Supplementary information for

Global-scale multidecadal variability in climate models and observations. Part II: The stadium wave

Sergey Kravtsov,[a, b] Andrew Westgate[a], Andrei Gavrilov[c]

[a] *University of Wisconsin-Milwaukee, Milwaukee, WI*

[b] *Shirshov Institute of Oceanology of Russian Academy of Sciences, Moscow, Russia*

[c] *Gaponov-Grekov Institute of Applied Physics of Russian Academy of Sciences, Nizhny Novgorod, Russia*

Corresponding author: Sergey Kravtsov, kravtsov@uwm.edu

Description of the contents

This online *Supplementary Information*, referenced in the main text, contains additional supporting figures and the accompanying text in the form of comments to the figures.

Data availability All MATLAB © scripts, input/output data, and the figures these scripts generated are available from the following manuscript's website.

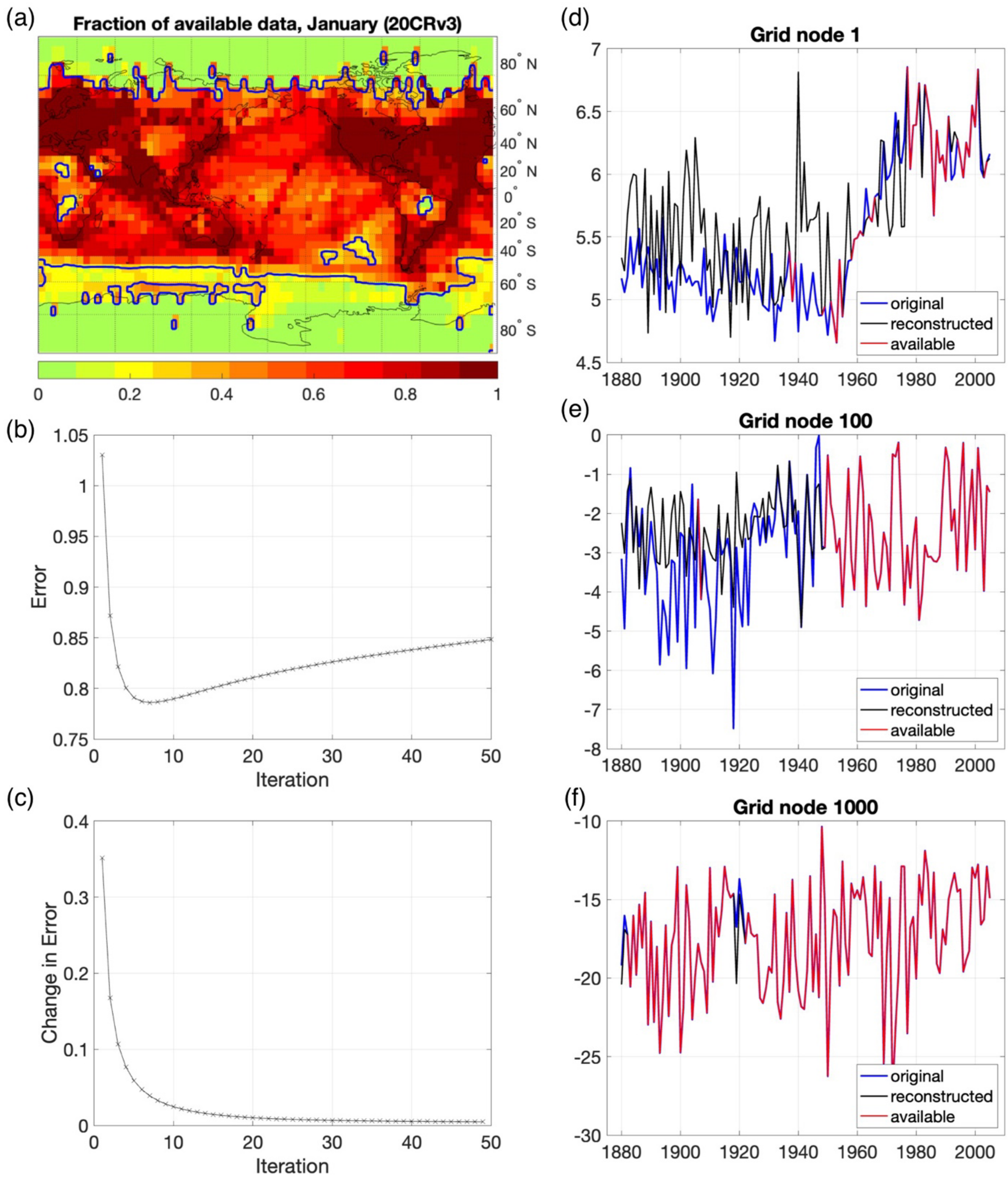

Fig. S1 Covariance-based imputation of the data gaps mimicking the actual historical coverage: (a) fraction of the available SAT 1880–2005 data for the month of January; (b) an example of the global root-mean-square error (ºC) of the SAT imputation as a function of iteration step for a fixed number *K* of the data EOFs used for reconstruction; (c) the magnitude of the error difference between the consecutive iteration steps; (d–f) examples of original and reconstructed data.

Comments:

(1) Only the locations with the data coverage exceeding 30% (regions bounded by thick blue contours in Fig. S1a) are used in the imputation procedure; in the final reconstructed product the SAT values within the 60ºS–60ºN region are obtained using the nearest-neighbor interpolation.
(2) The optimal parameters here (the number of EOFs *K* and the number of iterations *I*) were chosen to correspond to the absolute minima in the plots analogous to Fig. S1b (different for each month), thus assuming that we know the true value of the data we are trying to reconstruct at each grid node. When this is not the case, one could employ the Monte-Carlo procedures of generating synthetic analogues of the data, cross-validation procedures, and/or the information about the convergence based on the rate-of-change of the errors (such as Fig. S1c) to determine the optimal *K* and *I*.
(3) The reconstruction examples in (d–f) sample a range of behaviors when reconstructing the missing data over shorter or longer data gaps. It turns out that the fairly large reconstruction errors for the monthly subsets of data (e.g., all Januaries, as shown here) are greatly reduced when taking the annual averages of the reconstructed data (see, for example, Fig. S2).

