SUPPLEMENT TO

Efficient Estimation of Climate State and Its Uncertainty Using Kalman Filtering and Application to Policy Thresholds

J. Matthew Nicklas,[a] Baylor Fox-Kemper,[a] Charles Lawrence.[a]

[a] *Brown University, Providence, Rhode Island.*

Corresponding author: J. Matthew Nicklas, john_nicklas@brown.edu

Section A: RTS Smoother

$$\hat{\hat{K}}_n = \mathrm{P}_n \Phi_n (\mathrm{P}_{n|n-1})^{-1} \quad \text{back-updated Kalman gain} \quad \text{(SA1)}$$

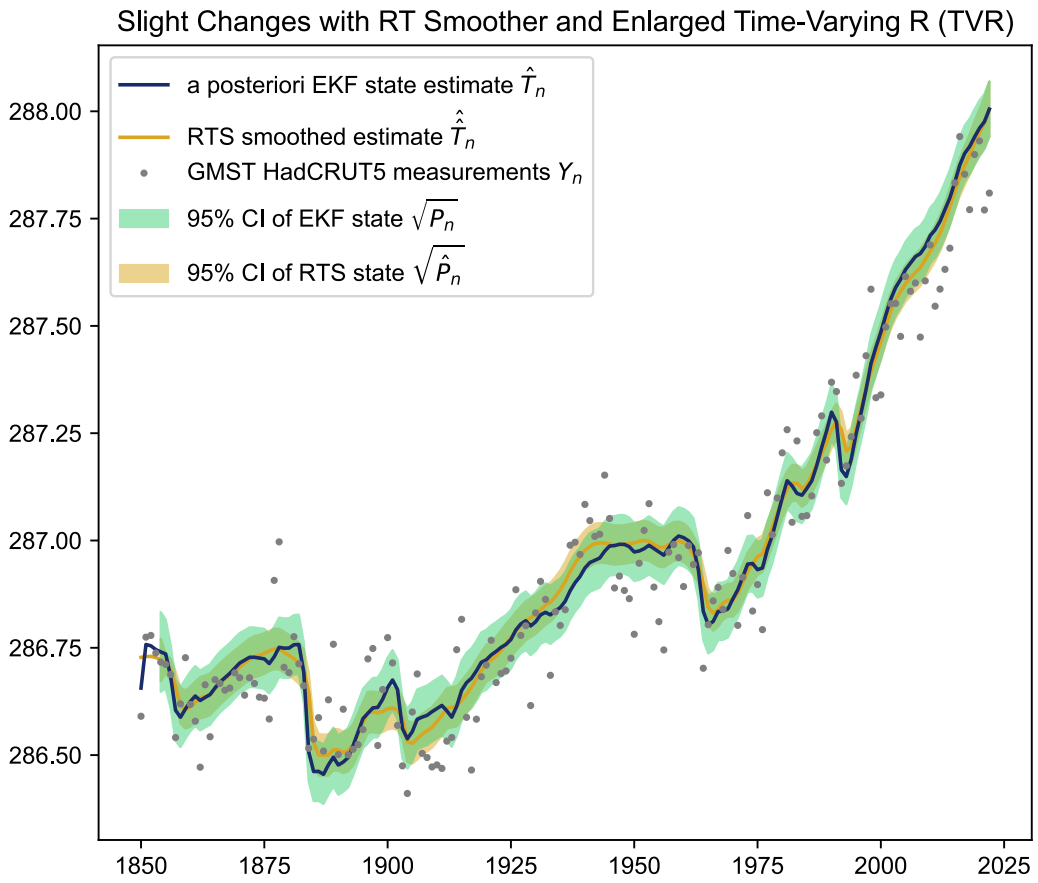
$$\hat{\hat{x}}_n = \hat{x}_n + \hat{\hat{K}}_n \left(\hat{\hat{\mathrm{x}}}_{\mathrm{n}} - \mathbf{F}(\hat{x}_n; u_{n+1}) \right) \quad \text{back-updated state estimate} \quad \text{(SA2)}$$

$$\hat{\hat{P}}_n = \mathrm{P}_n + \hat{\hat{K}}_n (\hat{\hat{P}}_{n+1} - \mathrm{P}_{n|n-1}) \hat{\hat{K}}_n^{\mathrm{T}} \quad \text{back-updated state covariance} \quad \text{(SA3)}$$

This RTS has a theoretical advantage of blending abrupt changes in the model state over greater time periods, while also slightly reducing the state covariance. For instance, if the measurements suddenly and persistently diverged from the blind, forward EBM (unrelated to a known volcanic eruption), an EBM-Kalman Filter model state would only react as these measurements diverge, whereas an EBM-RTS would slightly foreshadow this jump because it can see future as well as past measurements. This occurred in 1943 and 1944: even though the EBM-KF estimated state is trending up, the EBM-RTS state moves cooler to reflect the colder GMST measurements from 1945-1950. Generally, the EBM-RTS just provides a second "nudge" toward measurements. Also, between 1850 and 1860 the intentionally overestimated initial state uncertainty P_0 of 1K is reduced through successive filtering steps in the EBM-Kalman Filter and bi-directional smoothing steps within the EBM-RTS. However, for the purposes of this paper, these distinctions make little difference, as is demonstrated in Supp. Fig. 1 below.

