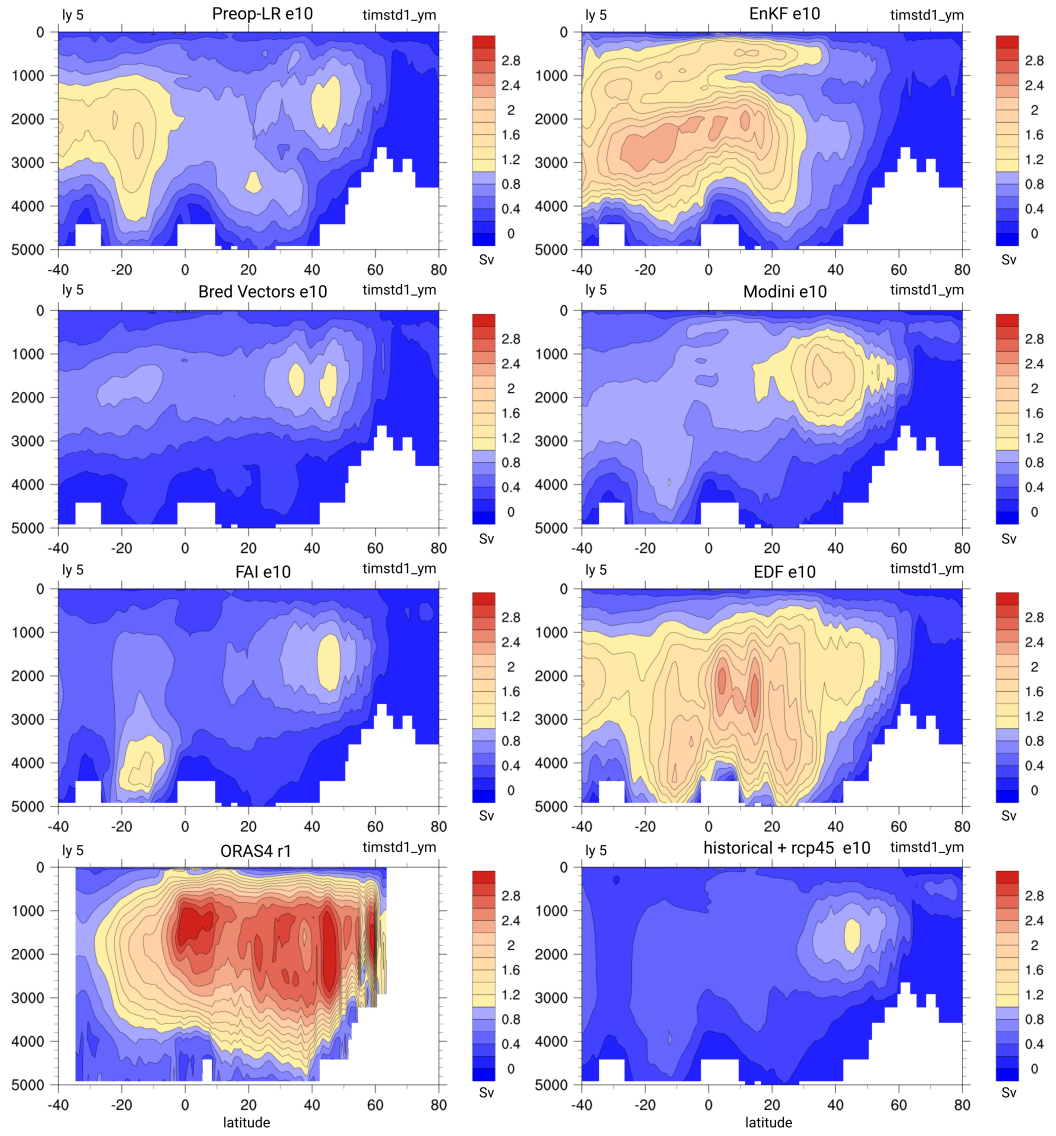


Figure S 5. The standard deviation of the Atlantic Meridional Overturning Circulation (in Sv) from the initialized hindcasts in the 5th lead year.

