PNAS

Supporting Information for: Internal variability and forcing influence model-satellite differences in the rate of tropical tropospheric warming

Stephen Po-Chedley*, John T. Fasullo, Nicholas Siler, Zachary M. Labe, Elizabeth A. Barnes, Céline J. W. Bonfils, Benjamin D. Santer

*To whom correspondence may be addressed. **Email**: pochedley1@llnl.gov.

This PDF file includes:

Supporting Information Text

Extended Methods

Model simulations. This analysis uses CMIP6 historical simulations, which are driven by estimates of external forcing over 1850 to 2014 (1). For the purposes of training ML models, we required that each GCM have at least 10 ensemble members with the same model physics configuration and forcing estimate, yielding data from 14 climate models (Table S1).

In addition to the climate model data used in training and validating our ML models, we contextualize our results with tropospheric temperature trend data from 45 other climate models (updated from reference 2), each of which has less than 10 ensemble members. When these models have multiple ensemble members, we require that each member have the same physics configuration and forcing estimate.

We also utilize the Community Earth System Model version 2 (CESM2) Large Ensemble, which includes 100 historical simulations (3). Within this ensemble half of the members prescribe the standard CMIP6 biomass burning emissions. We include these 50 simulations as part of our CMIP6 dataset. The other 50 simulations have temporally smoothed BB emissions (4; see below).

To explore the sensitivity to the analysis time period (Fig. S5), we extend CMIP6 historical simulations through 2021 using either the SSP3-7.0 or SSP5-8.5 experiment from ScenarioMIP (5). If a given historical simulation was extended with both SSP experiments, we used the simulation extended by SSP5-8.5. In practice, the emission scenario differences are small over 2015 to 2021, so the choice of SSP experiment for extension has little impact on TMT trends over 1979 to 2021 (6).

To compare satellite brightness temperature observations with model simulations, we compute synthetic brightness temperature values using temperature weighting functions that depend on surface characteristics (surface pressure, land fraction, and sea ice fraction) (7). These functions are used to produce a weighted value of atmospheric (and surface) temperature that approximates the satellite-observed brightness temperature. Our analysis is focused on the mid-tropospheric temperature channel (TMT). Since the TMT product includes a small contribution from the stratosphere, we remove influence from stratospheric cooling using a channel that measures the temperature of the lower stratosphere (TLS) (2, 8).

In addition to analysis over the historical era, we also consider pre-industrial control simulations from 13 of the 14 models used in the ML training and testing. The exception is for GISS-E2-1-G, for which sea ice output is unavailable to calculate the synthetic satellite brightness temperatures described above. We examine the distribution of unforced tropical TMT trends in these experiments by sampling 36-year trends offset by five years (e.g., years 0 - 35, 5 - 40, 10 - 45, …; Fig. S13). This yields a total of 2,205 control run samples. If multiple control simulations exist, we select the first ensemble member (e.g., r1i1p1f1).

Biomass burning sensitivity simulations. The standard CMIP6 biomass burning emissions dataset has a noticeable increase in interannual variability beginning in 1997 (4). To explore the impact of this discontinuity in variability, half of the CESM2 large ensemble used a modified biomass burning dataset in which the emissions were smoothed using an 11-year running mean filter. The smoothed emissions dataset largely conserved total BB emissions but reduced the year-to-year variability (3). We refer to this experiment as CESM2-SBB and compare the CESM2-

SBB simulations to the standard CESM2 simulations to quantify the sensitivity to the forcing dataset over the satellite era.

Machine learning approach. Our general approach is to use the spatial pattern of surface warming to predict the magnitude of: a) forced; and b) unforced tropical TMT trends over 1979 to 2014. In order to build up samples with which to train our ML model, we sample across historical time periods, climate models, and climate model ensemble members (see Fig. S1). Each sample includes a two-dimensional map of the surface temperature trend at each grid point (predictor) and the values of forced and unforced trends for spatially averaged tropical tropospheric temperature data (2 predictands).

Our analysis utilizes a "leave-one-out" cross-validation approach: we train our ML model on data from 13 of 14 climate models and apply the resulting ML fit to predict the forced and unforced components of the satellite-era tropical TMT trend for each ensemble member of the GCM not used in training. We then repeat this procedure for all 14 models, thus allowing us to predict the forced and unforced components of the satellite era (1979 – 2014) tropical TMT trend for each ensemble member of each GCM.

Simulated near surface air temperature (the CMIP "tas" variable) is regridded to a 2.5º by 2.5º grid (72 latitude by 144 longitude points) so that the predictor trend maps are on a common grid. Regridding was performed with the Earth System Modeling Framework (ESMF) bilinear regridder in the xarray Climate Data Analysis Tools (xCDAT v0.3.2) Python package (9). To train our ML model, we use ten ensemble members from each GCM and sample 25 different overlapping 36-year periods for each climate model simulation (1854 - 1889, 1889 - 1894, 1864 - 1900, …, 1974 - 2009). We use ten ensemble members for training, but we use all ensemble members when predicting the trends for the GCM not used in training. Note that the period of interest, 1979 - 2014, is omitted from training. For each GCM used in training, we have 10,368 spatial features (from the 72 latitude by 144 longitude grid) and 250 samples (25 time periods times 10 ensemble members). Since we use 13 GCMs for training, this leaves us with a predictor matrix, **X**, with size [3,250 x 10,368]. Our predictand matrix, **Y**, has two targets (the forced and unforced tropical TMT trend) for each sample, yielding a matrix of size [3,250 x 2]. Both the predictor and predictand matrices are standardized prior to fitting the ML model. We standardize each grid point by subtracting the grid point sample mean and dividing by the grid point sample standard deviation. We apply area weighting to predictor features in the PLS model by multiplying by the cosine of latitude.

In addition to near surface air temperature, we also prepare maps of sea surface temperature (SST) trends as a predictor of the forced and unforced components of tropical TMT change. These maps are regridded to a 2.5º x 2.5º grid using the conservative_normed method, but we only consider data equatorward of 40º in order to avoid the influence of sea ice. In the models, we use the surface skin temperature field over the ocean as a proxy for SST. If trend data was missing from any observed SST dataset, we excluded the grid cell in both the model and observational trend maps, resulting in a common mask for both model and observational data. Unlike the surface air temperature observations, we do not infill missing data. The SST trend data is standardized (as described above) prior to fitting our ML models.

In calculating the predictand values, we assume that the forced component of the tropical TMT trend is represented by the ensemble average trend; the forced trend is the same for all ensemble members for a given climate model and time period. Note that this means the forced component of the trend is influenced by both anthropogenic (e.g., greenhouse gas) and natural (e.g., volcanic) external forcing. The deviation from the forced trend in each ensemble member is the unforced component of the TMT trend (we variously refer to this as the internal or unforced trend). Consequently, the forced and unforced components of the tropical TMT trend sum to the total tropical TMT trend.

We use PLS regression as our primary ML model in order to separate and quantify the forced and unforced component of tropical TMT change. PLS regression is similar to singular value decomposition in that it reduces a large number of collinear predictors into a finite number of modes, which are linearly related to the predictand matrix. A benefit is that such modes are readily visualized (10, 11). PLS regression can also be reduced to coefficient matrices (i.e., fingerprint patterns), β, such that $\mathbf{Y} = \mathbf{X}\beta$. In our case, the coefficient matrices can be interpreted as the fingerprint patterns associated with forced and unforced tropical TMT change. We utilize scikit-learn (version 1.1.2) to fit and make predictions with the PLS Regression model (12).

To determine if our results are robust to application of other ML approaches, we also use ridge regression and an artificial neural network (ANN). Since we have a large number of correlated predictors, standard multiple linear regression would result in overfitting. Ridge regression helps to guard against overfitting by penalizing large regression coefficients with an added term in the loss function. We modulate the importance of this L_2 regularization term using a tunable parameter, α.

Encouraged by work that implements relatively shallow ANN architectures to shed light on the forced climate response (13–15), we implement a simple ANN architecture (using the MLPRegressor functionality in scikit-learn) consisting of two hidden layers with ten hidden units each. Our ANN uses a rectified linear unit activation function, L_2 regularization (as with ridge regression), a stochastic gradient descent solver, the squared error as the loss function and a constant learning rate of 0.001. Our ANN is trained using early stopping with 20% of the samples randomly set aside for validation; training terminates when the validation score fails to improve for 10 consecutive epochs. As with PLS Regression, we use scikit-learn to implement the ANN and ridge regression.

Parameter selection. Before fitting our ML models, we must select several fitting parameters. In PLS regression, we need to choose the number of modes, N_{modes} (these are referred to as "components" in the scikit-learn implementation of PLS regression). Both ridge regression and our ANN use an L_2 normalization parameter, α. The ANN can also be configured with many architectures. Since our goal is to test the robustness of the PLS regression result to other ML techniques, we chose to use a basic architecture (described above) that has been used for other climate applications, though we note that it may be possible to further improve upon our results with more complex ANN designs.