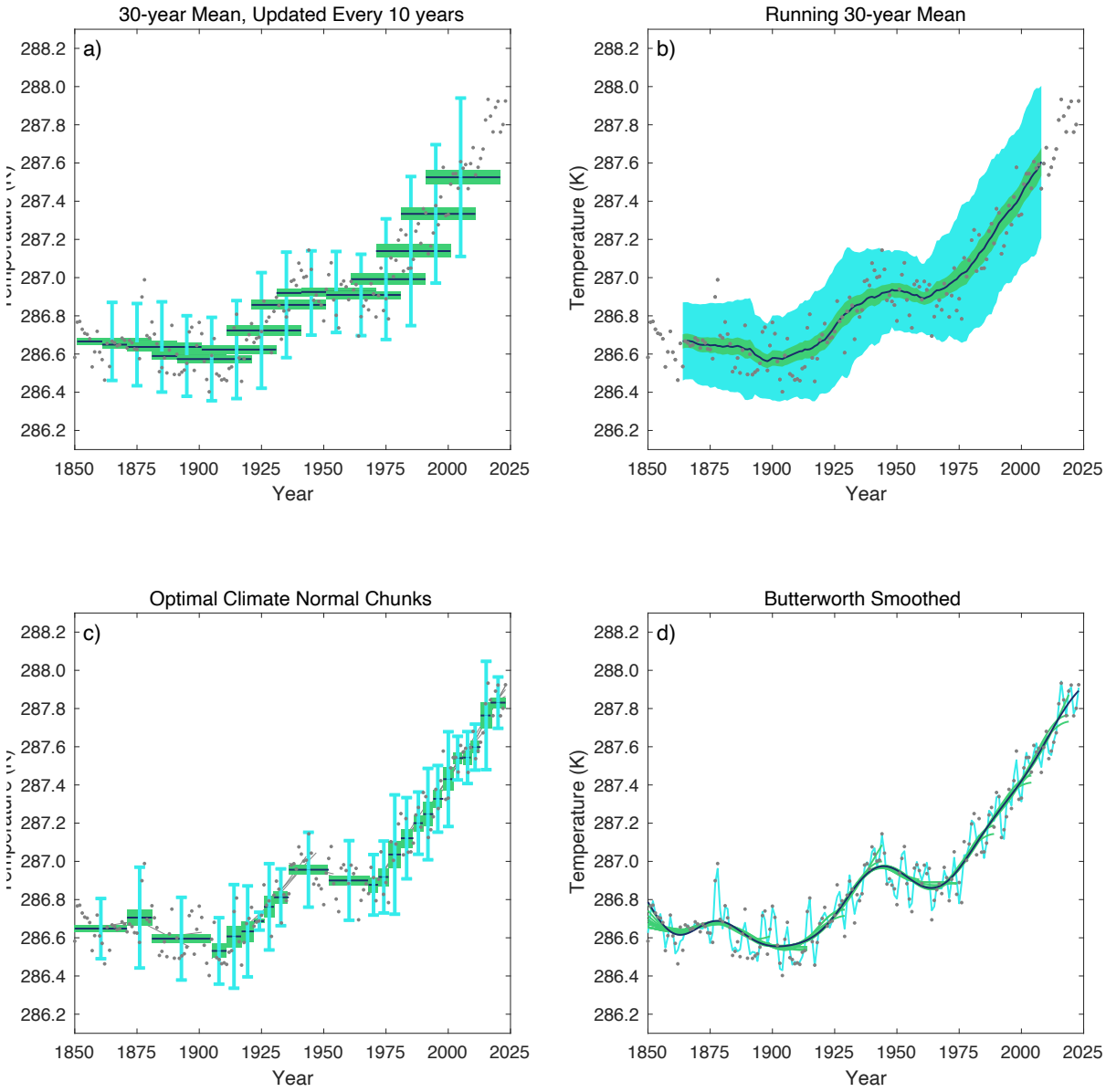
File generated with AMS Word template 2.0

