

Supplementary information to “Evaluating climate field reconstruction techniques using improved emulations of real-world conditions”

Jianghao Wang¹, Julien Emile-Geay¹,
Dominique Guillot², Jason E. Smerdon³, and Bala Rajaratnam²

¹University of Southern California, Los Angeles, California

²Stanford University, Stanford, California

³Lamont-Doherty Earth Observatory of Columbia University, Palisades, New
York

1 Methods

Assume a linear relationship exists between temperature T and proxy P :

$$\mathbf{T} = \mathbf{B}\mathbf{P} + \varepsilon, \quad (1)$$

where ε denotes the error term, following a standard multivariate normal distribution. Let \mathbf{P}_c be a truncated version of \mathbf{P} restricted to the calibration interval. One can show that when $\mathbf{P}_c^t \mathbf{P}_c$ is non-singular, i.e. \mathbf{P}_c is full rank, the estimate of \mathbf{B} is given by the ordinary least square estimates:

$$\hat{\mathbf{B}} = (\mathbf{P}_c^t \mathbf{P}_c)^{-1} \mathbf{P}_c^t \mathbf{T} \quad (2)$$

In paleoclimate reconstructions, however, it is often the case $\mathbf{P}_c^t \mathbf{P}_c$ is rank-deficient, making traditional regression methods no longer applicable. Four CFR techniques are used herein to solve the problem.

1.1 RegEM-TTLS (Schneider, 2001)

Reconstructions with RegEM-TTLS were performed using the standard algorithm described in Schneider (2001). In particular, the implementation of TTLS follows the description in Fierro et al. (1997). First, SVD is calculated on the augmented data matrix $A \equiv (\mathbf{T}_{\text{calib}}, \mathbf{P}_{\text{calib}})$, in which small singular values are treated as zeros, and $\mathbf{T}_{\text{calib}}, \mathbf{P}_{\text{calib}}$ refer to \mathbf{T}, \mathbf{P} over the calibration period. Regularization is achieved by retaining only the first k large eigenvalues of the SVD of A . The choice of k determines to what extent the SVD of A is regularized, and can be user-specified or chosen adaptively by an algorithm. In this study, k was selected by detecting the first break in the log eigenvalue spectrum of A (Mann et al., 2009, SOM) prior to executing the RegEM algorithm. On average, a value of 5 was selected in most SNR cases¹.

1.2 M09 implementation of RegEM-TTLS (Mann et al., 2009)

The M09 implementation of TTLS uses a hybrid version of RegEM-TTLS that treats low-frequency and high-frequency signals separately (a 20-year frequency split). The reconstruction is then performed in a forward “stepwise” approach century by century. For each century, only the first M leading modes of surface temperature are retained, determined by the number of degrees of freedom in the proxy network. M09 also uses a semi-adaptive choice for both low-frequency and high-frequency truncation parameters, k_l and k_h respectively. The method selects: 1) k_l to retain 33% of the low-frequency multivariate data variance, and 2) k_h by detecting the first break in the log eigenvalue spectrum of high-frequency multivariate data variance, which are also calculated for each century of reconstruction.

¹In rare cases, depending on the specific noise realization, $k = 6$.

1.3 CCA (Smerdon et al., 2010)

Canonical correlation analysis (CCA) is applied as described by Smerdon et al. (2010). The primary goal of CCA is to perform dimensional reductions on \mathbf{T} , \mathbf{P} and \mathbf{B} in Eq 1. CCA first decomposes \mathbf{T} and \mathbf{P} through singular value decomposition (SVD), using only the first few leading modes (d_p, d_t) to estimate the covariance matrix $\hat{\Sigma}$, and then further truncates the eigen decomposition of $\hat{\Sigma}$ (via the selection of d_{cca}) to estimate $\hat{\mathbf{B}}$. The truncation parameters are selected based on "leave-half-out" cross-validation (CV), as described by Smerdon et al. (2010). See Fig. S2 for an illustration of the implementation of CCA. Since the set of optimal (d_p, d_t, d_{cca}) is chosen through CV, model parameters for CCA are not set a priori, making the method fully data-adaptive.

Below are some technical notes on speeding up CCA and employing the method in a temporally variant network. Originally, in the study by Smerdon et al. (2010), CCA is applied on a temporally invariant network, which means that (d_p, d_t, d_{cca}) only needs to be chosen once over the entire reconstruction period. In reality (and so in our study), proxy availability declines back in time, so the truncation parameters should also be chosen adaptively. In this study, we define patterns of missing values and group proxies with the same pattern together. Here we provide a simple example to illustrate the process. Assume we know that 35 proxies are available until A.D. 1000, when estimating the set of truncation parameters, all temperatures and only these 35 proxies over the calibration period will be used. For this particular M08 proxy dataset, there are 282, 121 and 281 distinct patterns of missing values in the ideal SNR, local SNR and max SNR networks, respectively. The corresponding set of optimal truncation parameters (d_p, d_t, d_{cca}) is different for each pattern, and also depends on the specific noise ensemble, i.e. SNR level and noise term in the pseudoproxies. The set of (d_p, d_t, d_{cca}) changes with input dataset and the large number of chosen truncation parameters are therefore not provided herein.

To speed up the computation, we also modified the original code (predccabp.m) provided in the supplementary info of Smerdon et al. (2010). Instead of calculating SVD for \mathbf{T} and \mathbf{P} every time when performing a CCA reconstruction in the CV process, we apply SVD only once before using CV to choose the optimal set of parameters.

The code will be available online at <https://code.google.com/p/common-climate/>.

1.4 GraphEM (Guillot et al., submitted)

The GraphEM algorithm benefits from recent advances in high-dimensional statistics by using a Gaussian graphical model (GGM, a.k.a. Markov random field, Whittaker, 1990; Lauritzen, 1996) to estimate the covariance structure Σ – a critical ingredient of all CFR methods. GGMs provide a natural and flexible framework for modeling the inherent spatial heterogeneities of high-dimensional spatial fields, which would in general be more difficult with standard parametric covariance models. At the same time, they provide the necessary parameter reduction for obtaining precise and well-conditioned estimates of Σ , even when the sample size is much smaller than the number of variables (as is typically the case in paleoclimate applications).

In the case of multivariate normal data, two variables are conditionally independent given the rest of the variables if and only if the corresponding element of the inverse covari-

ance matrix $\Omega_{ij} = (\Sigma^{-1})_{ij}$ is equal to zero. Thus, when many variables are conditionally independent (as is typically the case in climate fields), the inverse covariance matrix Ω is sparse, and the number of parameters to estimate is greatly reduced.

In GraphEM, the conditional independence structure of the climate field is first estimated using an ℓ_1 -penalized maximum likelihood method (the “graphical lasso”, described in Friedman et al., 2008). Once the conditional independence structure is known, the covariance matrix Σ can be estimated in accordance with these conditional independence relations within the EM algorithm (see Fig S1 for an illustration of the GraphEM algorithm).

As described in Guillot et al. (submitted), when estimating the graphical structure of the field, three parameters need to be specified to determine the target sparsity of the graph for the temperature-temperature part (**TT**), proxy-proxy part (**PP**), and the temperature-proxy part (**TP**). For all three parameters, a value that is large enough has to be specified so that the true graph is contained in the estimated one. On the other hand, the sparsity has to be small enough for the covariance matrix to be well-conditioned. For this study, the parameters have been set to $(\mathbf{TT}, \mathbf{PP}, \mathbf{TP}) = (5\%, 5\%, 5\%)$, following the choices in (Guillot et al., submitted). Future studies will explore data-adaptive criteria for choosing these parameters.

2 Results

2.1 Spatial metrics

Figures S3 - S12 below plot the spatiotemporal variations of CE. The figures show the ensemble median, using the flat network.

Overall, CE scores are higher over more regions during the AD 1059–1149, 1150–1249, 1750–1849 intervals than during the other periods. Figures. S13–S14 show the MSE decomposed into bias and variance, which helps interpret the patterns of CE variations shown in Figs. 7 and 8. We find that both bias and variance have similar types of variations to CE: that is, during the aforementioned periods, both bias and variance are smaller than the rest periods, leading to a smaller MSE (i.e. a higher CE). Additionally, we note that, in the local SNR case (Figs. S13), the relative contribution of bias to MSE is, on average, higher than 60% across different periods. Therefore the patterns of MSE in Fig. S13 are largely driven by the patterns of bias. The relative contributions of bias and variance to MSE are comparable in the max SNR case (Fig. S14); their patterns of temporal variations are also similar to each other, therefore leading to a similar pattern in the total MSE. Various causes might be responsible for such patterns of bias, MSE and CE. In this study, we conjecture that the type and amplitude of climate variations may contribute to reconstruction skill. Figs. 7 and 8, and Figs. S13 - S14 here are compared with Fig. 1 in Ammann et al. (2007). The timing of high CE scores (low MSE and bias) is found to be coincident with the timing of high amplitude volcanic events and periods of anomalous solar irradiance. It remains to be seen whether this synchrony is found with other GCM targets, or in nature.

2.2 Global average

Table S1 and S2 summarize the weighted global mean ($\hat{\mu}$) verification statistics (CE, RE, R^2) for each century, with the estimated standard deviation ($\hat{\sigma}$) displayed in parentheses.

By definition, CE is the most stringent diagnostic. In Table S2 and S1, it is also evident that reconstruction skill based on CE indicates the worst results. The RE statistic, which is less stringent than CE, indicates that reconstructions are better than predictions of climatology (if $RE > 0$) for periods after 1300 AD in the local SNR case, and remain above zero during the entire reconstruction interval for all methods in the max SNR network. R^2 is the least strict statistic because it only accounts for similarity between two time series by trend. As in Table S2, $R^2 > 0$ for all methods in all cases. With R^2 as the judging criterion, CCA stands out in the max SNR network, which is consistent with the mechanisms of the CCA algorithm, i.e. maximize the cross-correlation between proxies and temperature time series.

2.3 Additional diagnostics

We also examined the relationship between:

1. CE vs. number of proxies per grid box (Figs. S15 and S16);
2. CE vs. average SNR per grid box (Figs. S17 and S18);
3. CE vs. sum of SNR per grid box (Figs. S19 and S20).

The results indicate that there are no simple apparent relationships between these variables of interest.

2.4 Error cancellation in index reconstructions

Following the approach of Guillot et al. (submitted), we consider the positive and negative contributions to the error in the reconstructed global mean temperature index (Figs. S21 and S22). For every spatial point s in a given year t , we compute the deviation from the true field $\hat{T}(t, s) - T(t, s)$, where \hat{T} is the reconstructed temperature and T is the target (here the GCM-simulated temperature). Such deviations are then partitioned into positive and negative contributions, and the area-weighted average of each is computed to form the timeseries displayed in Figs. S21 and S22.

This allows decomposition of the sum of reconstruction errors (black curves) into contributions stemming from regions of warm or cool biases. In an ideal reconstruction, the deviations should be small everywhere. In practice, however, reconstructions are often biased warm ($\hat{T}(t, s) > T(t, s)$), but in some locations are biased cold ($\hat{T}(t, s) < T(t, s)$). As long as the positive values are comparable with negative values, the value of the global mean deviation $\langle \hat{T}(t, s) - T(t, s) \rangle$, which is a simple linear combination of the positive and negative values, will be small, and so will the RMSE obtained from it.

As shown on Figs. S21 and S22, similar positive errors arise from all methods, but negative errors arising from RegEM-TTLS and M09 are much higher than those from CCA and GraphEM. Because the absolute values of RegEM-TTLS and M09's negative deviations

are comparable with their positive deviations, their overall sum is smaller than those arising from CCA and GraphEM. Thus we have the counterintuitive circumstance that larger spatial deviations result in smaller overall errors when composite indices (e.g. global or hemispheric means) are computed. This finding has important implications for interpreting CFR results, showing in particular that the global mean temperature reconstruction is a poor indicator of spatial skill.

3 Alternative designs of pseudoproxy networks

No proxy is ever solely indicative of local climatic conditions. However, even a proxy that is not (purely) temperature-sensitive can be useful in recording long-range temperature information through teleconnections. For instance, Fig. S23 depicts correlations between a given proxy (a moisture-sensitive tree-ring record from southwestern North America, black triangle) and all available instrumental temperatures. An ENSO-like pattern is evident in all four panels of this figure, suggesting that the records are indicative of ENSO teleconnections, as is well known (Cook et al., 2004, 2007). This indicates that proxies without high correlations to *local* temperature, but with high correlation with a more remote gridpoint, might still have value for reconstructing far-field temperature information, and motivated us to design two end-member networks: the *local SNR* network, which is reflective of local, temperature-only relations, and the *max SNR network*, modeling each proxy’s highest potential to capture information about a climate field. Neither choice is realistic, but each is a useful end-member of real-world conditions.

To more realistically model real-world conditions, we consider an intermediate design, which balances the search for optimal correlations with a constraint of localization. We do so by maximizing a distance-weighted correlation $|\rho_{ij}| \times e^{-\frac{d_{ij}}{d_0}}$, in which ρ_{ij} is the linear correlation between a given pair of (P_i, T_j) , d_{ij} is the corresponding P-T distance, and d_0 is a constant. The question then translates to finding the temperature gridpoint that gives the maximum absolute value in the vicinity of the proxy record. A potential problem is that no theoretical criterion exists to distinguish real teleconnections from spurious correlations. Here we set $d_0 = R_e$ (one Earth radius). Other values ($0.5R_e, 1.5R_e, 2R_e, 4R_e$) were also tried, and results were found to be insensitive to this choice (Fig. S23 and S25).

In the intermediate SNR network, the number of significant² proxies reduces to 854 (out of 1138). Although the average SNR (0.40) is lower than that in the max SNR network (0.45), the histogram remains very similar (Fig. S24). Looking at the joint distribution of P-T distance and $|\rho|$ (Fig. S25), we find that most of the significant correlations are *regional*, i.e. within 4000 km, which is more physically interpretable than the pattern shown in Fig. 2 (main text). Nevertheless, results based on the intermediate SNR network turn out to be very similar to those derived from the max SNR network.

For simplicity, we illustrate this point using only M09-TTLS and GraphEM. In Fig. S26, we plot the spatial patterns of CE, and compare results from the intermediate SNR network and the max SNR network. Comparisons are based on results of 20 noise realizations, thus the ensemble spread might be narrower compared with results based on the 100-realization

²The significance test is based on Ebisuzaki (1997), taking into account of the effect of multiple hypothesis test via calculating false discovery rate (Ventura et al., 2004).

ensemble in the original manuscript. For both methods, the spatial patterns between the intermediate and max SNR networks are very similar: the pattern correlation between CE patterns of the two SNR networks is 0.87 for M09-TTLS and 0.84 for GraphEM over A.D. 850 – 949, 0.98 for M09-TTLS and 0.76 for GraphEM over A.D. 1750 – 1849. Conclusions based on the global mean time series are also consistent (Fig. S27): both the ensemble medians and spreads are very similar to one another. Nevertheless, we emphasize that all these designs employ synthetic data, while real-world proxies have much more complex relationships to temperatures. Results based on pseudoproxies are not, by nature, directly comparable to real-world reconstructions, but we have provided reasonable end-member experiments demonstrating the dependence of these different assumptions about proxy performance.

4 Properties of the NCAR CSM1.4 simulation

As shown in Figs. 7 and 8, CE skill varies non-monotonically even when proxy availability is uniform (flat network). This suggests that data availability is not the only driving cause of temporal fluctuations in reconstruction skill. Here we show that intrinsic properties of the simulated climate are important as well.

We use the M09-TTLS reconstructions with the max SNR flat network to illustrate the point. In Fig. S28, we compare the temporal variations of globally averaged CE skill with temperature variability in the CSM1.4 model simulations. Temperature variability is quantified by the running standard deviation of the global mean temperature time series, averaged over sliding 100-year windows. As in Fig. S28, the global mean temperature displays pronounced variability during 1150AD - 1350 AD, which is broadly consistent with the period of high CE scores. Those turn out to be highly correlated with the running standard deviation ($R = 0.74$, $p < 0.05$). Linearly regressing the century-averaged CE onto the century-averaged running standard deviation yields a slope of 1.16 (Fig. S29), with a 95% confidence interval [0.30, 2.03] that excludes zero.

This analysis demonstrates that, in this model at least, reconstruction skill is higher when global mean temperature variability is higher – episodes of vigorous climate variability emerge more clearly from the noise in the pseudoproxies. This happens at the local level, and is then reflected in the global mean. Climate variability in CSM1.4, as in the climate system itself, is the result of both external drivers and internal climate oscillations. The NCAR CSM1.4, as acknowledged in Ammann et al. (2007), has relatively coarse resolution and therefore weak internal variability. In the simulation, the most pronounced cooling episodes are found to coincide with the timings of strong volcanically-forced events (e.g. 1258AD, 1453AD, 1815AD, see Ammann et al., 2007, their Fig. 1). This is partly because the response to hemispheric-scale volcanic events shows large spatial coherency (not shown), though our analysis cannot disentangle this effect from that associated with the higher amplitude of climate fluctuations at such times. Nevertheless, it does open the possibility that reconstruction skill may be partly dependent on the amplitude and spatial scales of temperature fluctuations. In nature, the ratio of internal to external variability is likely higher than in this coarse GCM, so the extent to which this effect matters for real-world reconstructions remains to be ascertained.