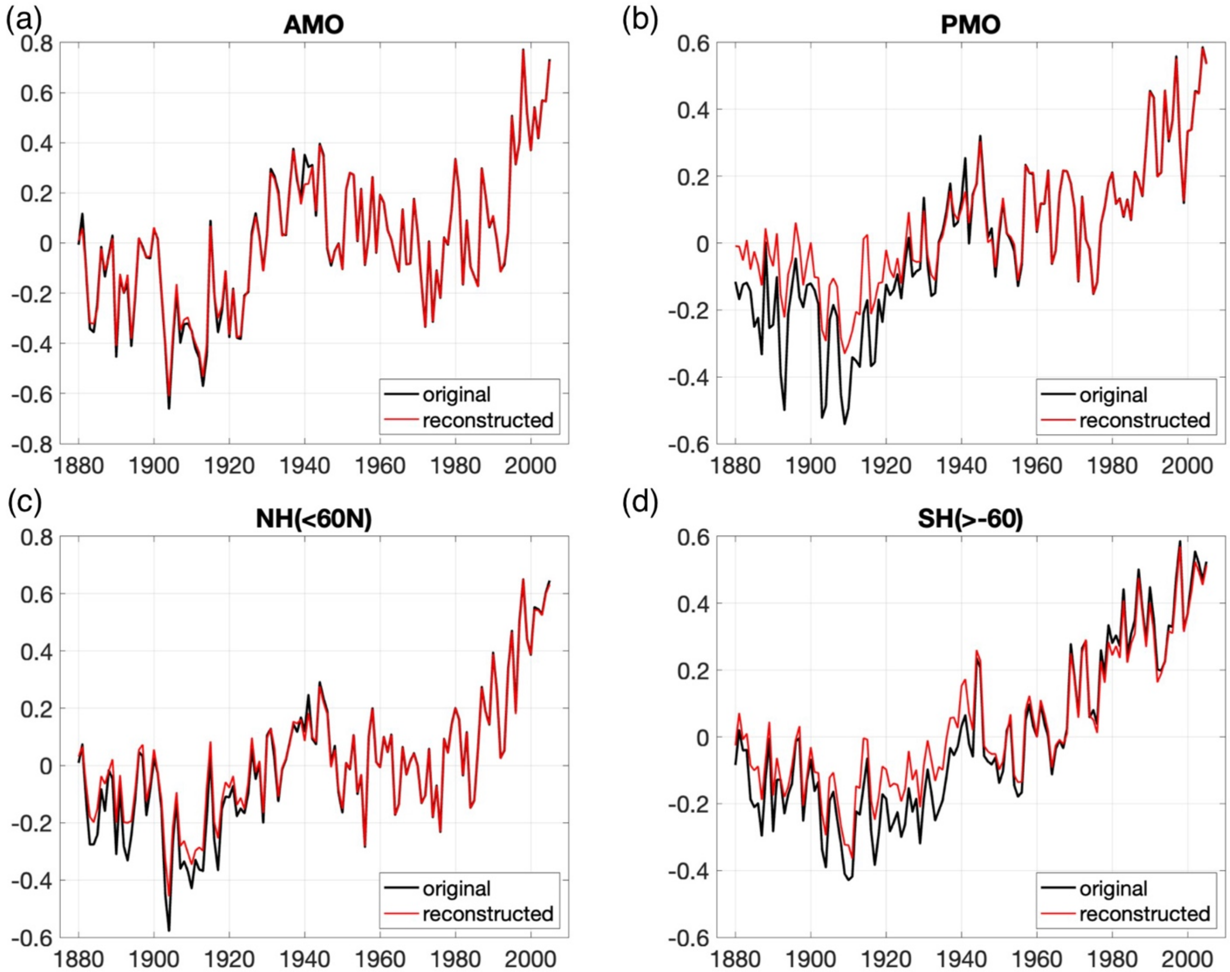

Fig. S2 Reconstructions of the annual-mean SAT indices (time series of regional averages): (a) Atlantic Multidecadal Oscillation (AMO) index (0–60ºN, 80–0ºW, over ocean); (b) Pacific Multidecadal Oscillation (PMO) index (0–60ºN, 120ºE–100ºW, over ocean); (c) Northern Hemisphere (NH0: 0–60ºN); and (d) Southern Hemisphere (SH0: 60ºS–0).

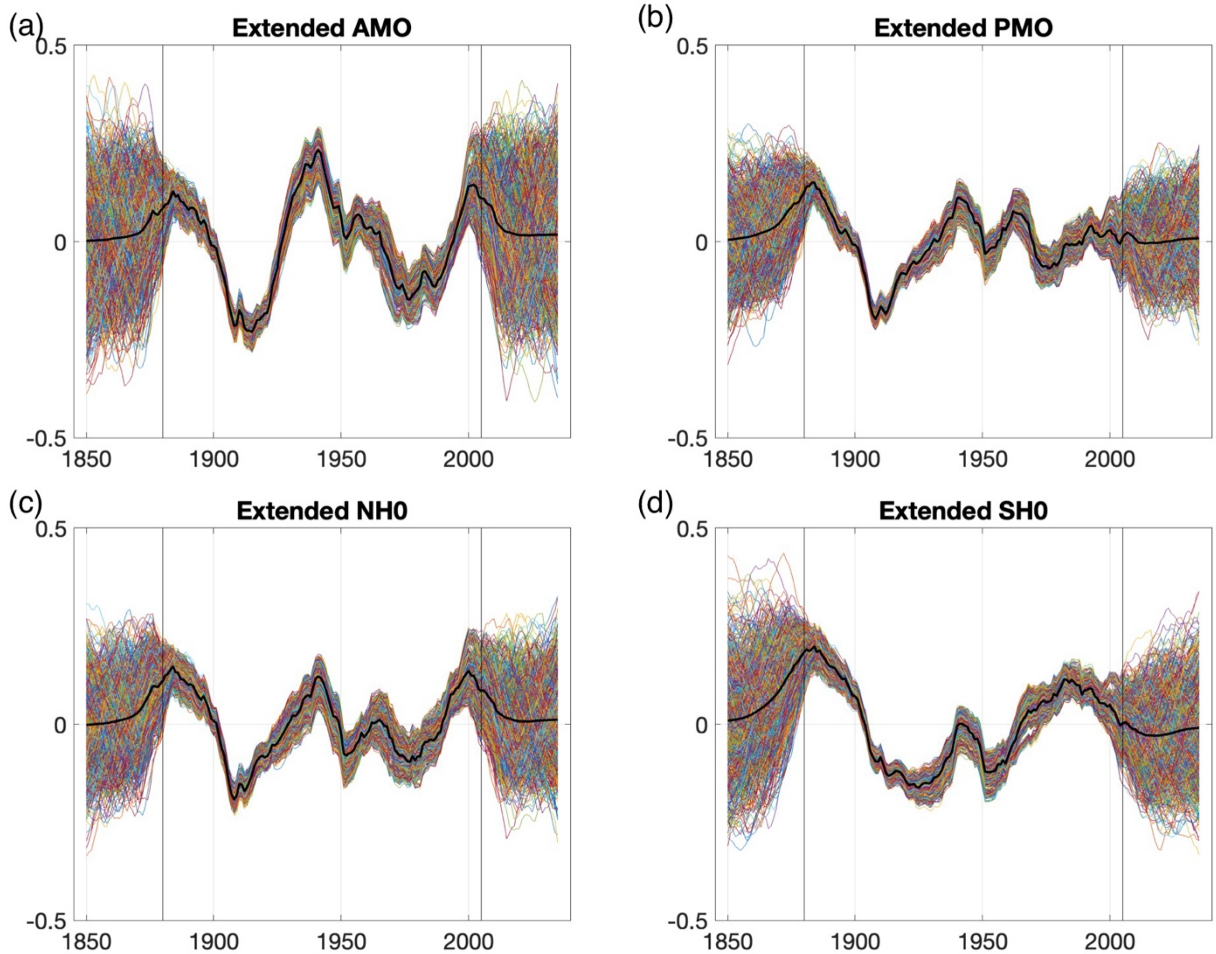

Fig. S3 Extrapolation of the 1880–2005 20CRv3 SAT data by 30 years forward and backward in time using linear inverse models (LIMs) estimated in the phase space of the thirty leading EOFs of the ensemble-mean 20CRv3 annual data. These models were initialized to extend each of the 80 available realizations of 20CRv3 using independent samples of random noise as forcing. Shown here are the same indices as in Fig. S2 but smoothed with the 11-point boxcar running mean smoother, for each extrapolated realization (spaghetti plots of thin colored curves), as well as their ensemble average (heavy black curves). The original 1880–2005 period is marked by vertical lines.

