Supplementary materials for 'An 'Observational Large Ensemble' to compare observed and modeled temperature trend uncertainty due to internal variability'

Karen A. McKinnon, Andrew Poppick, Etienne Dunn-Sigouin, and Clara Deser

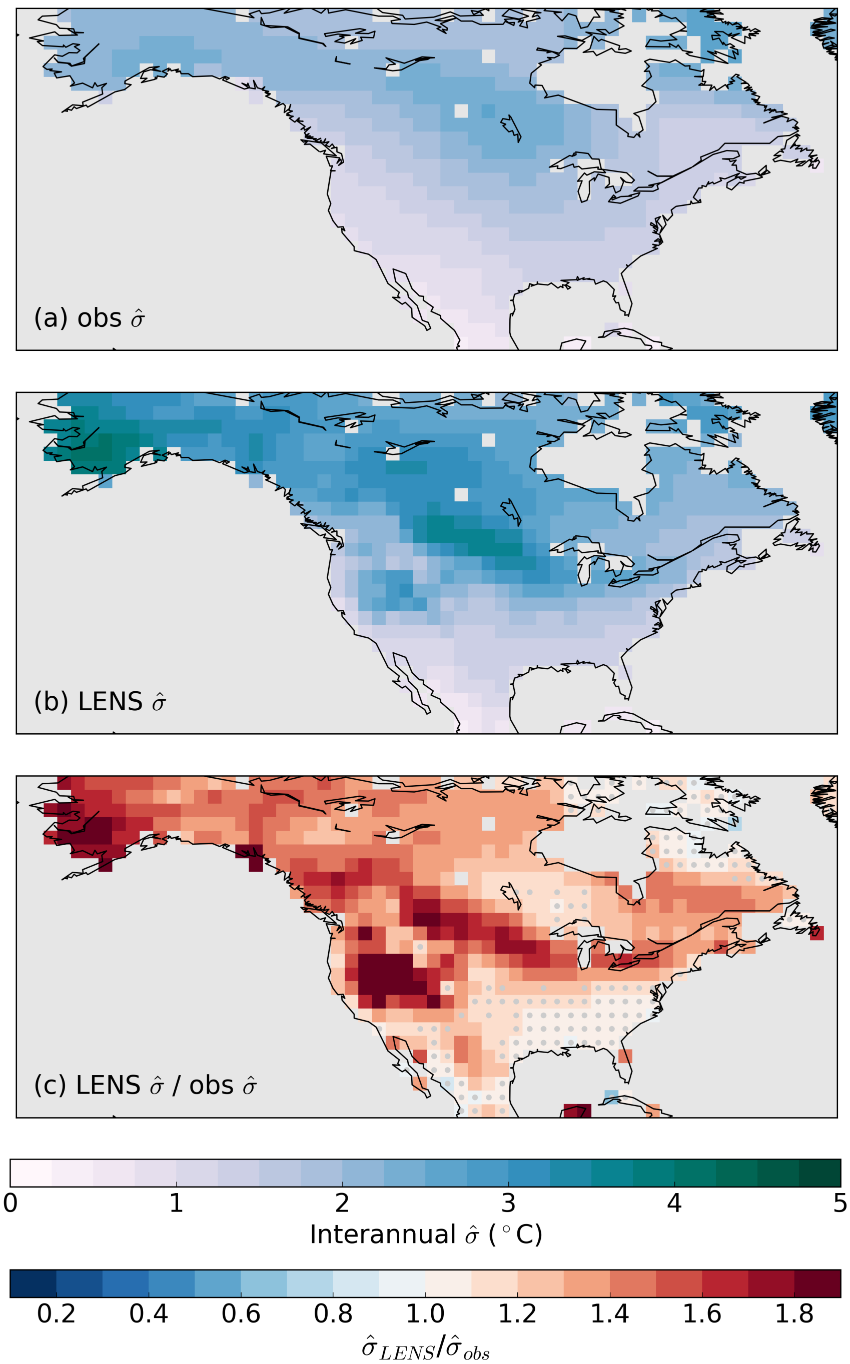

Figure S1: As in Fig. 1, but using the NASA GISTEMP dataset.