To guide our parameter selection, we tested the error associated with different values of N_{modes} and α, using the leave-one-out approach described above. For each parameter value, we trained on data from 13 (of 14) climate models and used that fit to predict the forced and unforced tropical TMT trend for the 14th model. We recorded the mean squared error (MSE) of these predictions across all 25 different 36-year time periods (1854 - 1889, …, 1975 - 2009) and all model ensemble members. We then repeated this procedure for each of the 14 models. This resulted in 14 MSE values for each candidate parameter value. We chose to use a leave-one-out sampling strategy and record the MSE across a range of time periods to ensure that our parameter selection is robust to a range of climate conditions (as represented by different models and time periods) and to guard against overfitting to our time period of interest (1979 - 2014).

We subsequently inspect the average MSE across all 14 models for the a) forced, b) unforced, and c) total (forced + unforced) components of the tropical TMT trend (Fig. S9). We selected the parameter that minimizes the average MSE across these three values. For our ML models, the MSE is minimized using 6 modes for PLS regression, α=17,500 for ridge regression, and α=30 for the ANN.

Observations of surface and tropospheric temperature change. Surface temperature datasets include the Berkeley Earth Surface Temperature dataset (BEST) (16), the Goddard Institute for Space Sciences Temperature Analysis (GISTEMP) version 4 (17), the most recent version (version 5) of the U.K. Met Office Hadley Centre/Climatic Research Unit global surface

temperature dataset (HadCRUT5) (18), an infilled version of the HadCRUT4 dataset (HadCRUT4-UAH) (19), and version 5 of the European Centre for Medium-Range Weather Forecasts reanalysis (ERA5) (20).

Sea surface temperature datasets include the Program for Climate Model Diagnosis and Intercomparison SST dataset (PCMDI) (21), version 2 of the Centennial In Situ Observation-Based Estimates of the Variability of SST and Marine Meteorological Variables (COBE2) (22), version 5 of the Extended Reconstructed Sea Surface Temperature (ERSST) (23), version 4 of the U.K. Met Office Hadley Centre's sea surface temperature data set (HadSST4) (24) and combined sea ice and sea surface temperature dataset (HadISST) version 1.1 (25).

Observations of TMT include version 4 of the Remote Sensing Systems (RSS) dataset (26), the University of Alabama in Huntsville (UAH) version 6 product (27), version 1 of a dataset from the University of Washington (UW) (28), and version 4.1 of the National Oceanic and Atmospheric Administration Center for Satellite Application and Research (NOAA) (29). As with the model data, we remove the effects of stratospheric cooling with data from the TLS channel. Since the UW dataset does not include TLS observations, we remove stratospheric effects from UW data using the NOAA TLS product.

Prediction for observations. Our ultimate goal is to quantify the influence of internal variability and external forcing on the observed tropical TMT trend. To produce this estimate, we use our ML models that were trained and validated on climate model data and apply them to the observed record. The observed predictions are based on five surface air temperature datasets. Each is regridded to the same 2.5^{o} x 2.5^{o} grid used for processing simulation output.

The HadCRUT5 dataset includes a 200-member observational ensemble that samples major sources of known uncertainty (18). For each member, we pass the surface temperature trend (1979 - 2014) to the ML model to produce a predicted forced and unforced tropical TMT trend. Since we use a leave-one-out validation approach, we generate predictions for 14 different ML models and 200 ensemble members (2,800 forced and unforced tropical TMT trend predictions).

The other surface temperature datasets that we consider consist of only one surface temperature trend map, yielding a single prediction for the forced and unforced TMT trend for each of the 14 ML models. In order to provide some representation of observational uncertainty for these datasets, we take the (200 member) distribution of HadCRUT5 predictions and center it so that the mean of the HadCRUT5 distribution matches the forced and unforced tropical TMT trend predictions for each of the other four surface temperature datasets. For example, if the GISTEMP dataset has an unforced tropical TMT trend prediction of -0.06 K decade^{-1} for one of the 14 ML models, we take the corresponding distribution of HadCRUT5 predictions and center the mean of this 200-member ensemble on -0.06 K decade^{-1}. This is repeated for each of the 14 ML models, yielding the same number of predictions as with the 200-member HadCRUT5 dataset. Each surface temperature dataset thus has 2,800 (14 ML models x 200 HadCRUT5 ensemble members) forced and unforced predictions, instead of 14 individual predictions. Although the uncertainty inferred from HadCRUT5 is not necessarily representative of the uncertainty in other datasets, this approach is preferable to assuming no observational error in individual datasets. Note that when using the observed SST pattern of warming as a predictor, we used the HadSST4 ensemble to estimate the observational uncertainty for each SST dataset (Fig. S4).

Aside from the ERA5 and HadCRUT4-UAH datasets, each surface temperature product has grid cells with missing data (Fig. S3). If a grid cell has missing data, we set the grid cell trend value to the zonal average trend (from grid cells without missing data) before predicting the forced and unforced tropical TMT trends. To test the error associated with this decision, we artificially mask the temperature trends in the ERA5 reanalysis dataset (which has no missing data) to match the data availability in the BEST, GISTEMP, and HadCRUT. We then infill these missing values using the zonal average trend to mimic our treatment of datasets with missing data. We compare the

predicted forced and unforced tropical TMT trend for the masked and unmasked version of the ERA5 surface temperature trends. The impact of masking is small. The average prediction difference (masked minus unmasked) across all observational datasets and all 14 PLS regression models is -0.001 and 0.001 K decade^{-1} for the forced and unforced trend prediction, respectively. The largest absolute bias across all PLS models is 0.004 K decade^{-1} (for the HadCRUT5 dataset). The results are similar for the ridge regression and the ANN. The small sensitivity to masking occurs because the predictions are most sensitive to surface temperature trends in the tropics and northern midlatitudes, but the largest areas of missing surface temperature data tend to be in the southern extratropics. We chose to include datasets with missing values to improve our representation of the structural uncertainty across the observational records.

The precision with which we can separate and quantify the forced and unforced tropical TMT trends ultimately relies on the accuracy of our training approach – learning from climate models and applying that knowledge to observations. Our leave-one-out sampling approach emulates the observational prediction process, because we iteratively train on data from 13 GCMs and predict the forced and unforced tropical TMT trends for each ensemble member of the 14th GCM. For each climate model, we can quantify the prediction error by comparing the predicted values to the actual trends calculated from each model ensemble. We then use the prediction errors as part of our observational estimate of the forced and unforced tropical TMT trend.

To incorporate prediction uncertainty, we use an approach typically used in emergent constraint studies (30, 31). We start by fitting a line to the climate model data in order to fit the actual (forced or unforced) satellite-era (1979 - 2014) tropical TMT trend, y_{actual}, using a constant offset, b, and the ML-predicted trend, $x_{predict}$, such that $y_{actual} = m \cdot x_{predict} + b$ (gray lines in Fig. 1a–b). We use orthogonal regression, which accounts for error in both the dependent and independent variables (31). We also weight the data so that each GCM contributes equally, irrespective of differences in ensemble size. We then use the observed values of $x_{predict,obs}$ in our linear regression equation to yield a final, bias-corrected estimate of the actual observed (forced or unforced) tropical TMT trend, $y_{actual,obs}$. To estimate the prediction error of $y_{actual,obs}$ from this regression, we follow Cox et al. (2018) to produce a probability distribution function (PDF) of the estimated (forced and unforced) tropical TMT trends given the observed prediction. We calculate this PDF for the 2800 predictions (see above) from each of the five surface temperature datasets. We then sum and normalize the resulting PDFs ($n = 5 \cdot 2800 = 14000$) to provide a single PDF of the forced and unforced tropical TMT trend that incorporates both the model-inferred prediction error and the structural uncertainty in the observations.

Text 1: PLS Modes and Robustness across ML Models

PLS regression fingerprints consist of contributions from several geophysical modes, which contribute to our unforced and forced fingerprint patterns (Fig. S7). For instance, modes 2 and 3 are the main modes contributing to the unforced fingerprint. Our analysis utilizes 6 modes as part of PLS regression, which minimizes the mean squared error in leave-one-out predictions across all time periods prior to 1979 to 2014 (Fig. S9). Recomputing our results across two to ten PLS modes (Fig. S10a), we find that the observation-based predictions vary by less than ±0.02 K decade^{-1} for the forced tropical TMT trend (with a range of 0.24 to 0.27 K decade^{-1}) and for the unforced tropical TMT trend (with a range of -0.09 to -0.07 K decade^{-1}).