Comments:

(1) While the uncertainties of the extrapolation are substantial, the ensemble-mean forecasts on both ends do suggest a non-zero decadal prediction skill (especially over the Atlantic and the Northern Hemisphere), apparently associated with the global stadium wave.

(2) Relatively mild end effects on the stadium-wave reconstruction within the reference period when using the extended SAT data (Figs. 9–11 of the main text) despite widely different extrapolated variability outside of this period also indicates the robustness of the global teleconnections underlying the stadium wave: in the opposite case, the extrapolated data (if characterized by different space–time teleconnections relative to the reference data) would be expected to skew the reference-period results much more drastically.

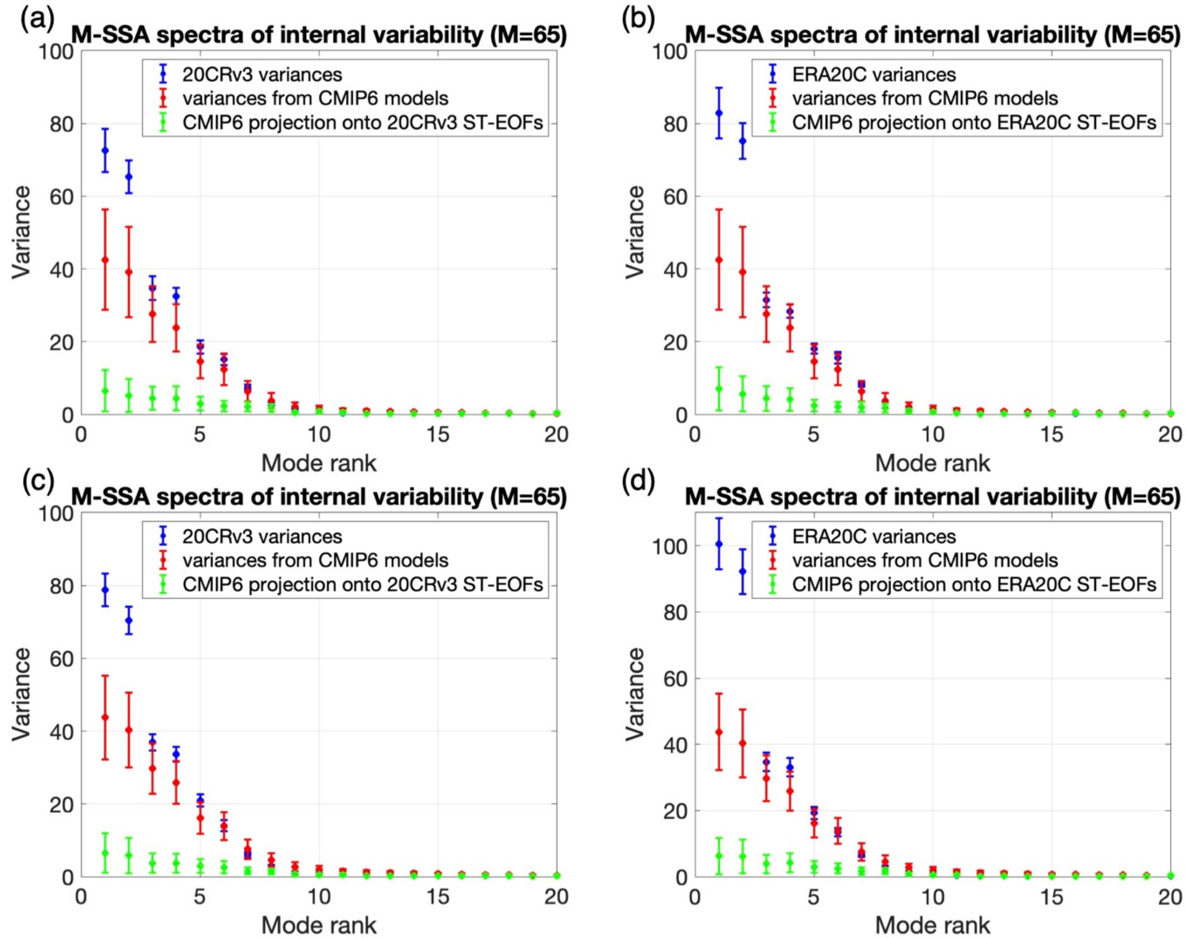

Fig. S4 Multi-channel Singular Spectrum Analysis (M-SSA) spectra associated with the estimates of the observed and CMIP6-simulated decadal+ internal variability in the ensembles of spatially extended near-surface atmospheric temperature (SAT) and sea-level pressure (SLP) annual time series (1880–2005), **for the non-outlier models (23 models and 151 simulations)**. Error bars show the average and the standard deviation of each M-SSA mode's variance across the ensemble of available estimates of internal variability (177–26=151 estimates for the CMIP6 non-outlier models, 21× 80 =1680 estimates for 20CRv3 data, and 21 estimates for the ERA-20C data) for: reanalysis data (blue), CMIP6 simulations (red), as well as the projection of the CMIP6 simulations onto the reanalysis M-SSA space–time patterns (green). Left column: 20CRv3 results; right column: ERA-20C results. Top row: SAT-only analysis; bottom row: common SAT/SLP analysis.

<u>Comments</u>:

(1) This figure complements analogous figures from the main text: Fig. 1, which shows the results based on CMIP5 simulations, as well as Figs. 2 and 3, which show the results for the entire 21-model, 177-simulation CMIP6 ensemble and the 3-model, 26-simulation outlier ensemble, respectively.