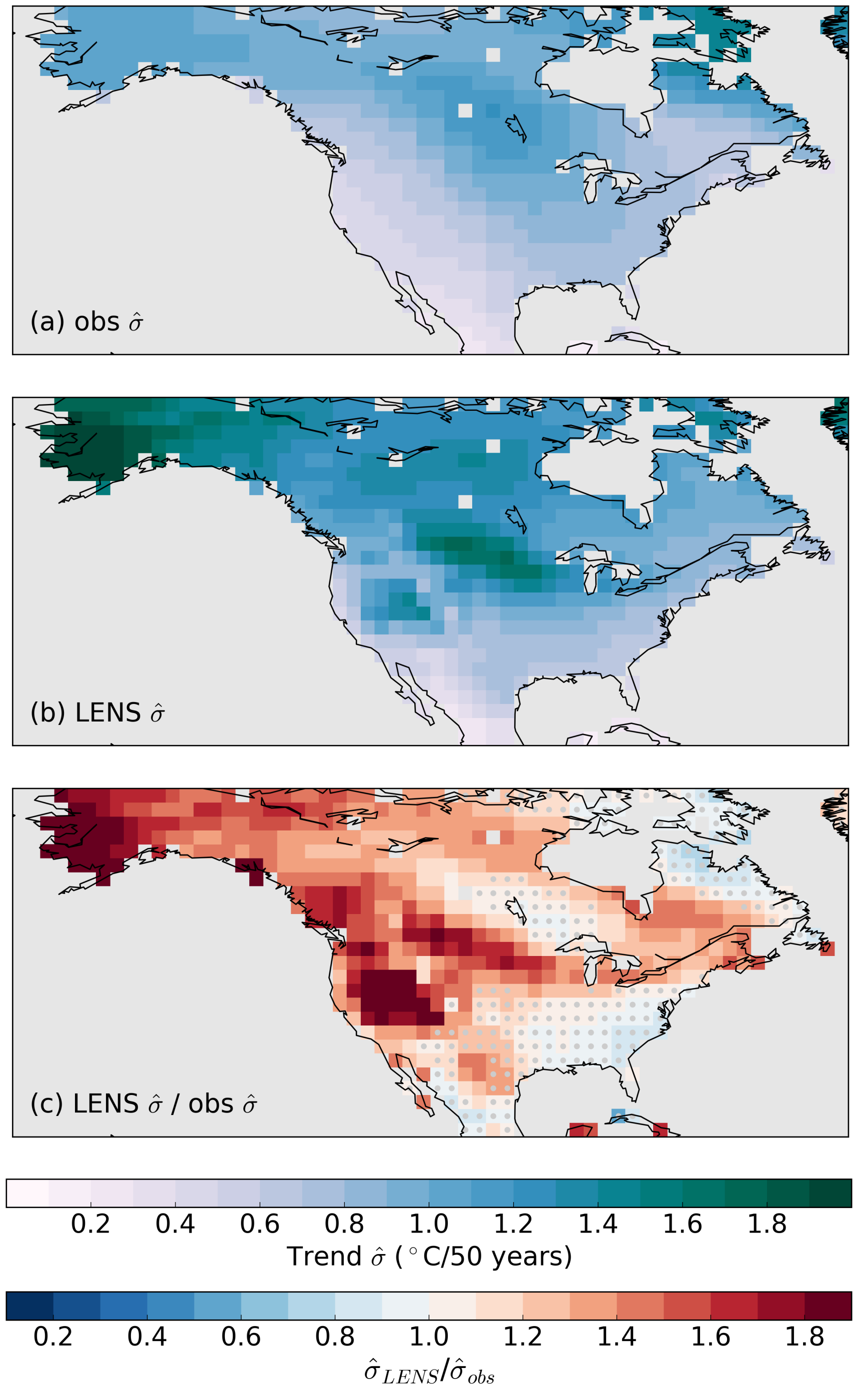

Figure S2: As in Fig. 5, but using the NASA GISTEMP dataset.

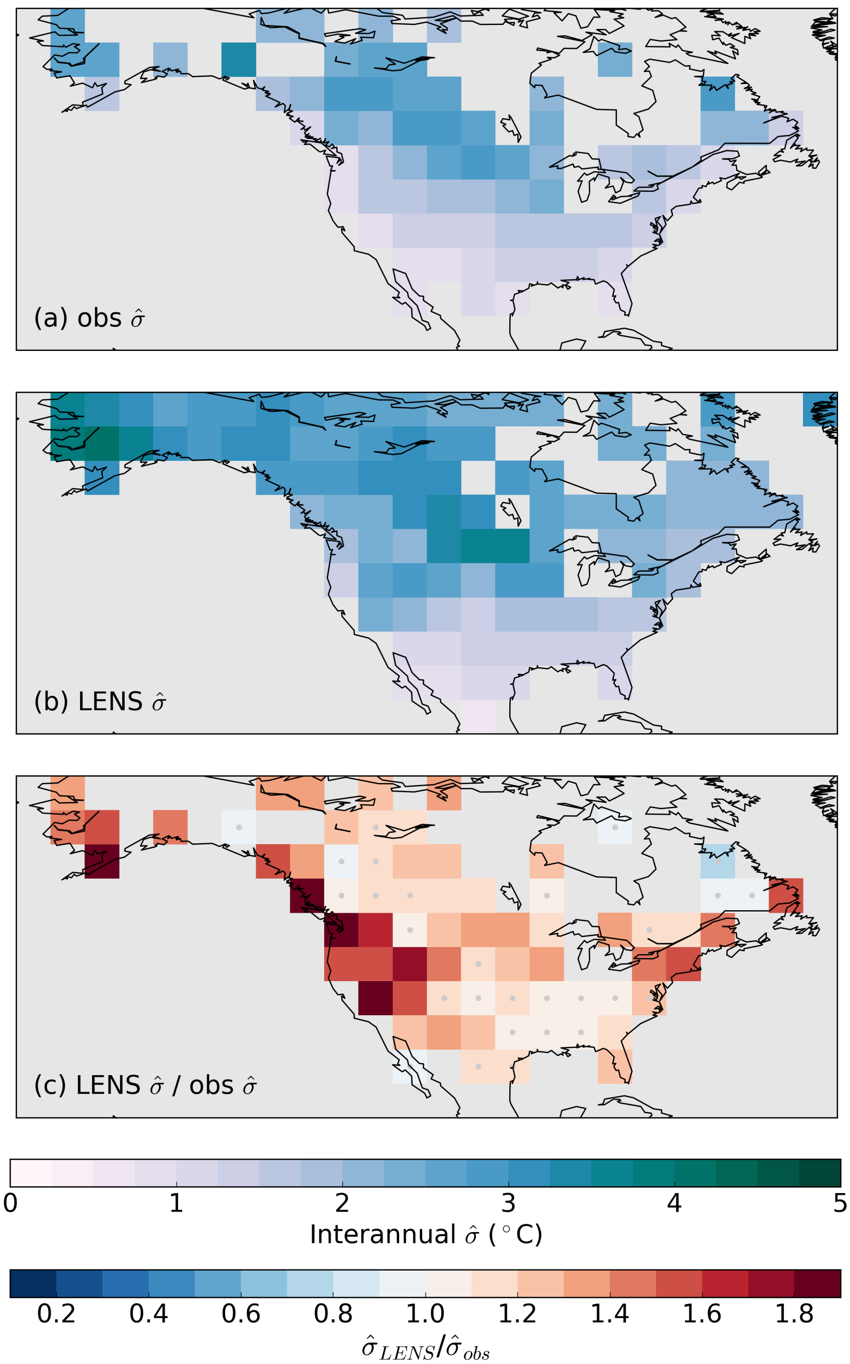

Figure S3: As in Fig. 1, but using the HadCRUT4 dataset.