Section B: Miscellaneous Additional Figures



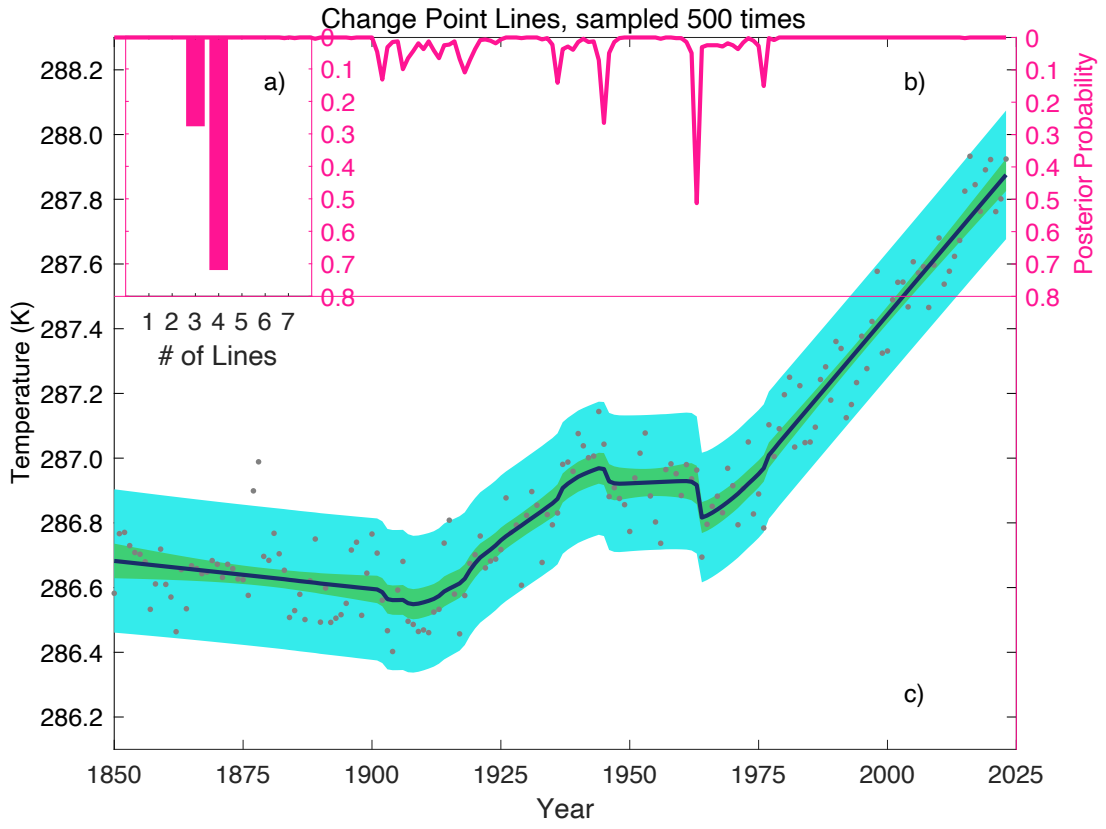
Supp. Fig. 1: Comparisons of the original EBM-Kalman Filtered climate state (navy blue line with green 1σ uncertainty window) with an EBM-RTS climate state (red line with red 1σ uncertainty window) and the effects of incorporating additional time-varying measurement uncertainty (green line with light blue 1σ uncertainty window). The addition of extra time-varying measurement uncertainty makes very little difference to the EBM-Kalman Filtered climate state, except from 1905-1930 when it lessens the deflection of repeated cooler GMST temperature measurements. In contrast, the EBM-RTS climate state doubly takes these annual temperature measurements into account, so it has a greater cooling deflection in this period, and many years that are warmer than the EBM-Kalman Filtered climate state after 1980, although even these differences are slight - at most 0.1K during years of volcanic activity. However, there is greater certainty in the state estimate with the EBM-RTS: $\hat{\hat{P}}_n$ shrinks relative to P_n (see Supp. Fig. 10) by factors of 2.25 and 2.84 for the GMST and OHCA components respectively. The off-diagonal heat-transfer uncertainty component of $\hat{\hat{P}}_n$ is negative and 29 times smaller than those of P_n.

File generated with AMS Word template 2.0



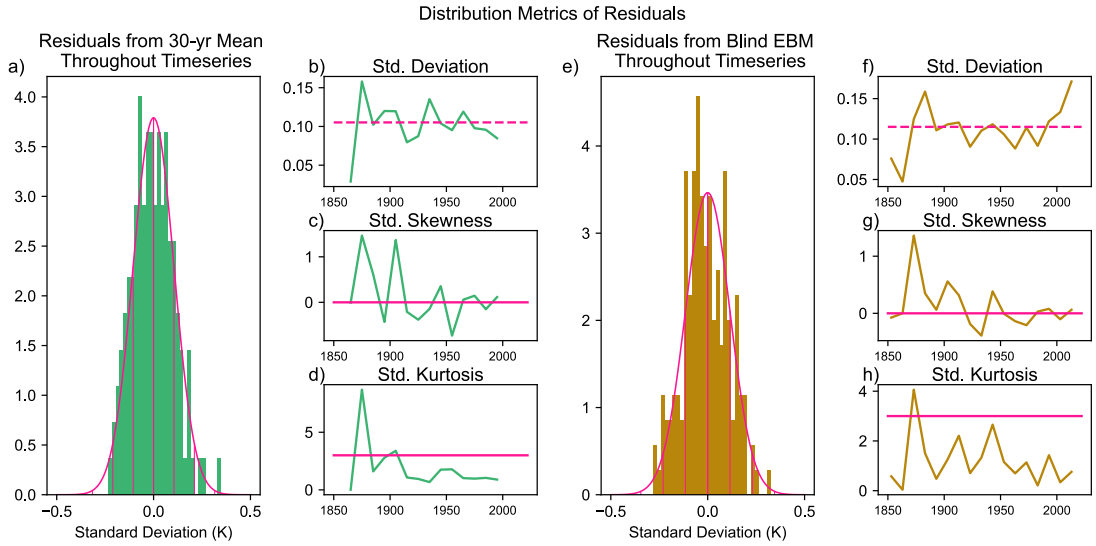
Supp. Fig. 2: Comparison of Prior Methods for Filtering or Smoothing the Climate as applied to the HadCRUT5 temperature dataset. (Morice, Kennedy et al. 2021) All metrics analogous to standard deviation are plotted at the 2σ level in light blue, and all metrics analogous to the standard error are plotted at the 1σ level in light green. a) The 30-year climate normals, updated every 10 years as per the World Meteorological Association in 1935. b) A running 30-year average. c) Adaptive periods of multiyear averages, known as the optimal climate normal (OCN). (Livezey, Vinnikov et al. 2007). Chunks became smaller as the rate of climate change increased in recent decades. d) The Butterworth Smoother applied to this temperature dataset. (Mann 2008) For the "standard error" highly smoothed lines, the lowpass adaptive, lowpass mean padded, and lowpass methods were applied to chunks of the timeseries data ranging from 50 to 170 years in increments of 15 years with a cutoff frequency of 1/30years. The black "best" line a lowpass adaptive curve extended to 2021. The blue "standard deviation" line is a lowpass mean padded filter with a cutoff frequency of 1/5years.