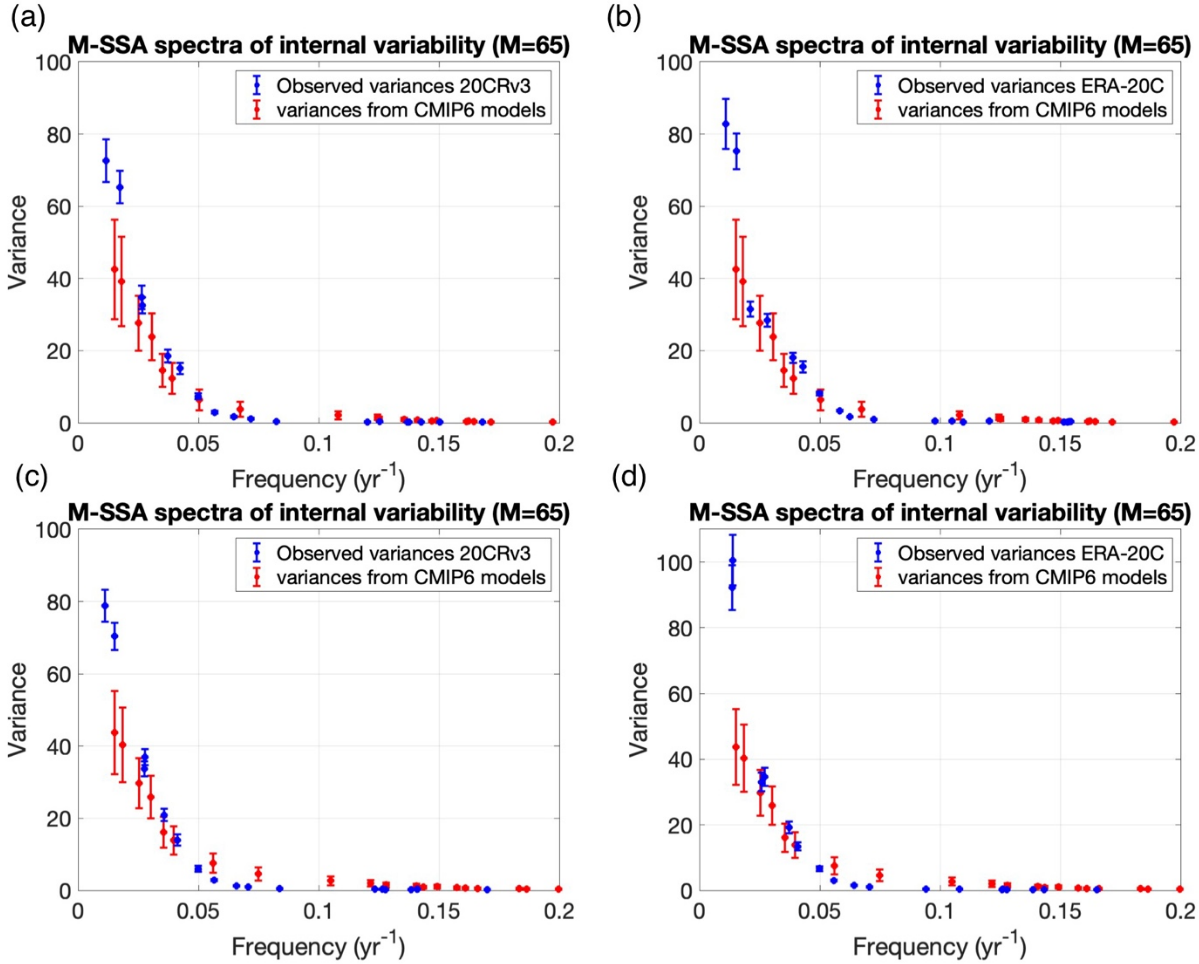

Fig. S5 The same as in Fig. S4, but for the (**non-outlier CMIP6**, 20CRv3 and ERA-20C) M-SSA variances plotted against the ensemble-averaged mode frequencies. The variances associated with projections of the simulated data onto the observed M-SSA space–time patterns are not shown here.

<u>Comments</u>:

(1) This figure complements analogous figures from the main text: Fig. 4, which shows the results based on CMIP5 simulations, as well as Fig. 5, which displays the results based on the CMIP6 outlier runs only.

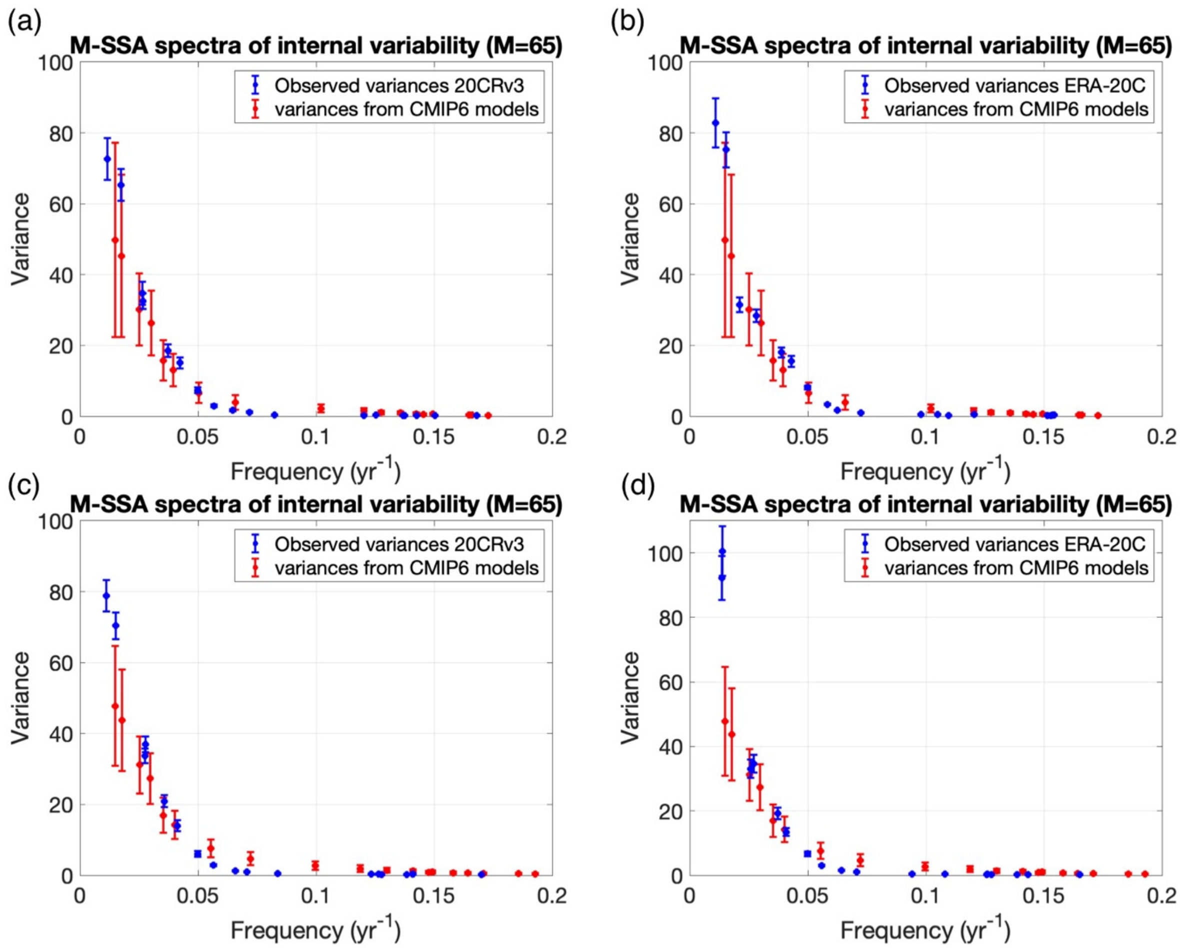

Fig. S6 The same as in Fig. S5, but for the (**entire CMIP6 ensemble**, 20CRv3 and ERA-20C) M-SSA variances plotted against the ensemble-averaged mode frequencies. The variances associated with projections of the simulated data onto the observed M-SSA space–time patterns are not shown here.

<u>Comments</u>:

(1) This figure is completely analogous to Fig. 4 of the main text (which is based on CMIP5 results).