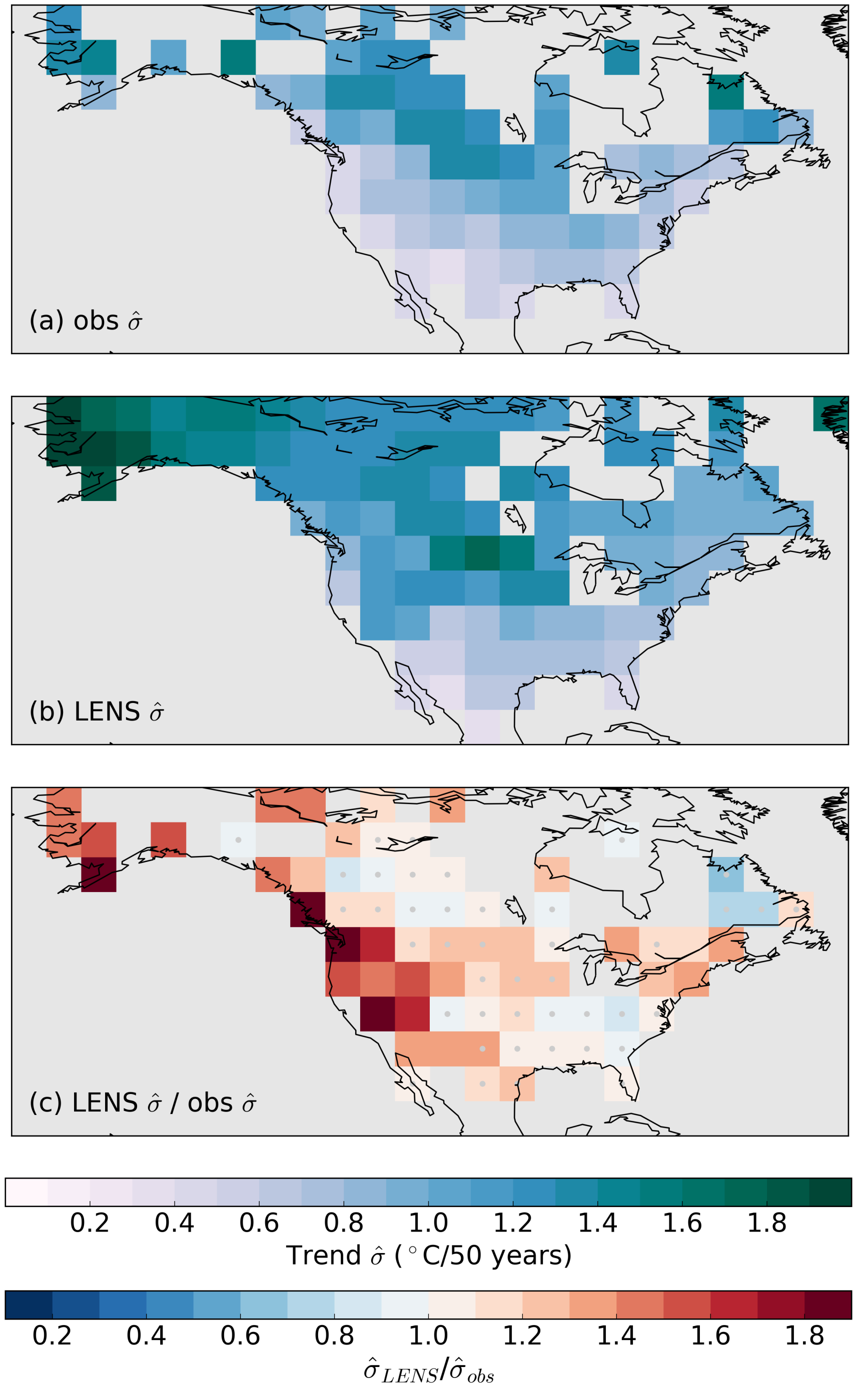

Figure S4: As in Fig. 5, but using the HadCRUT4 dataset.

