Extracting Modes of Variability and Change from Climate Model Ensembles

Robert C. Wills[1], David S. Battisti[1], Dennis L. Hartmann[1], Tapio Schneider[2]

Abstract—Ensembles of climate model simulations are commonly used to separate externally forced climate change from internal climate variability. However, much of the information gained from running large ensembles is lost in traditional methods of data reduction such as linear trend analysis or large-scale averaging. In this paper, we describe a statistical method to extract patterns of low-frequency variability and change from large ensembles. We demonstrate how this method characterizes modes of forced climate change (e.g., global warming) and low-frequency internal variability (e.g., the Pacific decadal oscillation) in the CESM large ensemble.

I. Motivation

Internal climate variability gives rise to uncertainty in long-term climate predictions [1]. Ensembles of climate model simulations are often used to quantify this uncertainty and to better understand the average response to external forcing [2], [3], [4]. Separating the forced response from the internal variability also helps to understand multi-decadal internal variability [5], which may lead to better decadal climate predictions [6]. However, most climate studies diagnose the spatial pattern of climate change by computing linear trends and/or diagnose the temporal behavior of climate variability by studying large-scale spatial averages. These methods of dimension reduction loose valuable information about the complex spatiotemporal structure of climate variability and change.

Principal component analysis (PCA) provides spatiotemporal information about the modes of variability that explain the most variance in a dataset. However, by maximizing variance, PCA can mix together physically distinct modes of variability such as global warming and the El Niño–Southern Oscillation (ENSO). One method to correct for this mode mixing is to look for linear combinations of the empirical orthogonal functions (EOFs) that maximize a particular type of variance

Corresponding author: R. C. Wills, rcwills@uw.edu [1]Department of Atmospheric Sciences, University of Washington, Seattle, WA [2]Department of Environmental Sciences and Engineering, California Institute of Technology, Pasadena, CA

representing a "signal" compared to "noise" that exists within internal variability or amongst realizations, so-called optimal filtering or signal-to-noise maximizing EOF analysis [7], [8], [9], [10], [11]. These methods take advantage of any spatial structure in the "noise" to optimally filter it out. Here, we use low-frequency component analysis (LFCA, [12]) to find patterns with the maximum ratio of low-frequency (signal) to high-frequency (noise) variance, correcting for mode mixing based on differences in time scale between physically plausible modes of variability.

II. Method

The basic assumption behind our approach is that externally forced climate change operates on longer time scales than most internal variability. We can thus isolate patterns of climate change by solving for spatial patterns that describe variability with the maximum ratio of low-frequency to total variance, where low-frequency variance is defined as the variance remaining after application of a lowpass filter. LFCA provides an algorithm to find such spatial patterns for a truncated basis of EOFs. This method orders modes by their ratio of low-frequency to total variance, providing orthogonal indices of climate variability that tend to be ordered by time scale. For example, it separates global warming, the Pacific Decadal Oscillation (PDO), and ENSO in observed Pacific SSTs [12]. Here, we generalize LFCA for application to climate model ensembles. Our *ensemble LFCA* method is as follows:

1. **Compute ensemble covariance matrix.** For an ensemble of n_E climate model simulations, each with $n \times p$ data matrix X_i, we compute the $p \times p$ covariance matrix C_i with respect to either (a) the ensemble-mean climatology vector $\mathbf{x}_E$ or (b) the individual-ensemble-member climatology vector $\mathbf{x}_i$. Option (b) discards differences in climatology between ensemble members, while option (a) does not.