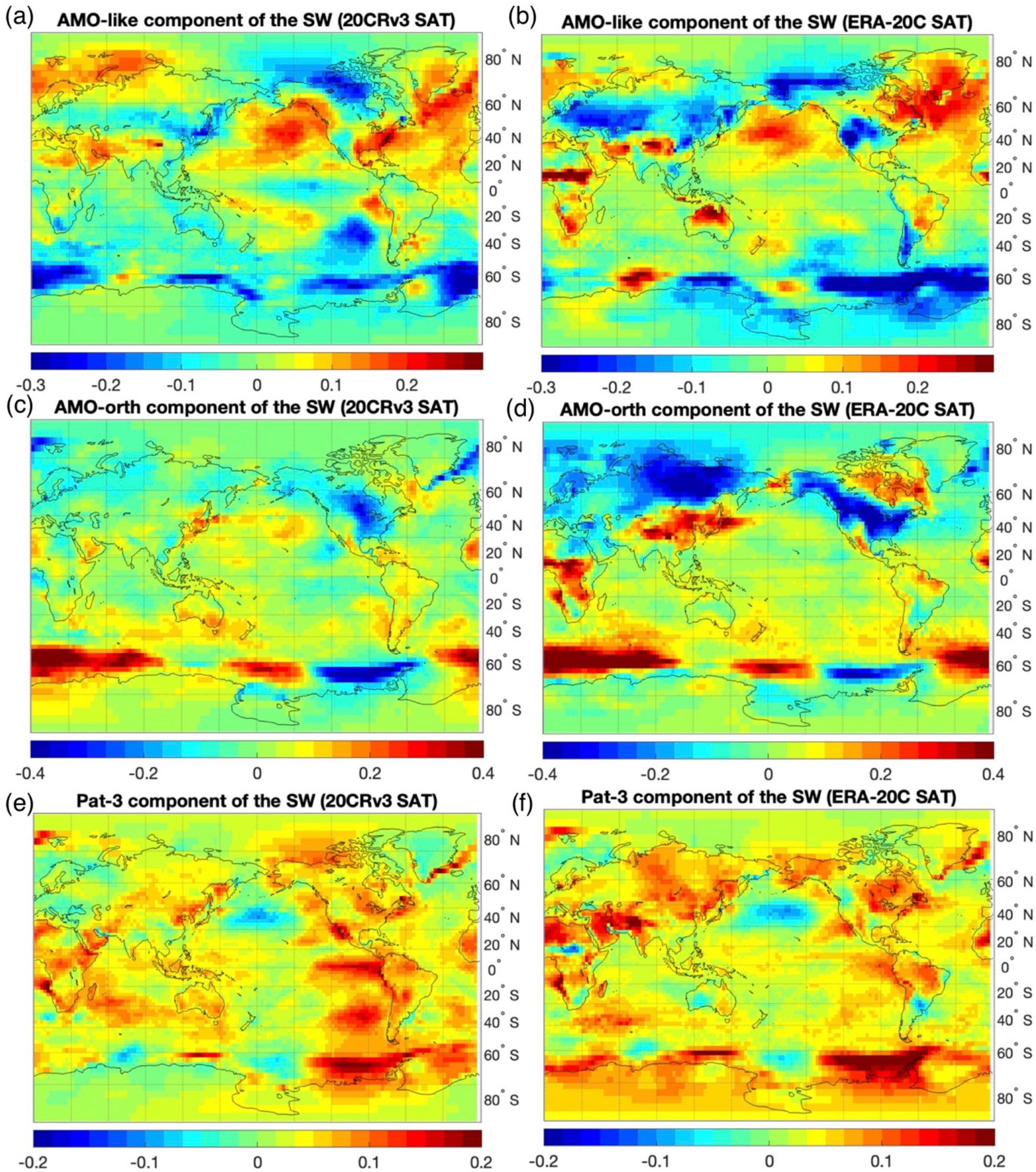

Fig. S7 The dimensional SAT patterns associated with the **SAT-only** analysis' rotated PCs shown in Fig. 6 of the main text, for the 20CRv3 (right) and ERA-20C (left) internal-variability estimates obtained using CMIP5 based forced signals. All patterns are weighted by the square root of the cosine of latitude.

<u>Comments</u>: This figure is completely analogous to Fig. 7 of the main text, which shows the results of the common SAT/SLP analysis.

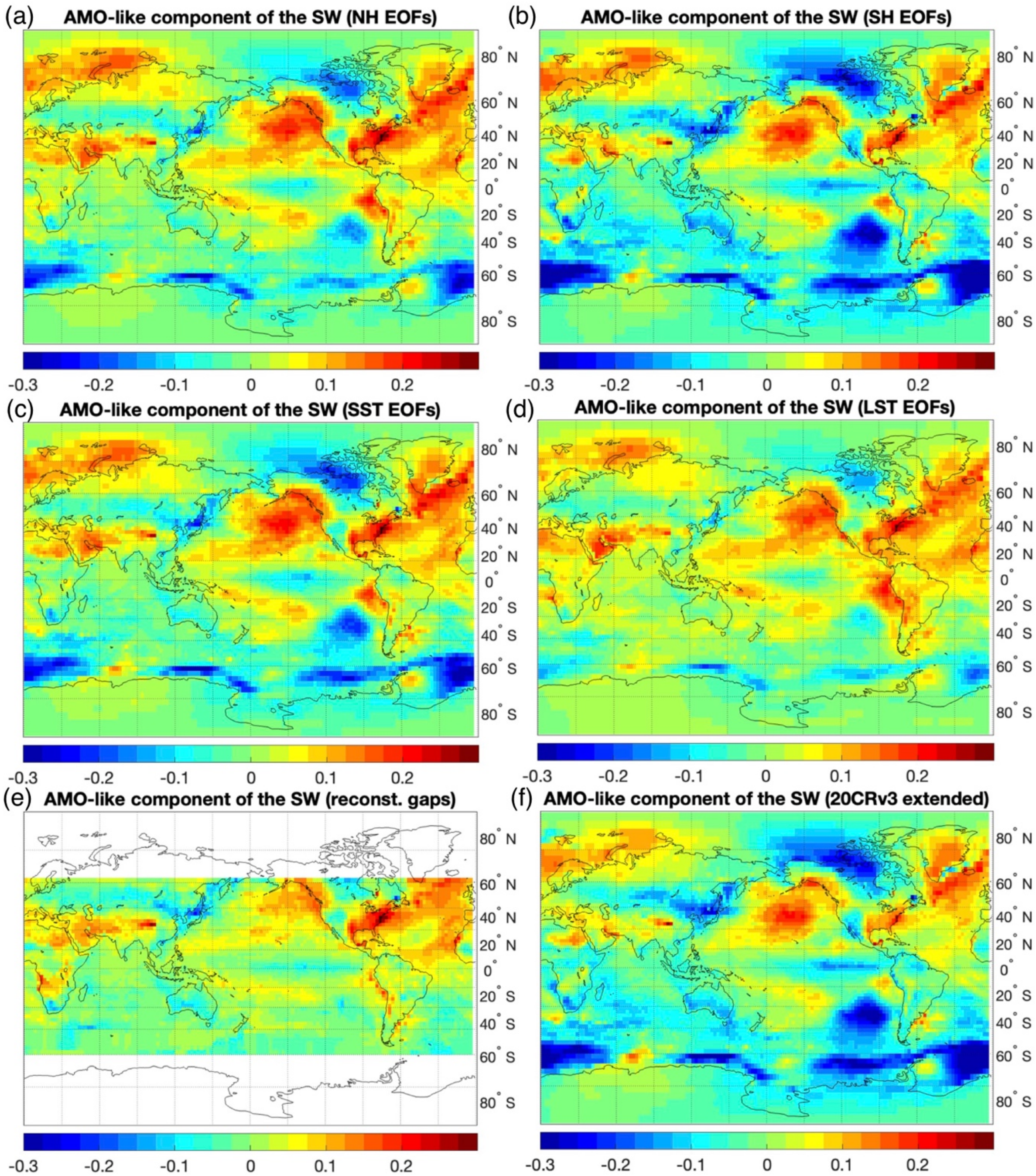

Fig. S8 AMO-like stadium-wave patterns stemming from different versions of 20CRv3 analysis, in particular, the ones based on: (a) Northern-Hemisphere EOFs; (b) Southern-Hemisphere EOFs; (c, d) EOFs computed using the 60ºS–60ºN SAT data over ocean (SST) or land (LST); (e) 60ºS–60ºN SAT data reconstructed by introducing and filling in (through a covariance-based imputation technique) the data gaps mimicking the actual historical coverage; and (f) for the extended 1850–2035 data set extrapolated by 30 years into the future and the past using forward and backward linear inverse models. All patterns are weighted by the square root of the cosine of latitude.