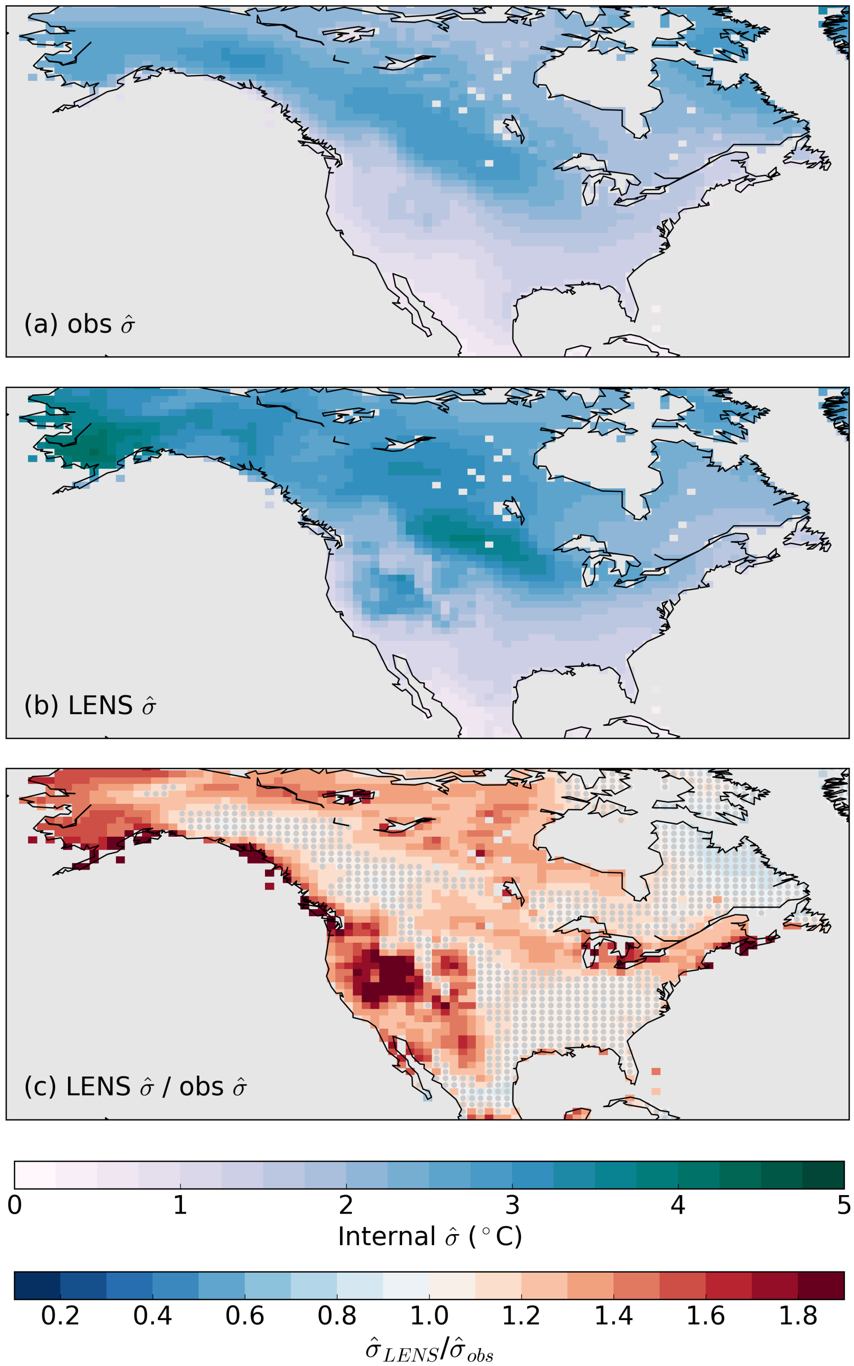

Figure S5: As in Fig. 1, but using the Dai et al. (2015) methodology to calculate the forced trend. Specifically, the forced trend at a given gridbox is calculated as the regression onto the global-mean ensemble-mean trend across the NCAR CESM1 Large Ensemble. Internal variability is estimated as the residual from this forced trend.

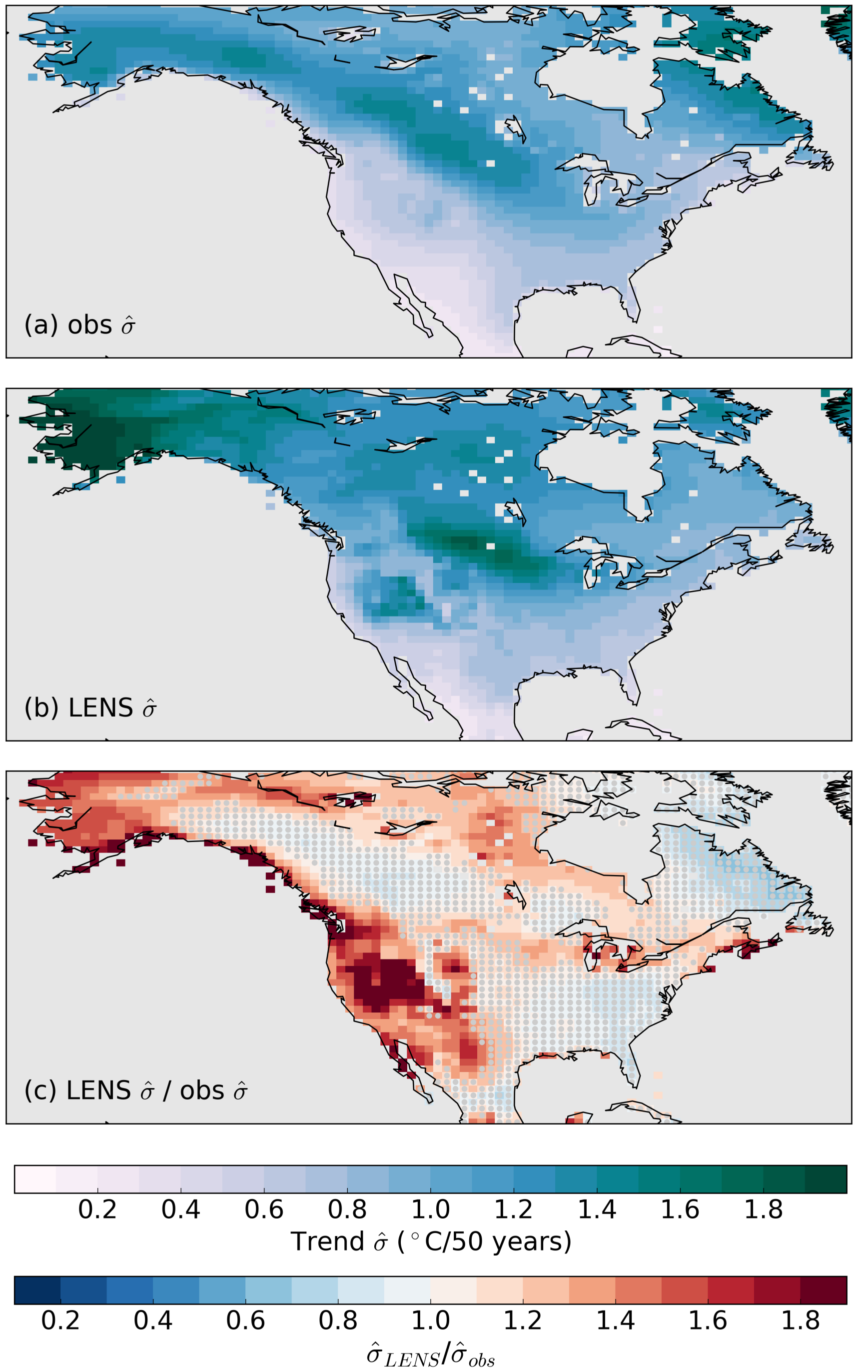

Figure S6: As in Fig. 5, but using the Dai et al. (2015) methodology to calculate the forced trend. For ease of comparison to Fig. 5, the trends shown remain linear in time, but here the forced trend at a given gridbox is calculated as the regression onto the global-mean ensemble-mean trend across the NCAR CESM1 Large Ensemble, and internal variability is estimated as the residual from this forced trend estimate. The synthetic ensemble is then created by resampling the residual and adding it back to the forced trend estimate.