2. **Ensemble EOF analysis.** We compute the EOFs $\mathbf{a}_k$, which are the eigenvectors of the ensemble covari-
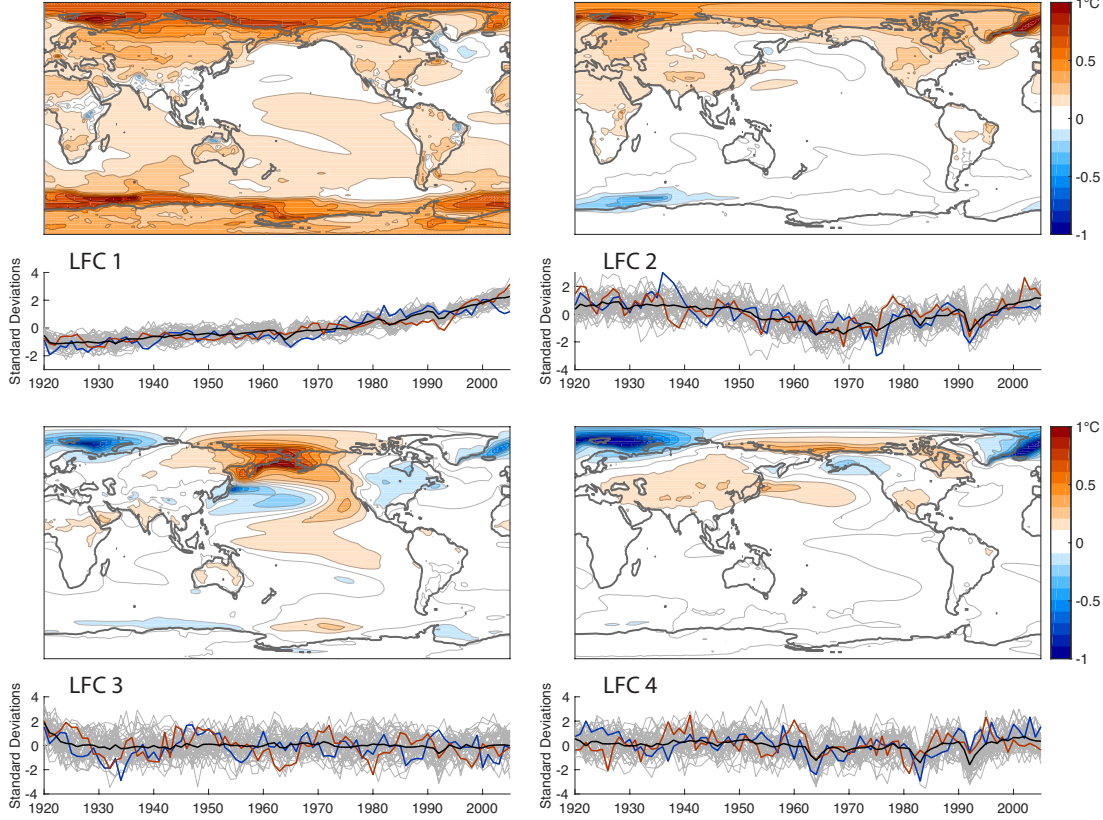
Fig. 1. Low-frequency patterns (LFPs) and components (LFCs) of the CESM large-ensemble historical simulation, with $N = 25$ EOFs retained and maximization of the variance remaining after application of Lanczos lowpass filter with cutoff $\tau = 10$ years. Orange (blue) lines show the ensemble member with the most (least) change in LFC 1 over the last 20 years. All other ensemble members are shown with grey lines. A black line shows the average of the LFC over all ensemble members.

ance matrix

$$C_E = n_E^{-1} \Sigma_{i=1}^{n_E} C_i. \tag{1}$$

The EOFs are normalized $||\mathbf{a}_k|| = 1$, such that the principal components have unit variance and the corresponding eigenvalue $\sigma_k^2 = \mathbf{a}_k^T C_e \mathbf{a}_k$ gives the variance associated with the kth EOF.

3. **Low-frequency component analysis.** We apply the LFCA algorithm [12] (see also [10]) to find the linear combination of the first N EOFs,

$$\mathbf{u}_k = \begin{bmatrix} \dfrac{\mathbf{a}_1}{\sigma_1} & \dfrac{\mathbf{a}_2}{\sigma_2} & ... & \dfrac{\mathbf{a}_N}{\sigma_N} \end{bmatrix} \mathbf{e}_k, \tag{2}$$

such that the ratio of low-frequency to total variance

$$r_k = \frac{\left(\widetilde{X}_E \mathbf{u}_k \right)^T \widetilde{X}_E \mathbf{u}_k}{(X_E \mathbf{u}_k)^T X_E \mathbf{u}_k} \tag{3}$$

is maximized when the data is projected onto it. Here, X_E is the full-ensemble data matrix, obtained by concatenating individual-ensemble-member data matrices

$X_E = [X_1^T \ X_2^T \ ... \ X_{n_E}^T]^T$, and $\widetilde{X}_E$ is the lowpass-filtered full-ensemble data matrix, obtained by concatenating lowpass-filtered data matrices (i.e., we do not filter over the discontinuities between ensemble members). In practice, the linear combination coefficients $\mathbf{e}_k$ are computed as the eigenvectors of the covariance matrix of the first N lowpass-filtered principal components, such that filtering only needs to be applied to an $n \cdot n_E \times N$ matrix of the leading principal components (see derivation in [12]).