References

- Ammann, C. M., Joos, F., Schimel, D. S., Otto-Bliesner, B. L., and Tomas, R. A.: Solar influence on climate during the past millennium: Results from transient simulations with the NCAR Climate System Model, *Proceedings Nat. Acad. Sc.*, 104, 3713–3718, doi:10.1073/pnas.0605064103, 2007.
- Cook, E. R., Woodhouse, C. A., Eakin, C. M., Meko, D. M., and Stahle, D. W.: Long-Term Aridity Changes in the Western United States, *Science*, 306, 1015–1018, doi:10.1126/science.1102586, 2004.
- Cook, E. R., Seager, R., Cane, M. A., and Stahle, D. W.: North American drought: Reconstructions, causes, and consequences, *Earth Science Reviews*, 81, 93–134, doi:10.1016/j.earscirev.2006.12.002, 2007.
- Ebisuzaki, W.: A Method to Estimate the Statistical Significance of a Correlation When the Data Are Serially Correlated, *Journal of Climate*, 10, 2147–2153, 1997.
- Fierro, R. D., Golub, G. H., Hansen, P. C., and O’ Leary, D. P.: Regularization by truncated total least squares, *SIAM J. Sci.Comput.*, 18, 1223–1241, 1997.
- Friedman, J., Hastie, T., and Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso, *Biostat*, 9, 432–441, doi:10.1093/biostatistics/kxm045, 2008.
- Guillot, D., Rajaratnam, B., and Emile-Geay, J.: Statistical Paleoclimate Reconstructions via Markov Random Fields, *Ann. Appl. Stat.*, doi:http://arxiv.org/abs/1309.6702, submitted.
- Lauritzen, S. L.: *Graphical Models*, Oxford: Clarendon Press., 1996.
- Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, *Proceedings Nat. Acad. Sc.*, 105, 13 252–13 257, doi:10.1073/pnas.0805721105, 2008.
- Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C., Faluvegi, G., and Ni, F.: Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly, *Science*, 326, 1256–1260, doi:10.1126/science.1177303, 2009.
- Schneider, T.: Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values., *J. Clim.*, 14, 853–871, doi:10.1175/1520-0442(2001)014<0853%3AAOICDE>2.0.CO%3B2, 2001.
- Smerdon, J. E., Kaplan, A., Chang, D., and Evans, M. N.: A Pseudoproxy Evaluation of the CCA and RegEM Methods for Reconstructing Climate Fields of the Last Millennium*, *J. Clim.*, 23, 4856–4880, doi:10.1175/2010JCLI3328.1, 2010.

Ventura, V., Paciorek, C. J., and Risbey, J. S.: Controlling the Proportion of Falsely Rejected Hypotheses when Conducting Multiple Tests with Climatological Data, *Journal of Climate*, 17, 4343–4356, doi:10.1175/3199.1, URL <http://dx.doi.org/10.1175/3199.1>, 2004.

Whittaker, J.: *Graphical Models in Applied Multivariate Statistics*, John Wiley and Sons, Chichester, United Kingdom, 1990.

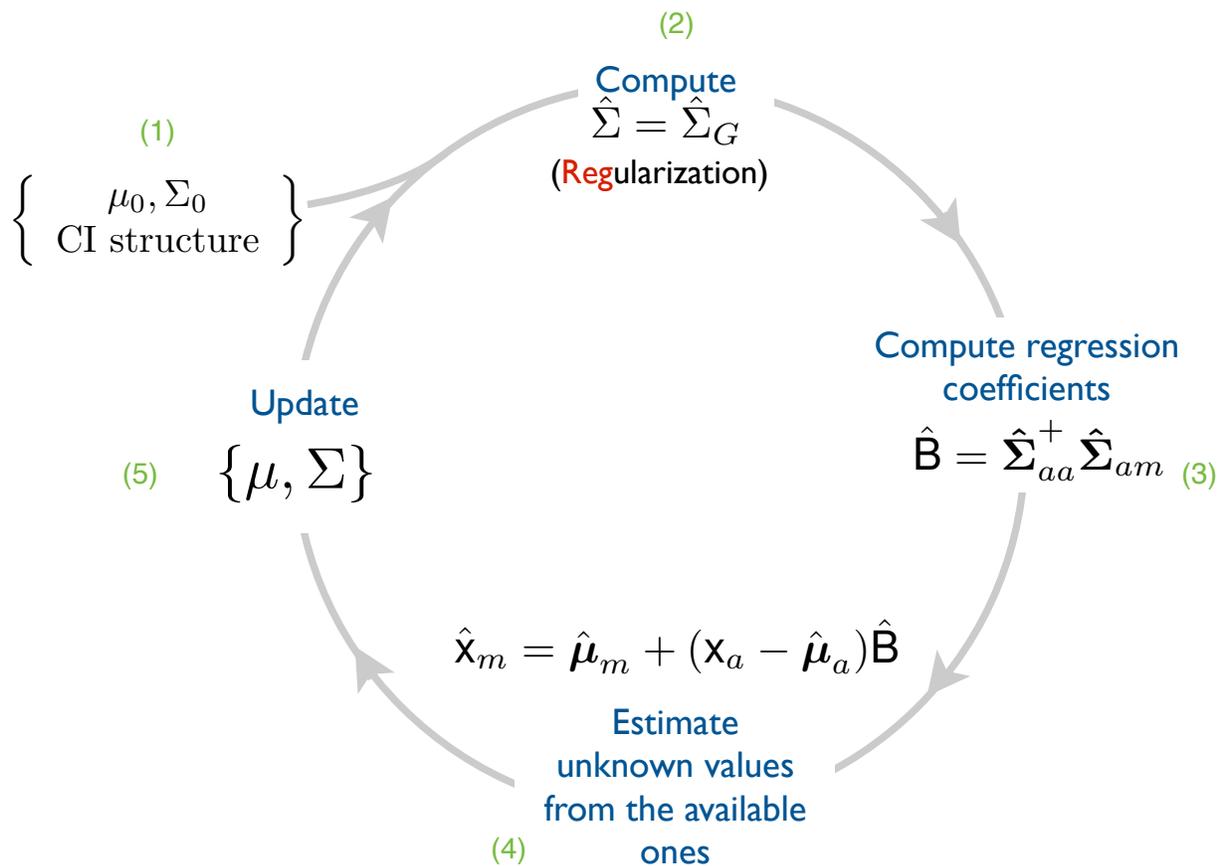


Figure S1: A summary of the GraphEM algorithm. Iterates until convergence. From Guillot et al. (submitted)

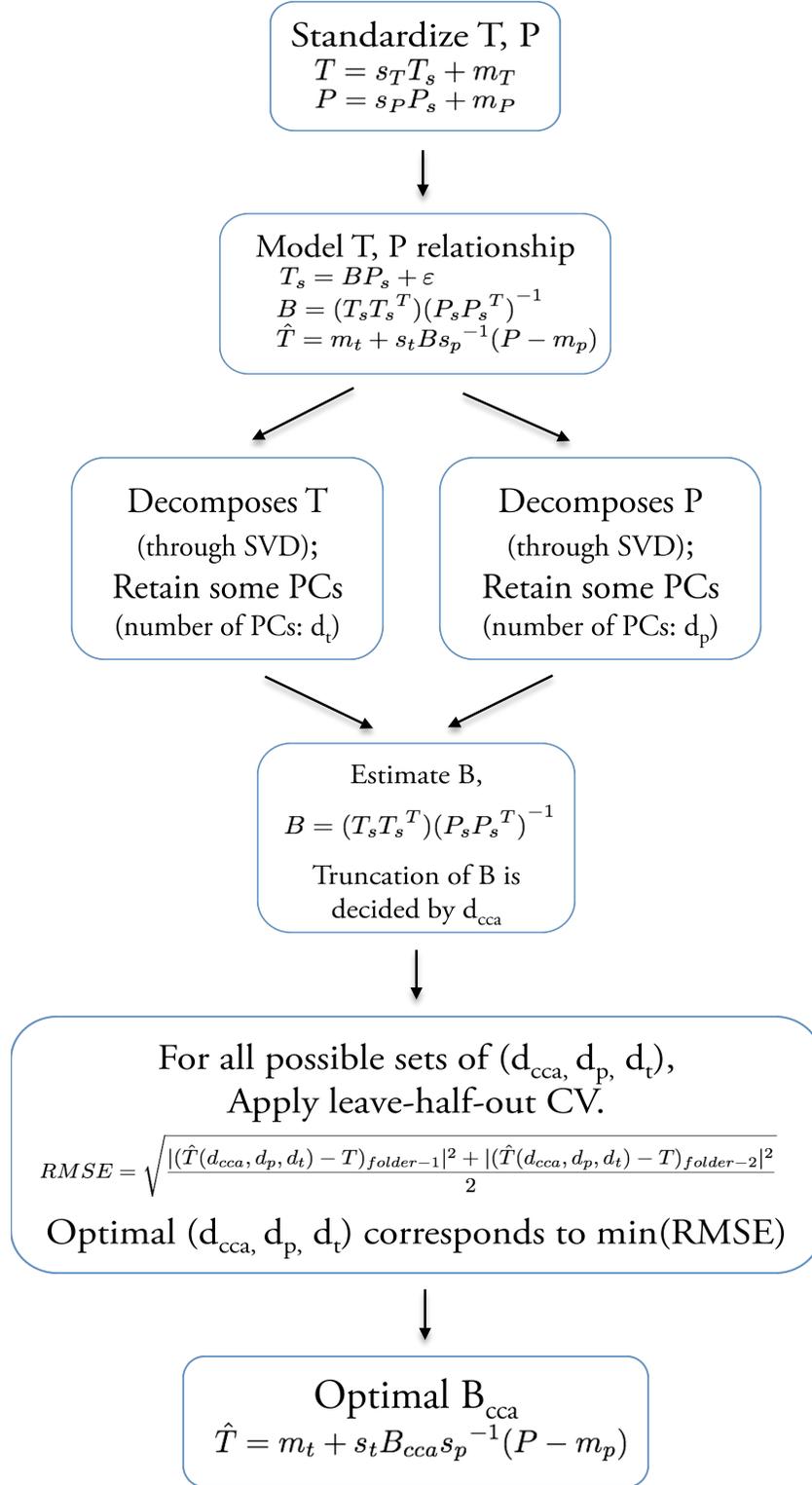


Figure S2: A summary of the CCA algorithm. Reproduced from Smerdon et al. (2010). T_s, P_s are temperature T and proxy P standardized over the temporal dimension spanning the calibration interval, with mean m_T, m_P and standard deviation s_T, s_P , respectively. B is the regression coefficient matrix.

RegEM-TTLS, flat network, local SNR

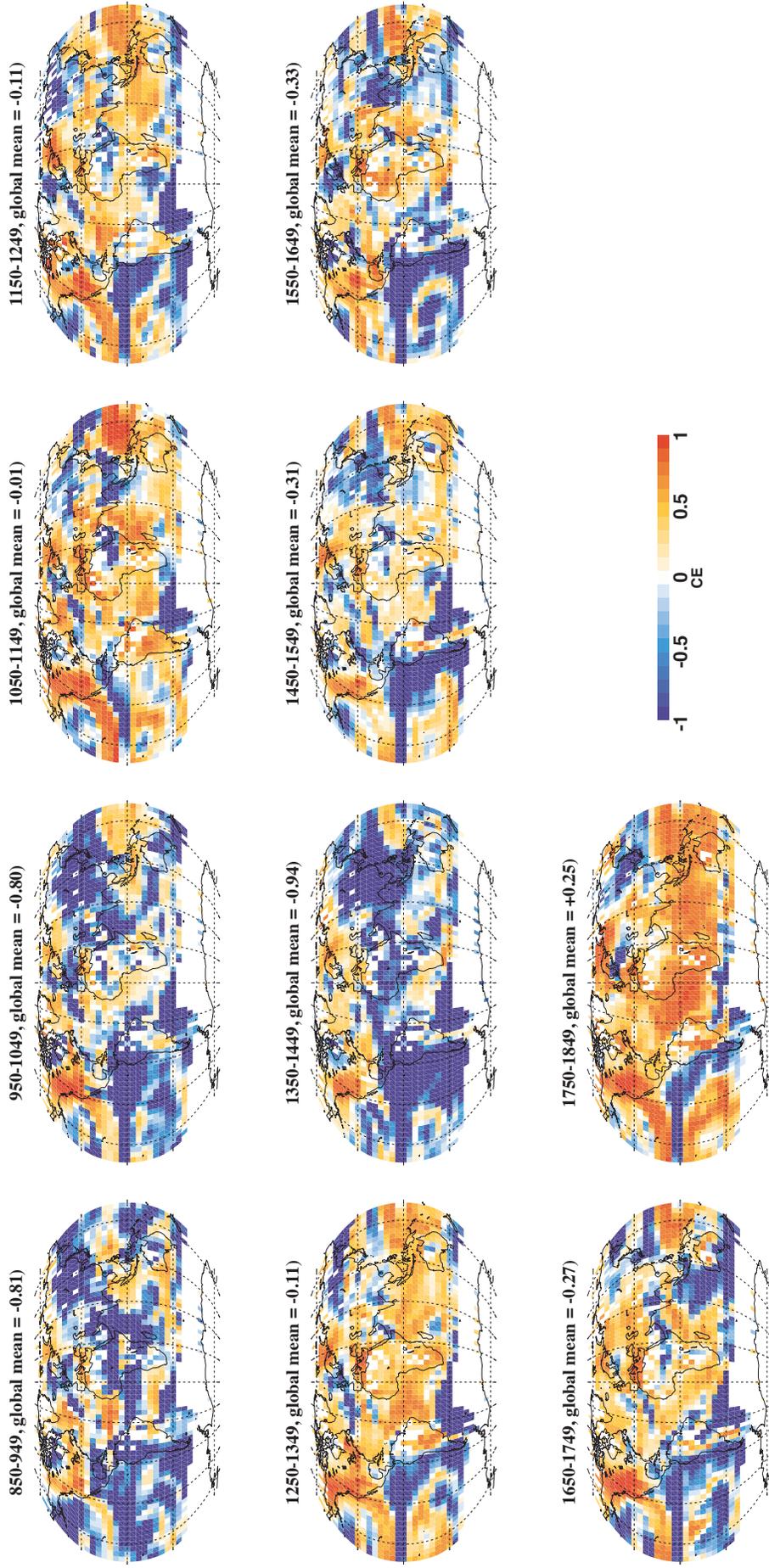


Figure S3: Spatiotemporal variation of RegEM-TTLS reconstructions, using the local SNR, flat network, based on the median of 100-member ensemble.

RegEM-TTLS, flat network, max SNR

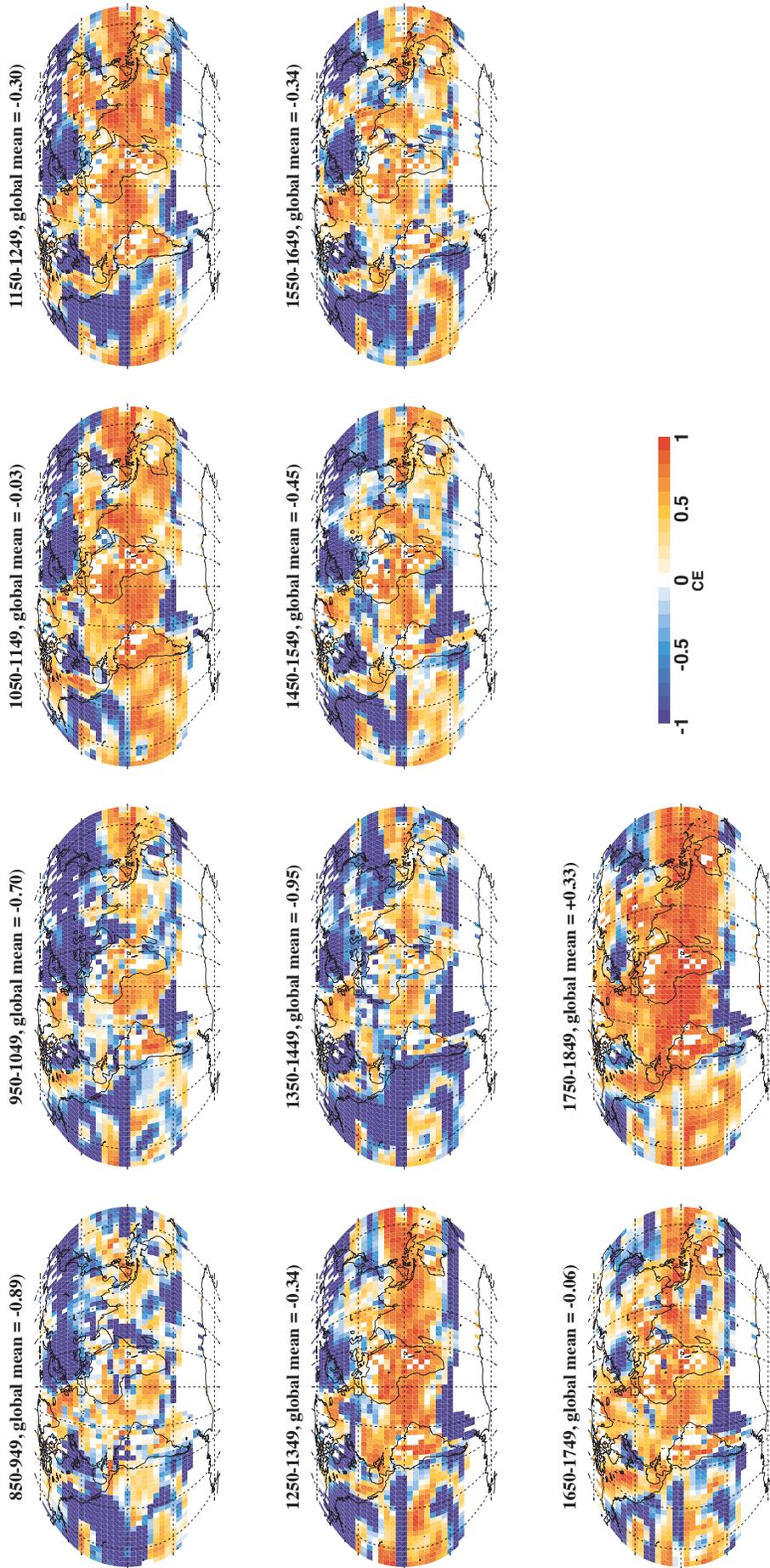


Figure S4: Spatiotemporal variation of RegEM-TTLS reconstructions, using the max SNR, flat network, based on the median of 100-member ensemble.