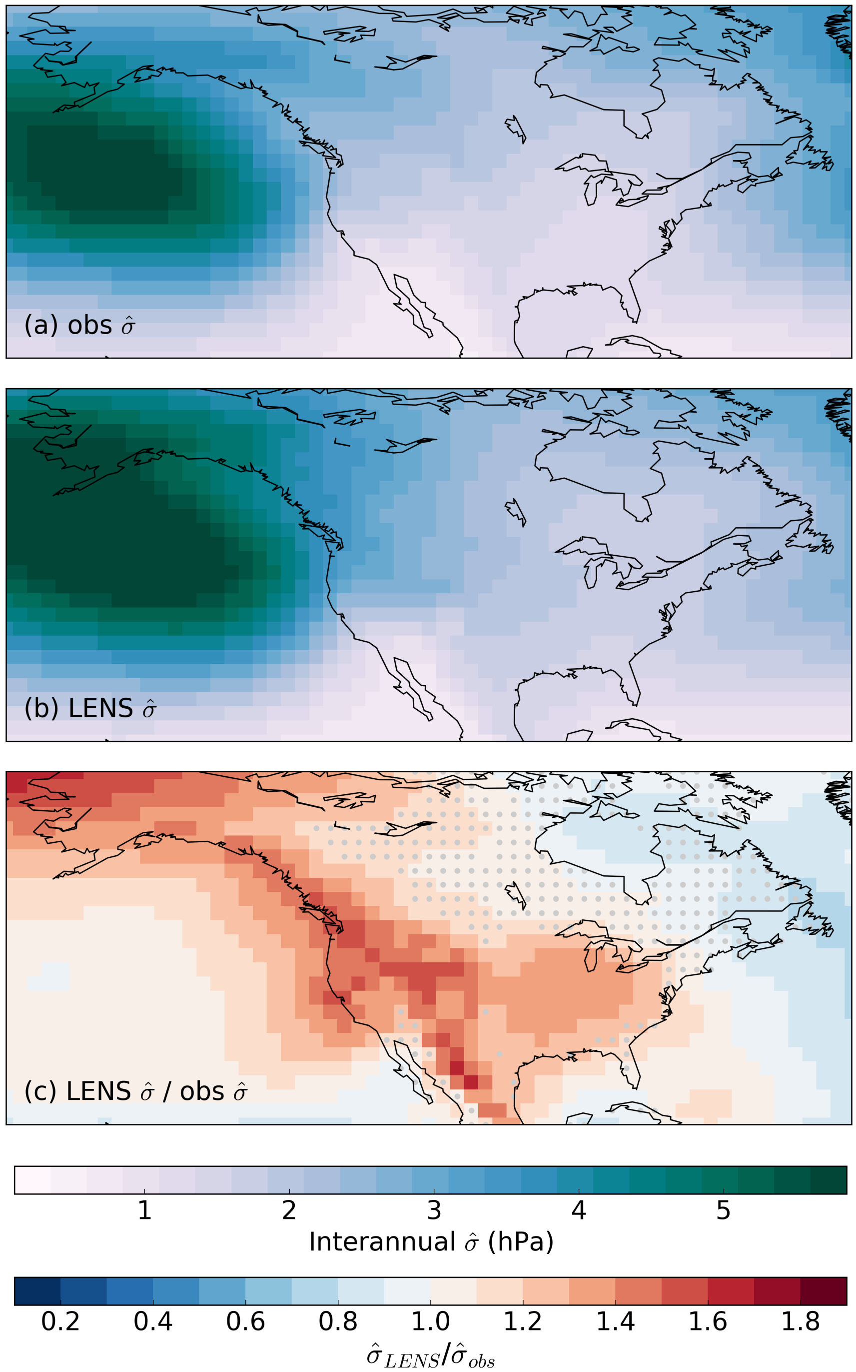

Figure S7: As in Fig. 2, but using SLP from the 20CRv2c reanalysis. Because the reanalysis ends in 2014, the variability is calculated using data from the 49-year period spanning 1966–2014.