4. **Visualizing results.** The result is low-frequency components (LFCs) given by

$$\text{LFC}_k = X_E \mathbf{u}_k \tag{4}$$

and low-frequency patterns (LFPs) given by

$$\mathbf{v}_k = X_E^T \text{LFC}_k = [\sigma_1 \mathbf{a}_1 \ \sigma_2 \mathbf{a}_2 \ ... \ \sigma_N \mathbf{a}_N] \mathbf{e}_k. \tag{5}$$

These are analogous to principal components and EOFs in PCA. The linear coefficients are normalized $||\mathbf{e}_k|| = 1$, such that the LFCs have unit variance and the LFPs show the spatial pattern associated with a 1-standard-deviation anomaly in the corresponding LFC.

III. RESULTS

We demonstrate our method by applying it to annual-mean surface temperatures from a 40-member ensemble of "historical" simulations with the Community Earth System Model (CESM) [4]. These simulations simulate climate from 1920 to 2005 based on historical forcing by greenhouse gasses, anthropogenic and volcanic aerosols, and ozone. The ensemble members differ only by machine-precession perturbations in their atmospheric initial condition in 1920, such that their climatology vectors $\mathbf{x}_i$ differ only as a result of internal variability. We include these climatology differences by using option (a) in step 1. We retain $N = 25$ EOFs in the LFCA and use a Lanczos filter with lowpass cutoff $\tau = 10$ years to focus on multi-decadal variability. The results are insensitive to the choice of cutoff for $\tau > 5$ years. However, there is no good criterion for choosing N, so in practice one must look for results that are robust across parameters (see discussion in [12]).

The first LFP shows a global warming pattern, with amplified warming over land and at high latitudes (Fig. 1). The associated LFC increases by 3 standard deviations from 1920 to 2005, emerging well beyond the ensemble spread. The second LFP/LFC shows cooling of the North Atlantic, Arctic, and Northern Hemisphere land through the 1950s and 60s and a subsequent recovery (Fig. 1), corresponding roughly to the time series of anthropogenic aerosol radiative forcing [13]. It also shows a large excursion in the ensemble mean due to the Mt. Pinatubo eruption in 1991 and a large variance amongst ensemble members, perhaps related to Atlantic multi-decadal variability [14]. The third LFP/LFC shows low-frequency internal variability associated with the PDO ([15], Fig. 1). There is only a small excursion in the ensemble mean, before 1930, related to spin-up from initial conditions. The fourth LFP/LFC shows low-frequency variability over the Barents-Kara Sea, Eurasia, and the North Pacific (Fig. 1). It shows primarily internal variability (i.e., there is little agreement amongst ensemble members), but there is a response to the volcanic eruptions in 1982 and 1991. The remaining LFCs 5-25 show internal variability with increasingly shorter time scales. One interesting aspect of these results is that modes can be a combination of internal variability and forced responses (e.g., LFC 2-4), whereas most other analysis methods assume that modes are either one or the other.

To quantify the number of ensemble members needed to obtain these results, we compute the pattern correlation of LFPs obtained from analyses with fewer ensemble members, with those of the 40-member analysis
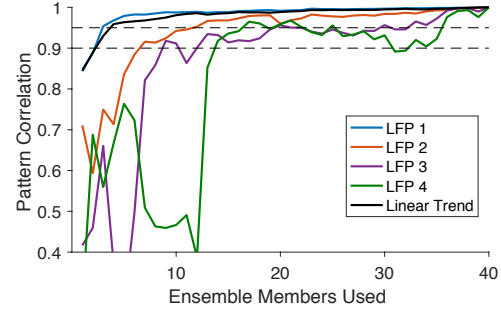

Fig. 2. Convergence of the first 4 LFPs as the ensemble size is increased, based on pattern correlation with results of the 40-ensemble-member analysis; comparison to convergence of the ensemble-mean linear trend. To make sub-ensembles, we pick the first n_E ensemble members from the 40-member ensemble. All ensemble members are identical in design, so we do not expect that our conclusions are sensitive to this sampling method.