M09, flat network, local SNR

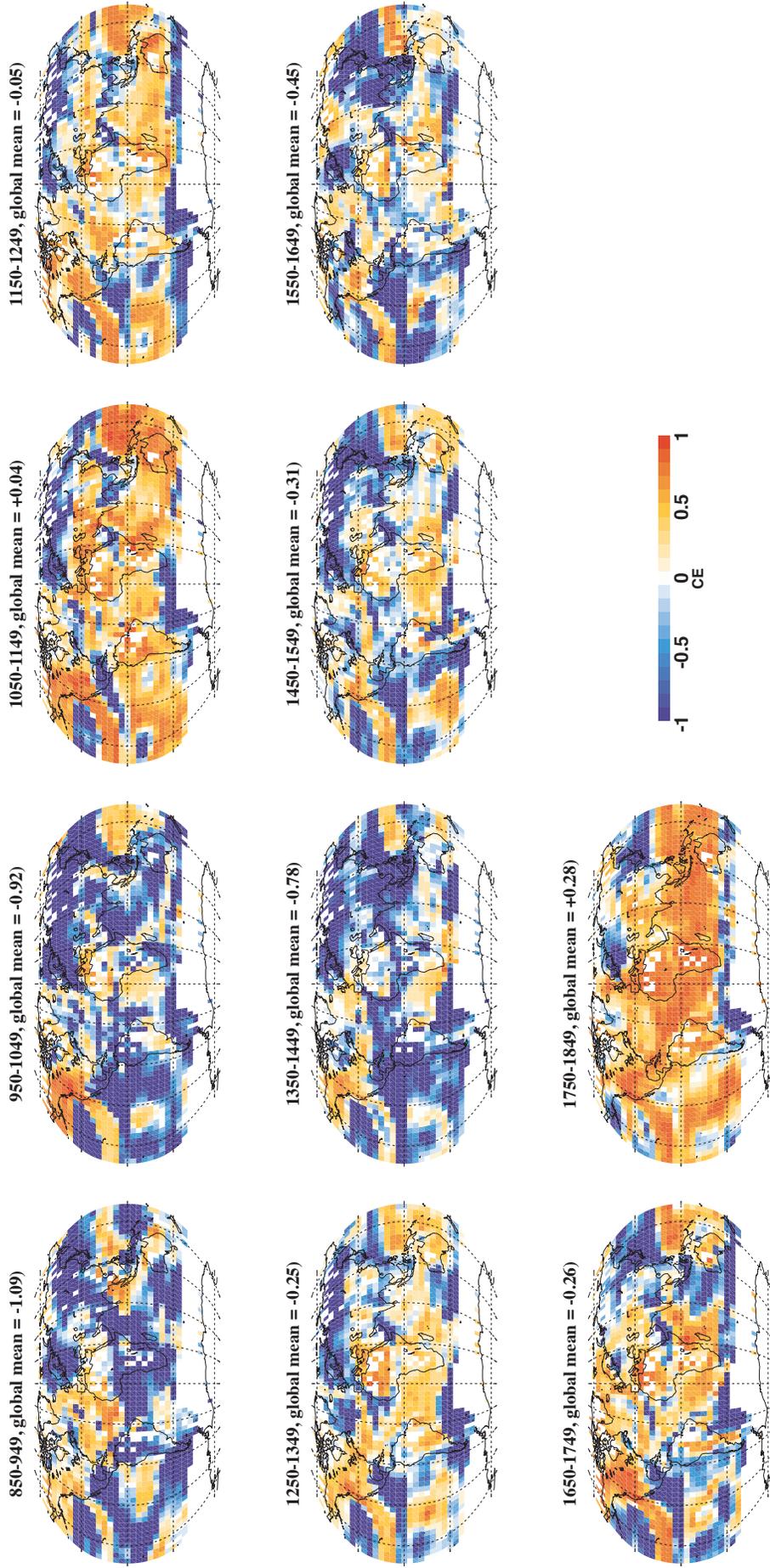


Figure S5: Spatiotemporal variation of M09 reconstructions, using the local SNR, flat network, based on the median of 100-member ensemble.

M09, flat network, max SNR

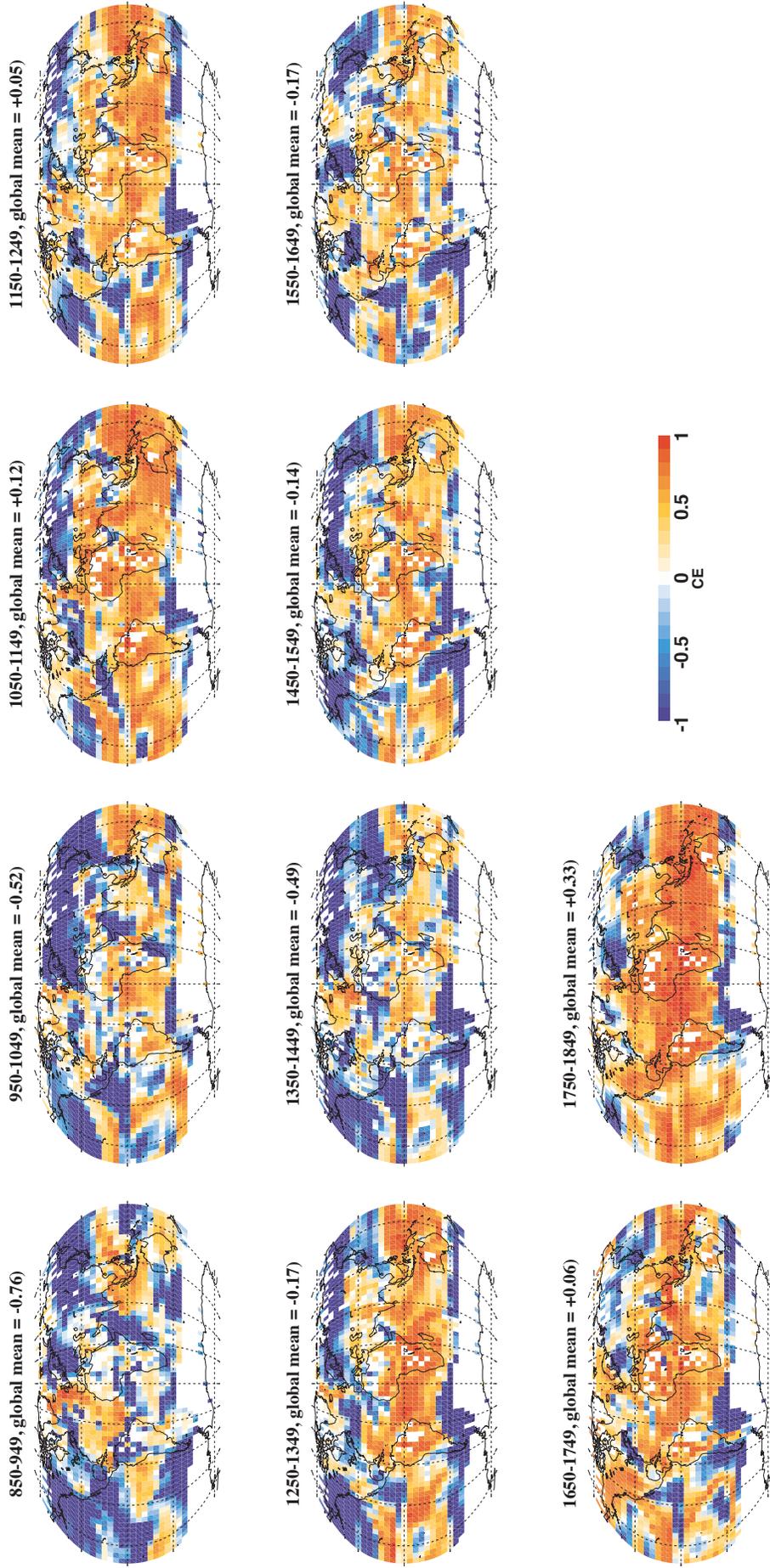


Figure S6: Spatiotemporal variation of M09 reconstructions, using the max SNR, flat network, based on the median of 100-member ensemble.

CCA, flat network, local SNR

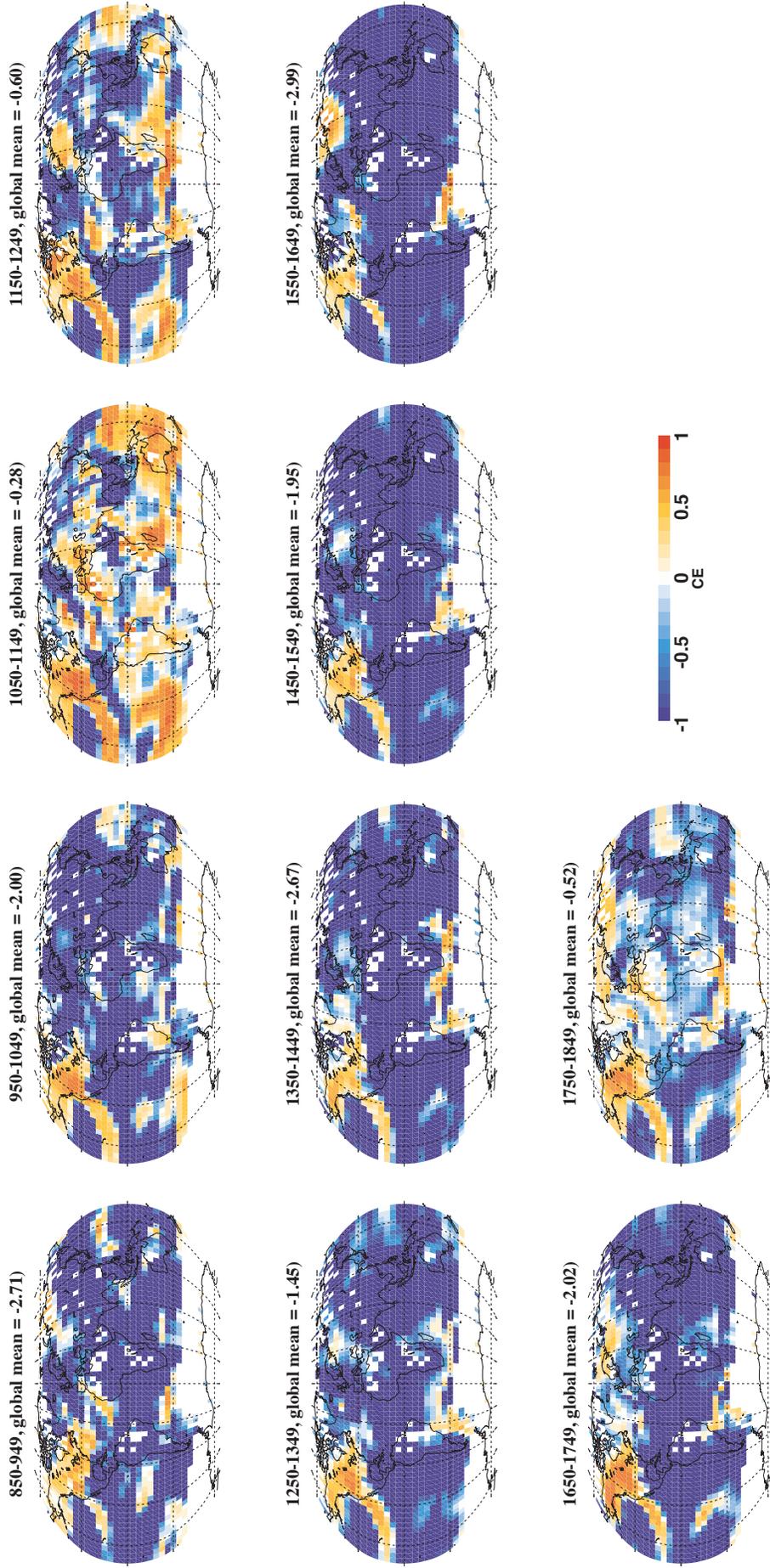


Figure S7: Spatiotemporal variation of CCA reconstructions, using the local SNR, flat network, based on the median of 100-member ensemble.

CCA, flat network, max SNR

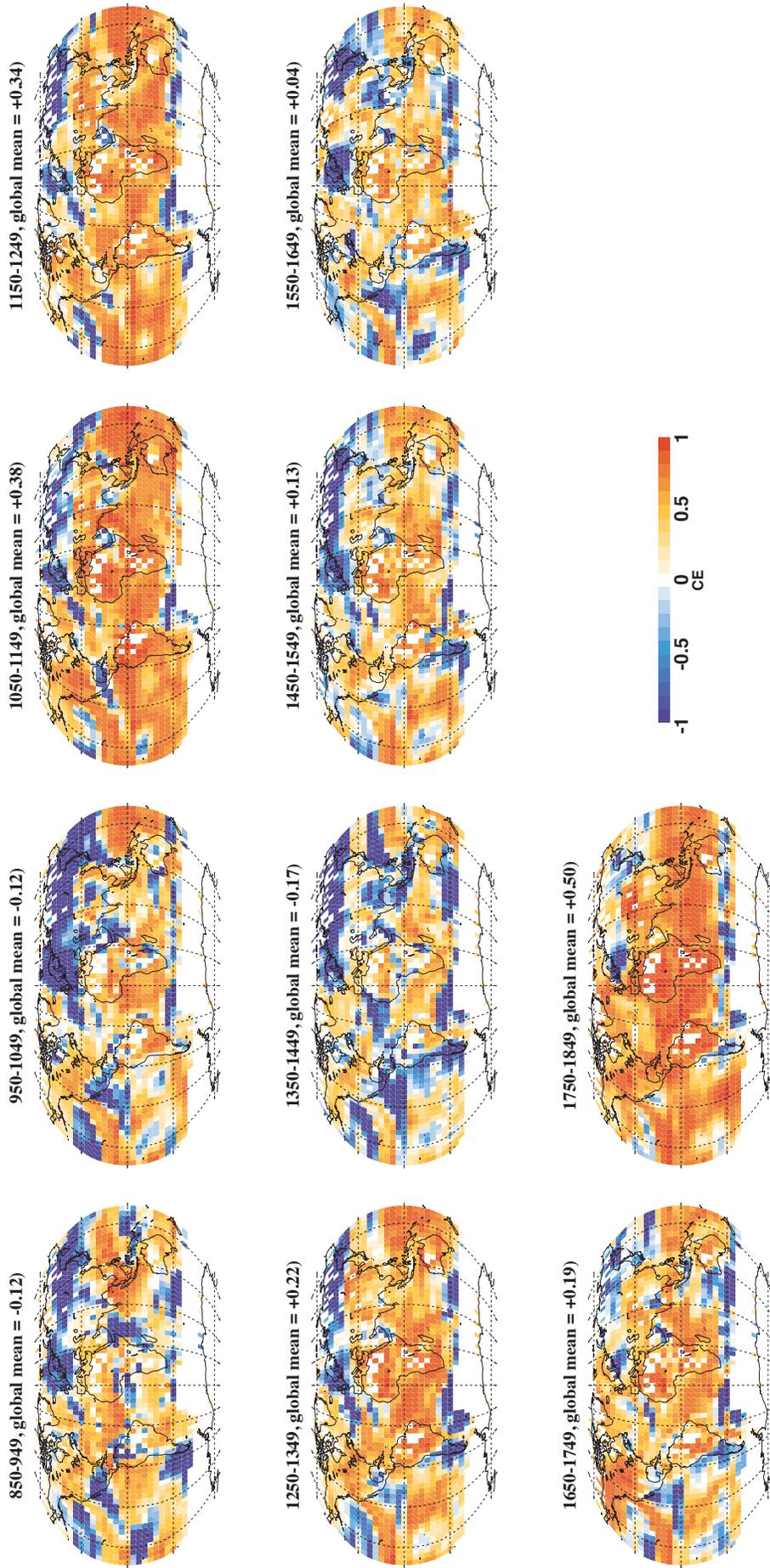


Figure S8: Spatiotemporal variation of CCA reconstructions, using the max SNR, flat network, based on the median of 100-member ensemble.