Another linear technique, ridge regression, also includes uncertainty that results from the selection of an L_2 regularization parameter (α) that is used to reduce overfitting. This uncertainty is small; in varying this parameter by three orders of magnitude (10^3 to 10^5) the estimated forced and unforced components of the TMT trend deviate by less than 0.02 K decade^{-1} (Fig. S10b). The L_2 parameter that minimizes MSE (α=17500, see Fig. S9), has a similar coefficient map compared to PLS regression (Fig. S11 and Fig. 2). Coefficient maps based on larger L_2 parameters are smoother and easier to interpret (Fig. S12). The predicted tropical TMT trends from the ANN are very similar to those from PLS regression, but the results are more sensitive to the L_2 parameter (Fig. S10c). Despite the larger parametric uncertainty, the ANN predictions are

consistent with the estimated TMT trends from PLS and ridge regression: across all L_2 parameter values, the unforced component of the trend is always negative (ranging from -0.04 to -0.08 K decade^{-1}) and the forced component of the trend is always substantial (ranging from 0.22 to 0.28 K decade^{-1}).

Table S1. Models used as part of ML disentanglement approach. List of models used as part of our ML disentanglement approach including their ensemble size (n) and ECS value.

Model	**n**	**ECS [K]**
ACCESS-ESM1-5	40	3.9
CESM2	50	5.1
CNRM-CM6-1	29	4.9
CanESM5	40	5.6
GISS-E2-1-G	12	2.7
GISS-E2-1-H	10	3.1
INM-CM5-0	10	1.9
IPSL-CM6A-LR	32	4.7
MIROC-ES2L	30	2.7
MIROC6	50	2.6
MPI-ESM1-2-HR	10	3.0
MPI-ESM1-2-LR	10	3.0
NorCPM1	30	3.0
UKESM1-0-LL	15	5.4

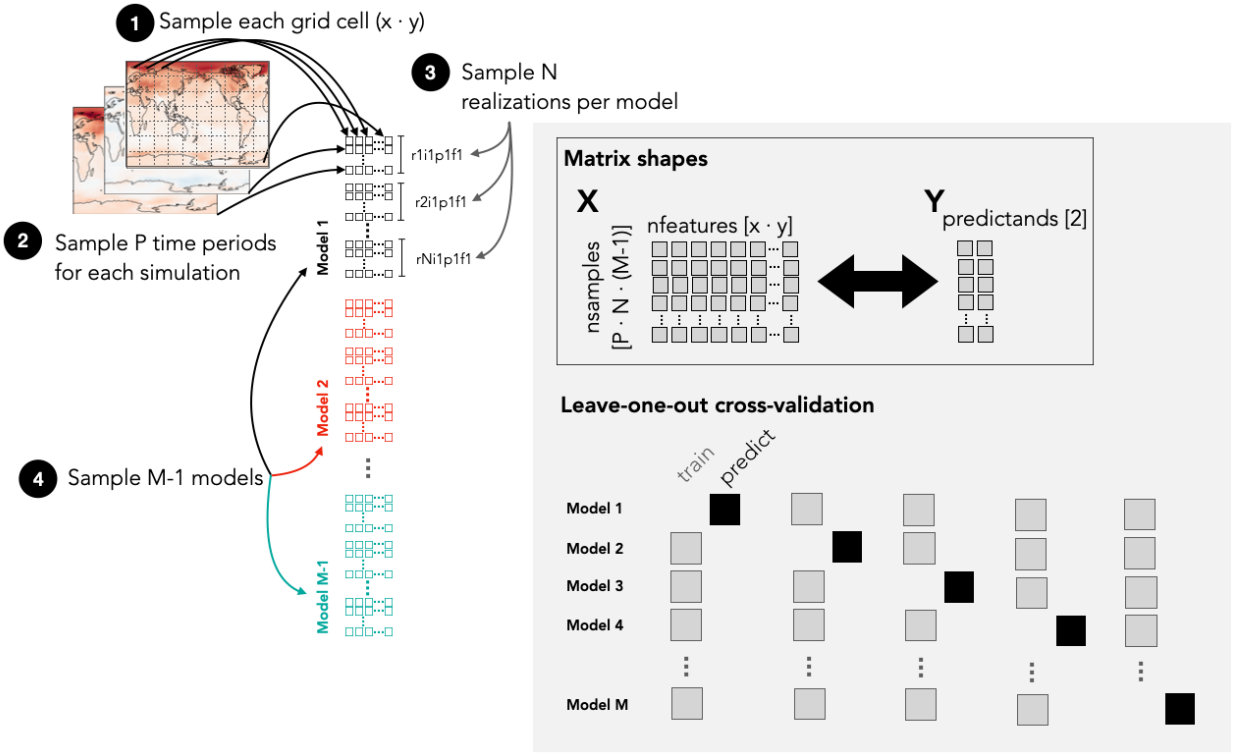

Fig. S1. Sampling design for machine learning approach. Schematic outlines the matrix shapes for the predictor (**X**) matrix, which consists of rows that contain the surface temperature trend values for each grid cell (a function of latitude, x, and longitude, y). Each row corresponds to a different map for P time periods, N model realizations, and (M-1) climate models. The predictand (**Y**) matrix has the same number of rows, but only has two columns: the forced and unforced component of the tropical TMT trend. We note here that the forced component of a model is repeated N times (for each realization of the ensemble), in order to obtain the same number of rows in the predictand matrix. Although we have M climate models, we use one less climate model (i.e., M-1) in training as part of our leave-one-out cross-validation approach. This means that we train our prediction model on all-but-one climate models and then make a prediction for the model not included in training. We perform this procedure iteratively so that a prediction is made for each of the M models. Climate model surface temperature trend maps were regridded to a 2.5º x 2.5º grid (x = 72 and y = 144). Over the historical period (1850 – 2014) we sampled 36-year trends with the start year spaced out in five-year increments (i.e., 1854-1889, 1859-1894, 1864-1900, …, 1974-2009); this yielded 25 different periods (P = 25). We included all 14 models (M = 14) that had at least 10 ensemble members (N=10). In all, this resulted in a predictor matrix, **X**, with shape [3250 x 10368] and a corresponding predictand matrix, **Y**, with shape [3250 x 2].

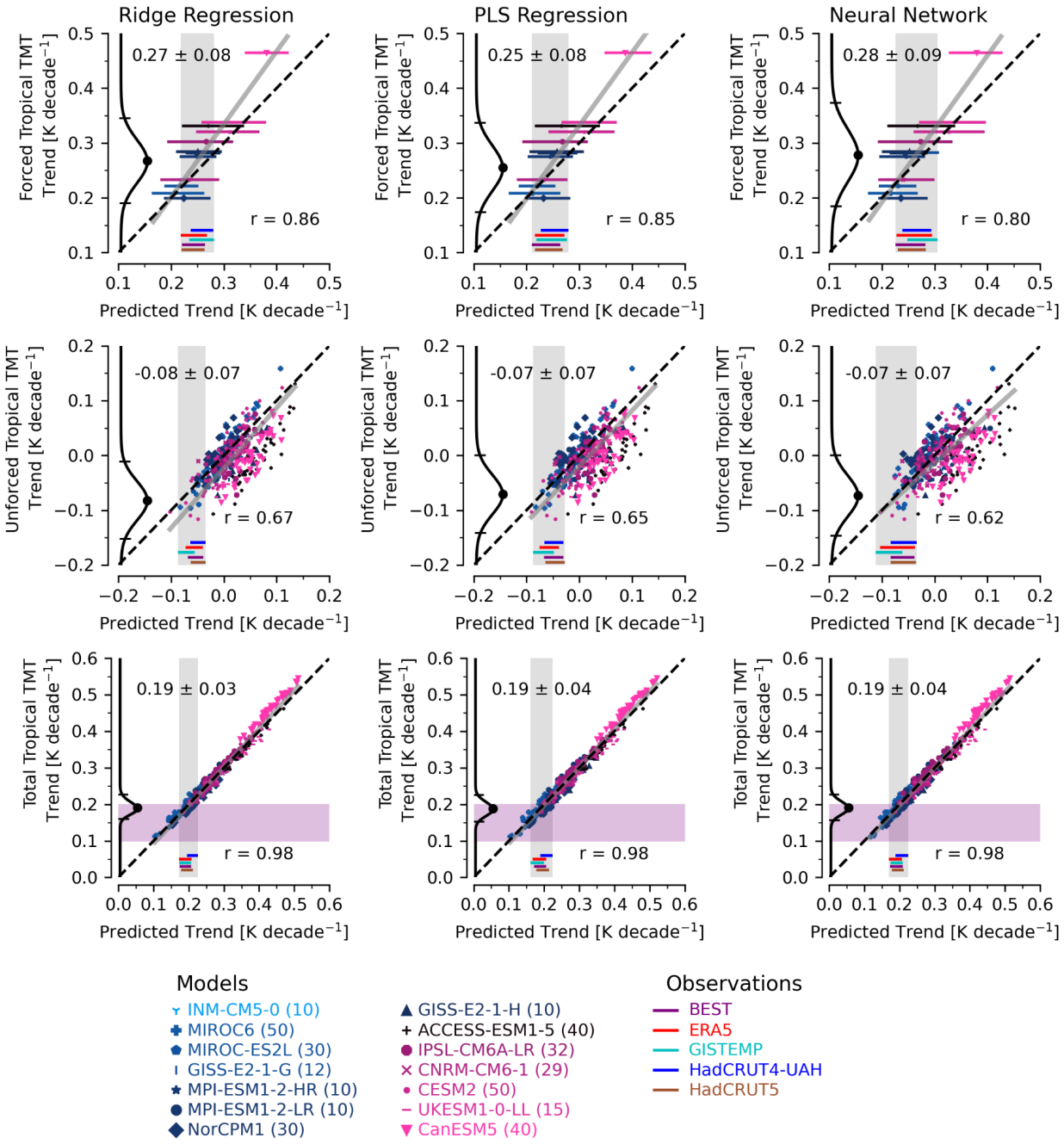

Fig. S2. Forced and unforced predictions for all machine learning methods. As in Figs. 1b (top row), 1a (middle row), and 3a (bottom) row, for Ridge Regression (left column), PLS Regression (middle column), and a Neural Network (right column).

