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

J. MATTHEW NICKLAS,[a] BAYLOR FOX-KEMPER,[a] AND CHARLES LAWRENCE[a]

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

(Manuscript received 22 September 2023, in final form 4 September 2024, accepted 1 October 2024)

ABSTRACT: We present the energy-balance model–Kalman filter (EBM–KF), a hybrid model projecting and assimilating the global mean surface temperature (GMST) and ocean heat content anomaly (OHCA). It combines an annual energy-balance model (difference equations) with 17 parameters drawn from the literature and a statistical extended Kalman filter assimilating GMST and OHCA, either observed time series or simulated by Earth system models. Our motivation is to create an efficient and natural estimator of the climate state and its uncertainty, which we believe to be Gaussian at a global scale. We illustrate four applications: 1) EBM–KF generates a similar estimate to the 30-yr time-averaged climate state 15 years sooner, or a model-simulated hindcasts' annual ensemble average, depending on whether volcanic forcing is filtered or not; 2) EBM–KF conveniently assesses annually likelihoods of crossing a policy threshold. For example, based on temperature records up to the end of 2023, $p = 0.0017$ indicates that the climate state was 1.5°C over preindustrial, but there is a 16% likelihood that the GMST in 2023 itself could have been over that threshold; 3) a variant of the EBM–KF also approximates the spread of an entire climate model large ensemble using only one or a few ensemble members; and 4) all variants of the EBM–KF are sufficiently fast to allow thorough sampling from non-Gaussian probabilistic futures, e.g., the impact of rare but significant volcanic eruptions. This sampling with the EBM–KF better determines how future volcanism may affect when policy thresholds will be crossed and what an ensemble with thousands of members exploring future intermittent volcanism reveals.

SIGNIFICANCE STATEMENT: The global average of Earth's historical climate over the past 150 years can be explained by a thermal/radiation physical equation involving a small number of constants, atmospheric carbon dioxide (CO_2) concentration, anthropogenic clouds, and volcanic emissions. Global mean surface temperature measurements vary around this climate state within a consistent normal distribution. This physical equation and statistical depiction allowed us to construct a simple model that can rapidly estimate the uncertainty in Earth's current climate, aid in policy discussions, and provide an alternative for some applications to expensive ensemble modeling.

KEYWORDS: Kalman filters; Time series; Uncertainty; Climate models; Ensembles; Interannual variability

1. Introduction

What is the uncertainty in Earth's climate? From a measurement standpoint, this issue was resolved many decades ago. The instantaneous measurement of global mean surface temperature (GMST) is currently performed with an average accuracy of 0.05°C (maximum 0.10°C) via arrays of infrared-sensing satellites and ground stations (Susskind et al. 2019). Both satellite and ground datasets extend back to 1981 (Merchant et al. 2019), and the yearly seasonal fluctuation is easy to smooth with a running annual average. However, this GMST still has significant dynamical and random stochasticity, from processes like the 2–7-yr quasi-periodic El Niño events (Hu and Fedorov 2017) and volcanic eruptions that intermittently affect climate for 1–2 years (Soden et al. 2002). Measurement errors also arise from sparse or inconsistently calibrated historical data and paleoproxies (Carré et al. 2012; PAGES2k Consortium 2017; Kaufman et al. 2020; McClelland et al. 2021). Internal variability dominates over climate-forced variability in most short-term signals, in both climate simulations and reality (Gulev et al. 2021; Kirtman et al. 2013; Lee et al. 2021; Marotzke and Forster 2015). By "simulations," we refer to computationally expensive global coupled models (and occasionally to numerical weather model predictions). Other climate variables reveal warming that is steadier than GMST (less "noisy" annual variability). One such steady climate variable is the ocean heat content anomaly (OHCA), where >90% of the anthropogenic energy anomaly is found (Cheng et al. 2017, 2022; Fox-Kemper et al. 2021; Gulev et al. 2021). Even pandemic-level reductions in global carbon dioxide (CO_2) emissions may not show an identifiable impact on GMST over a time scale of a few years (Szopa et al. 2021), posing a challenge for policy and assessment.

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-23-0580.s1.

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

DOI: 10.1175/JCLI-D-23-0580.1

In 1935, the International Meteorological Committee, which later became the World Meteorological Organization, began reporting the "standard climate normal" as surface temperature averages over an interval of 30 years $\overline{{}_{30}Y_t}$ (in this paper's notation). A 30-yr window was chosen to minimize most internal fluctuations (such as El Niño) and short-term forcings such as single volcanoes (Guttman 1989); the effect is similar to examining less noisy metrics of the climate system such as the OHCA. Figure 1 shows this metric and emphasizes the 30-yr span over which the average is taken. To generate continuous estimates of the climate, this 30-yr average can be updated annually, forming a running mean (Fig. 4b in the online supplemental material). While standard climate normals and running means are straightforward and widely accepted definitions of climate, they involve lag: The most current 30-yr unweighted average describes the average climate state of Earth over a window centered on 15 years ago. Weighted moving average can shift the center of this window closer toward the current year, but some lag always remains. A trailing average is a similar concept that will be discussed below. Moreover, anthropogenic climate change distorts standard statistical metrics: Most of the variance in recent 30-yr periods derives from the trend rather than internal variability (Fig. 1). Averaging filters (such as a running mean) remove high-frequency signals that reflect year-to-year variations in global weather, as do other statistical approaches better suited to removing frequencies above a particular cutoff (Smith 2003). For brevity, we will use the term "weather" in this paper to encompass all forms of subannual variability that contribute to the distribution of annual-mean properties, although this certainly includes much subseasonal-to-seasonal, coupled atmosphere–ocean modes, and other variability elsewhere described as "climate variability." The anthropogenic change in Fig. 1 is gradual enough to be mostly preserved by moving averages (running mean) or any low-pass filter/smoother, but this is not true in general: In a hypothetical (or extraterrestrial) climate where forcings undergo an impulse change, such as a quadrupling of CO_2 within 1 year as used to evaluate models in the Coupled Model Intercomparison Project (CMIP), the 30-yr running mean is an inadequate climate state indicator (see supplemental section B and supplemental Fig. 3). Other example applications to Earth's recent GMST of statistical, as opposed to physical, filters used in climate analysis are shown in supplemental section B (supplemental Figs. 4c,d and 5).

To directly fit the physical effect relating forcings to the climate (incorporating relaxation time), the multipattern fingerprint method was developed (Hasselmann 1997), leading to "attributable anthropogenic warming" (Otto et al. 2015) and a "real-time Global Warming Index" (Haustein et al. 2017). This methodology is statistically conservative, generating a wide 5%–95% confidence interval spanning ±0.1°C from 1980 to 2010 and a less certain 5%–95% confidence interval (CI) of ±0.15°C by 2017. Fingerprinting and basis function methods are examples of spatial filtering methods to reduce uncorrelated noise, and taking an ensemble mean is a filtering method generally thought to reduce internal variability. Filtering methods can be closely aligned with physical or dynamical rationales (as is the case where well-motivated basis functions are used) or purely statistical in nature (e.g., running means of time series). Another class of methods combines statistical and physical filters, by recognizing which physical patterns (e.g., phase of ENSO or Atlantic/Pacific variability mode) correspond to short-term deviations from the running mean (Foster and Rahmstorf 2011; Wills et al. 2020; Chen and Tung 2018; Sippel et al. 2019; Samset et al. 2022).