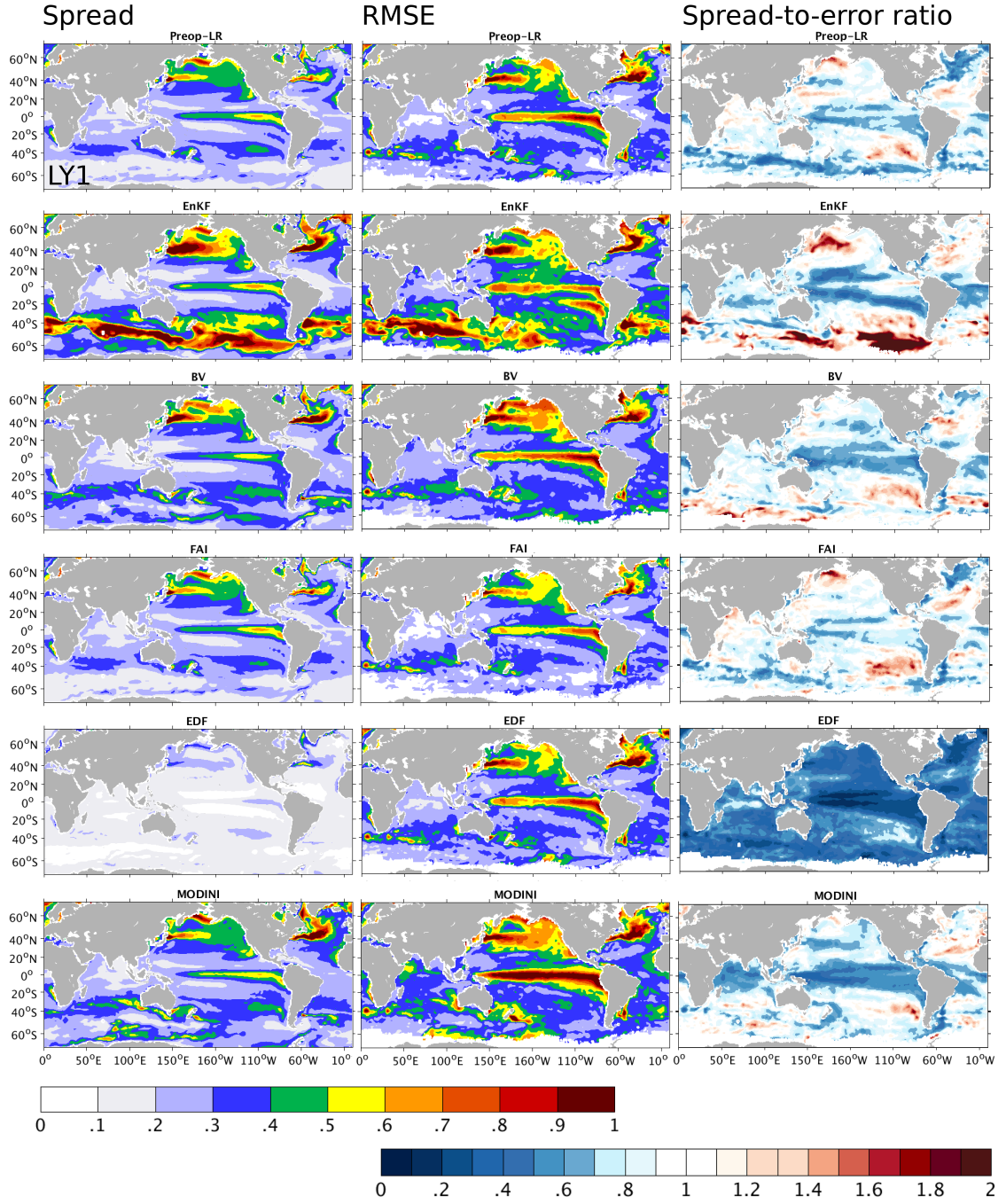


Figure S 6. SST spread (left panels, in degC), RMSE with respect to HadISST (middle panels, in degC) and spread-to-error-ratio (right panels) for lead year 1. The ratio is calculated as $\frac{\sqrt{(m+1)/m(\sigma^2)^{1/2}}}{RMSE}$, where $(\sigma^2)^{1/2}$ is the ensemble spread corrected with the ensemble size, m , dependent factor.