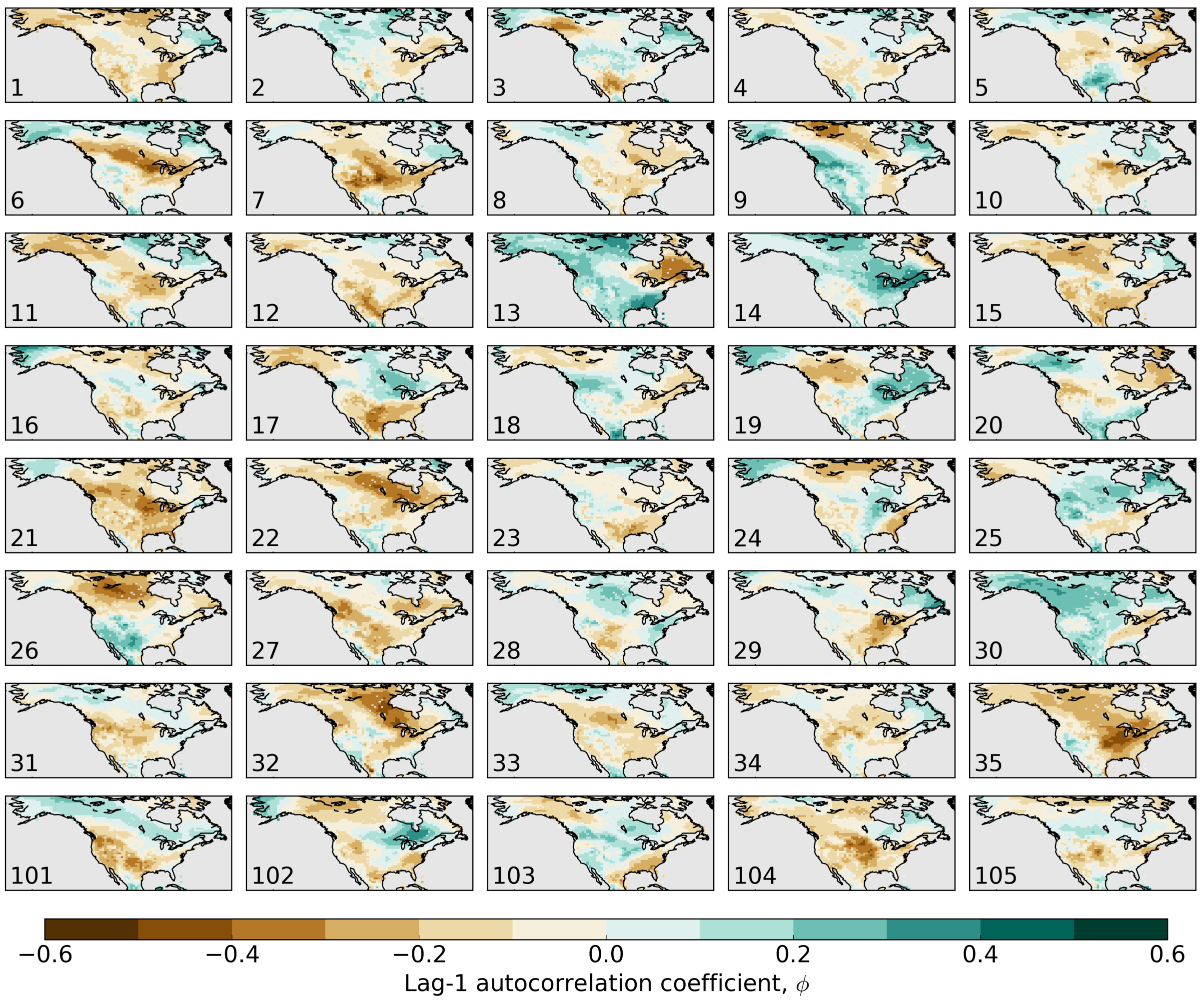

Figure S8: Empirical lag-one year autocorrelation coefficients for detrended DJF temperature in each member of the NCAR CESM1 Large Ensemble. The numbers in the lower right-hand corner of each subplot indicate the ensemble number; the last five begin at 101 because they were created on a different supercomputer from the first 35.

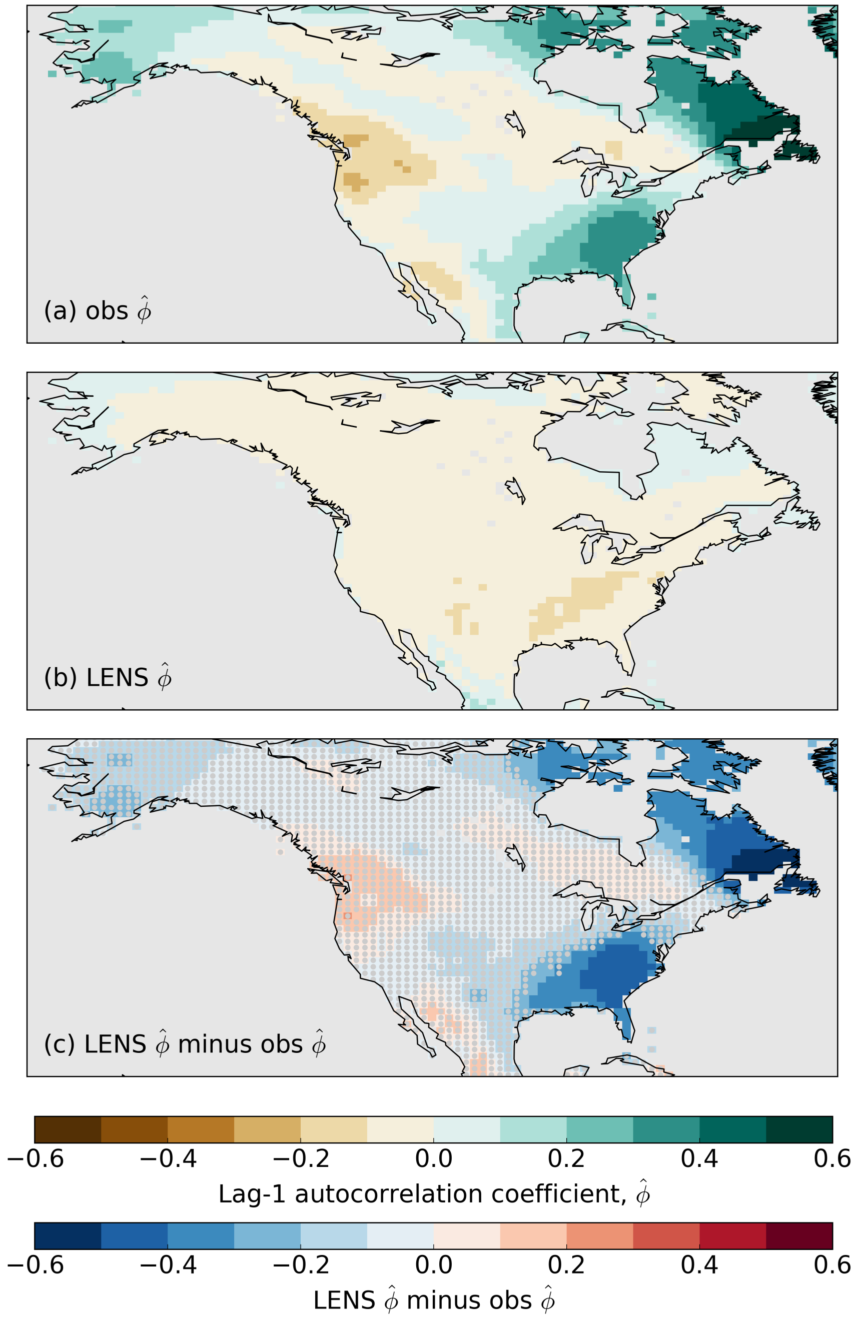

Figure S9: Empirical lag-one year autcorrelation, $\hat{\phi}$, of detrended DJF temperatures in (a) the BEST dataset and (b) the NCAR CESM1 Large Ensemble (LENS). In (b), the autocorrelation is calculated treating all ensemble members from LENS as identically distributed. (c) The difference between the lag-one autocorrelations in the observations and LENS, i.e. (b)-(a). Stippling indicates gridboxes that are not significant (see main text). Unlike in Fig. 1, nearly all gridboxes are found to be insignificant due to the variability of the estimator for autocorrelation coefficients from short time series (see Fig. S5).