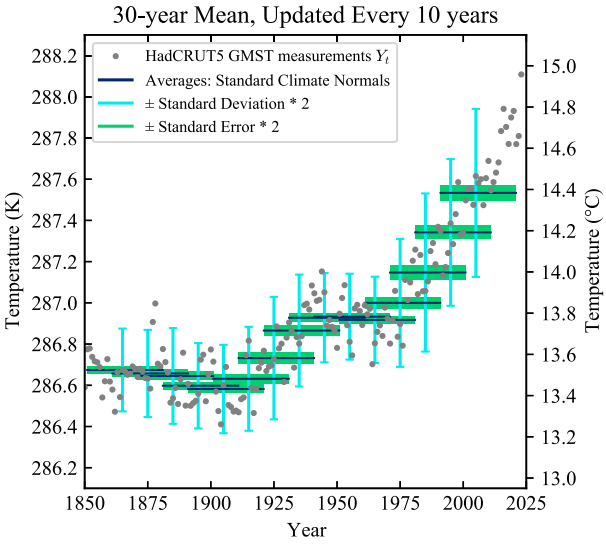

FIG. 1. Illustration of standard climate normals $\overline{{}_{30}Y_t}$ (blue horizontal lines in 10-yr overlapping bins) as applied to the HadCRUT5 GMST dataset (gray dots) (Morice et al. 2021). Twice the population standard deviation (cyan error bars) and two standard errors (green rectangles) are plotted. Note standard deviations widen due to the anthropogenic trend, and the last standard climate normal is cooler than recent GMST measurements.

Policy goals are often framed via climate change staying below a particular policy threshold (e.g., 1.5° or 2°C above preindustrial conditions as in the Paris Agreement). Using a 30-yr mean brings difficulty in determining exactly when or if a policy threshold is crossed (Lee et al. 2021). Policy thresholds are not system thresholds—temperature "tipping" points when the dynamics of the climate system are reorganized, often occurring abruptly or irreversibly—and so they are subject to vary under political choices. Relatedly, magnitudes and uncertainty ranges are meaningful only under specific averaging windows, e.g., "GMST increased by 0.85° (0.69°–0.95°C) between 1850–1900 and 1995–2014 and by 1.09° (0.95°–1.20°C) between 1850–1900 and 2011–20" (Gulev et al. 2021). Tools for assessing when a policy threshold has been crossed will be useful as future policy targets approach.

We use both $\mu \pm 2\sigma$ and $\mu(a–b)$ notation to refer to 95%CI, in contrast to $[a–b]$ notation with which we refer to finite or closed ranges. In this notation, μ is a point estimate, σ is a standard deviation, a is the minimum of the interval or range, and b is the maximum. Throughout, a 2σ or approximately 95% confidence interval is used, indicating the

extremely likely range in Intergovernmental Panel on Climate Change (IPCC) terminology.

As an alternative to the 30-yr running mean and to overcome limited observations sampling the real world, many climate studies instead investigate the climate system within globally coupled climate simulations ("coupled" refers to interacting submodels, such as atmosphere/ocean/land/ice components), also known as Earth system models (ESMs; e.g., Meehl et al. 2014). Typically, these simulations are forced using historical records and a range of scenarios for future projections including CO_2 emissions, other pollutants, land use, and volcanic eruptions (Lee et al. 2021). The chaotic nature of weather and varying initial conditions produce an ensemble of identically forced simulations that explore the span of outcomes consistent with forcing, such as for the CESM2 large ensemble (LENS2; Rodgers et al. 2021, supplemental Fig. 6). Unfortunately, each coupled ensemble member simulation is computationally expensive and deterministic, so one member does not accurately or transparently reflect the changing climate statistically, but only one realization of it including model errors. Combining such ensembles with real observations yields improvements, such as a more realistic possible spread (due to internal variability) of winter temperatures in North America from 1966 to 2015 (McKinnon et al. 2017). Betts et al. (2023) proposed avoiding the lag in climate state estimation by combining 10 years of previous observations with a subsequent 10 years forecasted by several ESMs, an approach named the "current global warming level." While useful, this technique oversimplifies some issues inherent to ESMs, such as whether some predictions should be weighted over others (Lehner et al. 2020; Sherwood et al. 2020) or how an ensemble of near-term projections should be initialized (e.g., Yeager et al. 2022).

We sought an efficient and natural estimator of the climate state and its uncertainty: the energy-balance model–Kalman filter (EBM–KF). We combined a nonlinear energy-balance difference equation (EBM) and a statistical observation equation (KF) that brings in the available measured GMST and OHCA data, yielding a hybrid physical model—statistical filter. This data-driven climate emulator (Forster et al. 2021) is vastly more computationally efficient than ensembles of ESMs that provide similar information about GMST and OHCA. Our emulator is interpretable as a global energy budget (and assimilates OHCA as well as GMST), benefits from the mathematical similarities between an energy-balance model and a Kalman filter, and allows access to proven methodologies for parameter estimation (Chen et al. 2018; Zhang and Atia 2020) and uncertainty quantification (Sætrom and Omre 2013). Our simple EBM has good skill at predicting the GMST and OHCA despite being by itself "blind" to all measurements (i.e., it is a "forward" model in numerical weather prediction terminology). The statistical component is an extended Kalman filter, which allows for the incorporation of current measurements to "course correct" under a well-understood mathematical framework, with time-varying weather and "climate state" uncertainty. While perhaps unconventional in data assimilation, this statistical climate state projection approach is directly analogous to the inference of some of our parameters: A handful of numbers were abstracted from the historical climate record using established statistical methods. Hybridizing the EBM with the extended Kalman filter yields statistical distributions of internal variability and a physical rationale for the filtered current climate state.

First, the EBM–KF is introduced within section 2 in phases: the EBM in section 2a and the structure of the extended Kalman filter in section 2b. An elaboration beyond fixed assumed measurement uncertainty is detailed in section 2c. The scope of EBM–KF is expanded to future projections including volcanic eruptions in section 2d. In section 3, variants of the EBM–KF are illustrated on four applications to historical and future climate. Section 3a shows that it estimates the 30-yr mean climate normal every year, including the latest observations and without lag. Section 3b shows how it can be used to assess the probability that a policy threshold has been crossed in any particular year. Section 3c shows how it can be used to estimate the ensemble mean of an ESM large ensemble from only one ensemble member. Section 3d shows that the EBM–KF is sufficiently fast to allow high-density sampling of non-Gaussian probabilistic futures, e.g., directly sampling over highly intermittent distributions of future volcanic eruptions. Section 4 discusses these results, some cautionary remarks, opportunities for extension, and application to policymaking. Section 5 concludes. The detailed EBM–KF code is available, and the equations as coded are provided in appendix A, and a glossary of mathematical symbols is provided in appendix C. Extensive appendices and supplemental material convey additional detail.

2. Methods

a. *Energy-balance model*

We constructed the energy-balance model (Fig. 2) by envisioning a uniform planet and capturing the principal atmospheric and surface energy fluxes (Budyko 1969; Sellers 1969). This model is blind with respect to observations and is inspired by other energy-budget models illustrating quantitative skill (Hu and Fedorov 2017; Kravitz et al. 2018; Millar et al. 2017; Smith et al. 2024) at approximating both GMST and the 30-yr running mean. The model includes two idealized layers, with each layer having homogenous temperature: a surface layer including thermally active soil and 86-m average ocean water depth (with temperature approximating GMST) and a deep ocean layer reaching (1141 + 86) m depth that exchanges energy (part of OHCA) with the surface layer (Geoffroy et al. 2013a; Gregory 2000; Held et al. 2010). These depths are chosen to select a two-state system that best represents the heat capacities of spatially complex heat uptake patterns in total into the global oceans (Newsom et al. 2023), rather than representing the heat uptake relative to depths associated with observational oceanographic traditions (e.g., 700 and 2000 m). As this EBM does not directly incorporate any spatial dimensions, it should be considered zero dimensional in the context of other ESMs with spatial gradients. Closely related variables to

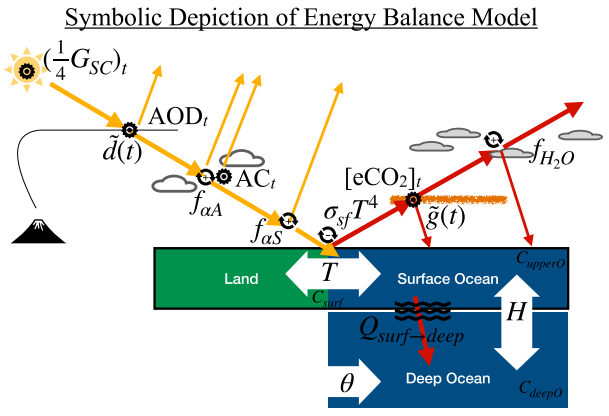

FIG. 2. Diagram of the EBM, with all major forcing functions, corresponding climate driver datasets, and feedback functions represented. All feedback functions are dependent on surface temperature T, but this is not written on the diagram. Shortwave incoming solar radiation is successively fractionated by various forcing and feedback functions (indicated by circling arrows), as is outgoing longwave radiation. These in net warm the surface layer, which in turn warms the deep ocean. Anthropogenic forcing effects are indicated by a small gear (or cog). See (1)–(5).