File generated with AMS Word template 2.0

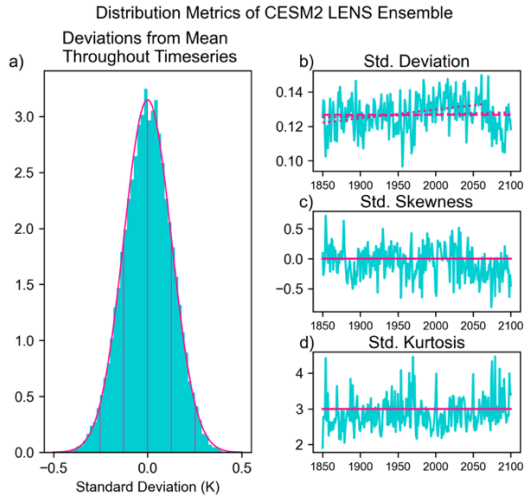


Supp. Fig. 3: Utilization of Bayesian Change Point on the HadCRUT5 data. (Ruggieri and Antonellis 2016) a) There are likely 4 trendlines with 72% of the posterior probability, and the remaining posterior probability on 3 trendlines. b) The posterior probability plot of where trendlines are most likely to occur: 51.2% of all samplings have a change point occur in 1963, and 26.4% of samplings have a change point occur in 1945. c) The posterior distribution of the trendlines in GMST, again with blue shading to indicate 2σ confidence interval of the data and green shading to indicate 2σ confidence interval of the mean trendline. These trend lines do not have to be continuous (note the dip at 1963), but over many samplings the average trend is smoothed.

File generated with AMS Word template 2.0

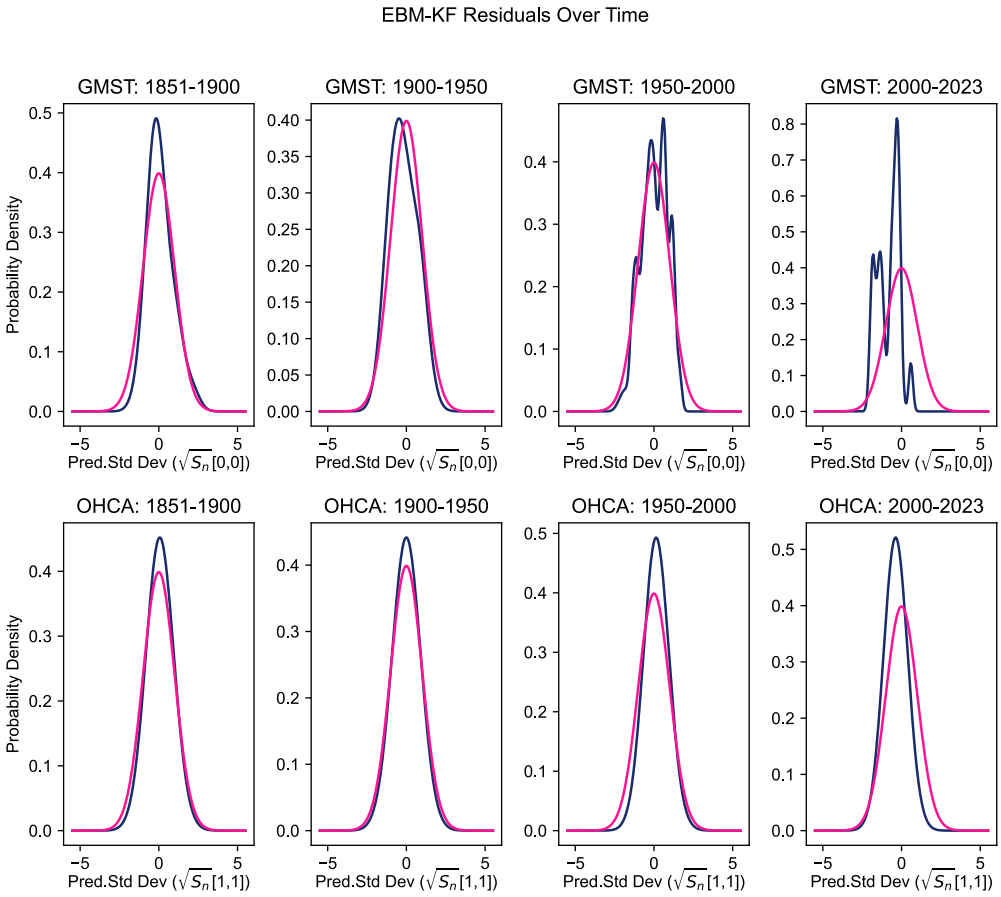


Supp. Fig. 4: Left panels show statistical features of the residuals between the HadCRUT5 measurements with respect to their 30-year running mean, which have a bias of -0.00339K. Pink lines in the histogram in (a) depict an ideal Gaussian distribution with standard deviation of 0.105K, and vertical lines drawn for each of these standard deviations. The dashed pink line (b) indicates the overall standard deviation. Solid pink lines for the skewness = 0.147 (c) and kurtosis = 1.904 (d) indicate the ideal values for a Gaussian distribution. Right panels show statistical features of the differences between the HadCRUT5 measurements with respect to the blind EBM, which have a bias of -0.00104K. Pink lines in the histogram in (e) depict an ideal Gaussian distribution with standard deviation of 0.115K, and vertical lines drawn for each of these standard deviations. The dashed pink line (f) indicates the overall standard deviation. Solid pink lines for the skewness = 0.123 (g) and kurtosis = 1.208 (h) indicate the ideal values for a Gaussian distribution.