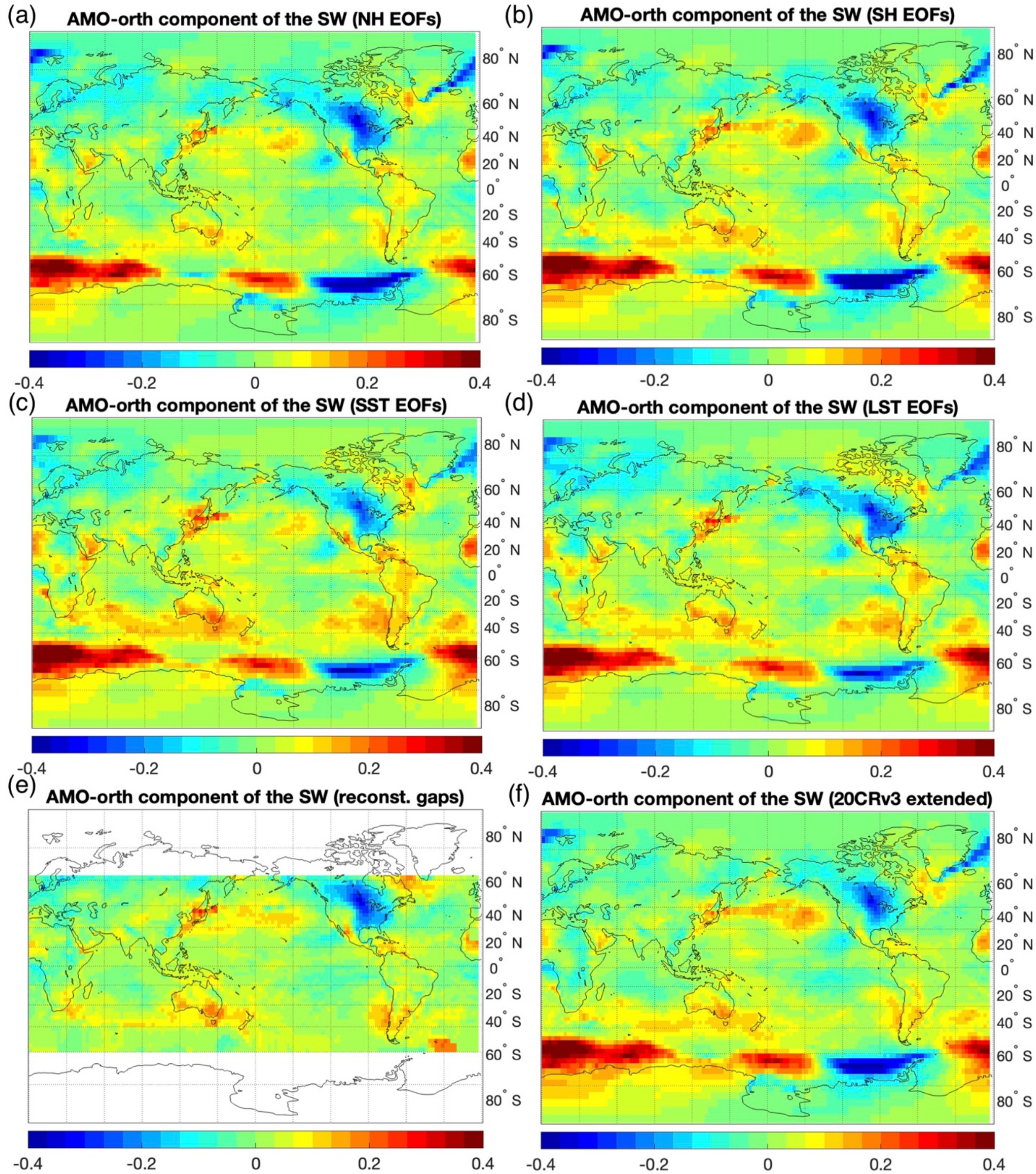

Fig. S9 The same as in Fig. S8, but for the AMO-orth stadium-wave component.

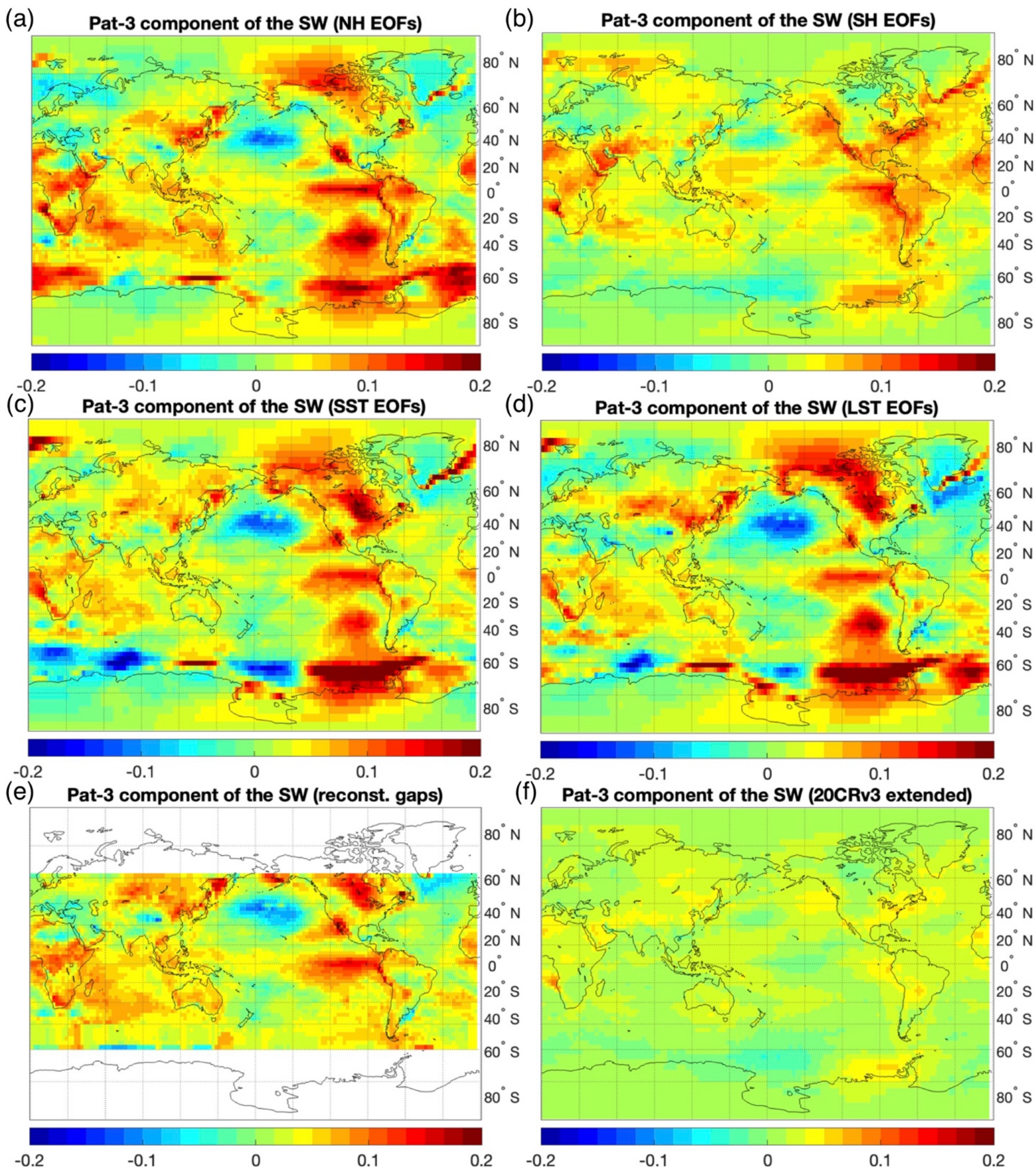

Fig. S10 The same as in Figs. S8 and S9, but for the third stadium-wave component.