GMST, such as global surface air temperature (GSAT), differ only from GMST by measurement and slightly in uncertainty (by less than our confidence intervals) but not systematically (Gulev et al. 2021).

The energy budgets for the EBM layers and the energy fluxes are (Fig. 2) as follows:

$$C_{\text{surf}}\frac{dT}{dt} = \mathscr{F}_{\text{SW}}(T,t) - \phi_{\text{LW}}(T,t) - \gamma(T-\theta-\zeta_0), \quad (1)$$

$$C_{\text{deepO}}\frac{d\theta}{dt} = \gamma(T-\theta-\zeta_0), \quad (2)$$

$$H = (T-T_{1850})\,C_{\text{upperO}} + (\theta-\theta_{1850})\,C_{\text{deepO}}, \quad (3)$$

$$\mathscr{F}_{\text{SW}}(T,t) = (\tfrac{1}{4}G_{\text{SC}})_t\,\tilde{d}(t)f_{\alpha A}(T,t)f_{\alpha S}(T), \quad (4)$$

$$\phi_{\text{LW}}(T,t) = \sigma_{sf}T^4\,\tilde{g}(t)f_{\text{H}_2\text{O}}(T), \quad (5)$$

where T is the GMST, whereas θ is the conservative temperature of the deep ocean in that same year, and H is the OHCA including both the deep ocean layer and the surface ocean (McDougall et al. 2021). The time variable t is the calendar year index and is often used as a subscript [e.g., T_{2000} is the modeled GMST in the year 2000, or $(\frac{1}{4}G_{\text{SC}})_t$ is a direct forcing record at index t], or t is used as an input to derived functions [e.g., $\tilde{d}(t)$]. On the right-hand side of the equation, both the shortwave radiative flux $\mathscr{F}_{\text{SW}}(T,t)$ and the longwave radiative flux $\phi_{\text{LW}}(T,t)$ take the same form: [source $\frac{1}{4}G_{\text{SC}}$ or $\sigma_{\text{sf}}T^4$] × [prescribed attenuation from forcing: $\tilde{d}(t)$ or $\tilde{g}(t)$] × [attenuation functions with feedback: $f(T)$ with various subscripts]. The attenuation function of clouds on shortwave radiation $f_{\alpha A}(T,t)$ contains both prescribed forcing and feedback. The overall surface heat capacity C_{surf} is 17 ± 7 W (year) m^{-2} K^{-1}, obtained from modeling/time-series analysis (Schwartz 2007), including 11.7 W (year) m^{-2} K^{-1} or 86 m of upper surface ocean C_{upperO}, while there is a separate deep ocean heat sink with capacity 155.7 W (year) m^{-2} K^{-1} or 1141 m C_{deepO} (Geoffroy et al. 2013a). The term $\frac{1}{4}G_{\text{SC}}$ is the total solar irradiance (TSI) normalized to Earth's surface area at $\frac{1}{4}\overline{G_{\text{SC}}} \approx 340.2\,\text{W m}^{-2}$. We elected to incorporate the record $(\frac{1}{4}G_{\text{SC}})_t$ of [340.06–340.48] from Coddington et al. (2017), but these variations are insignificant. The term $\tilde{d}(t)$ is the shortwave radiation attenuation [with the functional form defined by Schwartz et al. (2002) and Harshvardhan and King (1993) due to aerosol optical depth AOD$_t$ [from Sato et al. (1993), Vernier et al. (2011); NASA/Langley Research Center (LARC)/Science Directorate (SD)/Atmospheric Science Data Center (ASDC) 2018], and $f_{\alpha A}(T,t)$ is the additional atmospheric shortwave attenuation due to cloud albedo incorporating both feedback and anthropogenic cloud-nucleating aerosols AC$_t$, while $f_{\alpha S}(T)$ is the surface shortwave attenuation due to ground albedo. Infrared radiation emitted from the surface is $\sigma_{\text{sf}}T^4$, the ideal Planck blackbody radiation. The term $\tilde{g}(t)$ is the prescribed record of longwave attenuation due to CO_2 and other greenhouse gases combined as effective carbon dioxide concentration $[eCO_2]_t$ (square brackets are used for this forcing record to indicate a concentration in ppm). Lastly, $f_{\text{H}_2\text{O}}(T)$ is the additional atmospheric longwave attenuation due to water vapor parameterized as a function of GMST. For reference, appendix C tabulates all mathematical symbols used in this paper.

In discussion section 4a, we will discuss variants of the EBM–KF, constructed by prefiltering the input forcings. If we prefilter inputs (indicated by EBM–KF–ta, abbreviating "trailing average"), then the output estimated climate state more closely approximates the 30-yr running mean of GMST and OHCA. Without prefiltering (indicated by EBM–KF–uf, abbreviating "unfiltered"), the estimated climate state more closely resembles the ensemble mean of GMST and OHCA across members of a coupled ESM ensemble such as LENS2. Prefiltering is inconsequential for greenhouse gases which evolve slowly but is consequential for aerosol optical depth over the impulse changes during volcanic eruptions. All equations within the energy-balance model (section 2a) and the Kalman filter (section 2b) are used regardless of prefiltering. Thus, we will refer to EBM–KF when a statement is relevant to all variants, and in sections 3 and 4, we will specify the variant in each application.

Both AC$_t$ and $\tilde{g}(t)$ are taken from Forster et al. (2023). Several of the coefficients within the feedback functions f are defined to satisfy the constraints of the climate feedbacks presented in the IPCC AR6 (Forster et al. 2021; particularly Table 7.10), and all coefficients are based on observational and modeling literature values, typically with energy fluxes measured from satellites and temperature feedback coefficients determined from model results (full derivation in

appendix A and supplemental A). Because the Planck radiation requires absolute temperatures, we use degrees Kelvin in model calculations and convert them to degrees Celsius. OHCA is also approximately convertible to thermosteric sea level rise, via the 0.0121 cm ZJ^{-1} (ZJ = 10^{21} J) factor from analysis of 1995 to 2014 (Fox-Kemper et al. 2021). With this factor, the estimated thermosteric sea level rises we find are consistent with observations and projections. The two negative albedo attenuations $f_{\alpha A}(T, t) \times f_{\alpha S}(T)$ are expressed relative to Y_{2002} = 287.55 K (14.40°C), the inferred (see ζ_1 below) GMST measurement in 2002 (Jones and Harpham 2013; Morice et al. 2021).