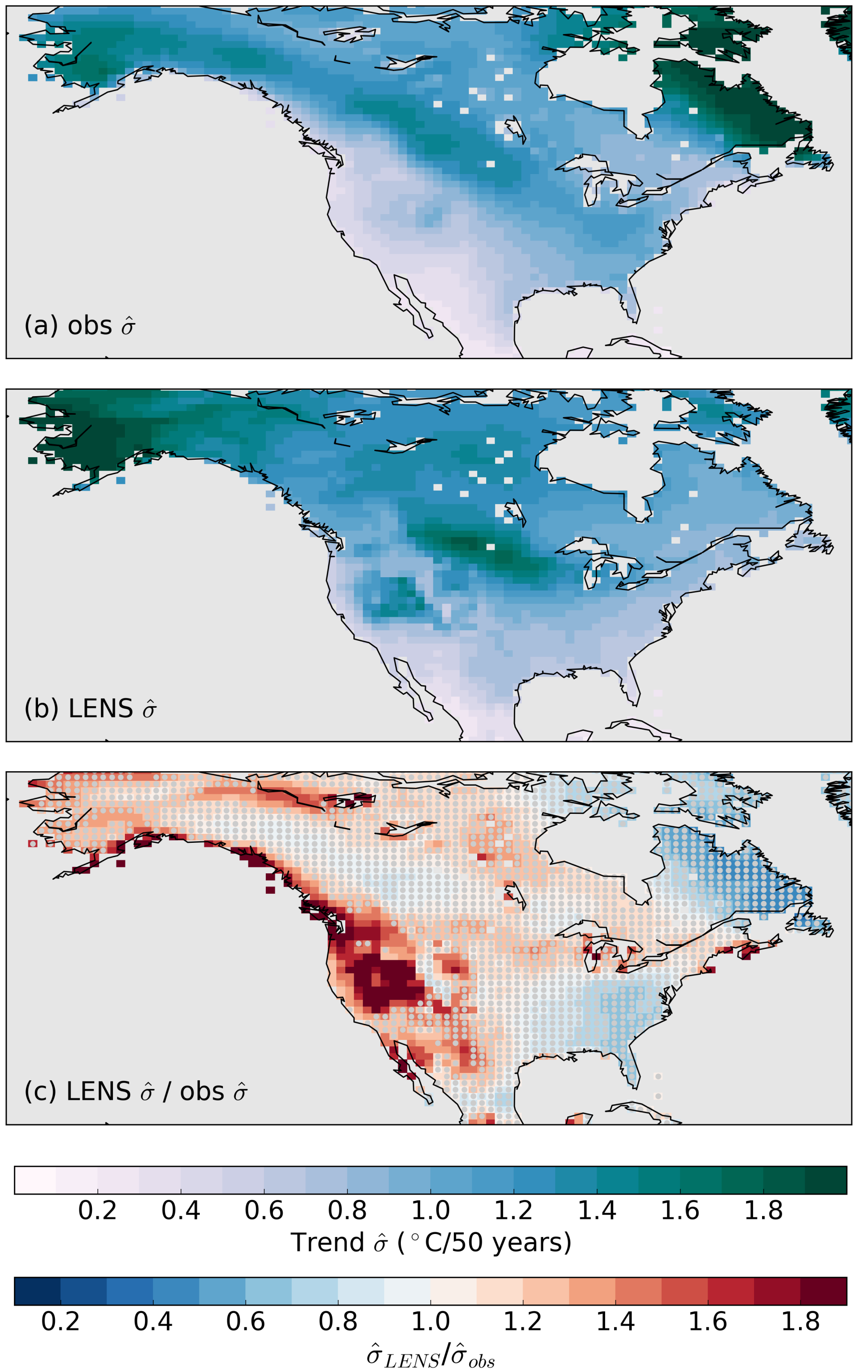

Figure S10: As in Fig. 5, but using the standard deviation inferred from the observations based on the methods of Thompson et al. (2015). Stippling indicates regions where the difference between the NCAR CESM1 Large Ensemble and the observations is identified as insignificant (see main text).