<u>Comments</u>: The patterns shown in Figs. S8–S10 are used to compute the Taylor diagrams in Fig. 9 of the main text.

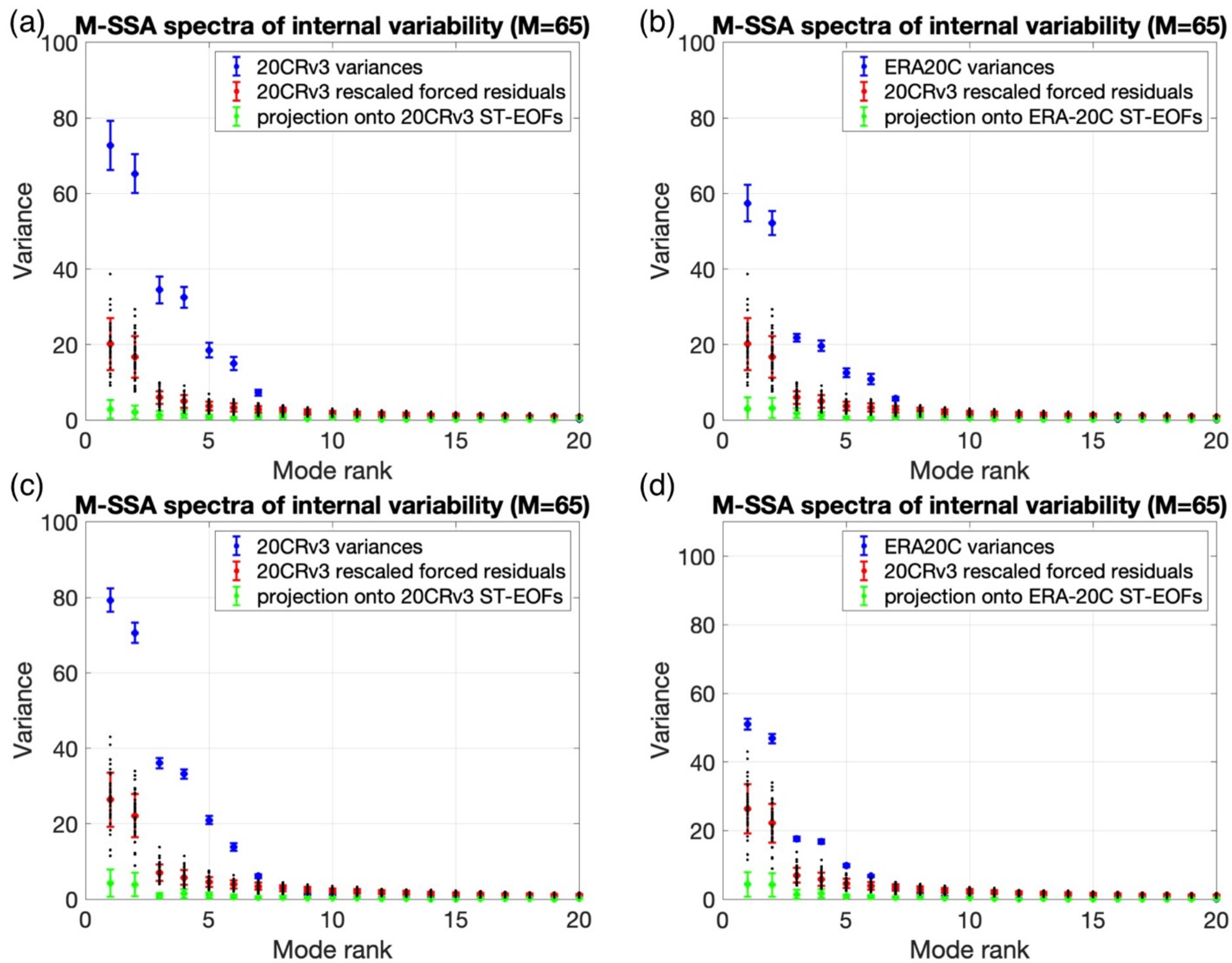

Fig. S11 M-SSA spectra associated with the estimates of the observed internal variability and the CMIP5/6 forced-signal residuals (that is, deviations of individual-model forced signals from the common evolution across the 38-member forced-signal ensemble; see text for details). Error bars show the average and the standard deviation of each M-SSA mode's variance across the ensemble of available estimates of residual variability (38 estimates for the forced-signal residuals based **on the 20CRv3-rescaled** CMIP5/6 data, 21 estimates for the 20CRv3 data (obtained by subtracting forced signals of 21 CMIP6 models from the reanalysis ensemble mean), and 21 analogous estimates for the ERA-20C data) for: reanalysis (blue), CMIP5/6 forced-signal residuals (red, black dots for individual estimates), as well as the projection of the CMIP residuals onto the reanalysis M-SSA space–time patterns (green). Left column: 20CRv3 results; right column: ERA-20C results. Top row: SAT-only analysis; bottom row: common SAT/SLP analysis.

<u>Comments</u>: This figure is completely analogous to Fig. 12 of the main text, but shows the results for the model residuals based on the 20CRv3-rescaled forced signals.