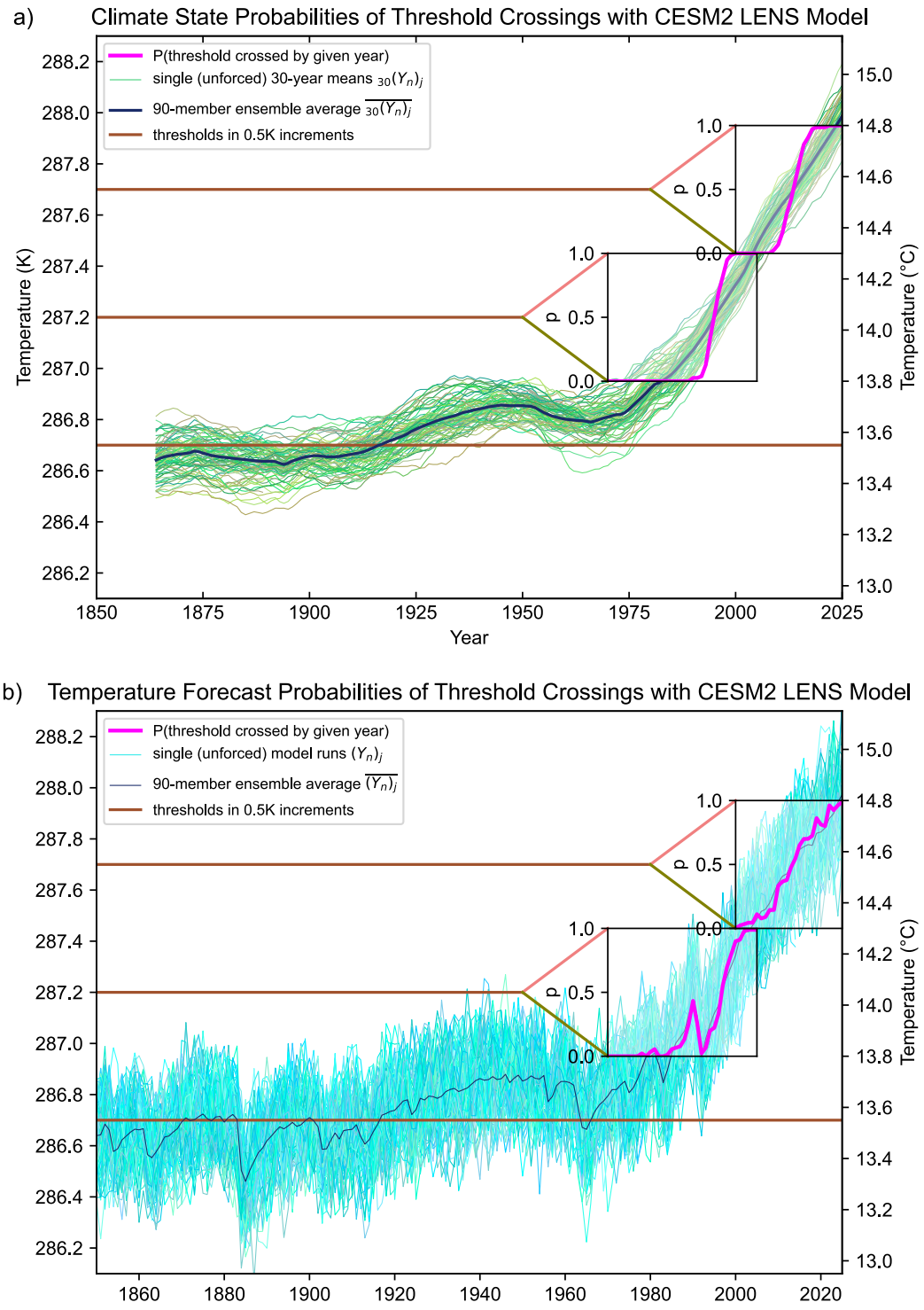
Supp. Fig. 5: Statistical Features of the CESM2 Large Ensemble. (Rodgers, Lee et al. 2021). Pink lines in the histogram in (a) depict an ideal Gaussian distribution with standard deviation of 0.127K, and vertical lines drawn for each of these standard deviations. The observed trend (b) up until 2065 and overall in the standard deviation over time is plotted in a dotted pink line, while the dashed line indicates the overall standard deviation of 0.127K. Solid pink lines for the skewness = -0.067 (c) and kurtosis = 2.871 (d) indicate the ideal values for a Gaussian distribution.