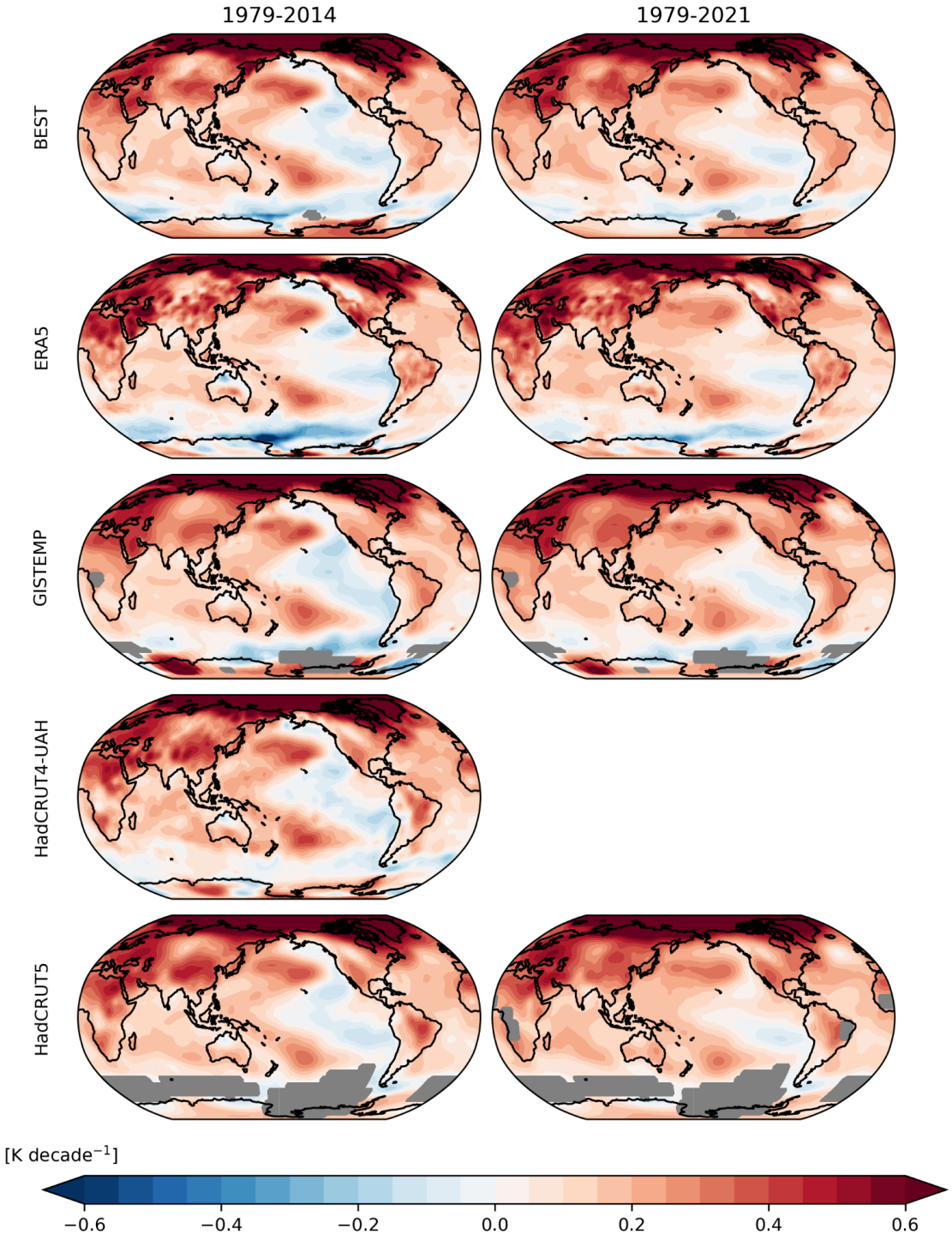

Fig. S3. Observed patterns of warming. Observed surface temperature trend patterns over 1979 - 2014 (left column) and 1979 - 2021 (right column) for five different observational datasets. Note that HadCRUT4-UAH is not displayed in the right-hand column because it does not extend through 2021.

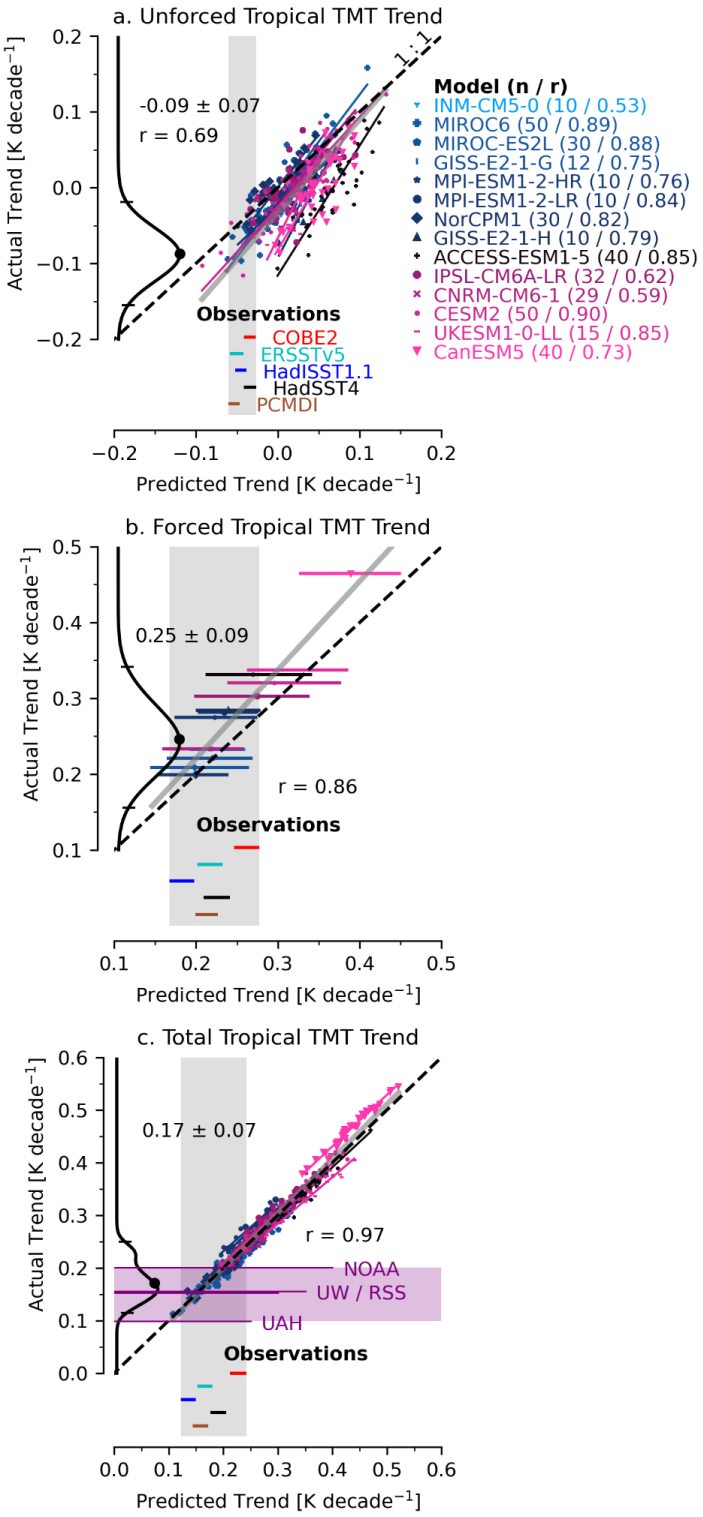

Fig. S4. Disentanglement results using SST trend maps as a predictor. As in Fig. 1 and Fig. 3a, but using SST (instead of land-ocean) trend maps as a predictor for the a) unforced, b) forced, and c) total tropical TMT trends.