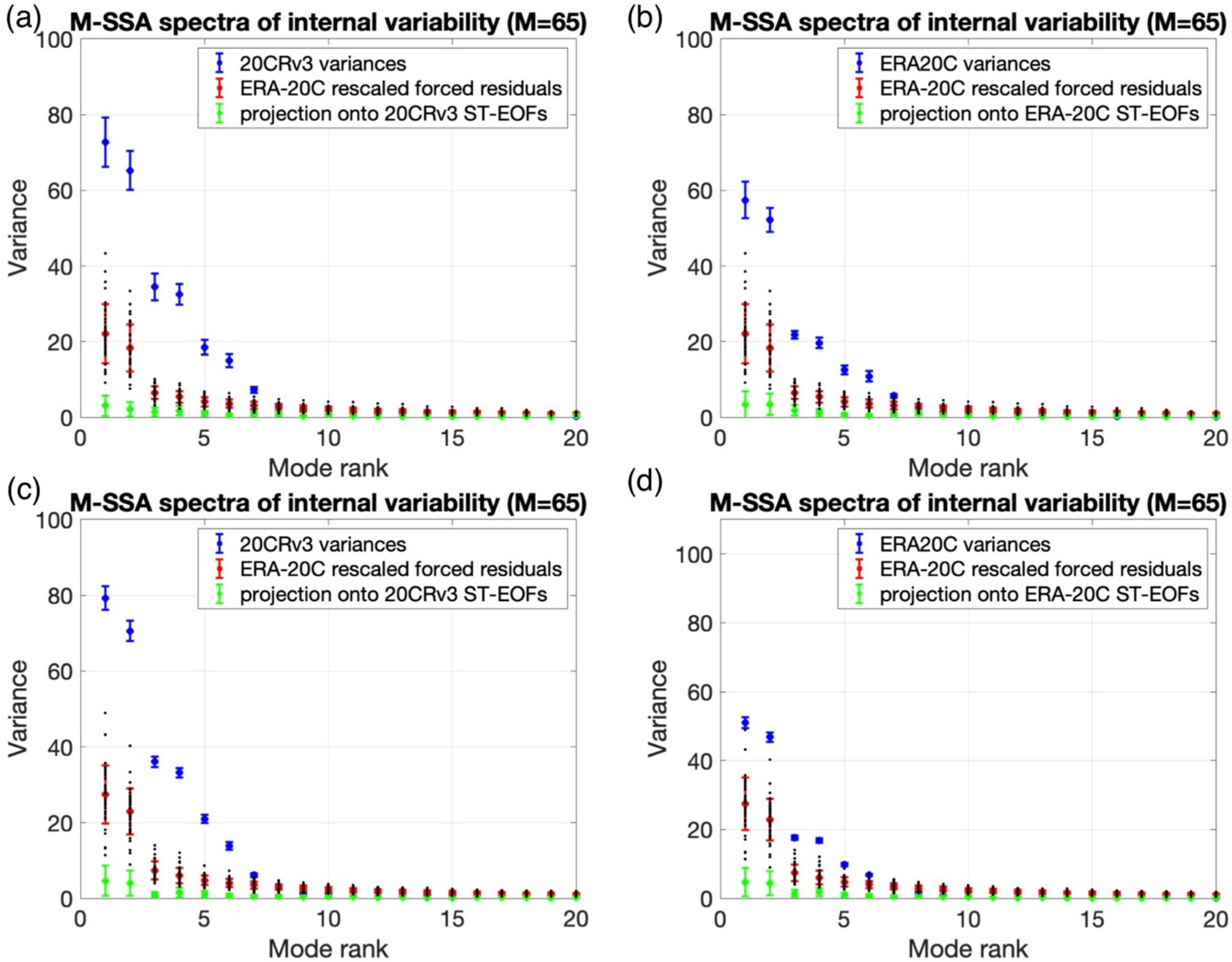

Fig. S12 The same as Fig. S11 above, except the forced-signal residuals are based **on the ERA20C-rescaled** CMIP5/6 data.

Comments: This figure is completely analogous to Fig. 12 of the main text but shows the results for the model residuals based on the ERA20C-rescaled forced signals.

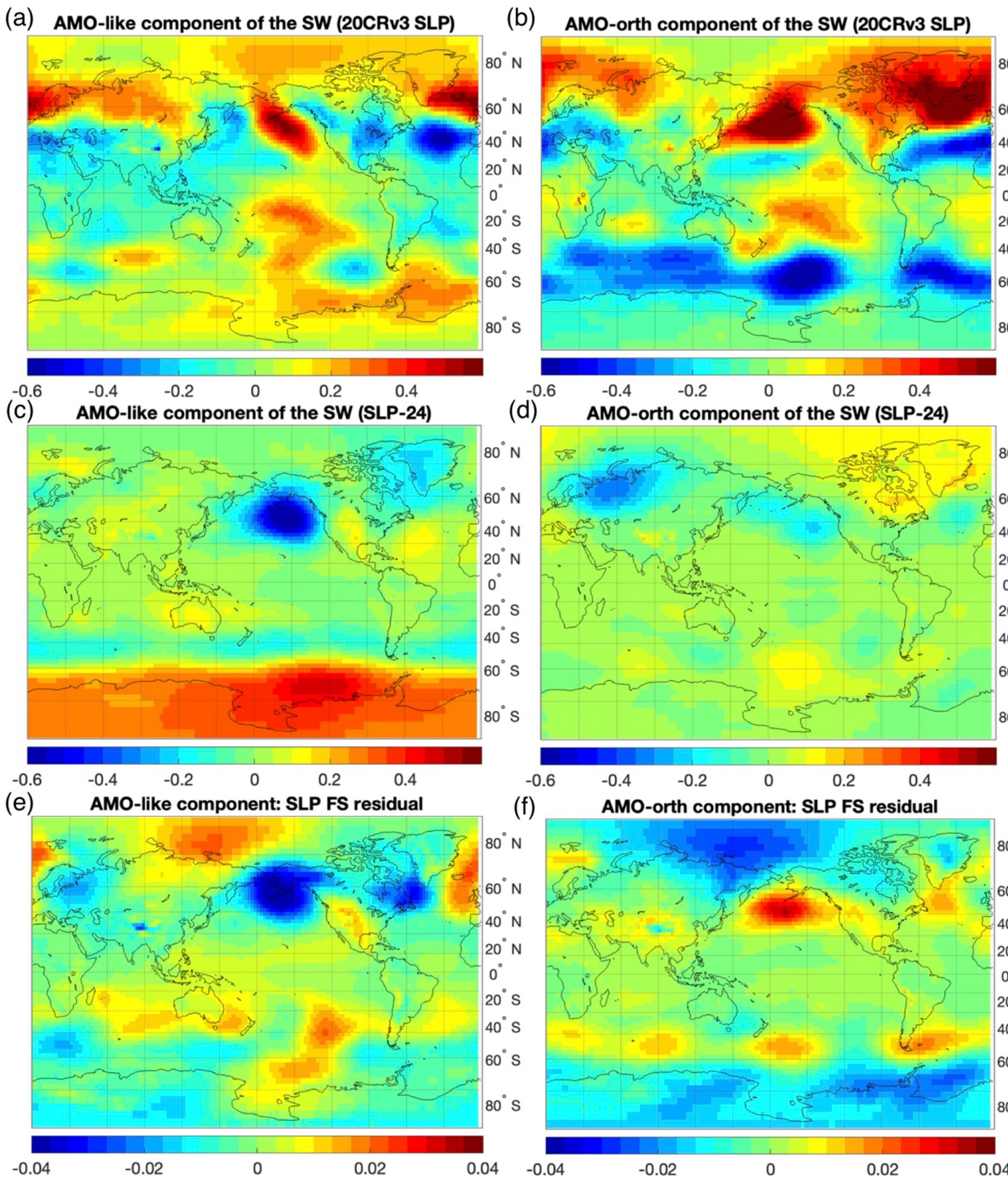

Fig. S13 Dimensional SLP patterns associated with the common SAT/SLP rotated PCs of the "stadium waves" based on the estimates of internal variability in reanalysis data and the CMIP5/6 forced-signal residuals. Top row: 20CRv3 patterns; middle row: model 24's forced-signal residual patterns; and bottom row: the ensemble-mean forced-signal residual ([non-existent] stippling would show the regions in which at least 29 out of 38 SLP residuals have the same sign). All patterns are weighted by the square root of the cosine of latitude.

Comments:

(1) This figure is analogous to Fig. 14 of the main text for SAT.

(2) Note that the ensemble-mean SLP patterns in the bottom row are essentially zero, without any regions of consistent SLP anomalies between the forced-signal residuals of different models.

(3) The individual-model forced-signal residuals can exhibit the magnitudes consistent with those of the observed stadium-wave patterns (see, for example, Fig. S13c), but appear to be random, without any systematic correlation with the observed patterns in the upper panels (see also Fig. 15 of the main text).