(Fig. 2). For LFP 1, we find a pattern correlation > 0.95 using only 3 ensemble members. This is slightly less than the 4 ensemble members needed for a pattern correlation > 0.95 between the ensemble-mean linear trend and the 40-member ensemble-mean linear trend. Few ensemble members are needed for a robust estimate of global warming. The second, third, and fourth LFPs take longer to converge, requiring 7, 9, and 14 ensemble members, respectively, to reach a pattern correlation > 0.9. Only with large ensembles (or long control runs) can we understand these higher-order LFCs.

IV. SUMMARY AND OUTLOOK

We have demonstrated that *ensemble LFCA* can identify modes of low-frequency variability and change that are robust across climate model ensembles. It needs only 3 ensemble members to identify the forced global warming pattern, and makes no assumptions on the linearity of the warming response.

We have also applied our method to multi-model ensembles, where it is beneficial to discard the large ensemble-member differences in climatology (option b in step 1). This method provides particular utility when there are multiple time scales of forced response, such as in simulations of the response to abrupt $4xCO_2$ forcing. This provides a powerful tool to visualize the dominant modes of low-frequency variability and change in large climate datasets.

ACKNOWLEDGMENTS

REFERENCES

[1] D. W. Thompson, E. A. Barnes, C. Deser, W. E. Foust, and A. S. Phillips, "Quantifying the role of internal climate variability in future climate trends," *J. Climate*, vol. 28, no. 16, pp. 6443–6456, 2015.

[2] C. Deser, R. Knutti, S. Solomon, and A. S. Phillips, "Communication of the role of natural variability in future North American climate," *Nat. Clim. Change*, vol. 2, no. 11, pp. 775–779, 2012.

[3] C. Deser, A. S. Phillips, M. A. Alexander, and B. V. Smoliak, "Projecting North American climate over the next 50 years: Uncertainty due to internal variability," *J. Climate*, vol. 27, no. 6, pp. 2271–2296, 2014.

[4] J. Kay *et al.*, "The Community Earth System Model (CESM) large ensemble project: A community resource for studying climate change in the presence of internal climate variability," *Bull. Am. Meteorol. Soc.*, vol. 96, no. 8, pp. 1333–1349, 2015.

[5] L. M. Frankcombe, M. H. England, M. E. Mann, and B. A. Steinman, "Separating internal variability from the externally forced climate response," *J. Climate*, vol. 28, no. 20, pp. 8184–8202, 2015.

[6] G. A. Meehl *et al.*, "Decadal prediction: can it be skillful?," *Bull. Am. Meteorol. Soc.*, 2009.

[7] M. R. Allen and L. A. Smith, "Optimal filtering in singular spectrum analysis," *Phys. Lett. A*, vol. 234, no. 6, pp. 419–428, 1997.

[8] S. Venzke, M. R. Allen, R. T. Sutton, and D. P. Rowell, "The atmospheric response over the North Atlantic to decadal changes in sea surface temperature," *J. Climate*, vol. 12, no. 8, pp. 2562–2584, 1999.

[9] T. Schneider and S. M. Griffies, "A conceptual framework for predictability studies," *J. Climate*, vol. 12, no. 10, pp. 3133–3155, 1999.

[10] T. Schneider and I. M. Held, "Discriminants of twentieth-century changes in earth surface temperatures," *J. Climate*, vol. 14, no. 3, pp. 249–254, 2001.

[11] M. Ting, Y. Kushnir, R. Seager, and C. Li, "Forced and internal twentieth-century SST trends in the North Atlantic," *J. Climate*, vol. 22, no. 6, pp. 1469–1481, 2009.

[12] R. C. Wills, T. Schneider, J. M. Wallace, D. S. Battisti, and D. L. Hartmann, "Disentangling global warming, multi-decadal variability, and El Niño in Pacific temperatures," *Proc. Natl. Acad. Sci, submitted*, 2017.

[13] D. T. Shindell *et al.*, "Radiative forcing in the ACCMIP historical and future climate simulations," *Atmos. Chem. Phys*, vol. 13, no. 6, pp. 2939–2974, 2013.

[14] D. B. Enfield, A. M. Mestas-Nuñez, and P. J. Trimble, "The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US," *Geophys. Res. Lett.*, vol. 28, no. 10, pp. 2077–2080, 2001.

[15] N. J. Mantua, S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, "A Pacific interdecadal climate oscillation with impacts on salmon production," *Bull. Am. Meteorol. Soc.*, vol. 78, no. 6, pp. 1069–1079, 1997.