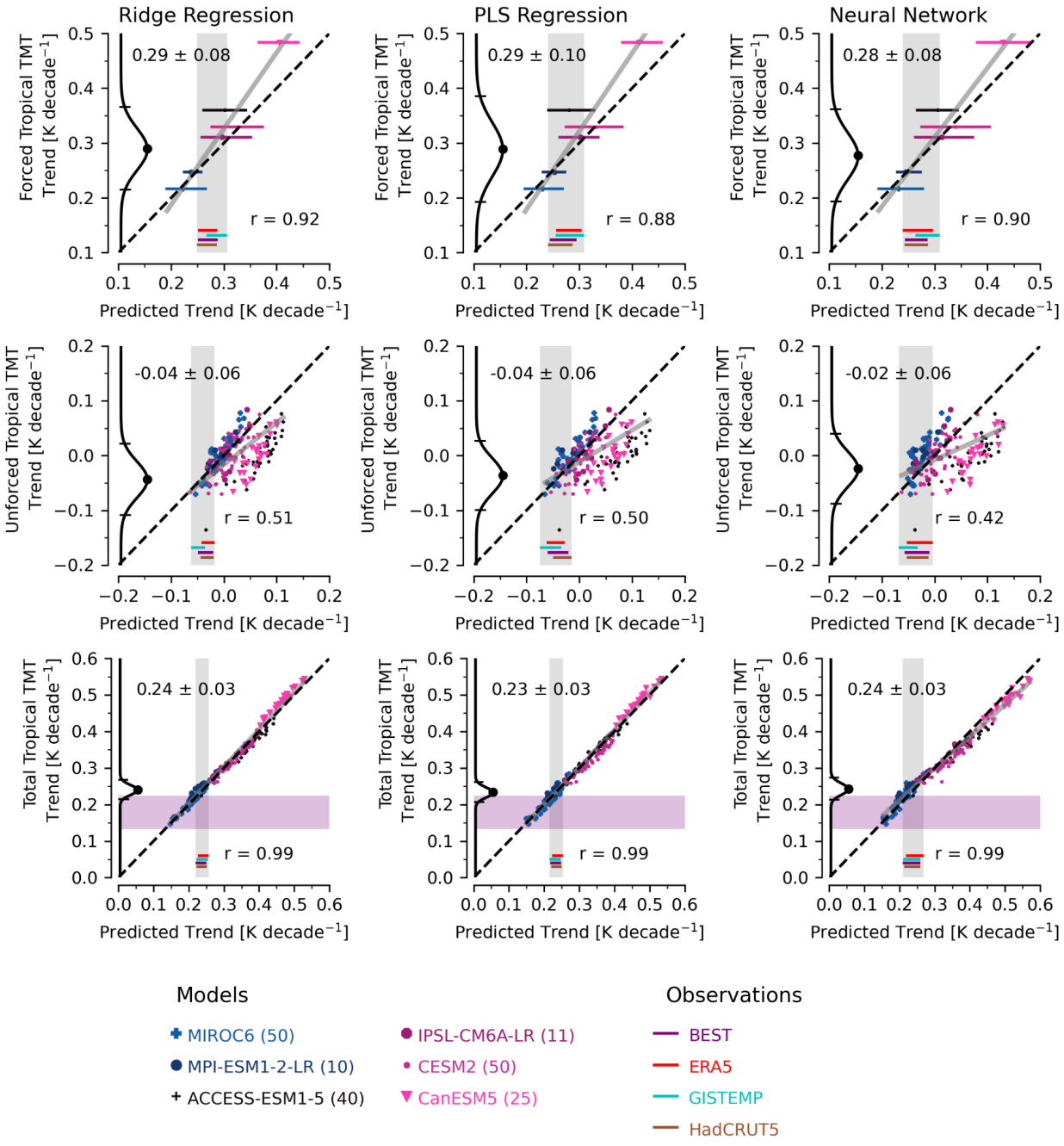

Fig. S5. Disentanglement results extended through 2021. As with Fig. S2, but for the period 1979 – 2021.

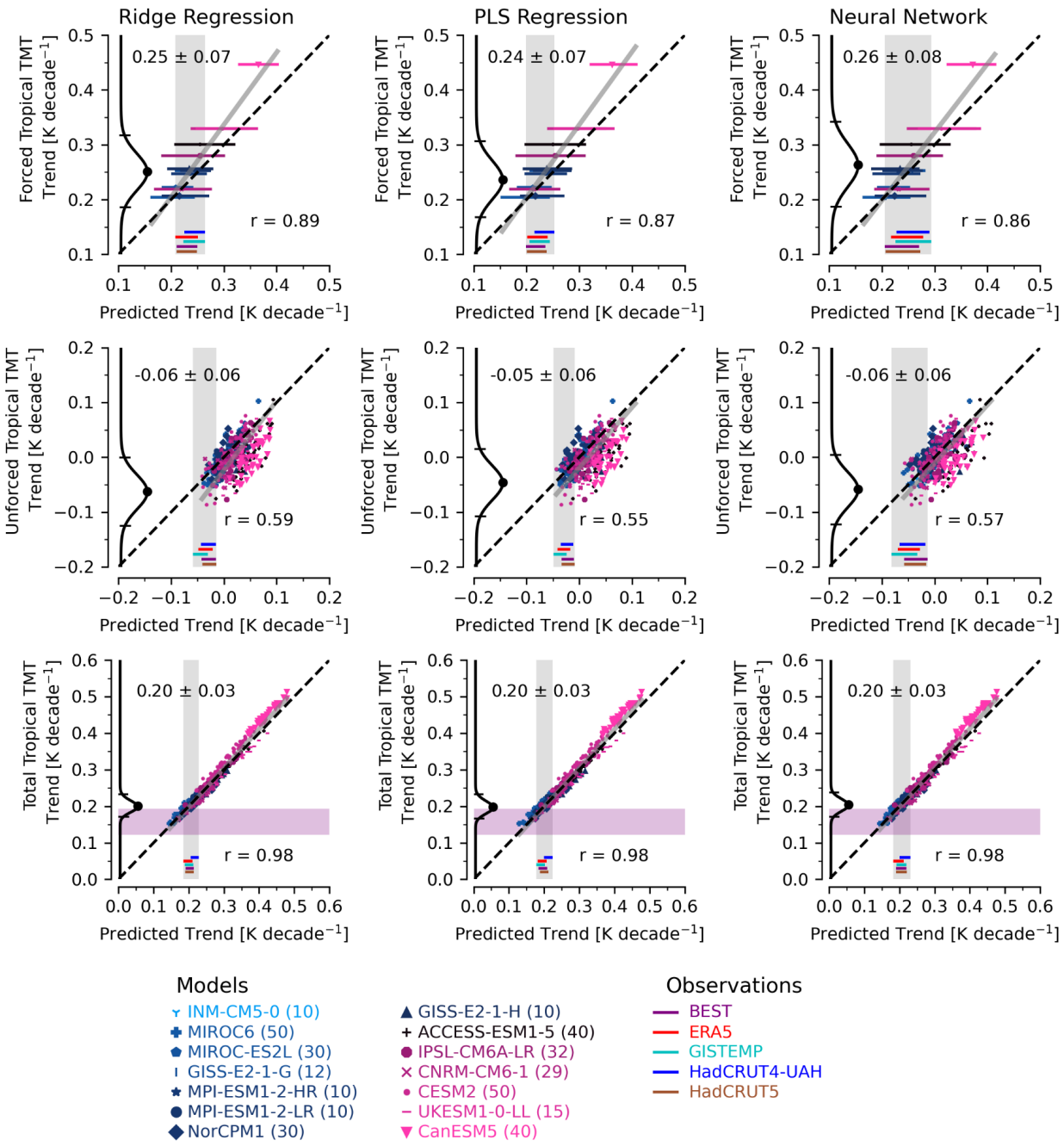

Fig. S6. Disentanglement results extended to the near-global scale. As for Fig. S2, but the predictions are for near-global (-82.5ºS – 82.5ºN) TMT trends.