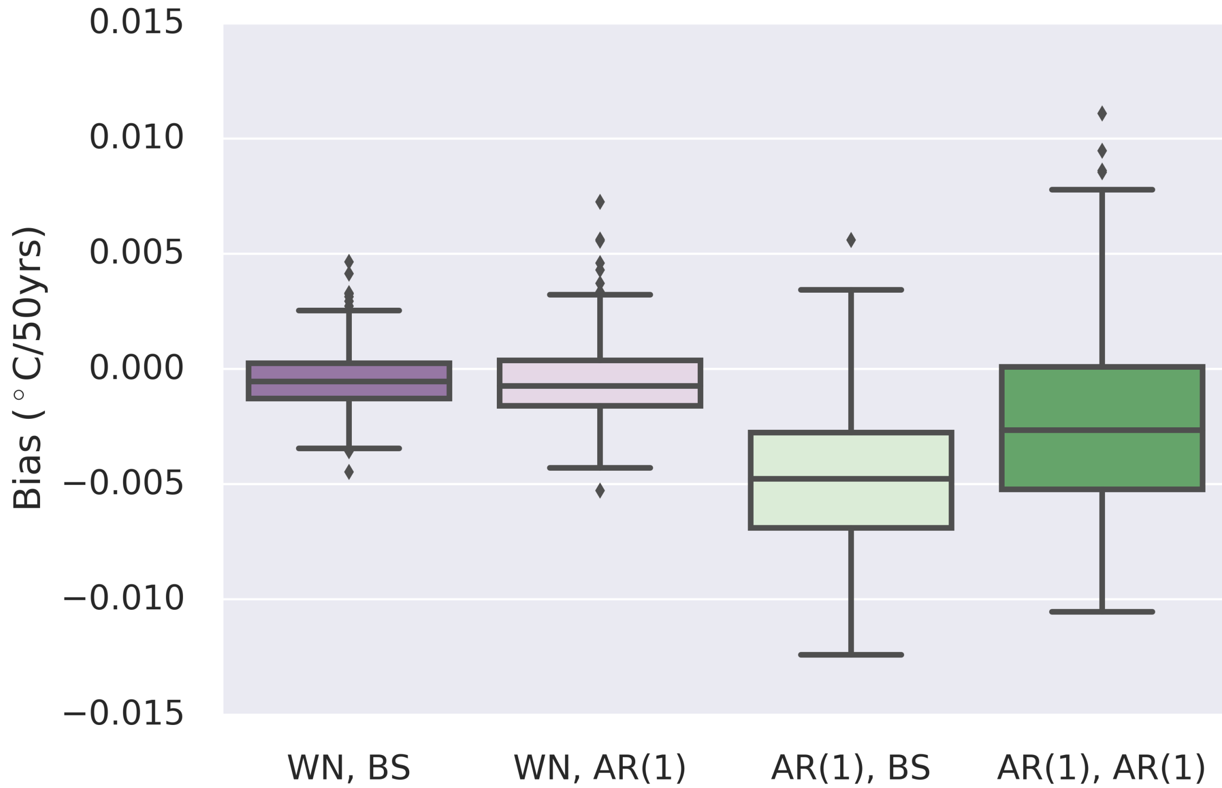

Figure S11: The bias in estimating the true uncertainty in a trend depending on the covariance structure of the noise and the method used to estimate the uncertainty. Bias is measured as the difference between the inferred and true standard deviation of the trend. The purple bars are for white noise (WN), whereas the green bars are for noise produced from an AR(1) model with a lag-one year autocorrelation coefficient of 0.5. For white noise, the use of a bootstrap (BS) with a block size of one leads to less biased and less variable estimates of the standard error of the trend than when the standard error is inferred from fitting an AR(1) model. For highly-autocorrelated data (green bars), however, the use of a bootstrap with an appropriate block size of five leads to a very biased estimate of the standard error in the trend. As such, the bootstrap is a better choice for data with small autocorrelation, but should not be used when data is highly-autocorrelated.

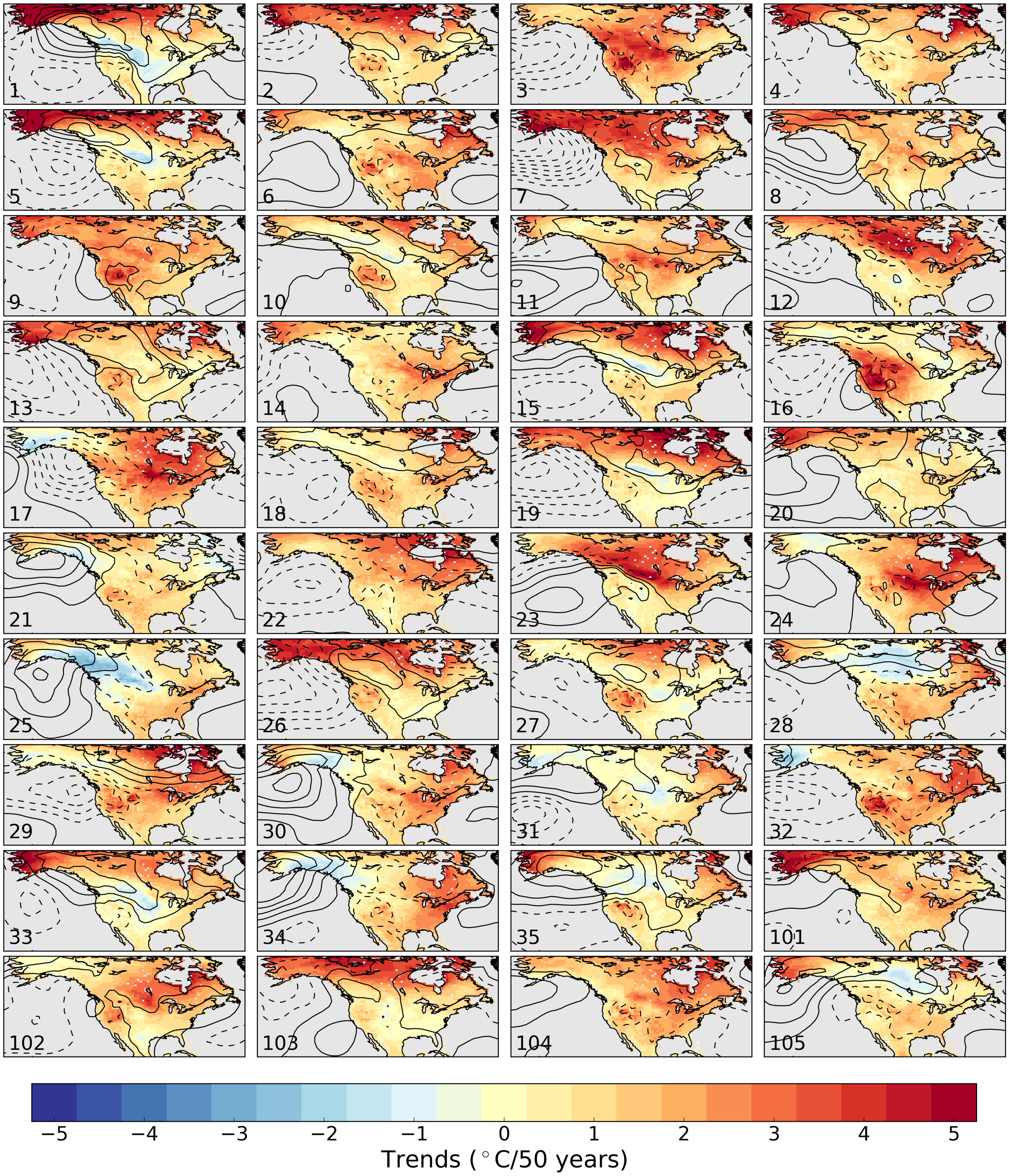

Figure S12: As in Fig. 6, but based upon the actual output from the NCAR CESM1 Large Ensemble. The numbers in the lower right-hand corner of each subplot indicate the ensemble number; the last five begin at 101 because they were created on a different supercomputer from the first 35.

References

Dai, A., J. C. Fyfe, S.-P. Xie, and X. Dai, 2015: Decadal modulation of global surface temperature by internal climate variability. *Nature Climate Change*, **5 (6)**, 555–559.

Thompson, D. W., E. A. Barnes, C. Deser, W. E. Foust, and A. S. Phillips, 2015: Quantifying the role of internal climate variability in future climate trends. *Journal of Climate*, **28 (16)**, 6443–6456.