GraphEM, flat network, local SNR

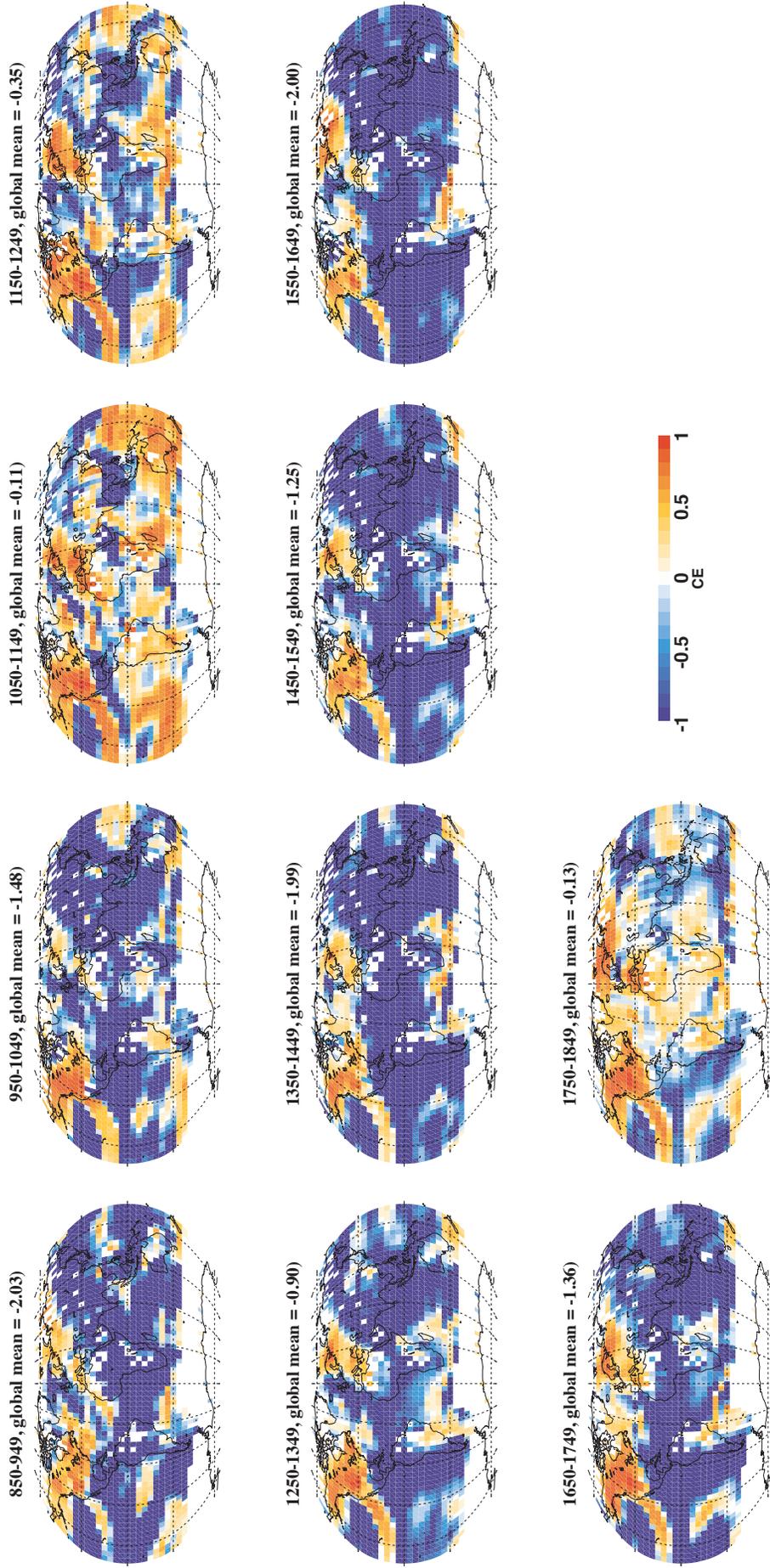


Figure S9: Spatiotemporal variation of GraphEM reconstructions, using the local SNR, flat network, based on the median of 100-member ensemble.

GraphEM, flat network, max SNR

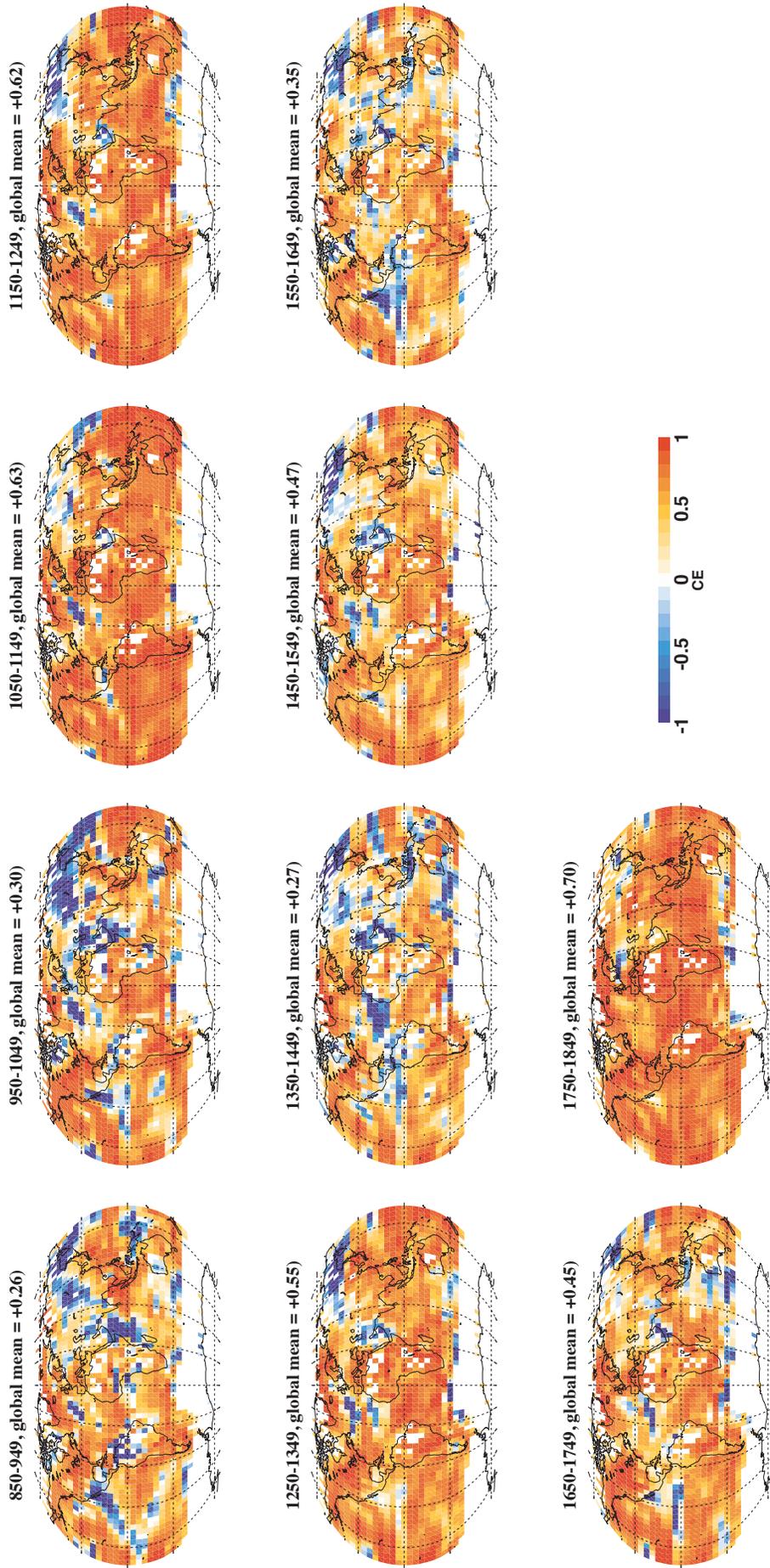


Figure S10: Spatiotemporal variation of GraphEM reconstructions, using the max SNR, flat network, based on the median of 100-member ensemble.

GraphEM, flat network, SNR = 0.25

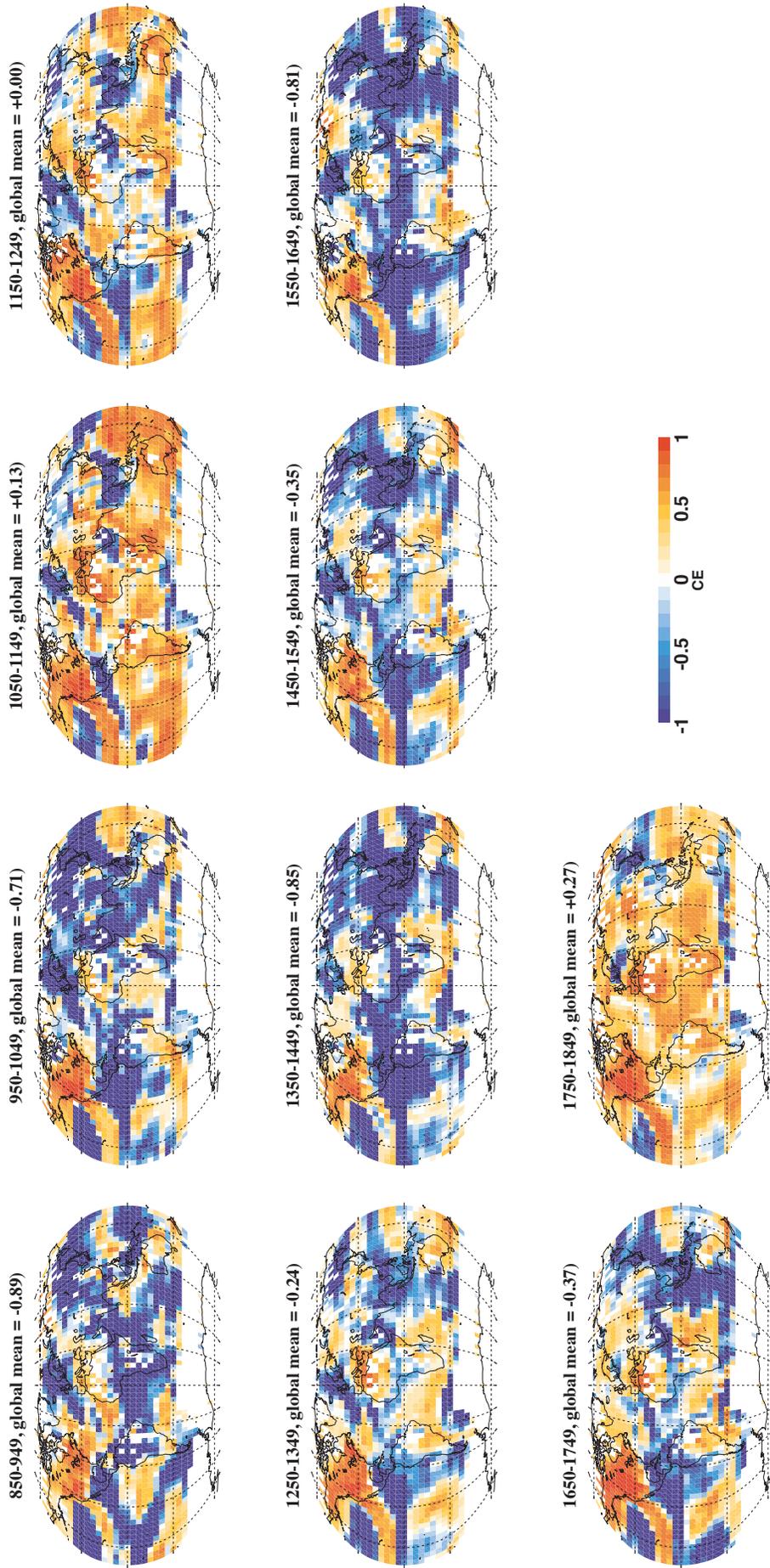


Figure S11: Spatiotemporal variation of GraphEM reconstructions, using the SNR = 0.25, flat network, based on the median of 100-member ensemble.

GraphEM, flat network, SNR = 0.5

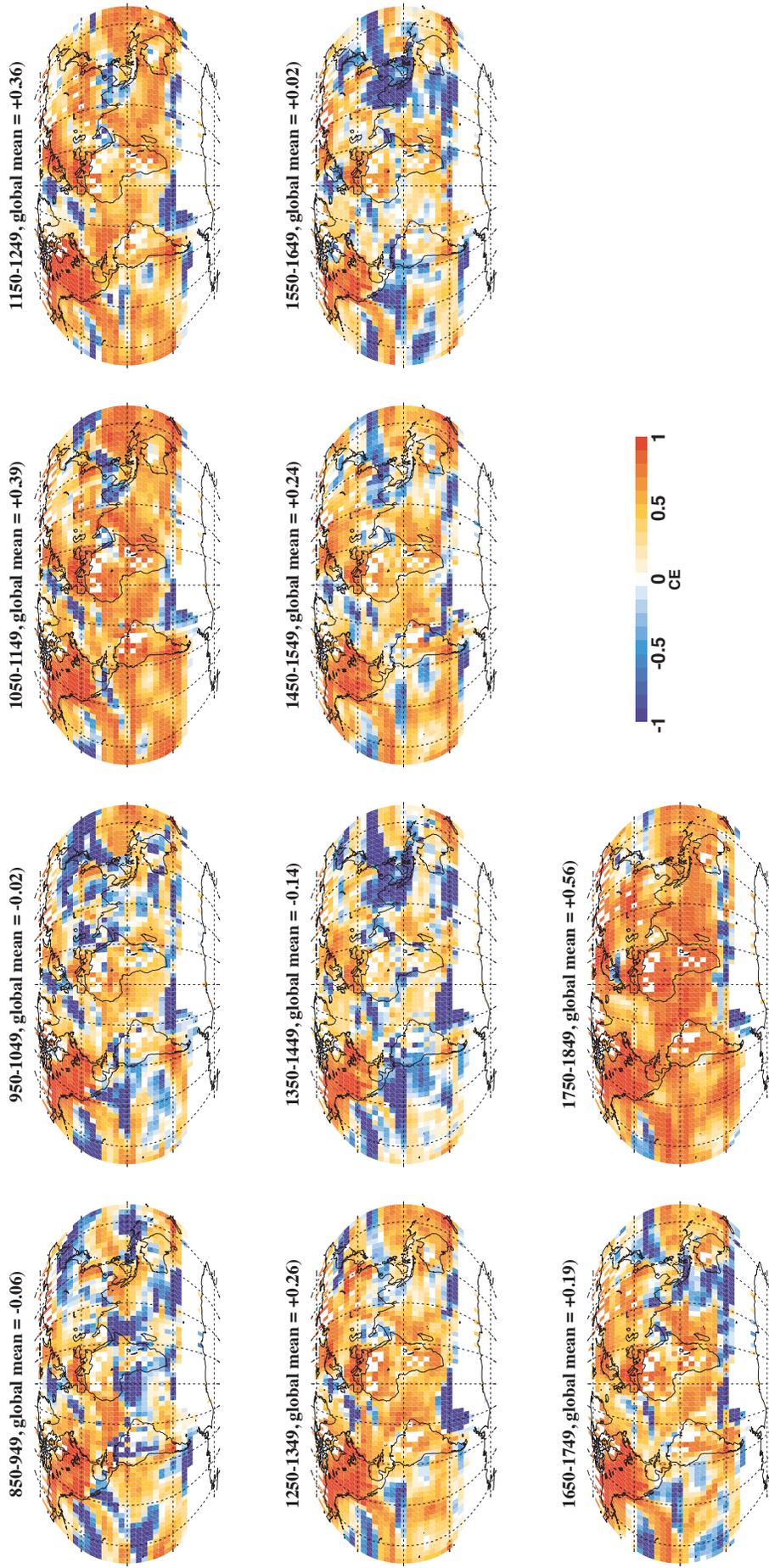


Figure S12: Spatiotemporal variation of GraphEM reconstructions, using the SNR = 0.5, flat network, based on the median of 100-member ensemble.

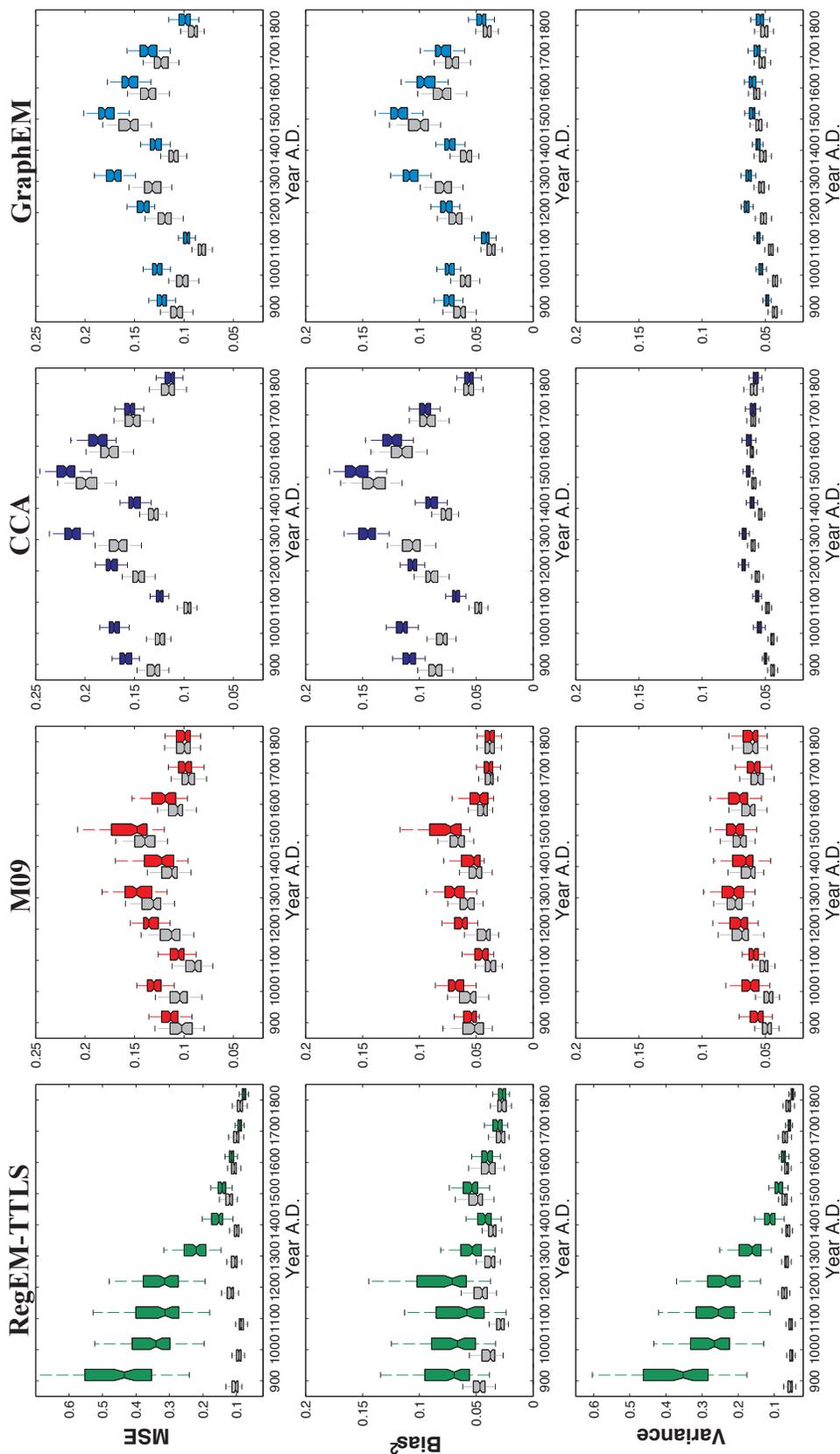


Figure S13: Temporal variation of globally averaged MSE decomposition as bias² + variance, within CFRs derived from the local SNR network. For each CFR method, grey and colored boxplots are reconstructions derived from the flat and staircase networks, respectively. Note that the y scales for RegEM-TTLS are different from the other methods.