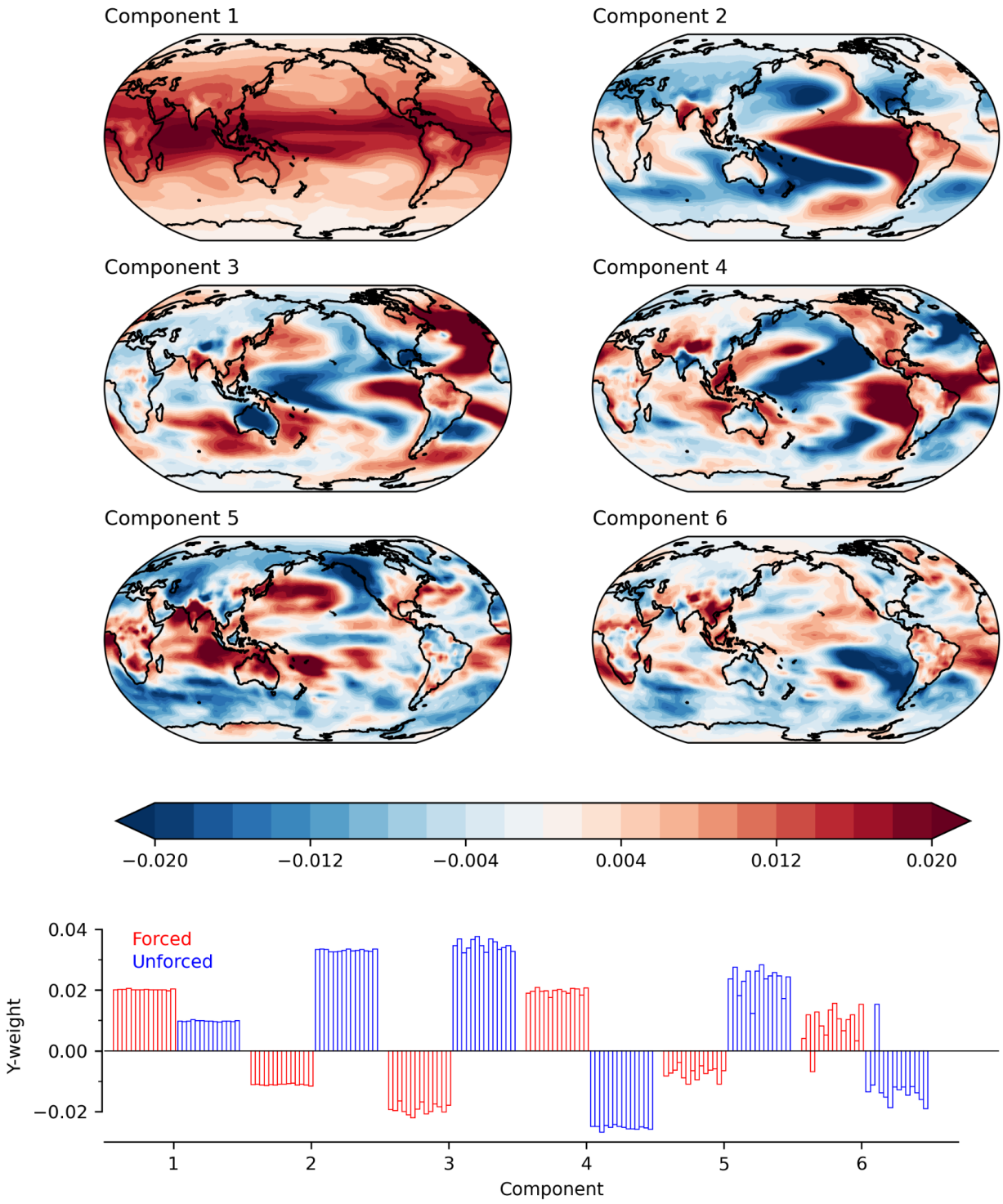

Fig. S7. PLS Regression weights for each mode. Spatial (**X**, top) and component (**Y**, bottom) weights (unitless) from PLS Regression. We display the Y-weights for each of the 14 leave-one-out ML models. The predictor weights are the average across models, which is very similar to each leave-one-out ML model for the leading modes (the spatial correlation between the mean map and individual maps exceeds 0.94 for the four leading modes).

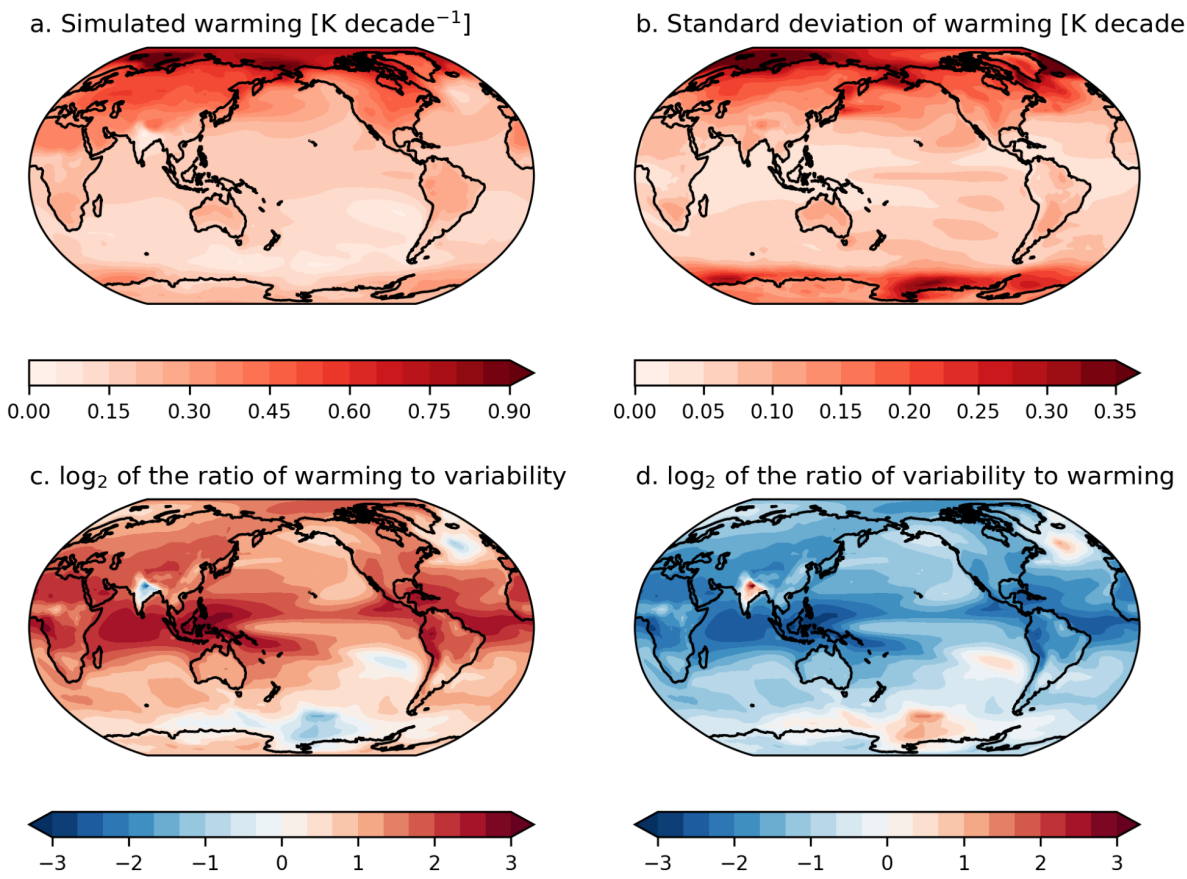

Fig. S8. Surface warming and variability. Multimodel mean a) surface warming over 1979 – 2014 and b) standard deviation of trends (across ensemble members) for the period 1979 – 2014. Panel c shows the $\log_2$ of the ratio of surface warming (from panel a) to the standard deviation of trends (from panel b). Panel d is the same, but for the reciprocal ratio.

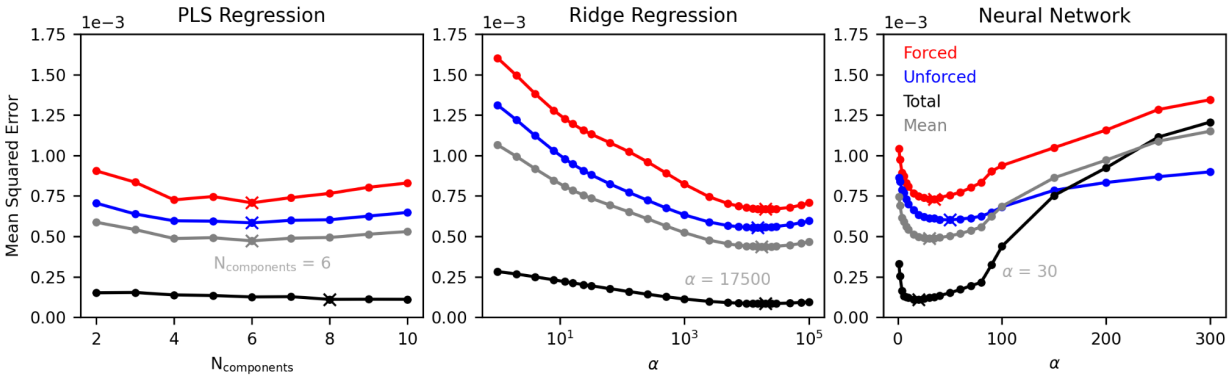

Fig. S9. MSE across parameter values. Mean squared error for the forced (red), unforced (blue), and total (black) TMT trends for (a) different numbers of modes in PLS regression and different α parameters for (b) ridge regression and (c) the ANN. The gray line is the average of the total, forced, and unforced MSE. The crosses indicate the parameter values that minimize the average MSE.