File generated with AMS Word template 2.0



Supp. Fig. 6: Deviation between the projected climate state (pink) and empirical PDFs of the Gaussian mixture of measurements with associated uncertainty (purple), plotted relative to the ideal distribution given by the innovation covariance. Each column indicates a different time window of the EMB-KF model's run length. The top row displays the empirical PDFs of the GMST HadCRUT measurements relative to the model's estimate of GMST state, whereas the bottom row displays empirical PDFs of the OHCA Zanna 2019 measurements relative to the model's estimate of OHCA state. Note the initial period begins at 1851 (and the 1850 measurement is excluded from main text Fig. 3 and 4) because this has comparison involves P_0, which was intentionally over-estimated (resulting in relatively too-narrow measurement kernel). Also note that the last period is less than half the time of the others, so the GMST empirical distribution is much more choppy.

File generated with AMS Word template 2.0

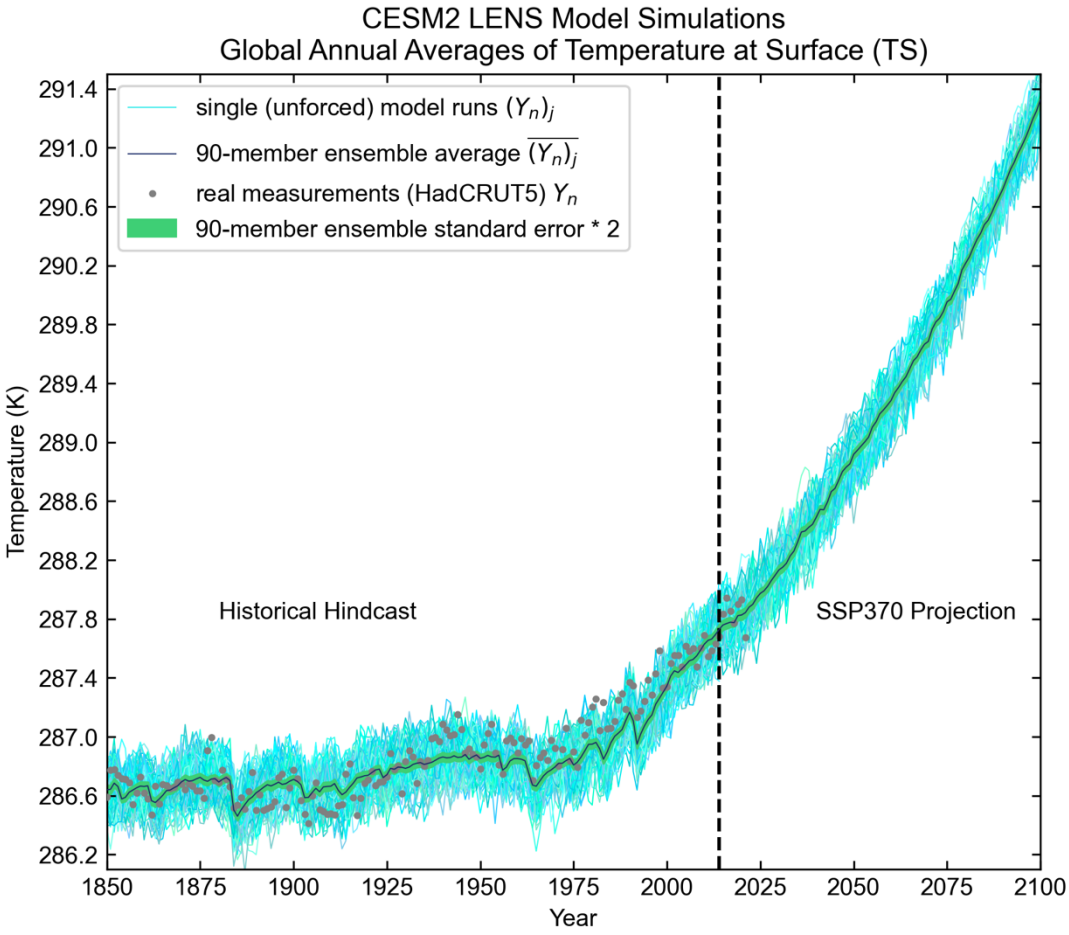


Supp. Fig. 7: As in Fig. 5 within the main text, except the threshold crossing calculations are performed on the LENS2 ensemble. a) climate state policy thresholds: the 21-year running means of individual simulations in light green lines, the two inset boxes indicate threshold crossing probability, given by the fraction of these light green lines that have crossed the indicated threshold. b) temperature forecast policy thresholds, showing a cloud of the possible next-year measurements in light blue from the simulations, and again the two inset boxes indicate the fraction of these light blue lines that have crossed the threshold.