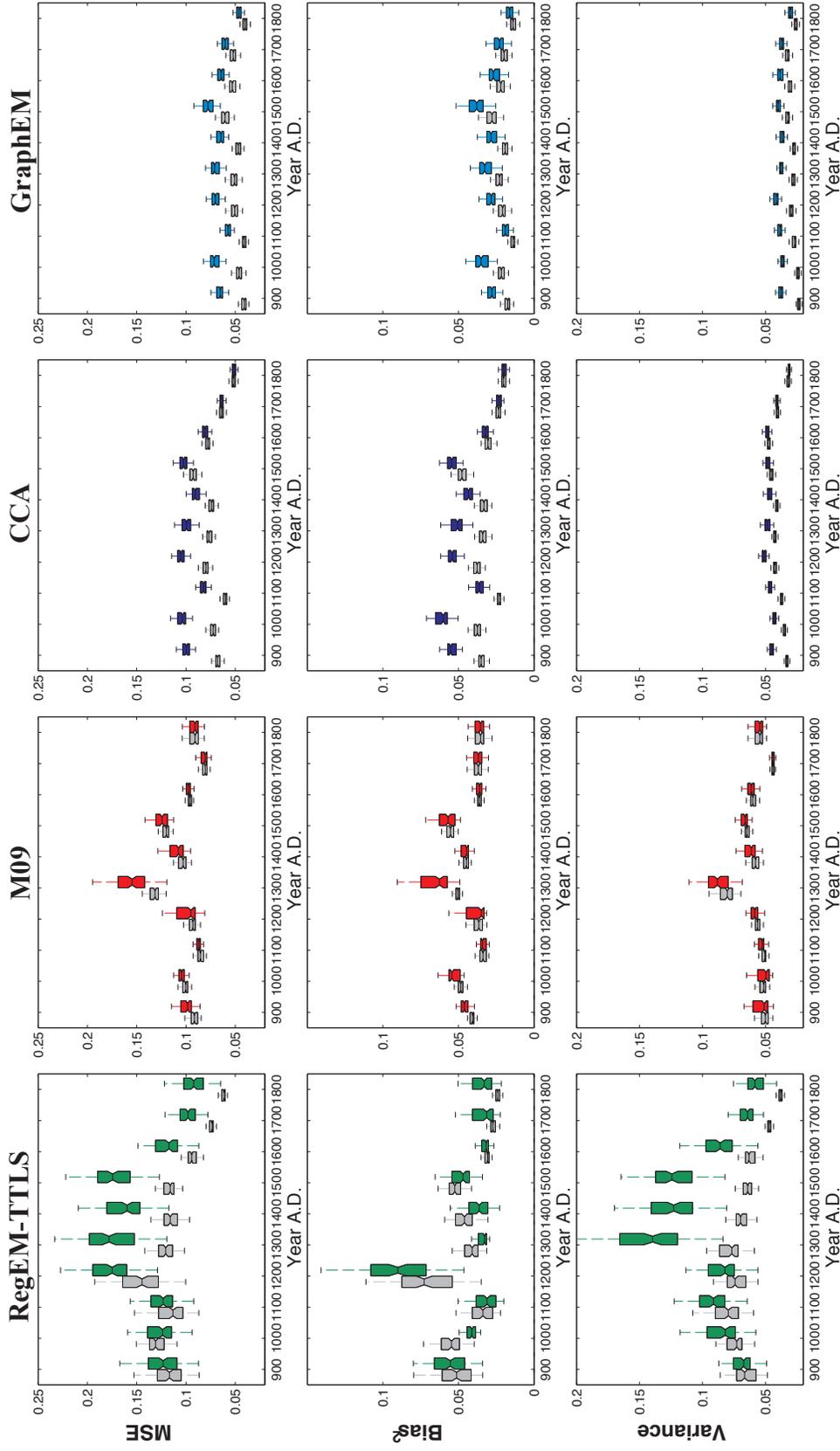


Figure S14: Temporal variation of globally averaged MSE decomposition as bias² + variance, within CFRs derived from the max SNR network. For each CFR method, grey and colored boxplots are reconstructions derived from the flat and staircase networks, respectively.

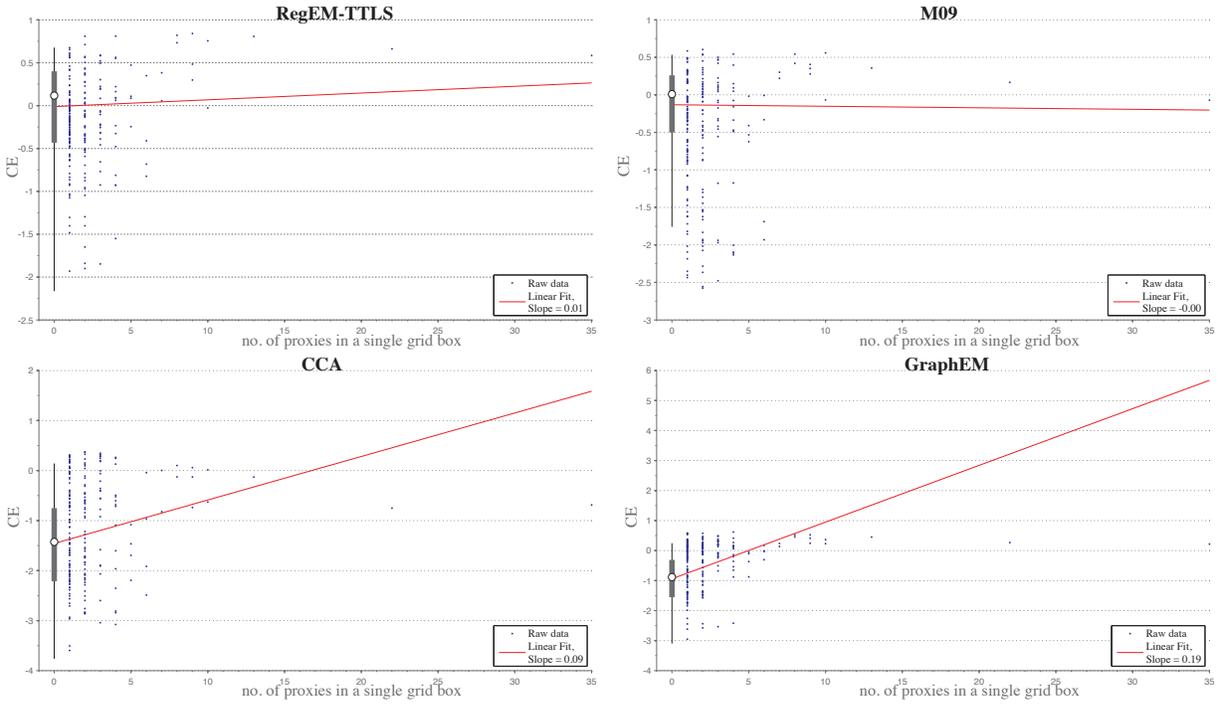


Figure S15: Relationship between CE and number of proxies per grid box, using the staircase network with local SNR.

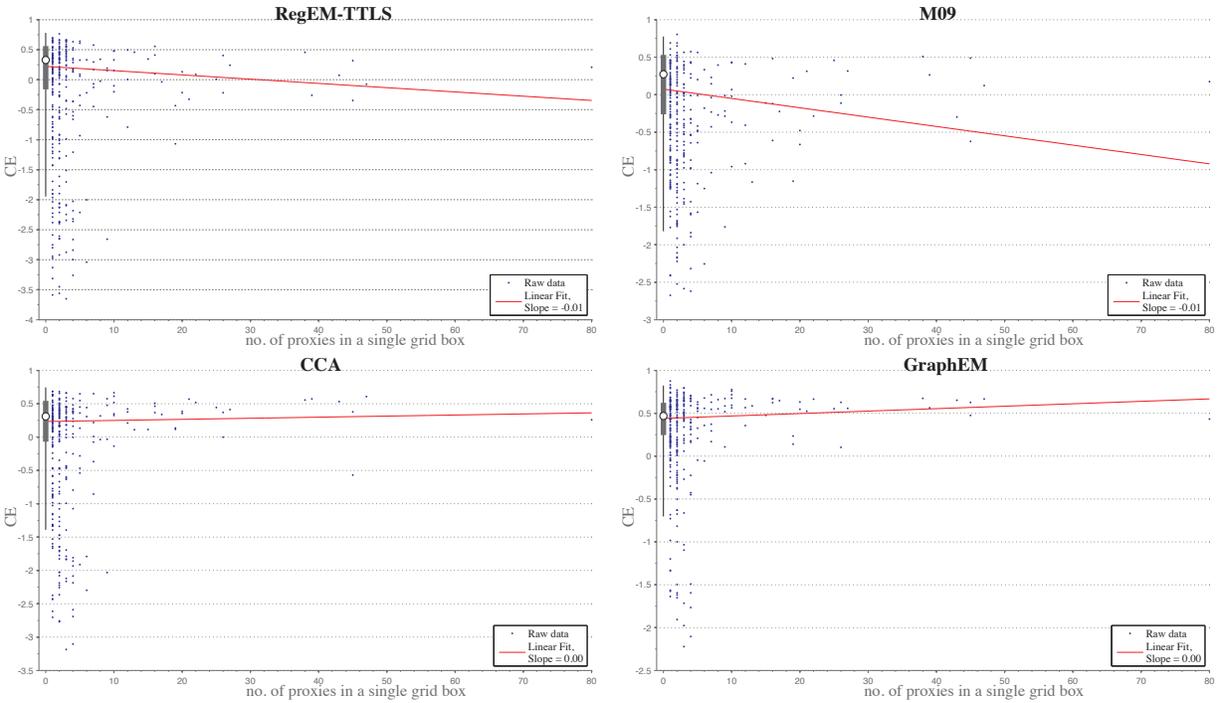


Figure S16: Relationship between CE and number of proxies per grid box, using the staircase network with max SNR.

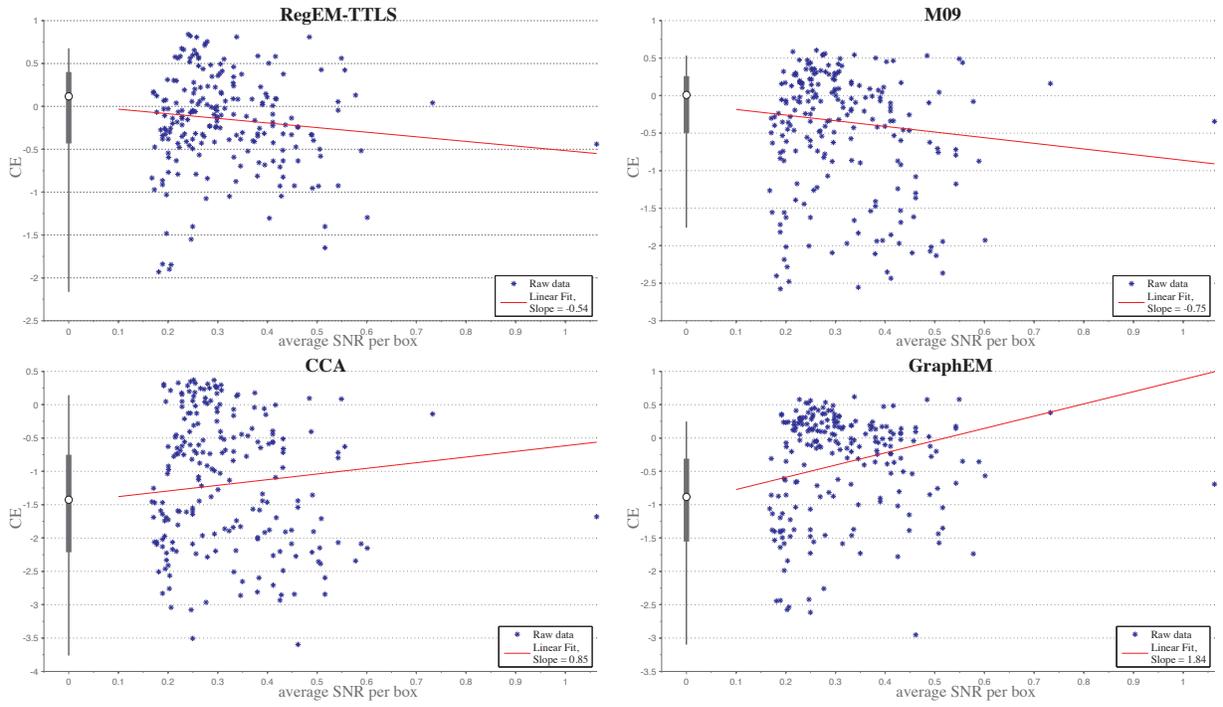


Figure S17: Relationship between CE and average SNR per grid box, using the staircase network with local SNR.

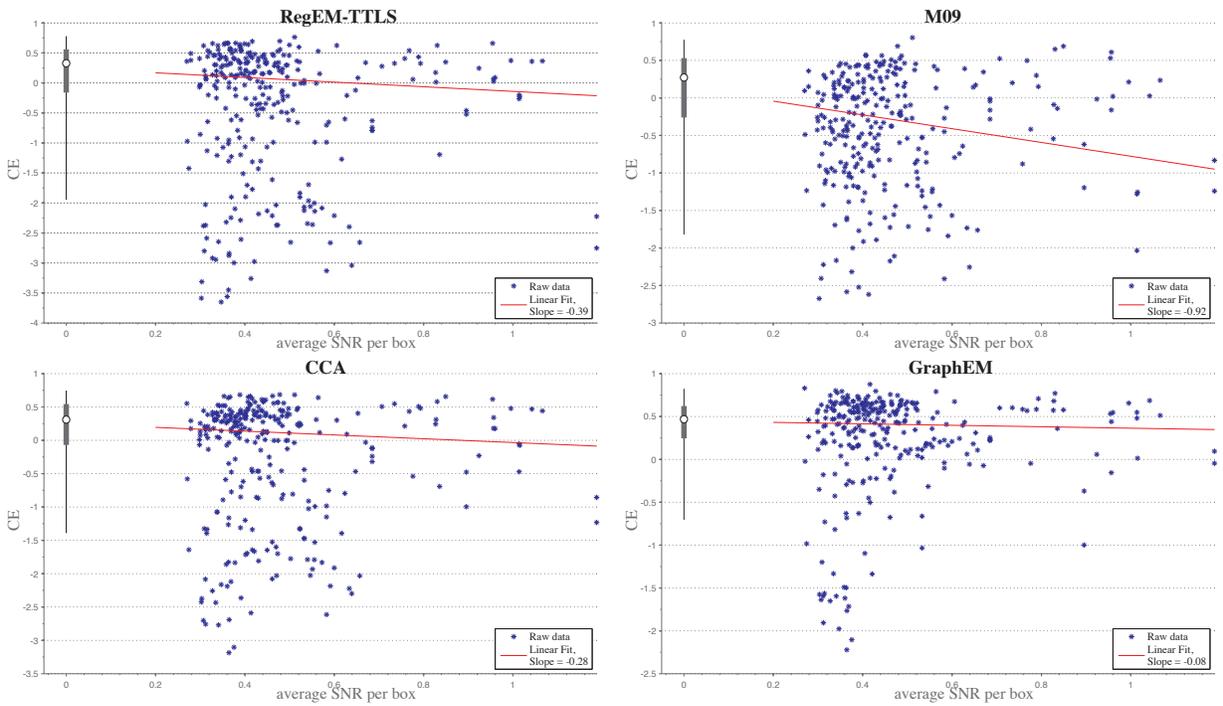


Figure S18: Relationship between CE and average SNR per grid box, using the staircase network with max SNR.

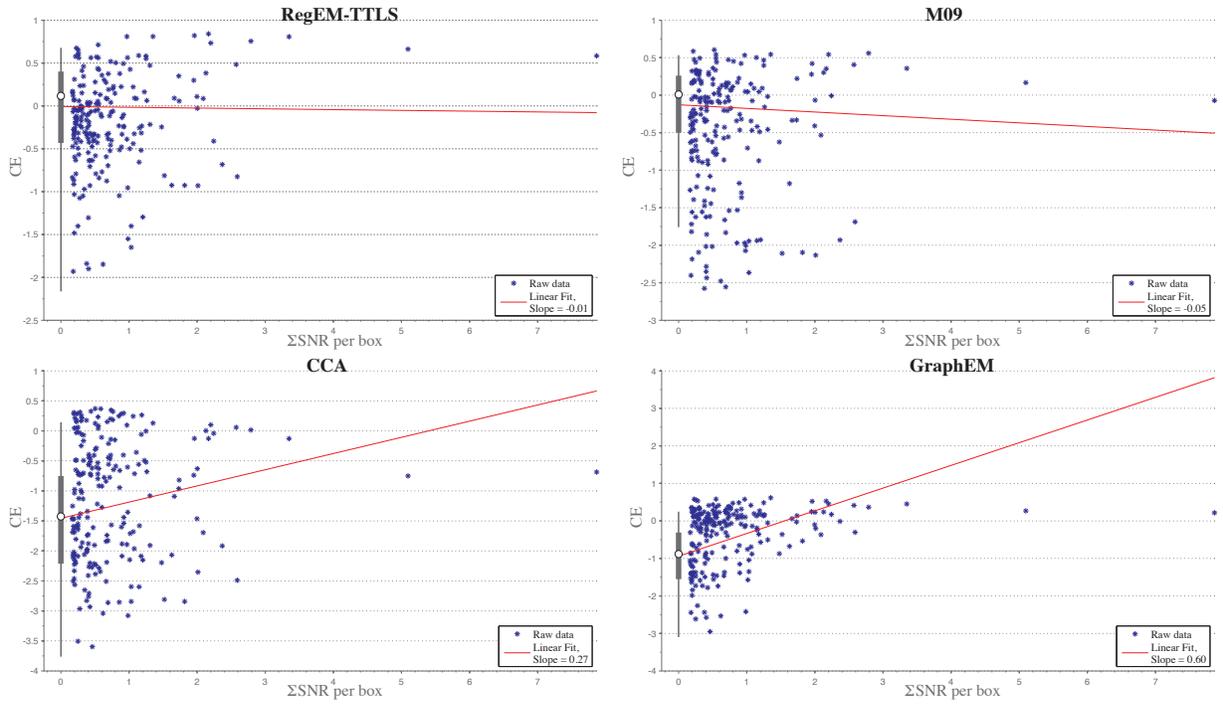


Figure S19: Relationship between CE and sum of SNR per grid box, using the staircase network with local SNR.