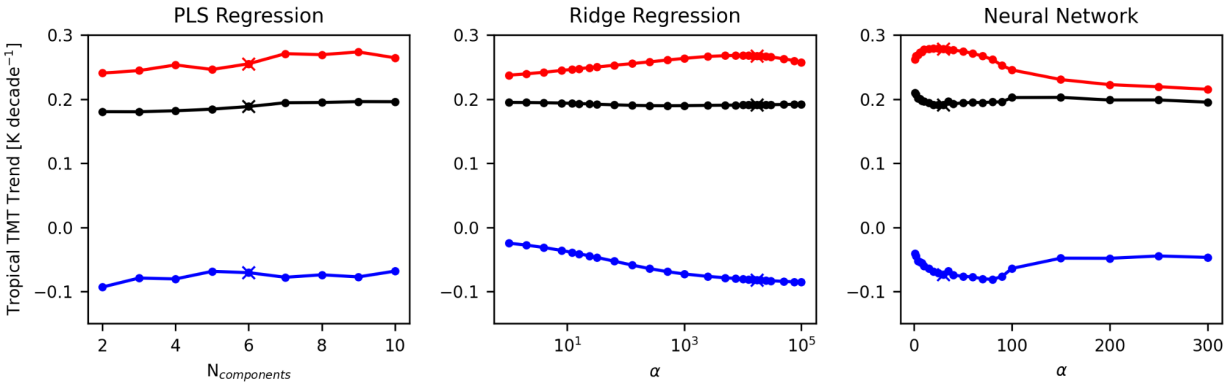

Fig. S10. Prediction sensitivity to parameter selection. Predicted forced (red), unforced (blue), and total (black) TMT trends for (a) different numbers of modes in PLS regression and different α parameters for (b) ridge regression and (c) the ANN.

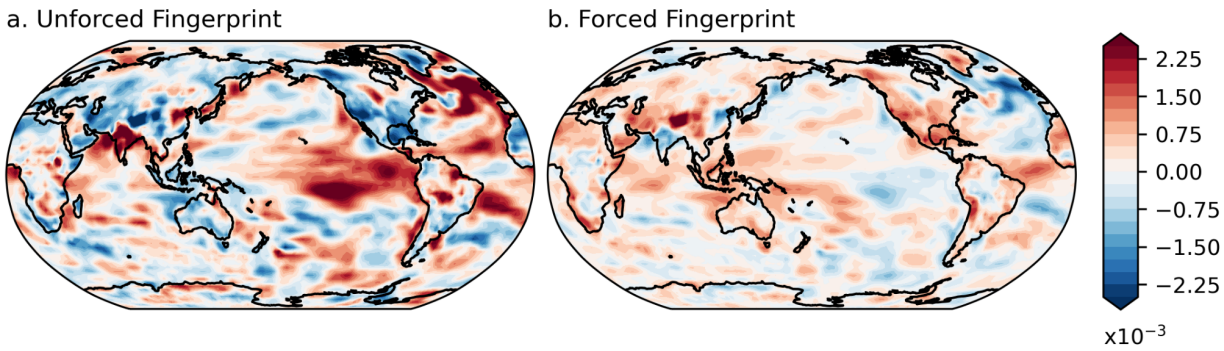

Fig. S11. Ridge regression fingerprint maps. As in Fig. 2a - b, but for ridge regression (α=17,500).

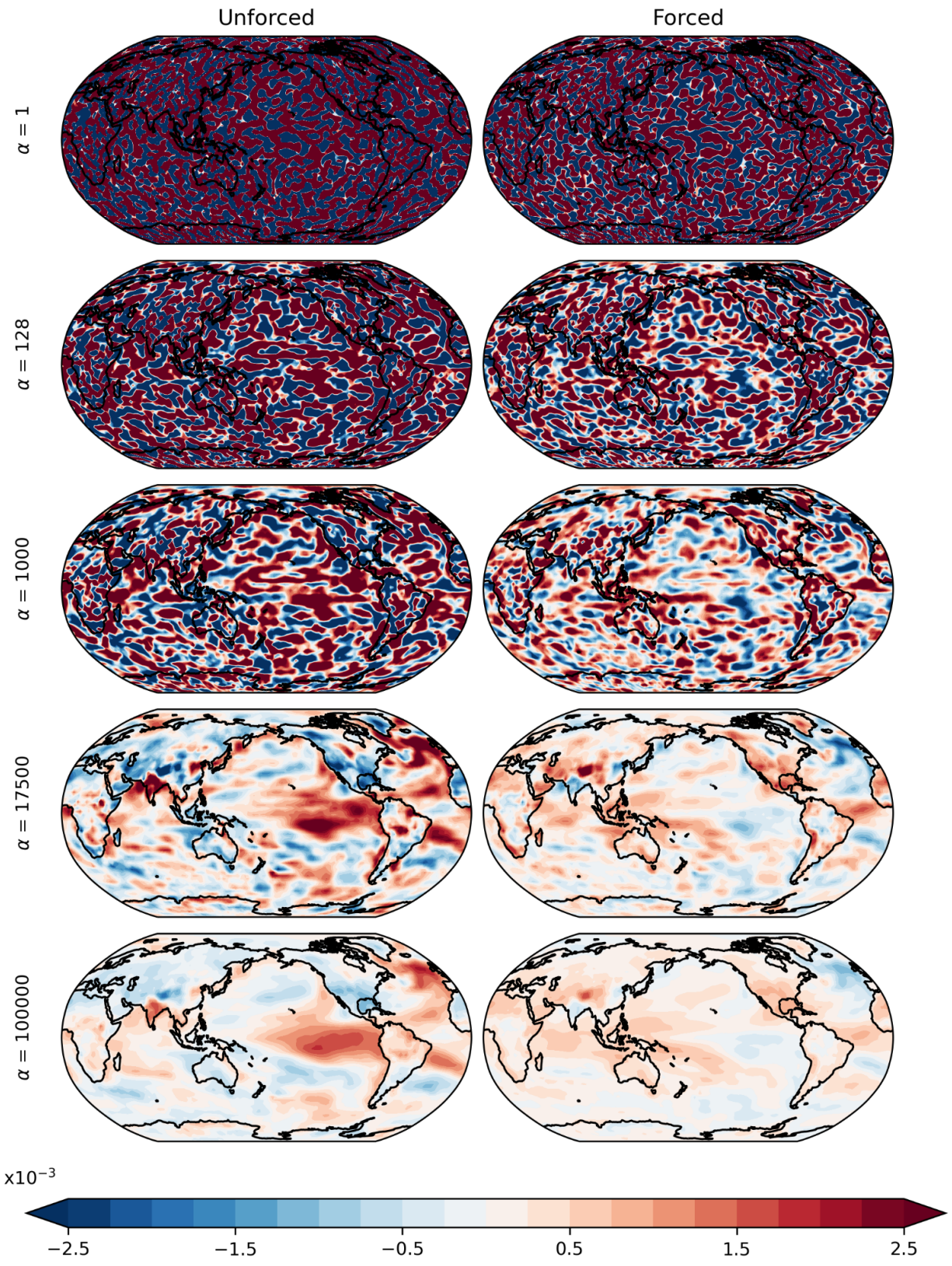

Fig. S12. Ridge regression fingerprint sensitivity to L_2 regularization parameter. Unforced (left) and forced (right) coefficient maps for different α parameters.