File generated with AMS Word template 2.0

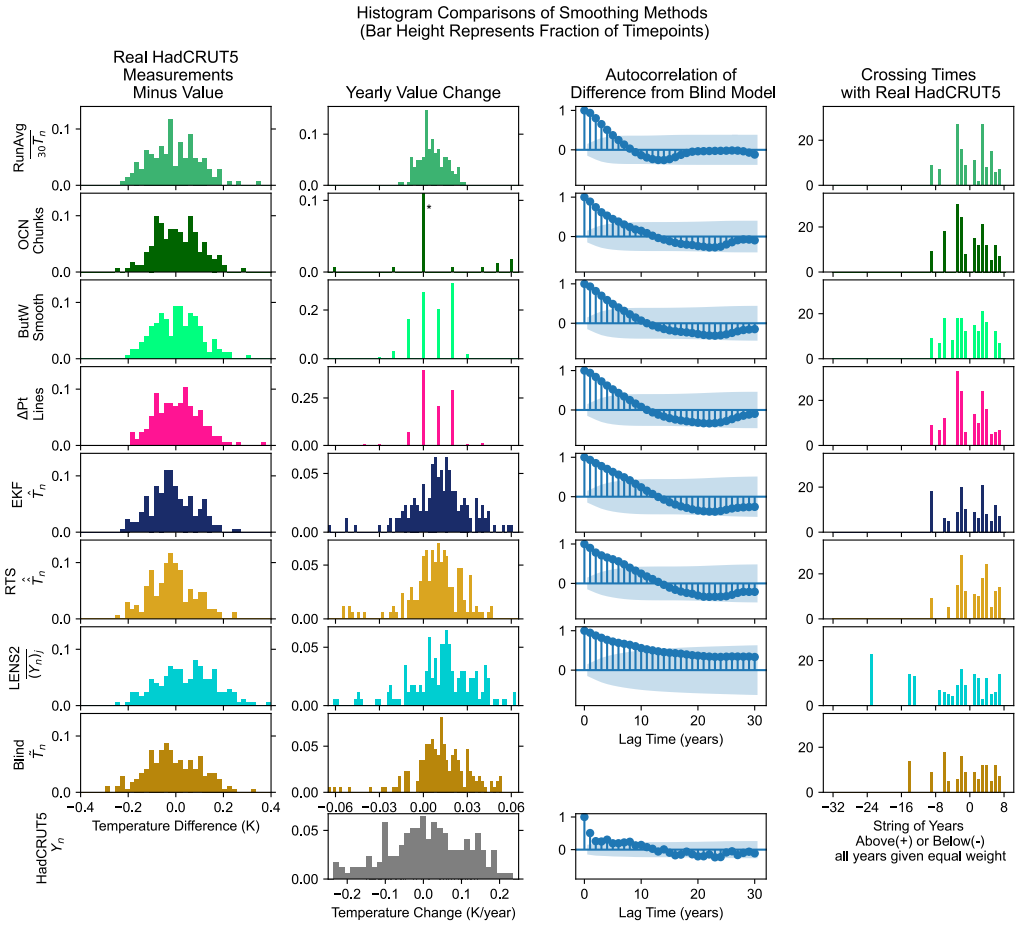
120

File generated with AMS Word template 2.0



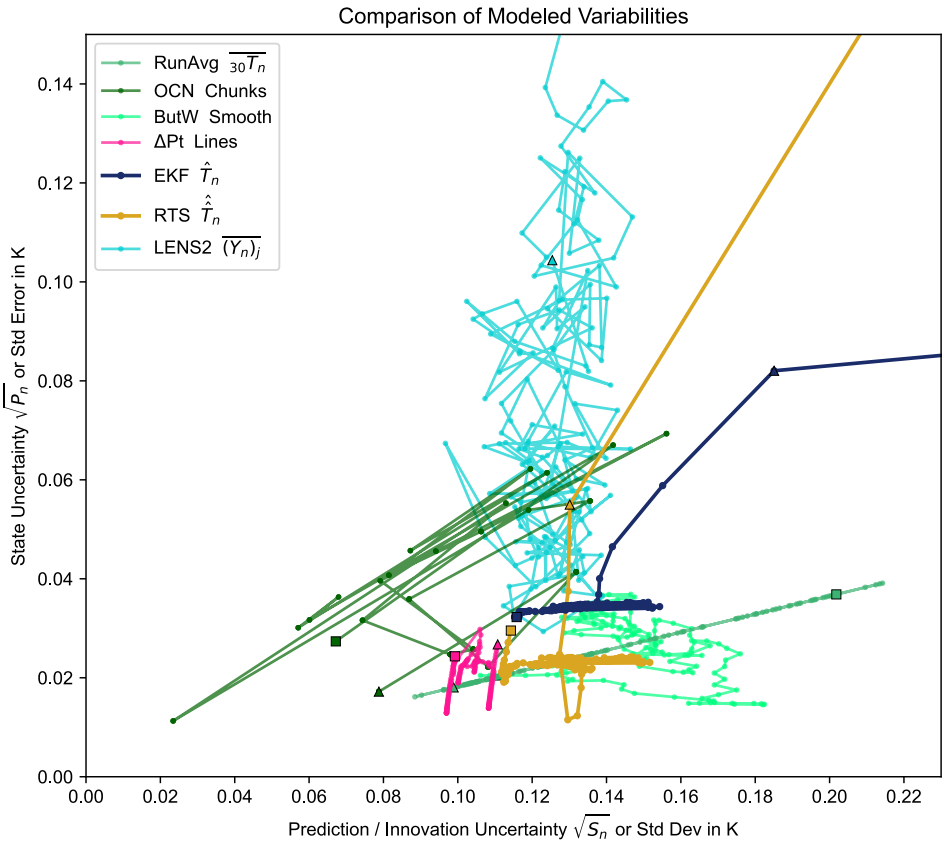
Supp. Fig. 8: Comparison of the CESM2 Large Ensemble (LENS2) GSAT (Rodgers, Lee et al. 2021) with HadCRUT5 GMST measurements. The various shades of thin light blue and turquoise lines represent each individual simulation $(Y_n)_j$ of the 90-member ensemble. The ensemble mean is plotted in a navy-blue line, and the ensemble mean standard error is plotted around this line in green. This standard error is twice the standard deviation divided by the square root of the number of ensemble members at that moment and shows the 2σ uncertainty in the yearly simulated climate is roughly 0.026K. The ensemble mean has $R^2 = 0.83$ relative to the HadCRUT5 measurements, lower than for the blind EBM (R^2=0.88). The dashed vertical line represents when LENS transitions from historical to future forcing (SSP3-7.0).