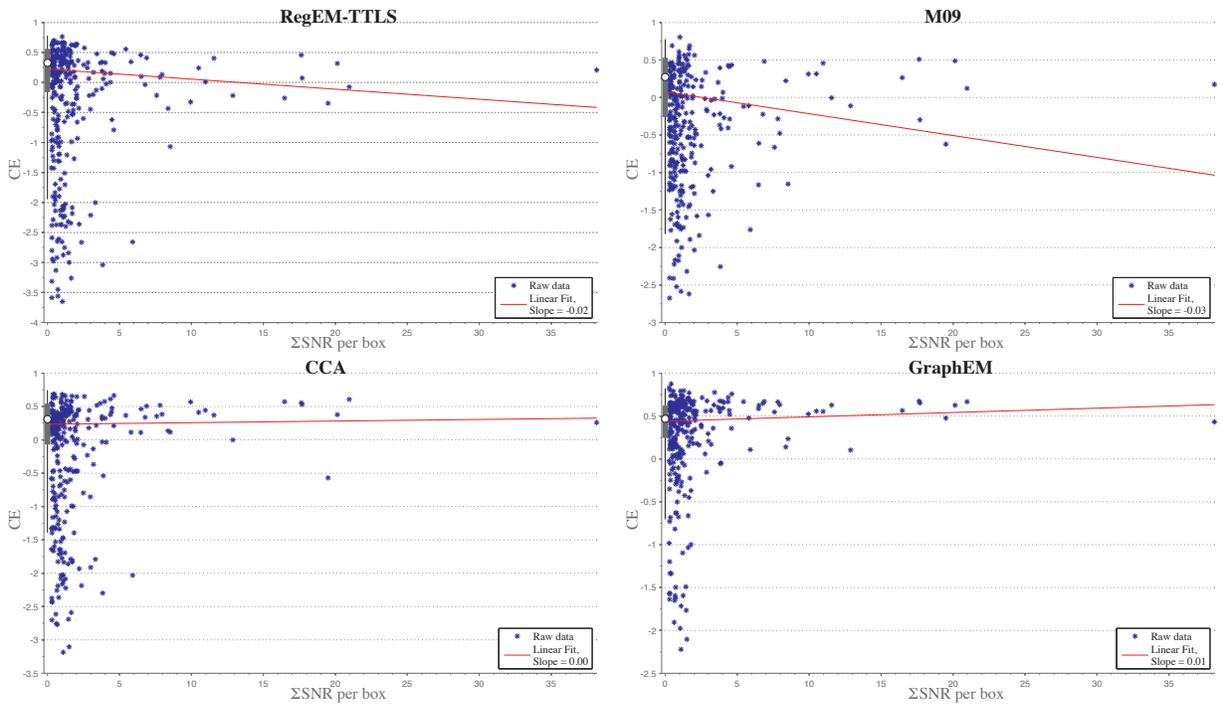


Figure S20: Relationship between CE and sum of SNR per grid box, using the staircase network with max SNR.

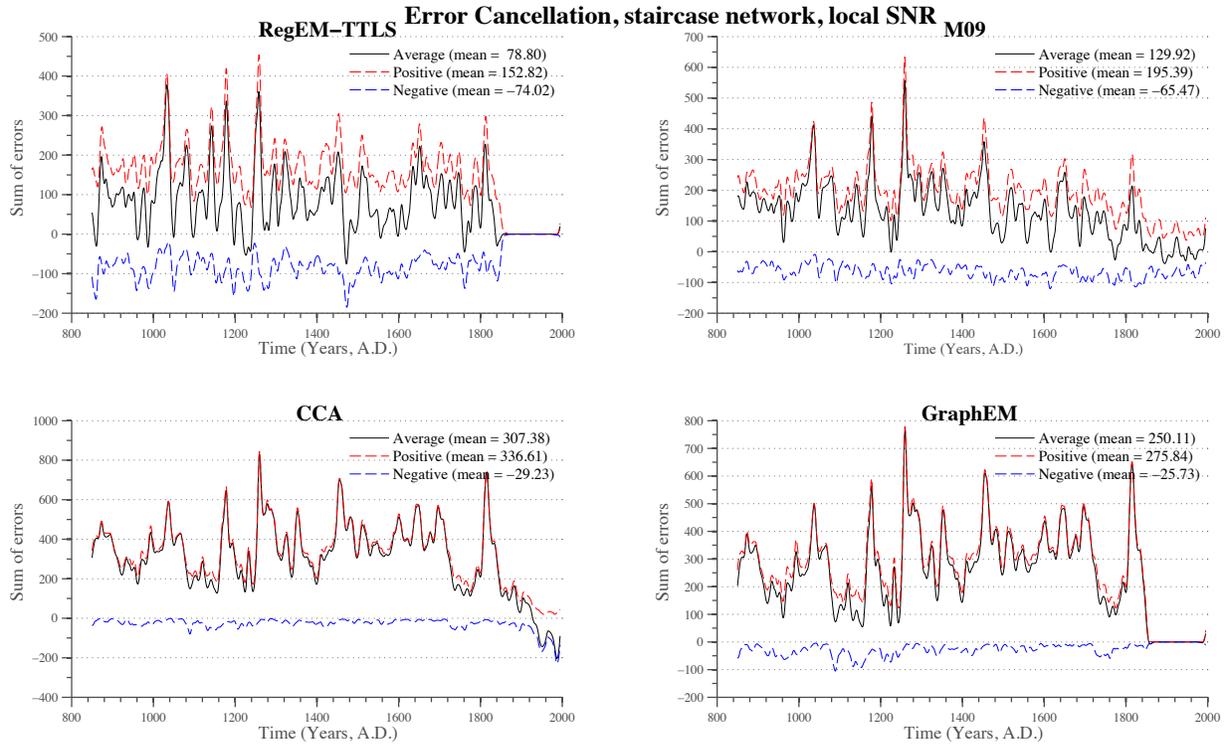


Figure S21: Error cancellation analysis, using the staircase network with local SNR.

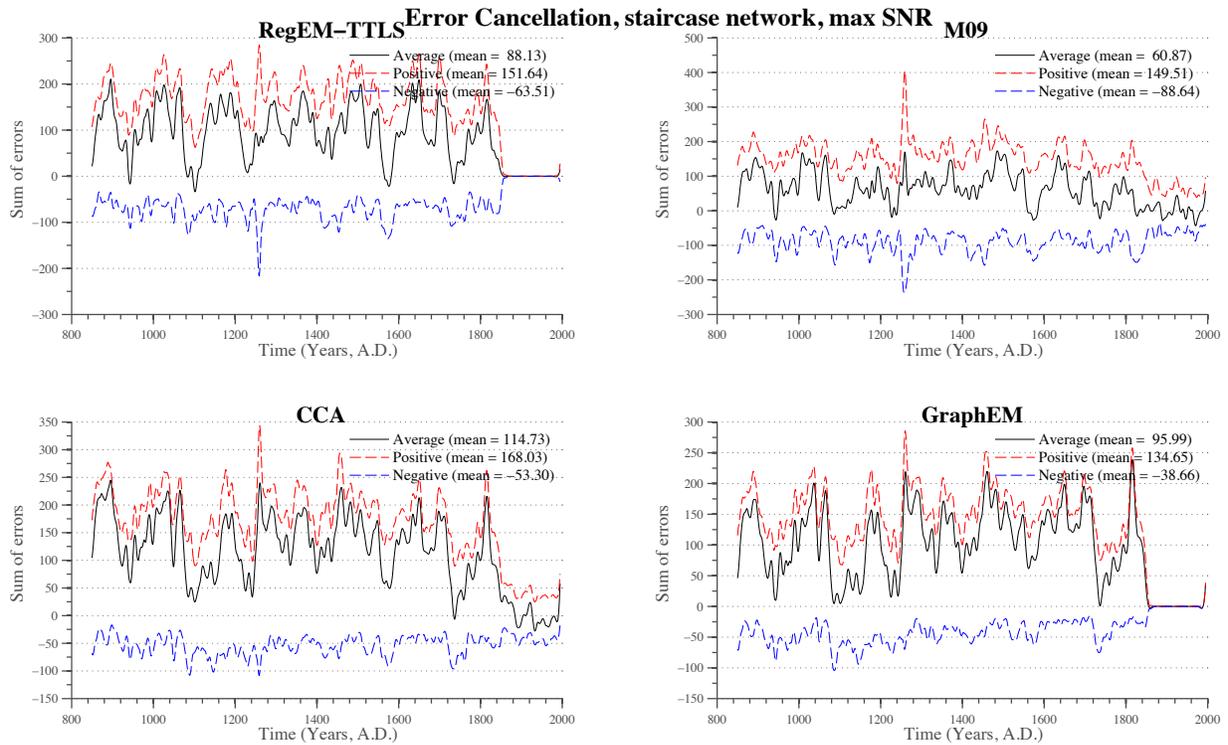


Figure S22: Error cancellation analysis, using the staircase network with max SNR.

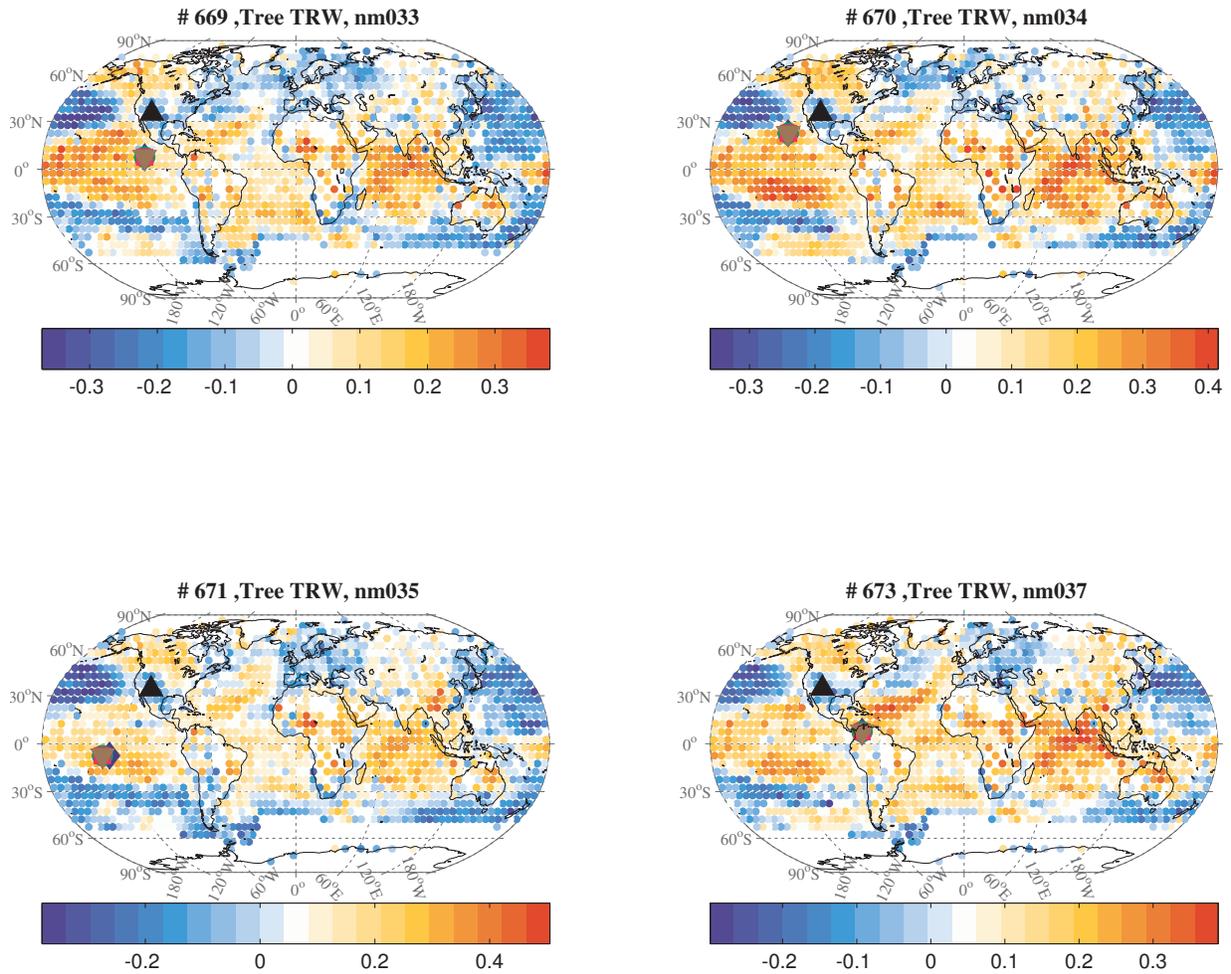


Figure S23: An example of maps of correlations for a given proxy. On each map, the black triangle represents the given proxy record. For instance, the top left map shows proxy No.669 in the Mann et al. (2008) network, known as the “nm033” tree ring width record. Colors represent the pairwise proxy-temperature correlations, with positive correlations represented by yellowish and reddish dots, negative correlations represented by blueish dots. Grey inverse triangles pinpoint the temperature grid points corresponding to the maximum distance-weighted correlations (absolute values).

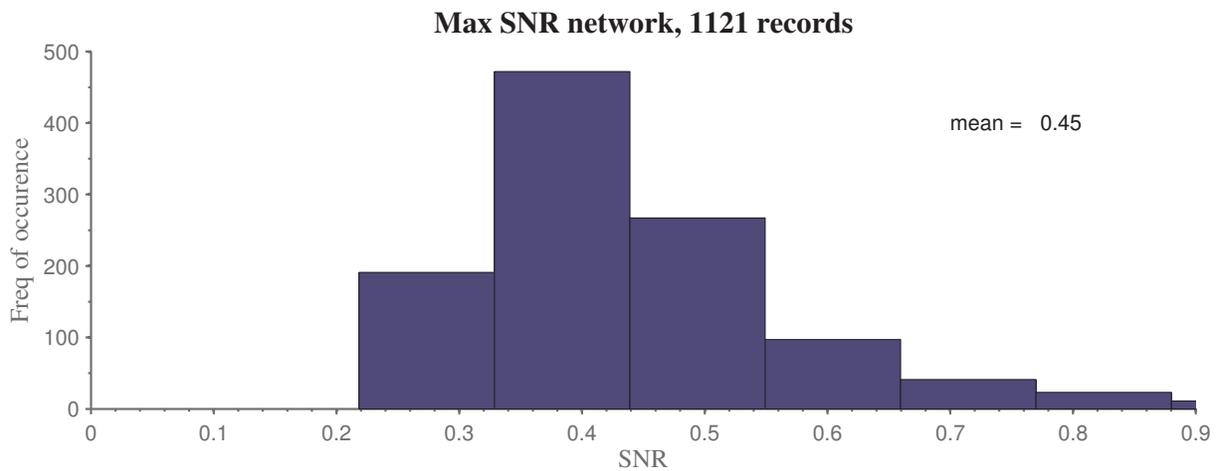
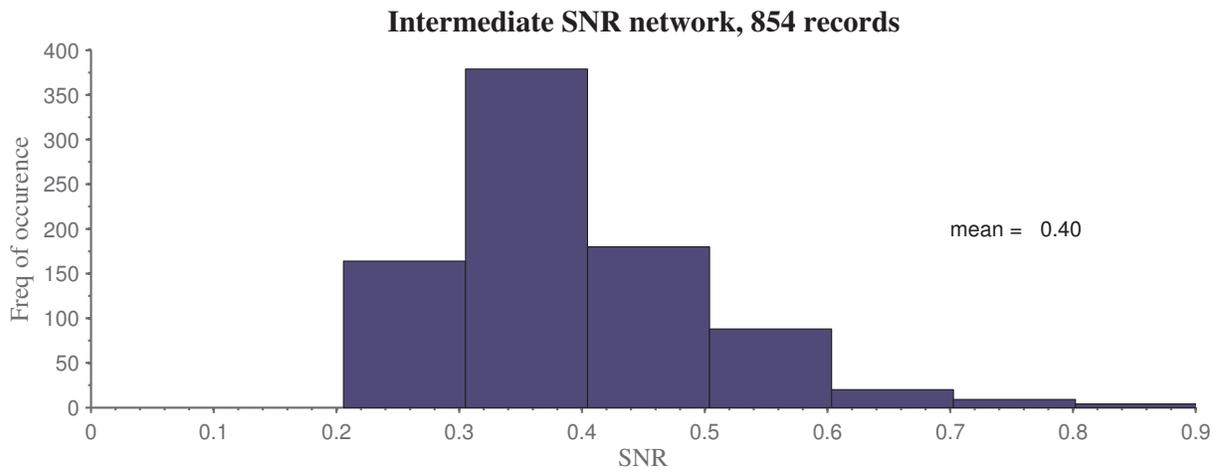
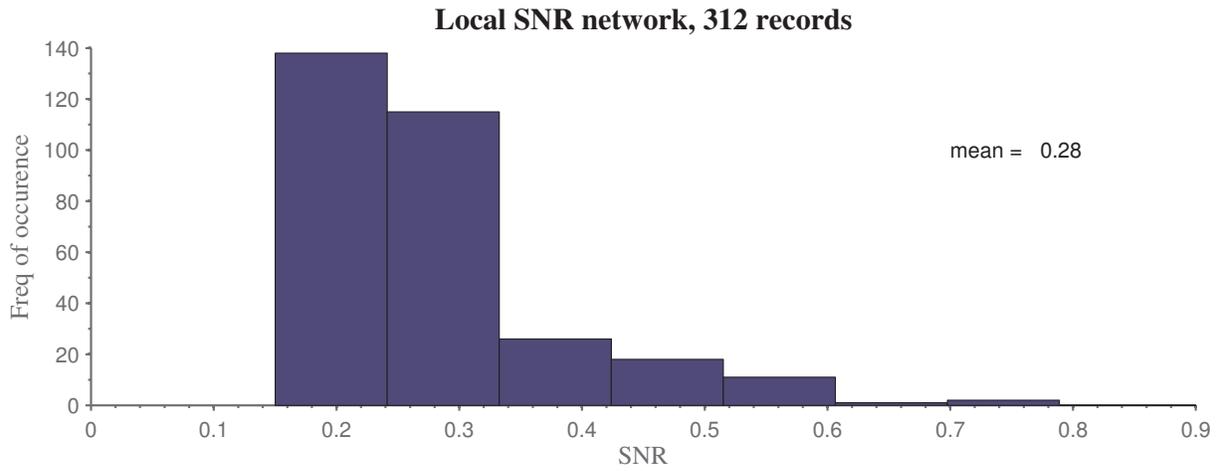