The term ζ_0 = 10°C is an equilibrium temperature difference between the surface layer (including the upper ocean) and the deep ocean, arising because the global ocean is thermally stratified. This realistic choice of ζ_0, explained below, does not affect either T or H, provided that T is in equilibrium at the model's preindustrial initialization (and thus, ζ_0 is often abstracted away in similar two-layer energy-balance models) (Geoffroy et al. 2013a; Gregory 2000; Held et al. 2010). The term γ is the thermal conductivity or "efficiency" between layers of the ocean, taken from Geoffroy et al. (2013b) to be 0.67 W m^{-2} K^{-1}, the average from the CMIP5 models. The form of this parameterization of deep ocean temperature exchange follows recent work in emulating ocean heat uptake, but ignoring "efficacy factor" heat loss (PAGES2k Consortium 2017; Geoffroy et al. 2013a; Gregory 2000; Palmer et al. 2018a; Winton et al. 2010).

We first obtained the baseline ζ_1 = 287.01 K of the HadCRUT5 GMST anomaly (Morice et al. 2021) to place the 1960–89 standard climate normal of absolute GMST HadCRUT5 measurements to fall at the center of the range given by Jones and Harpham (2013): 13.85° ± 0.15°C. We symbolize the HadCRUT5 GMST anomaly record as HCa_t. Measurements of surface temperature will later be assimilated as absolute temperatures: $Y_t = \zeta_1 + HCa_t$. Our model assumes energy fluxes were balanced before industrial forcings, which requires an equilibrium temperature. We set this preindustrial equilibrium temperature to the 1850–79 standard climate normal of 286.67 K (13.52°C) = T_{1850} and initialized the 1850 climate state to this temperature. This choice is important regarding the determination of many nonlinear feedback functions and coefficients affecting the surface layer [(7) below], particularly with respect to the Planck feedback. Inconsequentially to the EBM dynamics, the deep ocean conservative temperature θ was initialized to be 276.67 K (3.52°C) = θ_{1850}, such that current deep ocean conservative temperatures are ≈3.8°C, choices consistent with both recent and historical measurements of the globally averaged values (Abraham et al. 2013; Robinson and Stommel 1959). So, $\zeta_0 = T_{1850} - \theta_{1850}$ = 13.52°–3.52°C = 10°C. Initializing the deep ocean conservative temperature θ to another value would change ζ_0 correspondingly, such that the modeled heat flow into the deep ocean would be unchanged.

With the considerations above, (1)–(5) become (6)–(8) as follows, with coefficients specified in Table 1:

$$\theta_t = [H_t - (T_t - T_{1850}) \times C_{\text{upperO}}]/C_{\text{deepO}} + \theta_{1850}, \tag{6}$$

$$\begin{aligned} T_{t+1} = T_t &+ \frac{(\tfrac{1}{4}G_{\text{SC}})_t \times c_2}{(\text{AOD}_t + c_4)} \\ &\times \left[1 + \beta_2(T_t - Y_{2002}) + \frac{\text{AC}_t - \text{AC}_{2002}}{c_3}\right] \\ &\times [1 + \beta_3(T_t - Y_{2002})] \\ &- c_1(T_t)^{4-\eta}[1 - \beta_0 \log_{10}([\text{eCO}_2]_t)] \\ &- \frac{\gamma}{C_{\text{surf}}}(T_t - \theta_t - \zeta_0), \end{aligned} \tag{7}$$

$$H_{t+1} = H_t + (T_{t+1} - T_t) \times C_{\text{upperO}} + \gamma \times (T_t - \theta_t - \zeta_0). \tag{8}$$

Future projections along the shared socioeconomic pathways (SSPs) for the EBM–KF also require the concentrations of greenhouse gases including carbon dioxide $[\text{eCO}_2]_t$, aerosol optical depth due to volcanic and human emissions AOD_t, and the computed effect from anthropogenic clouds AC_t, which are clouds induced by human activity including both cloud feedbacks that are part of climate change and aerosol-induced clouds. ESMs simulate the carbon cycle and thus find an equivalent of $[\text{eCO}_2]_t$ from specified CO_2 and greenhouse gas emissions, but our EBM–KF does not have this capability. Future greenhouse gas concentrations and anthropogenic cloud forcings are instead taken from a conversion of anthropogenic fluxes by the MAGICC7.0 carbon cycle emulator (Meinshausen et al. 2020), reported as energy fluxes by Smith et al. (2021). The anthropogenic cloud energy fluxes were used directly, and the effective greenhouse gas concentrations were calculated by inverting (A14); see supplemental section A1 (lines 105–130). For instance, SSP1-2.6 and SSP3-7.0 are shown in Figs. 9 and 10, which flank the most likely result of current environmental policies (Pielke et al. 2022). Projection of anthropogenic forcings from Nazarenko et al. (2022) using the NASA GISS ESM yields very similar future curves (see supplemental Figs. 18 and 19).

Overall, the blind (forward) energy-balance model (orange dashed line in Fig. 2) has four yearly forcing inputs $[\text{eCO}_2]_t$, AOD_t, AC_t, $(\tfrac{1}{4}G_{\text{SC}})_t$ and 17 irreducible parameters (including one inferred exponent η, four inferred β coefficients, three heat capacities, and three reference temperatures). The time step of this iterative difference equation model is 1 year chosen arbitrarily to coincide with the calendar year. The deep ocean conservative temperature θ_t is recalculated at each time step from the GMST T_t and the OHCA H_t by (6), and then, these two terms are updated with (7) and (8). The measured temperature in the year 2002 Y_{2002} appears prominently in this model because that was the midpoint of the measurement window of the CERES satellite (Loeb et al. 2009; Wielicki et al. 1996), and all albedo-related feedbacks are expressed relative to these measurements. For this model, the OHCA (H_t) is calculated in units of watts year per square meter (W yr m^{-2}) on an average of Earth's surface and then converted to ZJ within the ocean by multiplying by a factor of

TABLE 1. Constants and climate driver datasets referenced in (6)–(8), in addition to temperature baselines. Equations referenced in the "Source" column are found in appendix A and supplemental SA1 and SA2.

Symbol	Value [range]	Derivation or definition	Source
ζ_1	287.01 K (13.86°C)	$13.85°\text{C} = \zeta_1 + \frac{1}{30}\sum_{t=1960}^{1989} \text{HCa}_t$	Jones and Harpham (2013)
Y_{2002}	287.55 K (14.40°C)	$\zeta_1 + \text{HCa}_{2002}$	Morice et al. (2021)
T_{1850}	286.67 K (13.52°C)	$\zeta_1 + \frac{1}{30}\sum_{t=1850}^{1879} \text{HCa}_t$	Morice et al. (2021)
θ_{1850}	276.67 K (3.52°C)	Approximate deep ocean temperature	Abraham et al. (2013)
ζ_0	10 K (10°C)	$T_{1850} - \theta_{1850}$	Abraham et al. (2013)
γ	0.67 W K^{-1} m^{-2}	Ocean heat conductivity/year	Geoffroy et al. (2013b)
C_{surf}	17 W K^{-1} m^{-2}	Heat capacity/year, Earth surface	Schwartz (2007)
C_{upperO}	11.7 W K^{-1} m^{-2} (86 m H_2O)	Heat capacity/year, upper ocean	Geoffroy et al. (2013b)
C_{deepO}	155.7 W K^{-1} m^{-2} (1141 m H_2O)	Heat capacity/year, deep ocean	Geoffroy et al. (2013b)
η	1.615	Degree (exponent) of H_2O feedback	(A26) and (A27)
β_0	0.046 60	Infrared reflection per $\log_{10}$ ppm CO_2	(A20), (A21), and (A35)
c_1	2.198 910^{-5} $K^{-3+\eta}$	$\sigma_{sf}\beta_1/C_{\text{surf}}$	(A22) and (A35)
β_2	0.001 63 K^{-1}	Atmospheric albedo temperature feedback	(A13) and (A28)
β_3	0.001 36 K^{-1}	Ground albedo temperature feedback	(A14) and (A29)
c_2	0.4044 K m^2 W^{-1}	$0.834 \times 0.909 \times 9.068/C_{\text{surf}}$	(A11), (A23), and (A24)
c_3	264.377 W m^{-2}	$^1/_4\overline{G_{\text{SC}}} \times \tilde{d}_{2002} \times 0.834$	(A23)
c_4	9.73 (unitless)	$\frac{2q'}{1-g}$	(A11), (supplemental SA1), Eq. (2) of Schwartz et al. (2002); Eq. (9) of Harshvardhan and King (1993)
AC_{2002}	−0.988 W m^{-2}	Anthropogenic cloud radiative forcing, 2002	(A23)
$(^1/_4G_{\text{SC}})_t$	[340.06–340.48] W m^{-2}	Top of the atmospheric total solar irradiance	Coddington et al. (2017)
AOD_t	[0.2–142.9]	Aerosol optical depth	Miller et al. (2014), NASA/LARC/SD/ASDC (2018)
AC_t	−[1.09–0.06] W m^{-2}	Anthropogenic cloud radiative forcing	Forster et al. (2023)
$[\text{eCO}_2]_t$	[287.9–563.4]	Effective CO_2 concentration, ppm	Forster et al. (2023)