File generated with AMS Word template 2.0



Supp. Fig. 9: Histogram comparisons of several aspects of many of the smoothing methods for generating a climate timeseries. The far-left column represents the absolute differences between the HadCRUT5 measurements and all the other models. All look similar in this respect. The center-left column shows the annual changes in the temperatures reported by each model. In this respect, the real HadCRUT5 measurements are the most spread out, because the stochastic change each year is large, whereas the in most years the OCN Chunks do not change. The center-right column shows an autocorrelation plot, which demonstrates that every other model aside from HadCRUT5 (and to a lesser extent the running average) are autocorrelated with the blind energy-balance model to similar degrees. The far-right column shows how many continuous years are spent above or below HadCRUT5: both the LENS2 ensemble average and the blind energy-balance model had >20 year spans for which they were colder than the “real” HadCRUT5 data, illustrating the benefit of data assimilation.

File generated with AMS Word template 2.0



Supp. Fig. 10: Comparisons of the state and prediction (or equivalent) uncertainties of the smoothing methods for generating a climate timeseries. The x-axis represents the state uncertainty (colored light green in all other figures), and the y-axis represents the prediction uncertainty (colored light blue and doubled in all other figures). As these quantities change over time, all points in these smoothing timeseries are connected with colored lines, with the triangle Δ representing the value of these quantities in 1850 or the first point that they entered the frame limits of this graph, and the square □ representing the value of these quantities in 2021 or the last point that they were within the frame limits. For instance, the running average draws a straight line because standard deviation and standard error are linearly correlated by a favor of $1/\sqrt{30}$, and latter points have larger quantities for each variability due to the changing climate. The Butterworth Smoother traces a curve roughly in this region, with both the standard deviations and standard errors being twice the 15-year running average of the maximum of the absolute value of differences between colored and black curve. The RTS and EKF methods rapidly converge to a state uncertainty of ~0.110K and ~0.03K. The Change Point Regression variance also fluctuate in this region, although this methods' standard error twice drops to 0.014K. Both the OCN and the LENS2 climates have standard errors that are above the other methods at most times. For LENS2, the standard deviation within the CESM2 ensemble generally remains between 0.11K and 0.14K, whereas the state uncertainty is taken to be the standard deviation of the 20 ensembles comprising CMIP6 in October 2021. (Meehl, Moss et al. 2014) These metrics have nothing to do with Figure 10 in the main text. Within CMIP6, the 20 ensembles are most in agreement in 1939, when the state uncertainty dipped down to only 0.029K between ensemble means, but this uncertainty was much greater at earlier and later time points, reaching 0.183K by 2014.

File generated with AMS Word template 2.0

Section C: Justification that the EKF is sufficient, will not diverge

The issue of nonlinearity arises not in the computation of $\hat{x}_{n|n-1}=F(\hat{x}_{n-1})$ but rather the covariance distribution P_n of points (infinitesimal probability masses) neighboring $\hat{x}_{n-1}$, which are assumed to scale linearly around this transformation to maintain a normal distribution. The OHCA part of the model can be ignored since it is purely linear. Nonlinear distortion may pile more probability density onto a state other than the transformed original projection $F(\hat{x}_{n-1})$, necessitating a new computation of $\hat{x}_{n|n-1}$ as the mean of this distorted PDF. Thus, for an arbitrary point that is z standard deviations away from $\hat{x}_{n-1}$, the remainder error R_1 (Lagrange mean-value form) induced in a single cycle is:

$$F\left(\hat{x}_{n-1}+z\sqrt{P_n};u_n\right) - F(\hat{x}_{n-1}) - \frac{\partial F(x;u_n)}{\partial x} z\sqrt{P_n} =$$

$$R_1\left(\hat{x}_{n-1}+z\sqrt{P_n}\right)=\frac{\partial^2 F(\xi_L;u_n)}{\partial \xi_L^2}\frac{\left(z\sqrt{P_n}\right)^2}{2} \quad \text{for} \quad \xi_L \in \left[\hat{x}_{n-1}-|z|\sqrt{P_n},\hat{x}_{n-1}+|z|\sqrt{P_n}\right] \tag{SC1}$$

$$=\left(\frac{0.441}{AOD_n+9.73}(0.00159)-(0.00005546)(1-0.0655\log_{10}([CO_2]_n))1.385\left(\xi_L\right)^{0.385}\right)\frac{z^2P_n}{2} \tag{SC2}$$

$$-0.5(10^{-5})\, z^2P_n<R_1\left(\hat{x}_{n-1}+z\sqrt{P_n}\right)< 0.5\,(10^{-5})\, z^2P_n \tag{SC3}$$

$$\left|R_1\left(\hat{x}_{n-1}+z\sqrt{P_n}\right)\right|<\ 10^{-5}\, z^2\, 0.5\,(0.032)^2<|z|*5*10^{-9} \tag{SC4}$$

This means that all probability masses that are within $|z|< 20$ standard deviations will have an one-step error of <0.000002K. Even if the error accumulates in the same direction in each cycle of the EKF, over the 173 year timeseries, the error will be within 0.0004K compared to a particle method such as the Unscented Kalman Filter. (Julier and Uhlmann 1997; Wan and Van Der Merwe 2000)

File generated with AMS Word template 2.0