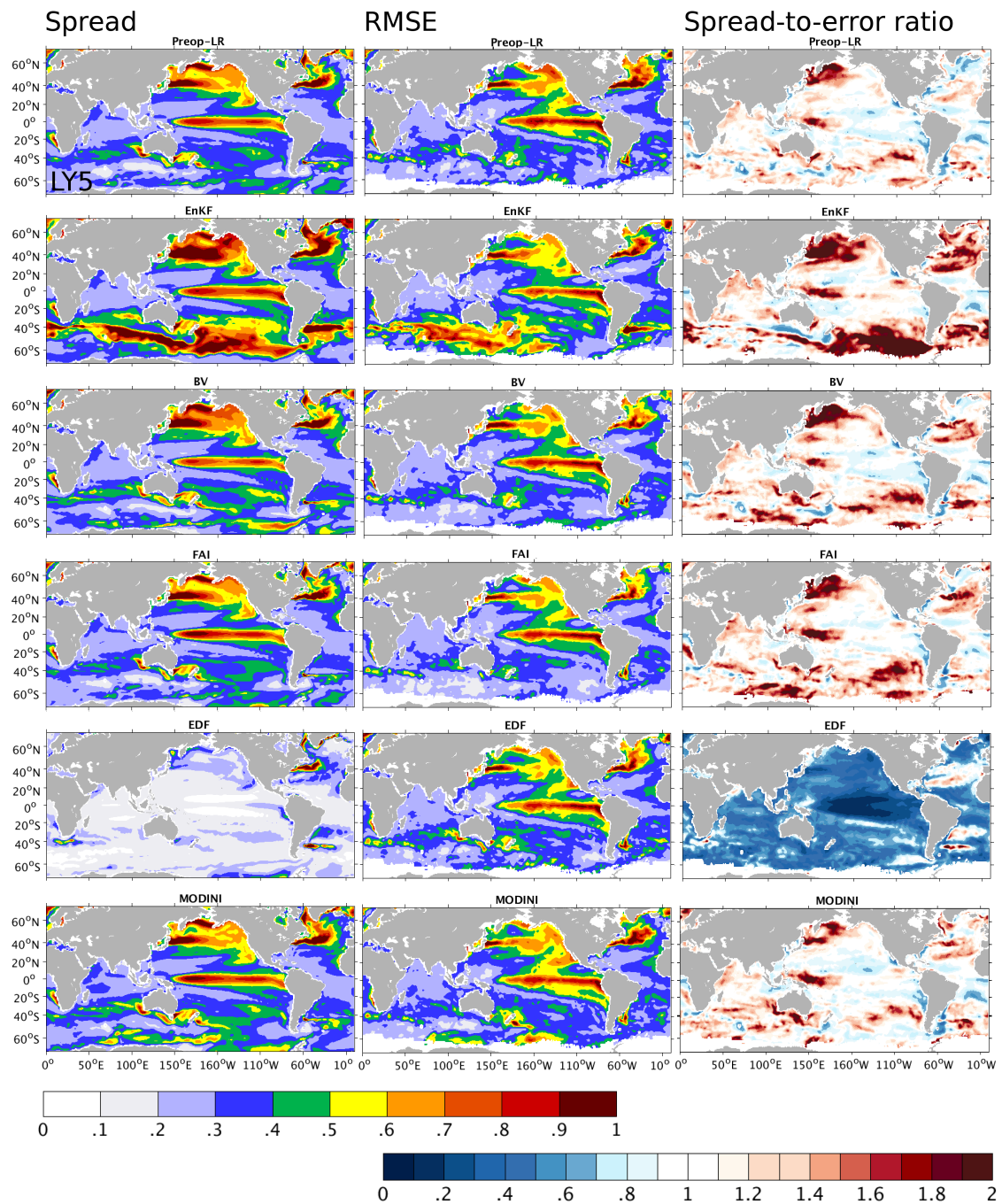


Figure S 7. SST spread (left panels, in degC), RMSE with respect to HadISST (middle panels, in degC) and spread-to-error-ratio (right panels) for lead year 5.

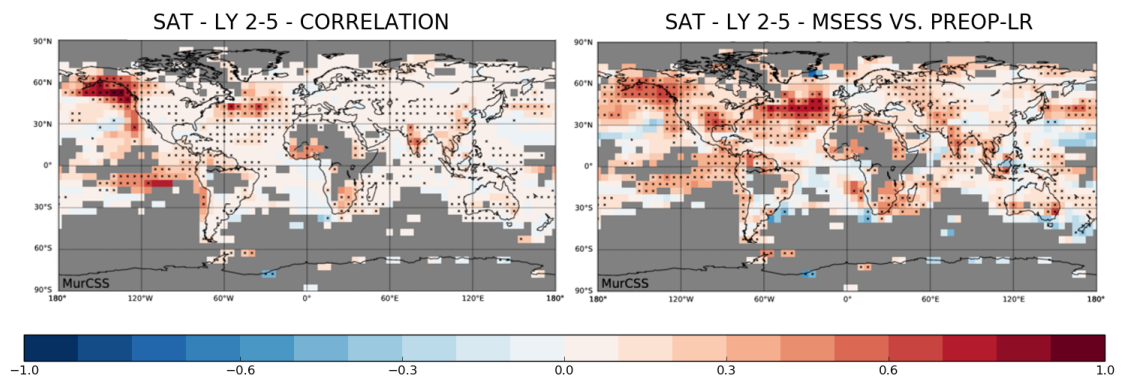


Figure S 8. Correlation skill difference between the multi-initialization ensemble mean and the Preop-LR (left) and MSESS w.r.t. Preop-LR (right) for SAT for lead years 2-5.