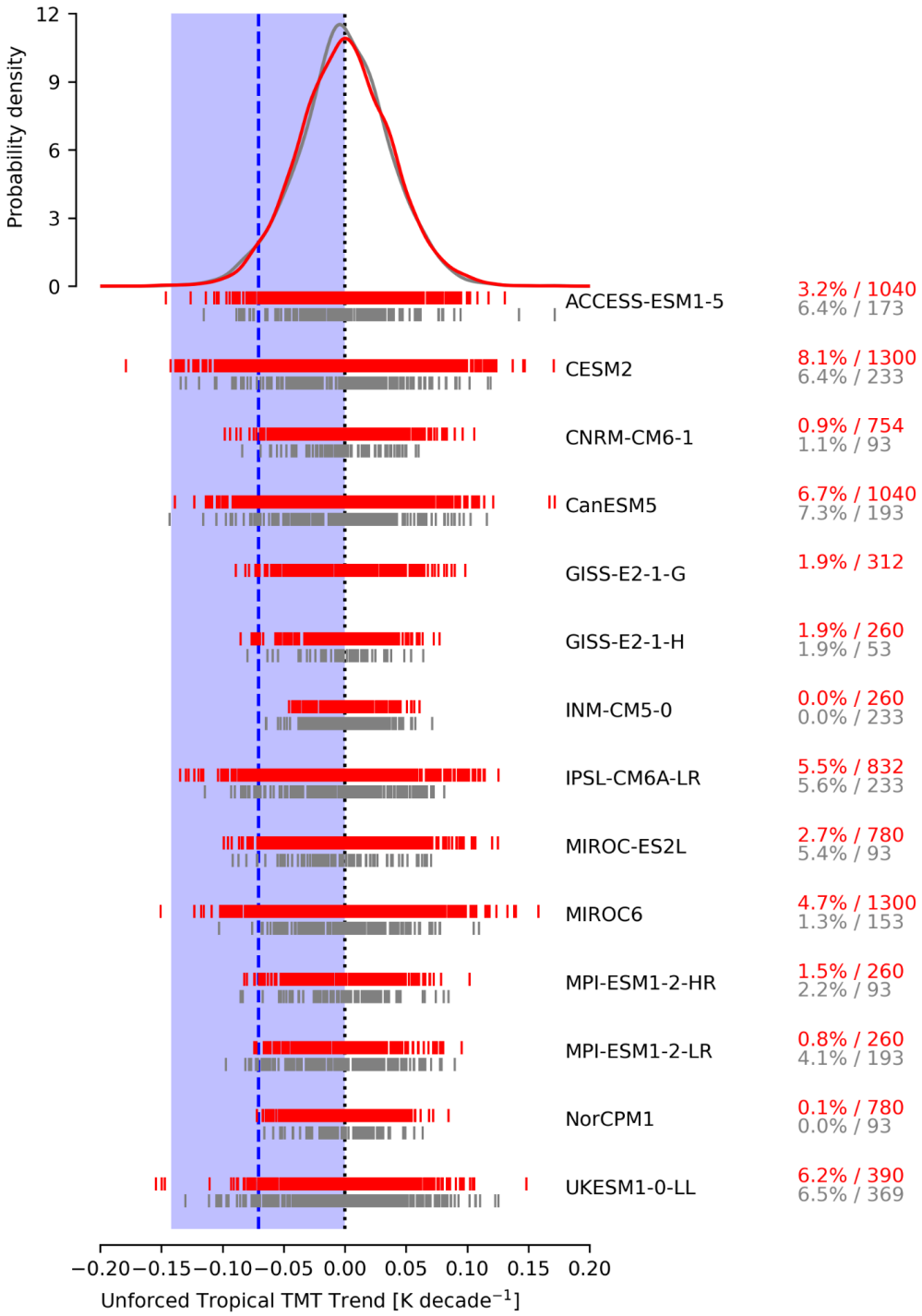

Fig. S13. Distribution of unforced tropical TMT trends. Probability distribution of the unforced tropical TMT trends from pre-industrial control simulations (gray) and historical simulations (red) along with our ML estimate of the observed unforced trend and its uncertainty (blue dashed line and purple shading). Below the probability distribution functions we also show the unforced trend values for each model (short vertical lines). We note both the percentage of unforced tropical TMT trends with values ≤-0.07 K decade^{-1} for the historical (red) and pre-industrial control simulations (gray) along with the total number of samples for each model and experiment. The black vertical dashed line is a reference line at 0.0 K decade^{-1}.

SI References

1. V. Eyring, *et al.*, Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).
2. S. Po-Chedley, *et al.*, Natural variability contributes to model–satellite differences in tropical tropospheric warming. *Proc. Natl. Acad. Sci.* **118**, e2020962118 (2021).
3. K. B. Rodgers, *et al.*, Ubiquity of human-induced changes in climate variability. *Earth Syst. Dyn.* **12**, 1393–1411 (2021).
4. J. T. Fasullo, *et al.*, Spurious Late Historical-Era Warming in CESM2 Driven by Prescribed Biomass Burning Emissions. *Geophys. Res. Lett.* **49**, e2021GL097420 (2022).
5. B. C. O'Neill, *et al.*, The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.* **9**, 3461–3482 (2016).
6. F. Lehner, *et al.*, Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth Syst. Dyn.* **11**, 491–508 (2020).
7. B. D. Santer, *et al.*, Comparing Tropospheric Warming in Climate Models and Satellite Data. *J. Clim.* **30**, 373–392 (2017).
8. Q. Fu, C. M. Johanson, S. G. Warren, D. J. Seidel, Contribution of stratospheric cooling to satellite-inferred tropospheric temperature trends. *Nature* **429**, 55–58 (2004).
9. T. Vo, *et al.*, xCDAT/xcdat: v0.3.2 (2022) https:/doi.org/10.5281/ZENODO.7216172 (October 17, 2022).
10. B. V. Smoliak, J. M. Wallace, M. T. Stoelinga, T. P. Mitchell, Application of partial least squares regression to the diagnosis of year-to-year variations in Pacific Northwest snowpack and Atlantic hurricanes. *Geophys. Res. Lett.* **37** (2010).
11. N. Siler, C. Proistosescu, S. Po-Chedley, Natural Variability Has Slowed the Decline in Western U.S. Snowpack Since the 1980s. *Geophys. Res. Lett.* **46**, 346–355 (2019).
12. F. Pedregosa, *et al.*, Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
13. E. A. Barnes, J. W. Hurrell, I. Ebert-Uphoff, C. Anderson, D. Anderson, Viewing Forced Climate Patterns Through an AI Lens. *Geophys. Res. Lett.* **46**, 13389–13398 (2019).
14. E. A. Barnes, *et al.*, Indicator Patterns of Forced Change Learned by an Artificial Neural Network. *J. Adv. Model. Earth Syst.* **12**, e2020MS002195 (2020).
15. Z. M. Labe, E. A. Barnes, Detecting Climate Signals Using Explainable AI With Single-Forcing Large Ensembles. *J. Adv. Model. Earth Syst.* **13**, e2021MS002464 (2021).
16. R. A. Rohde, Z. Hausfather, The Berkeley Earth Land/Ocean Temperature Record. *Earth Syst. Sci. Data* **12**, 3469–3479 (2020).
17. N. J. L. Lenssen, *et al.*, Improvements in the GISTEMP Uncertainty Model. *J. Geophys. Res. Atmospheres* **124**, 6307–6326 (2019).
18. C. P. Morice, *et al.*, An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set. *J. Geophys. Res. Atmospheres* **126**, e2019JD032361 (2021).
19. K. Cowtan, R. G. Way, Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Q. J. R. Meteorol. Soc.* **140**, 1935–1944 (2014).
20. H. Hersbach, *et al.*, The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **146**, 1999–2049 (2020).
21. P. J. Durack, K. E. Taylor, S. Ames, S. Po-Chedley, C. Mauzey, PCMDI AMIP SST and sea-ice boundary conditions version 1.1.7 (2022) https:/doi.org/10.22033/ESGF/input4MIPs.16485.
22. S. Hirahara, M. Ishii, Y. Fukuda, Centennial-Scale Sea Surface Temperature Analysis and Its Uncertainty. *J. Clim.* **27**, 57–75 (2014).
23. B. Huang, *et al.*, Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *J. Clim.* **30**, 8179–8205 (2017).
24. J. J. Kennedy, N. A. Rayner, C. P. Atkinson, R. E. Killick, An Ensemble Data Set of Sea Surface Temperature Change From 1850: The Met Office Hadley Centre HadSST.4.0.0.0 Data Set. *J. Geophys. Res. Atmospheres* **124**, 7719–7763 (2019).
25. N. A. Rayner, *et al.*, Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res. Atmospheres* **108**

(2003).
26. C. A. Mears, F. J. Wentz, Sensitivity of Satellite-Derived Tropospheric Temperature Trends to the Diurnal Cycle Adjustment. *J. Clim.* **29**, 3629–3646 (2016).
27. R. W. Spencer, J. R. Christy, W. D. Braswell, UAH Version 6 global satellite temperature products: Methodology and results. *Asia-Pac. J. Atmospheric Sci.* **53**, 121–130 (2017).
28. S. Po-Chedley, T. J. Thorsen, Q. Fu, Removing Diurnal Cycle Contamination in Satellite-Derived Tropospheric Temperatures: Understanding Tropical Tropospheric Trend Discrepancies. *J. Clim.* **28**, 2274–2290 (2015).
29. C.-Z. Zou, W. Wang, Intersatellite calibration of AMSU-A observations for weather and climate applications. *J. Geophys. Res. Atmospheres* **116** (2011).
30. P. M. Cox, C. Huntingford, M. S. Williamson, Emergent constraint on equilibrium climate sensitivity from global temperature variability. *Nature* **553**, 319–322 (2018).
31. N. Siler, S. Po-Chedley, C. S. Bretherton, Variability in modeled cloud feedback tied to differences in the climatological spatial pattern of clouds. *Clim. Dyn.* **50**, 1209–1220 (2018).