$$11.42\frac{\text{m}^2\ \text{ZJ}}{\text{year W}} = \frac{3.154\times10^7\ \text{s}}{1\ \text{year}} \times \frac{5.101\times10^{14}\ \text{m}^2}{\text{Earth surface area}} \times \frac{\text{ZJ}}{10^{21}\ \text{J}} \times \frac{0.71\ \text{m}^2\ \text{(ocean)}}{\text{m}^2\ \text{(total area)}}.$$

This time-step function [(6)–(8)] and its partial derivative (see appendix A, section c) will become critical parts of our Kalman filter [(9) and (10)] below.

This blind EBM has good skill at predicting the GMST with $r^2 = 0.908$ when compared to the HadCRUT5 GMST time series (Morice et al. 2021), and OHCA with $r^2 = 0.910$ when compared with the inferred history (Zanna et al. 2019), as is demonstrated by the dashed orange lines in Fig. 3. The blind EBM has a comparably high correlation ($r^2 = 0.923$) with the 30-yr running mean (i.e., the climate normal) of the HadCRUT5 GMST, indicating that this forward energy-balance model also has skill in reproducing the climate state as determined by standard approaches, with departures due to volcanic eruptions. Thus, most observed climate change can be explained by the literature-based blind, forward EBM with records of emissions (greenhouse gases and anthropogenic clouds) and measurements at the top of the atmosphere of aerosol optical depth. The distribution of residuals in the GMST record from either the 30-yr running mean or the EBM has small bias and skewness and a kurtosis slightly less than Gaussian (see supplemental Fig. 9). So, the 30-yr running mean's weather or "noise" empirical probability density function combining residuals and measurement uncertainty is very nearly Gaussian and thus amenable to treatment by a Kalman filter framework (see section 2b). The Fig. 3 forward model comparisons were made without any assimilated data, illustrating that the EBM physics alone has skill in reproducing aspects of the GMST and OHCA records. Tuning the EBM parameters may further improve skill, but the EBM is only the forward projection component of the data-assimilating Kalman filter hybrid model described in the next section. The combined system is the focus of this paper.

b. *EBM–Kalman filter: A weighted average of energy balance and measurements*

While similar algorithms were developed in the 1880s by Thorvald Nicolai Thiele (Lauritzen 2002, 1981), Kalman filtering rose to prominence due to its use in the Apollo navigation computer as proposed by Stratonovich (1959, 1960), Swerling (1959), Kalman (1960), Bucy (Kalman and Bucy 1961) and implemented by Schmidt (1981). Versions of this statistical filter are universally used in aerospace guidance systems (Grewal and Andrews 2001), aspects of numerical weather prediction (Houtekamer and Mitchell 1998; Kalnay 2003), and recently in climate science as ensemble Kalman filters, which use a Monte Carlo approximation via simulations in high-dimensional space (Compo et al. 2011; Hakim et al. 2016). We need to disambiguate ensemble Kalman filters from

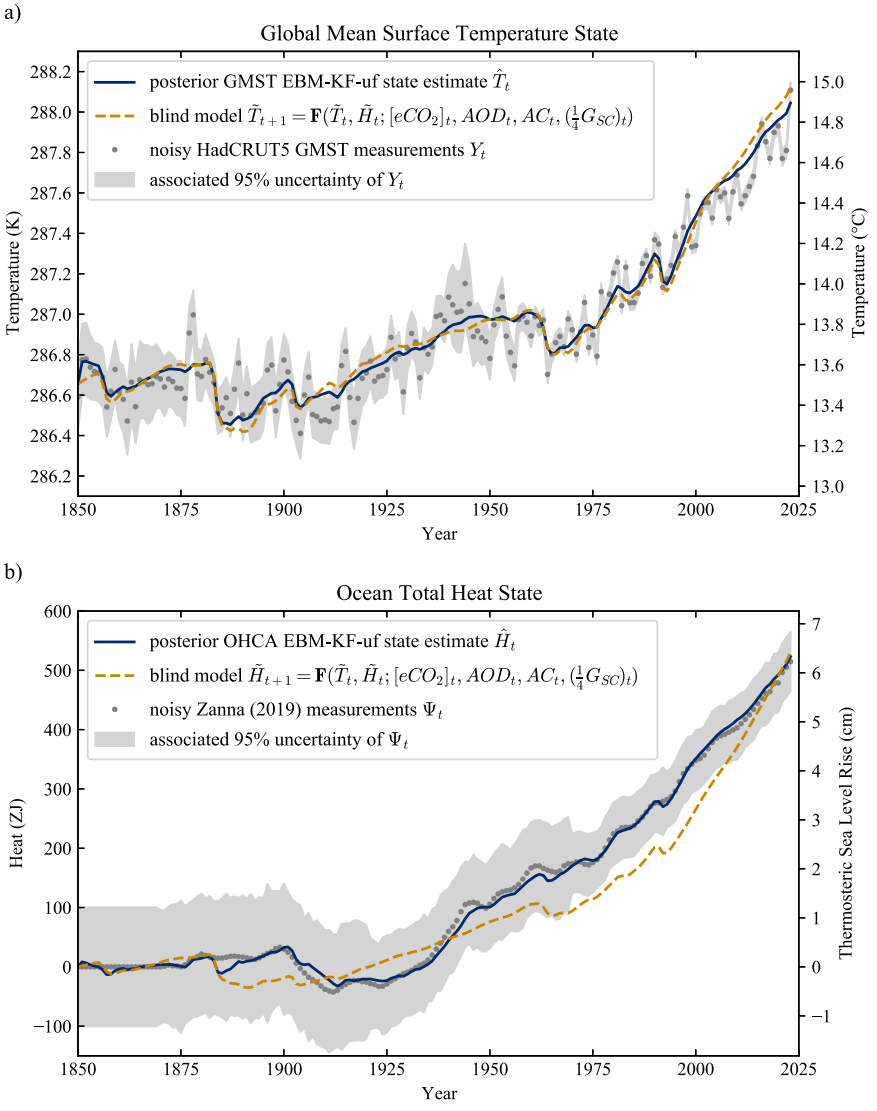

FIG. 3. Behavior of the EBM–KF–uf state in relation to blind EBM projections and the stochastic measurements of GMST and OHCA. (a) GMST prediction and (b) the OHCA prediction. The blind model (dashed orange) and KF state estimate (navy blue) use EBM dynamics to project from the previous state to the current state, but the state estimate also assimilates observations with uncertainty weighting [gray dots; GMST from HadCRUT5 (Morice et al. 2021) and OHCA from Zanna et al. (2019)]. Incorporation of these observations makes only small modifications to the EBM–KF's GMST state in (a), whereas in (b), there is an impressive difference between the blind EBM's OHCA projections and the EBM–KF's OHCA state—the latter sticks close to observations.