Figure S24: Histograms of different designs of realistic SNR networks. Top: local SNR network, middle: intermediate SNR network, bottom: max SNR network

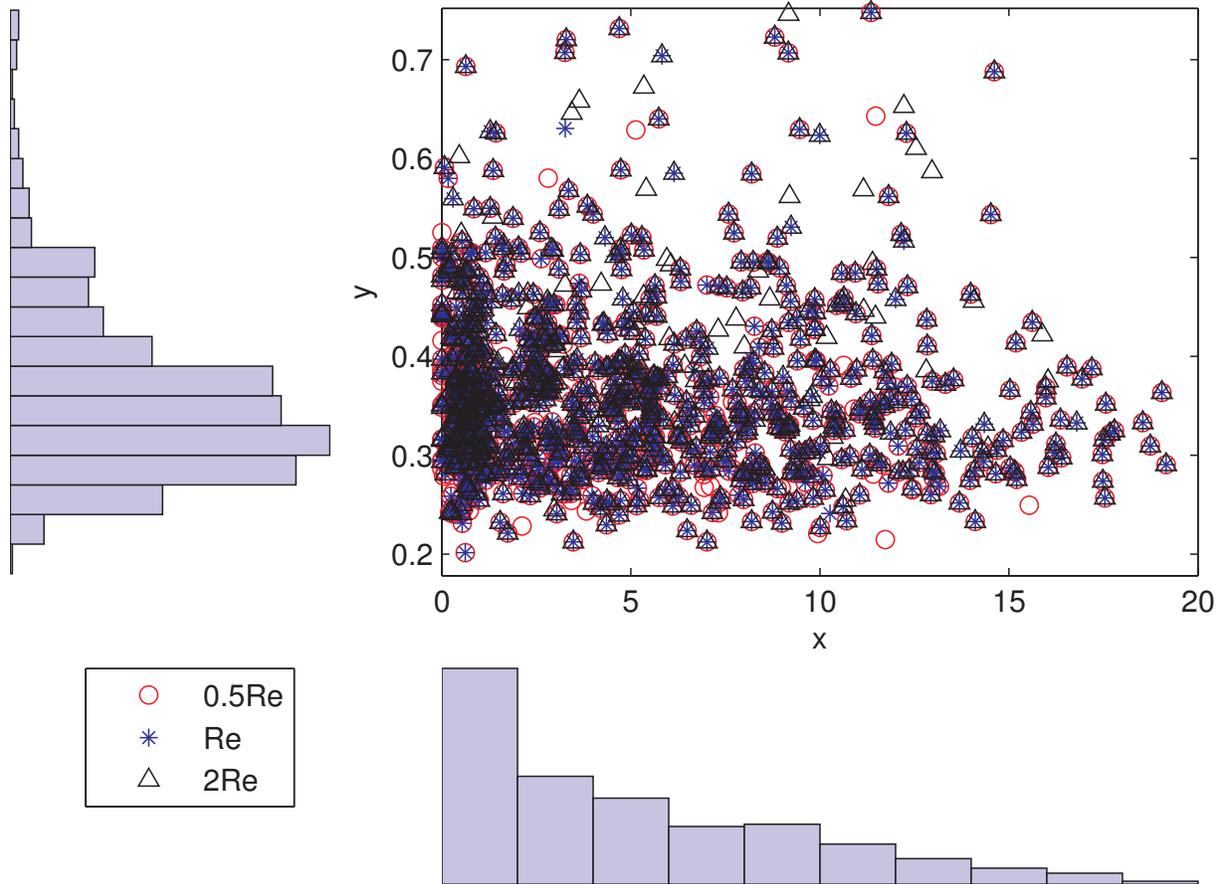


Figure S25: Maximum absolute correlation coefficient $|\rho|$ between M08 proxies and the HadCRUT3v grid point temperatures vs. the corresponding distance between the proxy location and the grid point, with a distance constraint of Re (one Earth radius). On the y-axis is the histogram of the maximum distance-weighted $|\rho|$; on the x-axis is the histogram of distance between each proxy \mathbf{P}_i and the corresponding temperature grid \mathbf{T}_j that gives the highest distance-weighted $|\rho|$.

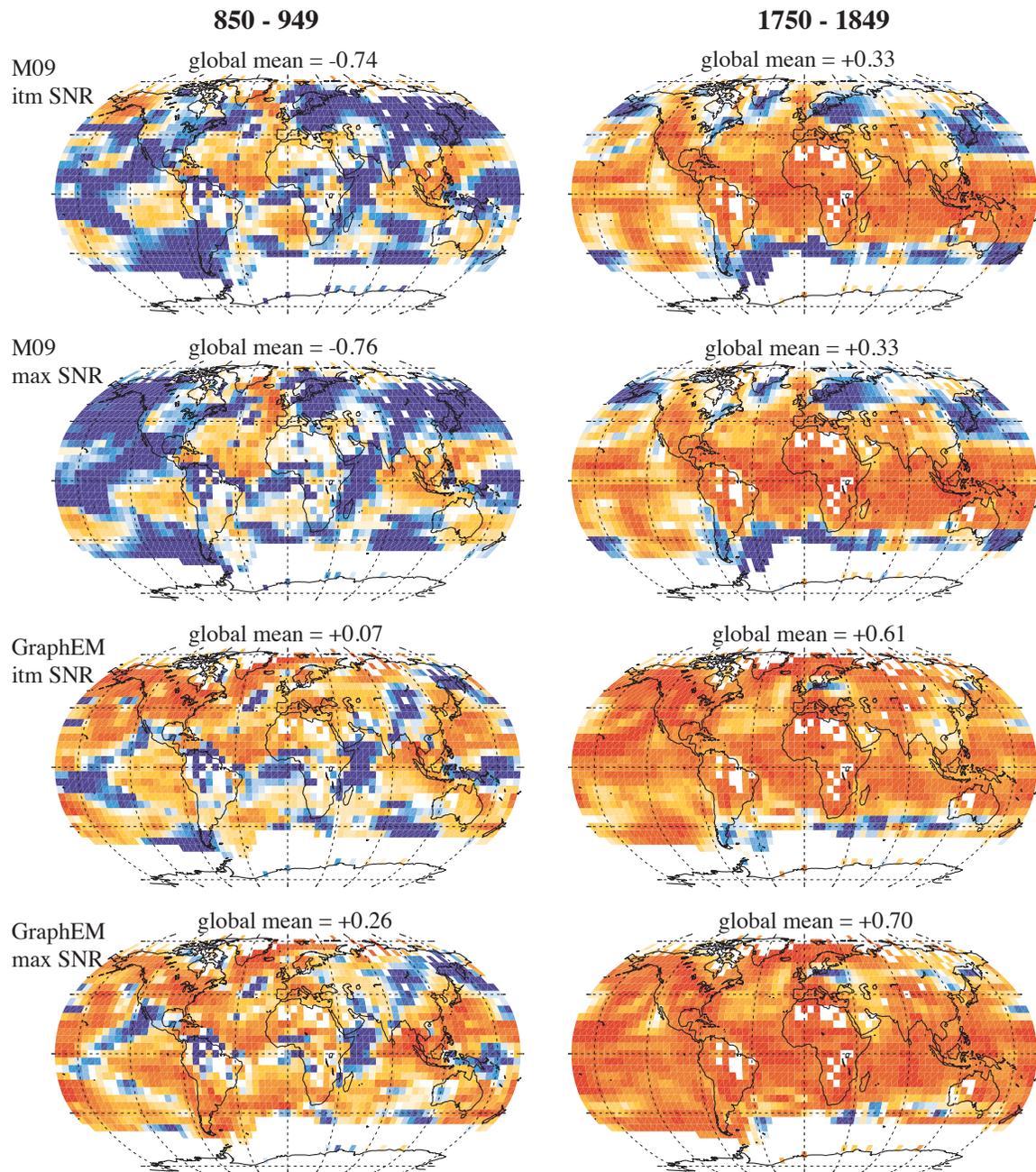


Figure S26: Spatial Pattern of CE, a comparison between the intermediate and max SNR flat network. 850 - 949 AD and 1750–1849 AD represent periods with the minimum and maximum global mean CE over the entire reconstruction interval, respectively. Comparisons are based on the median of a 20-realization ensemble.

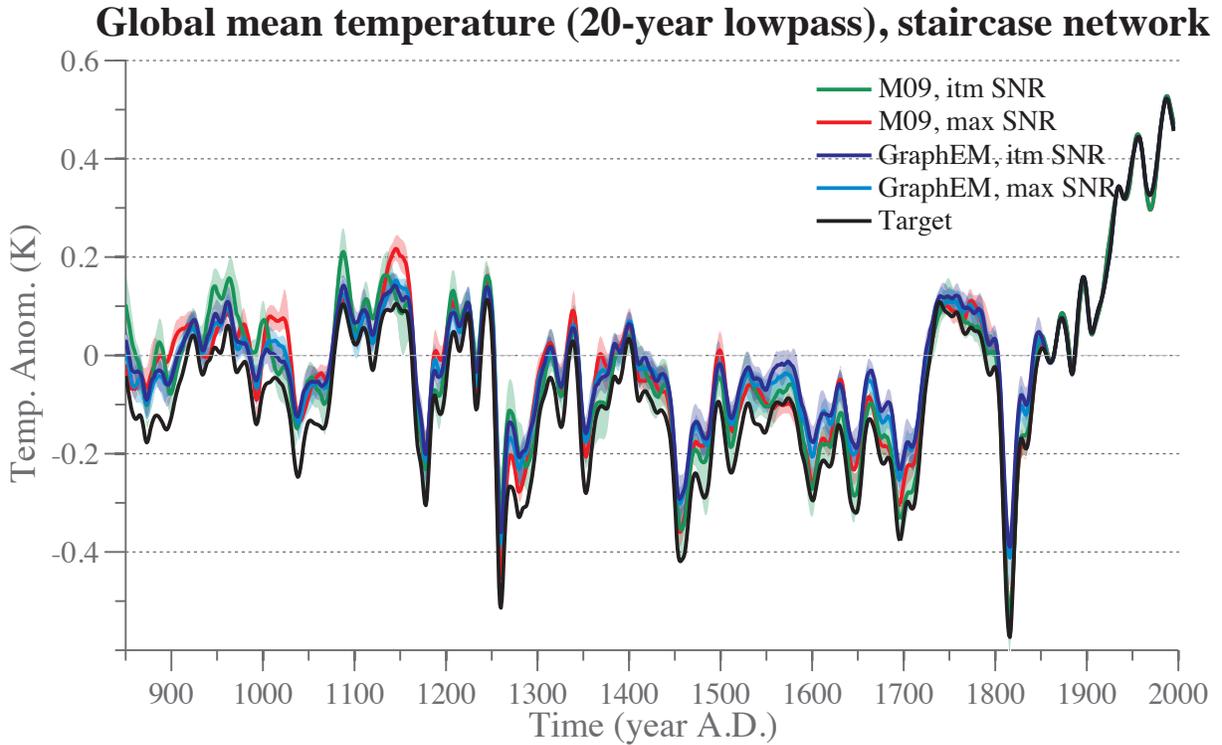


Figure S27: Area-weighted global mean time series comparison of M09-TTLS and GraphEM, with the staircase intermediate and max network. Only the low-frequency (20-year lowpass) component is plotted. Black line: target temperature from the CSM1.4 model output; colored lines: reconstructed temperature from median of the reconstruction ensembles; shaded areas: [2.5%,97.5%] quantiles derived from the reconstruction ensembles. Comparisons are based on results of 20 realizations.

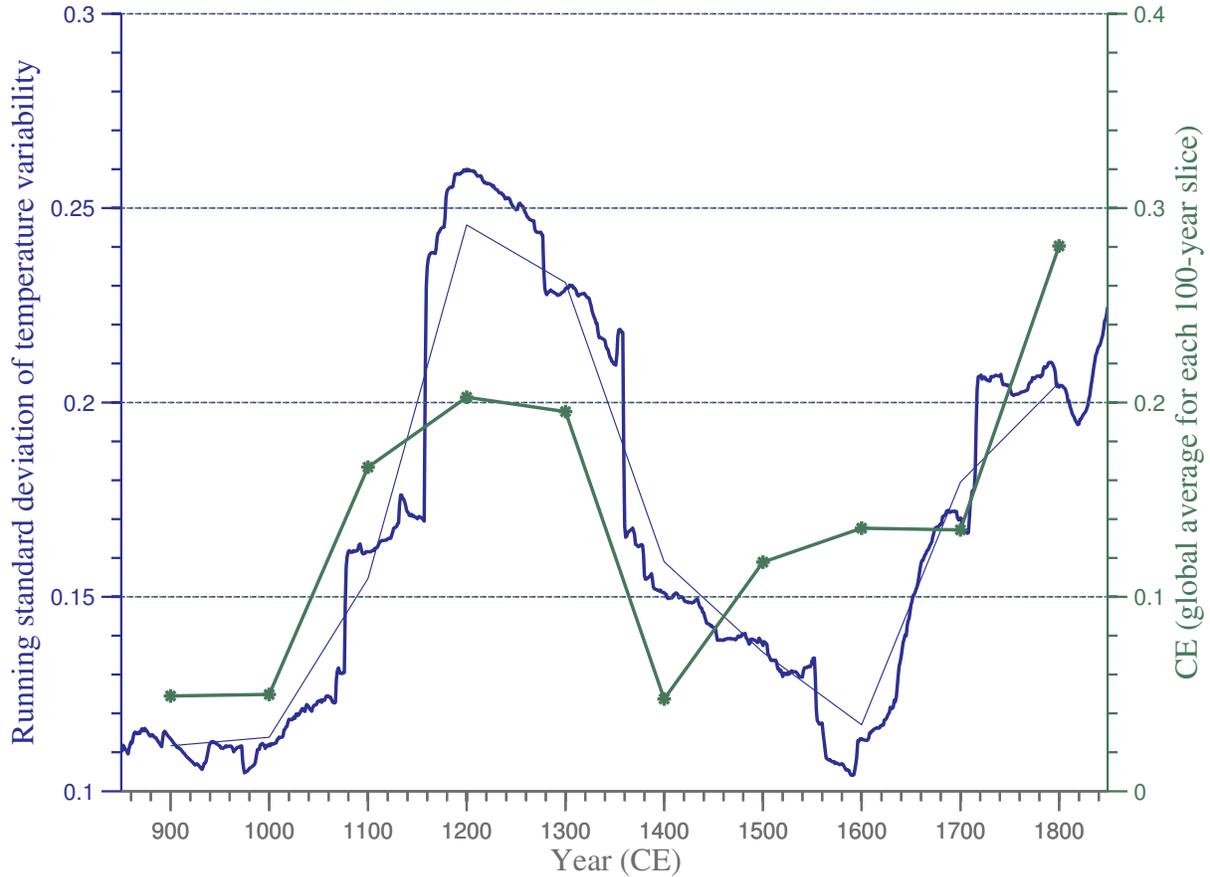


Figure S28: Temperature variability and CE skill comparison. Temperature variability is quantified as the running standard deviation using sliding 100-year windows for the simulated global mean temperature in NCAR CSM1.4, shown as the blue thick curve; the blue thin curve corresponds to the thick one's 100-year average. The CE scores here are the same as values in the M09-TTLS case in Fig. 8 using the flat network, shown as the green curve. The time series (blue thin curve vs. green curve) have a correlation of 0.74 ($p < 0.05$).