extended Kalman filters. In our field, the ensemble Kalman filter is a data assimilation technique based on weighted averages of GCM ensemble members. The extended Kalman filter is a method to apply the linear Kalman filter methods to nonlinear dynamical equations based on local linearization of the dynamical equations. Despite the success of ensemble Kalman filters, extended Kalman filters are inapplicable as the sole data assimilation tool for regional weather patterns (Bouttier 1996), because local weather processes do not sample from a Gaussian of extended Kalman filters. The multidimensional extended Kalman filter assumes

$$\mathbf{x}_t = \mathbf{F}(\mathbf{x}_{t-1}; u_t) + \mathbf{w}_t$$
$$\text{climate state: } \mathbf{x}_t,$$
$$\text{model ("dynamical") uncertainty: } \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}), \quad (9a)$$

$$\mathbf{y}_t = \mathbf{x}_t + \mathbf{v}_t$$
$$\text{weather state: } \mathbf{y}_t,$$
$$\text{additional ("measurement") uncertainty: } \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t). \quad (9b)$$

Bold type indicates state vectors. In this case of global GMST and OHCA, an extended Kalman filter works because both

model (or "dynamical") and additional (or "measurement") noise are approximately Gaussian (by the central limit theorem expectation[1] verified in section 3 and supplemental Fig. 9), and because the energy-balance equation [section 2a, (1)–(5)] has a continuous and bounded gradient of $\mathbf{F}(\mathbf{x}_{t-1}; u_t)$ [see (6)–(8) and supplemental section D], so it can be locally linearized.[2] This approximate linearity means that more complex realizations of the Kalman filter, particularly the unscented Kalman filter (Julier and Uhlmann 1997; Wan and Van Der Merwe 2000), are not necessary (see supplemental section D). This approach has already proven successful using a one-(spatial)-dimensional (north–south) energy-balance model, with time steps of decades (or longer), and optimized for use in paleoclimate research (García-Pintado and Paul 2018). Thus, for a variety of reasons, an EBM-KF can be built from an extended Kalman filter combined with an (annual, 0-spatial-dimensional) energy-balance model.

In-depth derivations and tutorials for constructing Kalman filters have been published elsewhere (Benhamou 2018; Thacker and Lacey 1998; Miller 1996; Ogorek 2019; Särkkä 2013; Kim and Bang 2018). Here, we describe enough for basic intuition, and we refer readers to Kalnay (2003), p. 281, for a more detailed explanation with alternative notation. We use the term "forecast" where other authors use "prior," and we avoid the use of "measurement error" in a manner that would be ambiguous and confusing in this application. Our equations for the extended Kalman filter (the KF part of the EBM–KF) are

$$\boldsymbol{\Phi}_t = \left.\frac{\partial \mathbf{F}(\mathbf{x}; u_t)}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}_{t-1}} \quad \text{local linearization at timepoint } t, \tag{10}$$

$$\hat{\mathbf{x}}_{t|t-1} = \mathbf{F}(\hat{\mathbf{x}}_{t-1}; u_t) \quad \text{forecast("prior")state estimate,} \tag{11}$$

$$\mathbf{P}_{t|t-1} = \boldsymbol{\Phi}_t \mathbf{P}_{t-1} \boldsymbol{\Phi}_t^* + \mathbf{Q} \quad \text{forecast("prior")covariance,} \tag{12}$$

$$\mathbf{z}_t = \mathbf{y}_t - \hat{\mathbf{x}}_{t|t-1} \quad \text{innovation residual,} \tag{13}$$

$$\mathbf{S}_t = \mathbf{P}_{t|t-1} + \mathbf{R}_t \quad \text{innovation covariance,} \tag{14}$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}(\mathbf{S}_t)^{-1} \quad \text{Kalman gain,} \tag{15}$$

$$\hat{\mathbf{x}}_t = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t \mathbf{z}_t \quad \text{posterior state estimate,} \tag{16}$$

$$\mathbf{P}_t = (\mathbf{I} - \mathbf{K}_t)\mathbf{P}_{t|t-1} \quad \text{posterior state covariance.} \tag{17}$$

We proceed through this mathematical algorithm [(10)–(17)] as follows. Initially, there is some estimated state vector (GMST and OHCA within this paper) $\hat{\mathbf{x}}_{t-1}$ and a Gaussian uncertainty envelope around this vector defined by a state covariance matrix $\mathbf{P}_{t-1}$. In the basic setup of a Kalman filter, the state vector is transformed (or projected) 1 year into the future using a dynamic model Jacobian matrix $\boldsymbol{\Phi}_t$ into a forecast state $\hat{\mathbf{x}}_{t|t-1} = \boldsymbol{\Phi}_t \hat{\mathbf{x}}_{t-1}$, a transformation that may depend on time-varying control parameters u_t. For our climate system, this linear projection is extended to the nonlinear function $\hat{\mathbf{x}}_{t|t-1} = \mathbf{F}(\hat{\mathbf{x}}_{t-1}; u_t)$ in (11), which is just the forward energy-balance model [(6)–(8)], where u_t represents the collection of climate forcings: $[\mathrm{eCO_2}]_t$, AOD_t, AC_t, and $(^1/_4 G_{\mathrm{SC}})_t$. This simple extension to nonlinearity is the meaning of "extended" Kalman filter. The state covariance $\mathbf{P}_{t-1}$ is projected to the next year using the local linear approximation of the dynamic model Jacobian matrix $\boldsymbol{\Phi}_t$ [(10)] and enlarges by an additional assumed model error covariance $\mathbf{Q}$, yielding $\mathbf{P}_{t|t-1}$ as the forecast covariance [(12)]. To arrive at a posterior, information from a measurement vector $\mathbf{y}_t$ is considered [(13)].[3] The probabilistic range of anticipated discrepancies between $\hat{\mathbf{x}}_{t|t-1}$ and $\mathbf{y}_t$ is given by the innovation covariance matrix $\mathbf{S}_t$, which is the sum of $\mathbf{P}_{t|t-1}$ and an assumed measurement covariance $\mathbf{R}_t$ [(14)]. The posterior estimate of the state $\hat{\mathbf{x}}_t$ is found by taking a weighted average of $\hat{\mathbf{x}}_{t|t-1}$ and $\mathbf{y}_t$ [(16)], with the weight on $\mathbf{y}_t$ given by $\mathbf{P}_{t|t-1}(\mathbf{S}_t)^{-1}$, a product known as the Kalman gain $\mathbf{K}_t$ [(15)]. To reflect the greater certainty in the state vector because of this correction $\mathbf{P}_t$, the posterior covariance matrix is $\mathbf{P}_{t|t-1}$ shrunk by the Kalman gain $\mathbf{I} - \mathbf{K}_t$ per (17). Within the context of climate modeling, this "posterior state estimate" $\hat{\mathbf{x}}_t$ is somewhat analogous to a climate reanalysis product, as both combine observations and models. Within the context of Bayesian probability, the prior (forecast) distribution is given by projecting $\mathcal{N}(\hat{\mathbf{x}}_{t-1}, \mathbf{P}_{t-1})$ into the future using the Jacobian matrix $\boldsymbol{\Phi}_t$, which is multiplied by the marginalized likelihood of $\mathbf{y}_t$ to give a posterior distribution $\mathcal{N}(\hat{\mathbf{x}}_t, \mathbf{P}_t)$. Note that $\boldsymbol{\Phi}_t^*$ in (12) above indicates matrix transposition.

The true climate state $\mathbf{x}_t$ is the two-entry vector underlying GMST and OHCA, filtering out weather and internal variability: $\mathbf{x}_t = [T_t, H_t]$. Throughout this paper, $[a, b]$ indicates a two-dimensional vector with components a and b. The noisy measurements $\mathbf{y}_t = [Y_t, \psi_t]$ are the yearly time series of GMST and OHCA, and $\hat{\mathbf{x}}_t = [\hat{T}_t, \hat{H}_t]$ is the estimate of the unknown two-dimensional climate state, expressed in kelvins and watts year per square meter (convertible to ZJ by the factor 11.42 $\mathrm{m^2\ ZJ\ W^{-1}\ yr^{-1}}$). The energy-balance model's $\mathbf{F}(\hat{\mathbf{x}}_{t-1}; u_t)$ in (10) governing T and H is nonlinear (as described above with T^2 and $T^{2.385}$ terms due to albedo, Planck, and water vapor feedbacks) (Friedrich et al. 2016), which necessitates linearization. In our extended Kalman filter, the forecast state $\hat{\mathbf{x}}_{t|t-1}$ [(11)] is given by (6)–(8) above and $\boldsymbol{\Phi}_t$ and the forecast

[1] The central limit theorem states that taking the average of many independent samples from the same non-Gaussian distribution with bounded moments will produce a mean that approximates a Gaussian distribution (Montgomery and Runger 2016). This is the case for the detrended annual GMST, a climate state variable composed of the average of many non-Gaussian regional and daily weather patterns (Hu and Fedorov 2017; Loikith and Neelin 2019; Perron and Sura 2013; Sura and Sardeshmukh 2008). Likewise, while annual OHCA is largely constrained by the subtropical pycnocline depth (Newsom et al. 2023), it too is comprised of numerous regional and seasonal patterns (Cheng et al. 2017; Huguenin et al. 2022; Hummels et al. 2013). Many dynamical components of the global oceans are non-Gaussian, such as velocity (Smith and Gille 1998) and sea surface height (Sura and Gille 2010; Nieto-Reyes 2022).

[2] Careful construction of the EBM with T^2 in the shortwave term and $T^{2.39}$ in the counteracting longwave term in (1) and (5) ensures the derivative [(A21)–(A24)] does not change significantly over the relevant range of temperatures [286–291] K, $[\mathrm{eCO_2}]_t$ effective $\mathrm{CO_2}$ concentrations [278–2000] ppm, AOD_t aerosol optical depths [0–0.15], and AC_t anthropogenic cloud forcing [−1–0] $\mathrm{W\ m^{-2}}$.

[3] If $\mathbf{y}_t$ is an indirect measurement of the hidden state vector $\mathbf{y}_t$, an observation (or emission) matrix $\mathbf{H}$ further complicates the procedure (details in the references above). Here, we consider only direct "observations" of GMST and OHCA making mapping and interpolation errors implicit and the observation matrix $\mathbf{H} = \mathbf{I}$, the identity matrix.

covariance projection $\mathbf{P}_{t|t-1}$ [(12)] is a time-varying linearization [(A21)–(A24)]. This energy-conserving difference equation thus resembles a first-order Taylor series approximation of a differential energy-balance model (if discretization errors are considered part of the tendency) or the integral form of a conservative discretization in time (if shortwave and longwave fluxes are taken as a model for their time-integrated value), and the Kalman filter reapproximates a GMST and OHCA climate state every year. The initial estimated state uncertainty is intentionally overestimated at

$$\mathbf{P}_{1850} = \begin{bmatrix} 1\,\mathrm{K}^2 & 1\,\mathrm{K}\dfrac{\mathrm{W\ yr}}{\mathrm{m}^2} \\ 1\,\mathrm{K}\dfrac{\mathrm{W\ yr}}{\mathrm{m}^2} & 20\left(\dfrac{\mathrm{W\ yr}}{\mathrm{m}^2}\right)^2 \end{bmatrix}.$$

Then, $\mathbf{P}_t$ rapidly converges in the EBM–KF–uf (and EBM–KF–ta) to

$$\mathbf{P}_{1865} = \begin{bmatrix} 0.0017\ \mathrm{K}^2 & 0.035\ \mathrm{K}\dfrac{\mathrm{W\ yr}}{\mathrm{m}^2} \\ 0.035\ \mathrm{K}\dfrac{\mathrm{W\ yr}}{\mathrm{m}^2} & 4.0\left(\dfrac{\mathrm{W\ yr}}{\mathrm{m}^2}\right)^2 \end{bmatrix}.$$

Therefore, it continues to slowly shrink with time as more accurate measurements are made. We form confidence intervals for the GMST climate state [(19)] and OHCA climate state by taking twice the square root of the respective diagonal elements of $\mathbf{P}_t$ [(18a)]:

$$[\hat{p}_t^T, \hat{p}_t^H] = \mathrm{diag}(\mathbf{P}_t), \tag{18a}$$

$$[\hat{s}_t^T, \hat{s}_t^H] = \mathrm{diag}(\mathbf{S}_t). \tag{18b}$$

For example,

95% CI of estimated GMST state, 1865:

$$\hat{T}_{1865} \pm 2\sqrt{\hat{p}_{1865}^T} = 286.66\ \mathrm{K} \pm 2\sqrt{0.0017\,\mathrm{K}^2}$$
$$= 286.66\ \mathrm{K} \pm 0.07\ \mathrm{K}. \tag{19}$$

We give both diagonal elements their own symbols and similarly for $\mathbf{S}_t$ [(18b)] noting that here superscripts T and H are labels not exponentiation nor transposition. Similar to (19), we use the diagonal elements of $\mathbf{S}_t$ to form confidence intervals of next-year measurements about $\hat{\mathbf{x}}_{t|t-1}$. These confidence intervals are functions of time, as indicated by the subscript.

The extended Kalman filter implicitly assumes that Gaussian model ("dynamical") noise is added to this climate state at each time step [(9a)], and additionally, the climate state emits annual "weather variability" (other authors refer to this as year-to-year climate variability) from a yet wider Gaussian noise distribution [(9b)] quantified by $\mathbf{S}_t$, which is additional ("measurement") uncertainty $\mathbf{R}_t$ combined with the forecast covariance projection $\mathbf{P}_{t|t-1}$ [(14)]. We use the term weather variability to clearly distinguish from annual variations in the climate state $\hat{\mathbf{x}}_t = [\hat{T}_t, \hat{H}_t]$. Whereas we interpret global annual weather to be the noisy measurements $\mathbf{y}_t = [Y_t, \psi_t]$, weather variability is observed via innovation residuals $\mathbf{z}_t$:

$$z_t^T = Y_t - \hat{T}_{t|t-1}, \tag{20a}$$

$$z_t^H = \psi_t - \hat{H}_{t|t-1}. \tag{20b}$$

These innovation residuals have components $\mathbf{z}_t = [z_t^T, z_t^H]$, and the Kalman filter expects them to come from an unbiased Gaussian noise distribution with covariance $\mathbf{S}_t$ (not $\mathbf{R}_t$ because the Kalman filter does not have knowledge of the true climate state $\mathbf{x}_t$).

The EBM–KF climate state $\hat{\mathbf{x}}_t$ and state covariance $\mathbf{P}_t$ are causal in time at each year t, so they only access information from the measurements taken prior to and at year t: $\{\mathbf{y}_{1850}, \mathbf{y}_{1851}, \ldots, \mathbf{y}_t\}$. This past-to-present Kalman filter incorporated into the EBM–KF [(10)–(17)] could be further extended into a Rauch–Tung–Striebel (RTS) smoother (Rauch et al. 1965) by additional steps (see supplemental SA3), which meld information from all measurements in the time window: $\{\mathbf{y}_{1850}, \mathbf{y}_{1851}, \ldots, \mathbf{y}_{2023}\}$ into each reestimated posterior state $\hat{\hat{\mathbf{x}}}_t$ and posterior state covariance $\hat{\hat{\mathbf{P}}}_t$ by running backward from the latest EBM–KF state estimates $\hat{\mathbf{x}}_{2023}$ and $\mathbf{P}_{2023}$. However, in the 1850-to-present application, this extension has little effect on $\hat{\mathbf{x}}_t$ (supplemental Fig. 2), with the only impacts being greater certainty in the smoothed state at the cost of violation of causality. Defining as in (18a), $[\hat{\hat{p}}_t^T, \hat{\hat{p}}_t^H] = \mathrm{diag}(\hat{\hat{\mathbf{P}}}_t)$, for the GMST uncertainty $\hat{p}_t^T \approx 2.25 \times \hat{\hat{p}}_t^T$, and for the OHCA uncertainty, $\hat{p}_t^H \approx 2.84 \times \hat{\hat{p}}_t^H$ (within the EBM–KF–uf). Overall, we deemed this extension not worth the added complications and retained the past-to-present, causal approach.