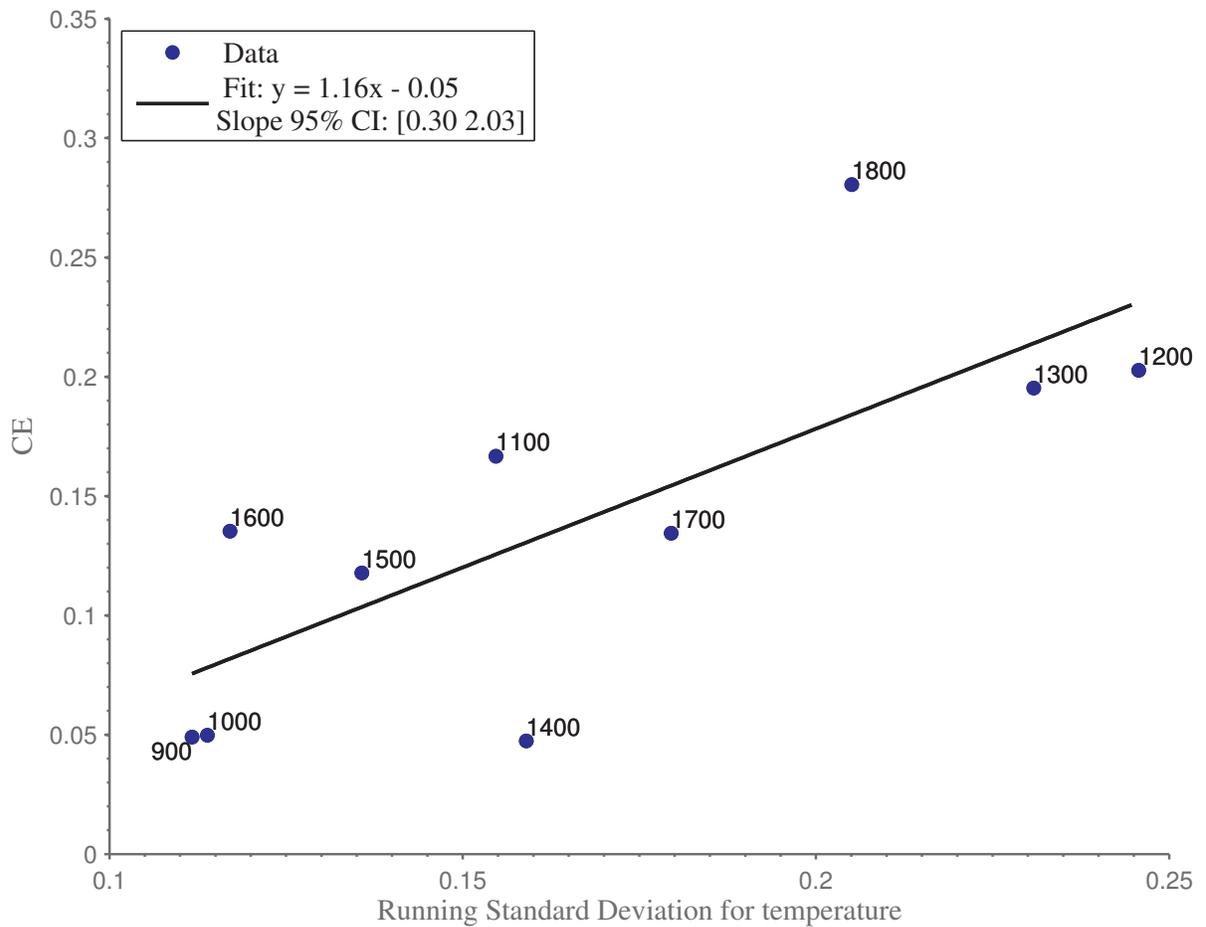


Figure S29: Regression fitting of globally averaged CE onto temperature variability of the global mean time series. Blue dots: (temperature variability, CE) pairs for each 100-year time slice; black curve: linear regression fit via OLS.

SNR	Metric	900	1000	1100	1200	1300	1400	1500	1600	1700	1800
RegEM - TTLS											
local	CE	-1.51 _(0.49)	-1.31 _(0.34)	-0.31 _(0.19)	-0.46 _(0.23)	-0.42 _(0.18)	-1.48 _(0.35)	-0.67 _(0.23)	-0.82 _(0.28)	-0.66 _(0.26)	+0.02 _(0.14)
	RE	+0.42 _(0.10)	+0.35 _(0.08)	+0.09 _(0.16)	+0.13 _(0.14)	+0.54 _(0.06)	+0.38 _(0.07)	+0.53 _(0.06)	+0.65 _(0.04)	+0.61 _(0.05)	+0.59 _(0.05)
	R ²	+0.20 _(0.03)	+0.19 _(0.03)	+0.28 _(0.04)	+0.31 _(0.04)	+0.39 _(0.06)	+0.19 _(0.03)	+0.21 _(0.04)	+0.25 _(0.03)	+0.47 _(0.04)	+0.53 _(0.05)
max	CE	-1.34 _(0.48)	-1.01 _(0.22)	-0.33 _(0.18)	-0.65 _(0.30)	-0.65 _(0.21)	-1.60 _(0.38)	-0.80 _(0.26)	-0.60 _(0.18)	-0.27 _(0.13)	+0.16 _(0.10)
	RE	+0.35 _(0.12)	+0.39 _(0.08)	+0.07 _(0.13)	-0.00 _(0.17)	+0.45 _(0.07)	+0.27 _(0.10)	+0.46 _(0.07)	+0.64 _(0.04)	+0.63 _(0.04)	+0.63 _(0.05)
	R ²	+0.24 _(0.03)	+0.29 _(0.02)	+0.36 _(0.02)	+0.43 _(0.03)	+0.49 _(0.02)	+0.30 _(0.02)	+0.36 _(0.02)	+0.29 _(0.02)	+0.47 _(0.02)	+0.63 _(0.02)
M09											
local	CE	-1.76 _(1.08)	-1.47 _(0.94)	-0.29 _(0.37)	-0.49 _(0.45)	-0.56 _(0.37)	-1.30 _(0.60)	-0.65 _(0.40)	-0.77 _(0.41)	-0.47 _(0.21)	+0.09 _(0.15)
	RE	+0.45 _(0.16)	+0.37 _(0.17)	+0.24 _(0.19)	+0.19 _(0.20)	+0.53 _(0.08)	+0.44 _(0.08)	+0.58 _(0.07)	+0.65 _(0.06)	+0.63 _(0.04)	+0.61 _(0.06)
	R ²	+0.16 _(0.02)	+0.20 _(0.04)	+0.29 _(0.05)	+0.30 _(0.07)	+0.33 _(0.08)	+0.15 _(0.04)	+0.22 _(0.06)	+0.19 _(0.04)	+0.48 _(0.05)	+0.57 _(0.07)
max	CE	-0.80 _(0.13)	-0.55 _(0.07)	+0.09 _(0.03)	+0.02 _(0.06)	-0.21 _(0.06)	-0.55 _(0.07)	-0.19 _(0.06)	-0.21 _(0.05)	+0.03 _(0.03)	+0.30 _(0.04)
	RE	+0.55 _(0.03)	+0.59 _(0.02)	+0.43 _(0.02)	+0.41 _(0.02)	+0.58 _(0.02)	+0.53 _(0.02)	+0.64 _(0.02)	+0.71 _(0.01)	+0.71 _(0.01)	+0.68 _(0.02)
	R ²	+0.20 _(0.01)	+0.34 _(0.01)	+0.40 _(0.01)	+0.48 _(0.01)	+0.56 _(0.01)	+0.37 _(0.01)	+0.42 _(0.01)	+0.31 _(0.01)	+0.58 _(0.01)	+0.68 _(0.01)
CCA											
local	CE	-2.76 _(0.40)	-2.05 _(0.28)	-0.32 _(0.10)	-0.63 _(0.15)	-1.47 _(0.21)	-2.71 _(0.35)	-1.98 _(0.30)	-3.02 _(0.45)	-2.04 _(0.30)	-0.54 _(0.14)
	RE	+0.38 _(0.06)	+0.36 _(0.05)	+0.34 _(0.04)	+0.32 _(0.05)	+0.41 _(0.04)	+0.38 _(0.04)	+0.42 _(0.05)	+0.43 _(0.05)	+0.43 _(0.05)	+0.42 _(0.05)
	R ²	+0.18 _(0.03)	+0.20 _(0.03)	+0.29 _(0.04)	+0.33 _(0.04)	+0.38 _(0.06)	+0.17 _(0.04)	+0.22 _(0.05)	+0.19 _(0.04)	+0.46 _(0.04)	+0.54 _(0.05)
max	CE	-0.18 _(0.05)	-0.17 _(0.05)	+0.35 _(0.02)	+0.31 _(0.02)	+0.19 _(0.03)	-0.22 _(0.06)	+0.09 _(0.04)	-0.01 _(0.05)	+0.15 _(0.04)	+0.47 _(0.02)
	RE	+0.72 _(0.01)	+0.70 _(0.01)	+0.61 _(0.01)	+0.61 _(0.01)	+0.75 _(0.01)	+0.69 _(0.01)	+0.75 _(0.01)	+0.79 _(0.01)	+0.78 _(0.01)	+0.78 _(0.01)
	R ²	+0.41 _(0.01)	+0.47 _(0.01)	+0.53 _(0.01)	+0.60 _(0.01)	+0.68 _(0.01)	+0.49 _(0.01)	+0.56 _(0.01)	+0.43 _(0.02)	+0.64 _(0.01)	+0.76 _(0.01)
GraphEM											
local	CE	-2.29 _(0.37)	-1.70 _(0.27)	-0.24 _(0.10)	-0.49 _(0.13)	-1.04 _(0.19)	-2.23 _(0.30)	-1.41 _(0.22)	-2.23 _(0.34)	-1.54 _(0.24)	-0.24 _(0.10)
	RE	+0.44 _(0.05)	+0.41 _(0.05)	+0.36 _(0.04)	+0.35 _(0.04)	+0.50 _(0.04)	+0.44 _(0.04)	+0.51 _(0.04)	+0.53 _(0.04)	+0.52 _(0.04)	+0.52 _(0.04)
	R ²	+0.18 _(0.02)	+0.21 _(0.03)	+0.29 _(0.03)	+0.33 _(0.04)	+0.37 _(0.05)	+0.18 _(0.03)	+0.24 _(0.03)	+0.20 _(0.03)	+0.43 _(0.03)	+0.52 _(0.04)
max	CE	-0.02 _(0.06)	+0.06 _(0.06)	+0.49 _(0.02)	+0.46 _(0.03)	+0.38 _(0.05)	-0.00 _(0.07)	+0.28 _(0.05)	+0.10 _(0.07)	+0.25 _(0.06)	+0.57 _(0.03)
	RE	+0.77 _(0.01)	+0.76 _(0.01)	+0.68 _(0.02)	+0.71 _(0.02)	+0.83 _(0.01)	+0.78 _(0.01)	+0.82 _(0.01)	+0.83 _(0.01)	+0.82 _(0.01)	+0.82 _(0.01)
	R ²	+0.47 _(0.02)	+0.54 _(0.02)	+0.58 _(0.02)	+0.65 _(0.02)	+0.72 _(0.01)	+0.55 _(0.02)	+0.62 _(0.02)	+0.50 _(0.02)	+0.65 _(0.01)	+0.78 _(0.01)

Table S1: Verification statistics summary for the weighted global average, with the flat network. CE , RE and R^2 are computed for each 100-year time slice of interest. All numbers given outside of parentheses are the mean of the 100-member ensemble; numbers in parentheses are the corresponding standard deviation.

SNR	Metric	900	1000	1100	1200	1300	1400	1500	1600	1700	1800
RegEM - TTLS											
local	CE	-10.98 _(6.52)	-6.31 _(2.35)	-3.34 _(1.33)	-2.68 _(1.12)	-1.58 _(0.56)	-2.56 _(0.75)	-0.84 _(0.29)	-0.75 _(0.29)	-0.42 _(0.19)	0.21 _(0.08)
	RE	-1.85 _(1.57)	-0.87 _(0.58)	-1.61 _(0.82)	-0.93 _(0.57)	0.24 _(0.16)	0.15 _(0.16)	0.47 _(0.08)	0.65 _(0.04)	0.66 _(0.03)	0.66 _(0.03)
	R ²	0.12 _(0.02)	0.12 _(0.03)	0.17 _(0.05)	0.18 _(0.05)	0.25 _(0.08)	0.14 _(0.04)	0.21 _(0.05)	0.20 _(0.04)	0.48 _(0.03)	0.59 _(0.04)
max	CE	-1.11 _(0.32)	-1.03 _(0.18)	-0.18 _(0.13)	-0.27 _(0.19)	-0.23 _(0.10)	-0.84 _(0.14)	-0.23 _(0.08)	-0.30 _(0.08)	-0.00 _(0.04)	0.39 _(0.02)
	RE	0.48 _(0.08)	0.46 _(0.05)	0.25 _(0.08)	0.29 _(0.10)	0.62 _(0.03)	0.53 _(0.03)	0.65 _(0.02)	0.73 _(0.01)	0.72 _(0.01)	0.75 _(0.01)
	R ²	0.22 _(0.02)	0.29 _(0.02)	0.33 _(0.03)	0.41 _(0.02)	0.51 _(0.02)	0.30 _(0.02)	0.39 _(0.02)	0.32 _(0.01)	0.54 _(0.01)	0.70 _(0.01)
M09											
local	CE	-1.67 _(0.61)	-1.42 _(0.54)	-0.35 _(0.20)	-0.35 _(0.20)	-0.76 _(0.31)	-1.44 _(0.50)	-0.79 _(0.36)	-0.83 _(0.40)	-0.51 _(0.35)	0.09 _(0.15)
	RE	0.47 _(0.08)	0.42 _(0.11)	0.28 _(0.11)	0.30 _(0.08)	0.50 _(0.07)	0.44 _(0.08)	0.51 _(0.09)	0.63 _(0.06)	0.62 _(0.07)	0.61 _(0.06)
	R ²	0.13 _(0.04)	0.14 _(0.05)	0.23 _(0.06)	0.22 _(0.10)	0.31 _(0.09)	0.14 _(0.05)	0.16 _(0.05)	0.17 _(0.04)	0.46 _(0.05)	0.57 _(0.07)
max	CE	-0.68 _(0.22)	-0.54 _(0.08)	0.04 _(0.07)	0.03 _(0.10)	-0.45 _(0.16)	-0.65 _(0.13)	-0.24 _(0.09)	-0.23 _(0.06)	0.01 _(0.04)	0.29 _(0.04)
	RE	0.56 _(0.07)	0.55 _(0.02)	0.39 _(0.06)	0.34 _(0.06)	0.48 _(0.05)	0.46 _(0.05)	0.63 _(0.02)	0.71 _(0.01)	0.71 _(0.01)	0.68 _(0.02)
	R ²	0.19 _(0.02)	0.31 _(0.02)	0.35 _(0.03)	0.46 _(0.02)	0.52 _(0.03)	0.35 _(0.02)	0.40 _(0.01)	0.30 _(0.01)	0.58 _(0.01)	0.68 _(0.01)
CCA											
local	CE	-3.72 _(0.43)	-2.88 _(0.27)	-0.65 _(0.10)	-0.95 _(0.12)	-1.94 _(0.21)	-3.04 _(0.35)	-2.32 _(0.29)	-3.39 _(0.39)	-2.16 _(0.24)	-0.54 _(0.11)
	RE	0.23 _(0.06)	0.20 _(0.05)	0.20 _(0.05)	0.21 _(0.04)	0.30 _(0.05)	0.32 _(0.05)	0.35 _(0.05)	0.39 _(0.05)	0.42 _(0.04)	0.42 _(0.04)
	R ²	0.12 _(0.03)	0.12 _(0.04)	0.22 _(0.06)	0.17 _(0.09)	0.32 _(0.09)	0.12 _(0.03)	0.17 _(0.06)	0.18 _(0.04)	0.45 _(0.04)	0.56 _(0.04)
max	CE	-0.79 _(0.15)	-0.62 _(0.10)	0.15 _(0.04)	0.12 _(0.04)	-0.00 _(0.06)	-0.45 _(0.09)	-0.02 _(0.05)	-0.07 _(0.05)	0.15 _(0.04)	0.48 _(0.02)
	RE	0.63 _(0.02)	0.61 _(0.02)	0.53 _(0.02)	0.51 _(0.02)	0.70 _(0.01)	0.65 _(0.01)	0.73 _(0.01)	0.78 _(0.01)	0.78 _(0.01)	0.78 _(0.01)
	R ²	0.22 _(0.03)	0.35 _(0.02)	0.39 _(0.03)	0.50 _(0.02)	0.61 _(0.02)	0.40 _(0.02)	0.50 _(0.02)	0.39 _(0.02)	0.64 _(0.01)	0.76 _(0.01)
GraphEM											
local	CE	-2.40 _(0.32)	-1.87 _(0.22)	-0.28 _(0.10)	-0.59 _(0.11)	-1.48 _(0.17)	-2.54 _(0.33)	-1.67 _(0.24)	-2.57 _(0.36)	-1.78 _(0.28)	-0.35 _(0.11)
	RE	0.42 _(0.04)	0.38 _(0.04)	0.36 _(0.04)	0.32 _(0.03)	0.40 _(0.04)	0.40 _(0.04)	0.47 _(0.04)	0.49 _(0.04)	0.48 _(0.04)	0.48 _(0.04)
	R ²	0.14 _(0.03)	0.14 _(0.05)	0.23 _(0.05)	0.22 _(0.07)	0.31 _(0.07)	0.15 _(0.04)	0.19 _(0.04)	0.19 _(0.03)	0.40 _(0.04)	0.48 _(0.05)
max	CE	-0.28 _(0.10)	-0.20 _(0.07)	0.35 _(0.03)	0.30 _(0.04)	0.19 _(0.06)	-0.24 _(0.09)	0.10 _(0.06)	-0.08 _(0.09)	0.12 _(0.08)	0.50 _(0.03)
	RE	0.71 _(0.01)	0.70 _(0.01)	0.61 _(0.02)	0.62 _(0.02)	0.77 _(0.01)	0.72 _(0.01)	0.78 _(0.01)	0.80 _(0.01)	0.79 _(0.01)	0.79 _(0.01)
	R ²	0.29 _(0.02)	0.41 _(0.02)	0.45 _(0.03)	0.54 _(0.02)	0.64 _(0.01)	0.44 _(0.02)	0.53 _(0.02)	0.42 _(0.02)	0.61 _(0.02)	0.74 _(0.01)

Table S2: Same as Table S1, but with the staircase network.