In summary, the extended Kalman filter projects forward 1 year into the future based on the unbalanced fluxes of the energy-balance model equation and then takes a weighted average of this projection with the annual GMST measurement (the data assimilation increment). Thus, even though the EBM conserves energy (by construction), the combined EBM–KF does not, unlike other alternative data assimilation approaches (Wunsch and Heimbach 2007). The state estimates from this EBM–KF (in navy blue in Fig. 3) often lie between the blind EBM (in dashed orange in Fig. 3) and the annual temperature measurements (scattered gray dots in Fig. 3). These data assimilation corrections can be seen most clearly within the GMST measurements in Fig. 3a during 1900–45 and within the OHCA measurements in Fig. 3b during 1940–70. It is possible for the EBM–KF state estimates to escape these bounds for a short time, for instance, during 1945–50 in Fig. 3a or after 2007 in Fig. 3b. These "escape periods"

may reflect bias in the measurements, such as warm-biased World War II (WWII)–era measurements of (sea) surface temperature (Chan and Huybers 2021) or bias in the Zanna et al. (2019) OHCA product (which may be indicated by this product having less heat uptake since 2005 than all but 1 of 19 other OHCA estimates: Gulev et al. 2021). Both the blind EBM predictions $[\tilde{T}_{t+1}, \tilde{H}_{t+1}] = \mathbf{F}(\tilde{T}_t, \tilde{H}_t; u_t)$ and EBM–KF–uf state estimates $\hat{\mathbf{x}}_t = [\hat{T}_t, \hat{H}_t]$ dip down with each major volcanic eruption within the AOD record (see Fig. 10 in the discussion and section 4). These volcanic dips are far more pronounced for the GMST component than for OHCA (Fig. 3) and are present only as flat spots in the deep ocean conservative temperature curve (supplemental Fig. 11).

c. *Selection of model uncertainty and time-varying measurement uncertainty*

Figure 3 also demonstrates the accuracy associated with each of the temperature measurements. The uncertainty in the climate state $\mathbf{P}_t$ automatically responds to unexpected values of the measured temperature (Wunsch 2020). The HadCRUT5 GMST decreases in reported measurement standard deviation from 0.079 K in the 1850–79 window to 0.017 K in the 1990–2019 window (Morice et al. 2021), a 78% reduction primarily reflecting a lack of observations in the Southern Hemisphere before the satellite age. The inferred deep ocean heat content taken primarily from a hybrid model–observation reconstruction (Zanna et al. 2019) had a very wide confidence interval before the introduction of modern sampling methods in the 1970s. We use the Zanna et al. (2019) hybrid product due to its long record of OHCA estimates (based on surface forcing in early years) rather than the shorter direct measurement products (Ishii et al. 2017), although both could be assimilated simultaneously within EBM–KF if desired (as discussed in section 4c). The additional increase in OHCA after 2018 was provided from a separate NCEI dataset (Levitus et al. 2017). Our EBM–KF incorporates these known physical measurement uncertainties in the HadCRUT5 measurements of GMST and the OHCA reconstruction as $\mathbf{R}_t^{\text{var}}$. The total assumed measurement covariance $\mathbf{R}_t$ [(14)] is composed of two components: the time-varying physical measurement uncertainty $\mathbf{R}_t^{\text{var}}$ and the constant uncertainty $\mathbf{R}^{\text{const}}$ reflecting internal variability due to dynamical chaos: primarily ENSO and other climate oscillations with limited predictability. Both forms of measurement noise are added onto the underlying climate signal via the random vector υ_t to generate annual observations [(9b)]. We assume that $\mathbf{R}_t^{\text{var}}$ is diagonal and simply sum the two variance matrices to obtain a time-varying value:

$$\mathbf{R}_t = \mathbf{R}_t^{\text{var}} + \mathbf{R}^{\text{const}}. \tag{21}$$

Both realizations of our EBM–KF also have a measurement uncertainty $\mathbf{R}^{\text{const}}$ that is constant in time and based on the [HadCRUT5's GMST, Zanna et al. (2019) OHCA] residual covariance with respect to their 30-yr running means. In other words, we computed

$$\begin{aligned}\mathbf{R}^{\text{const}} &= \text{Cov}(\mathbf{y}_t - \overline{{}_{30}\mathbf{y}_t}) \\ &= \begin{bmatrix} 0.01099\ \text{K}^2 & 0.04523\ \text{K}\ \dfrac{\text{W yr}}{\text{m}^2} \\ 0.04523\ \text{K}\ \dfrac{\text{W yr}}{\text{m}^2} & 1.12991\left(\dfrac{\text{W yr}}{\text{m}^2}\right)^2 \end{bmatrix} \\ &= 30 \times \mathbf{Q}. \end{aligned} \tag{22}$$

The assumed model covariance $\mathbf{Q}$ used in (12) is set to $\mathbf{R}^{\text{const}}/30$ to emulate the 30-yr running average definition of climate state (Guttman 1989). That is, we assume that the random noise contained within the climate model has a variance that is 1/30th as large as the variance in the weather measurements and assume yearly anomalies are uncorrelated. By this simple method, the data-assimilating EBM–KF is tuned to match the standard climate normal, as any 30-member uncorrelated sample average has a variance 1/30th as large as the annual measurements' variance. Variance in these annual measurements arises both from chaos within the climate system and time-varying physical measurement uncertainty, and this $\mathbf{R}^{\text{const}}$ contribution to the model and measurement uncertainty quantifies the chaotic internal variability and would exist even if all measurements could be made with arbitrary accuracy.

d. *Non-Gaussian future projection and sampling of volcanic activity*

The EBM–KF can project 1 year into the future, given greenhouse gas and aerosol concentrations, without any new measurements using only the forward model to obtain forecast estimates [(11)–(12)]. To project farther into the future, the posterior state and posterior covariance are set equal to the forecast state and forecast covariance, i.e., a posterior unaffected by any new observations: $\hat{\mathbf{x}}_t = \mathbf{F}(\hat{\mathbf{x}}_{t-1})$ and $\mathbf{P}_t = \mathbf{\Phi}_t \mathbf{P}_{t-1} \mathbf{\Phi}_t^* + \mathbf{Q}$. While these far-future state estimates behave equivalently to a blind model, the covariance grows over time, either sublinearly or exponentially (see section 3d).

While the SSPs are used for most forcing variables, future volcanic eruptions require modeling as well. Volcanic eruptions determining AOD_t are inherently stochastic, but the time intervals between eruptions can be approximated using exponential distributions (Papale 2018). In standard ESM SSP forcing, future volcanism is usually included by a steady "background" volcanism, neglecting volcanism's intermittency and the associated exponential distributions. Even though the EBM–KF assumes Gaussian error and thus cannot include exponential distributions in the same way as measurement and internal chaotic variability, it is so computationally inexpensive that it can be rerun to sample repeatedly over non-Gaussian distributions. This ability to include future volcanoes illustrates a major advantage of this system: Thousands of future scenario inputs can be generated and utilized within minutes on a laptop, while each ESM of the LENS2 ensemble took over a week to run on a supercomputer (roughly a billion times more effort in core hours per ensemble member) which limits the ensemble size and thus motivates using only a background constant level of volcanism. No single exponential distribution fits well to the observed series of volcano eruption intervals, so an exponential mixture with two components was

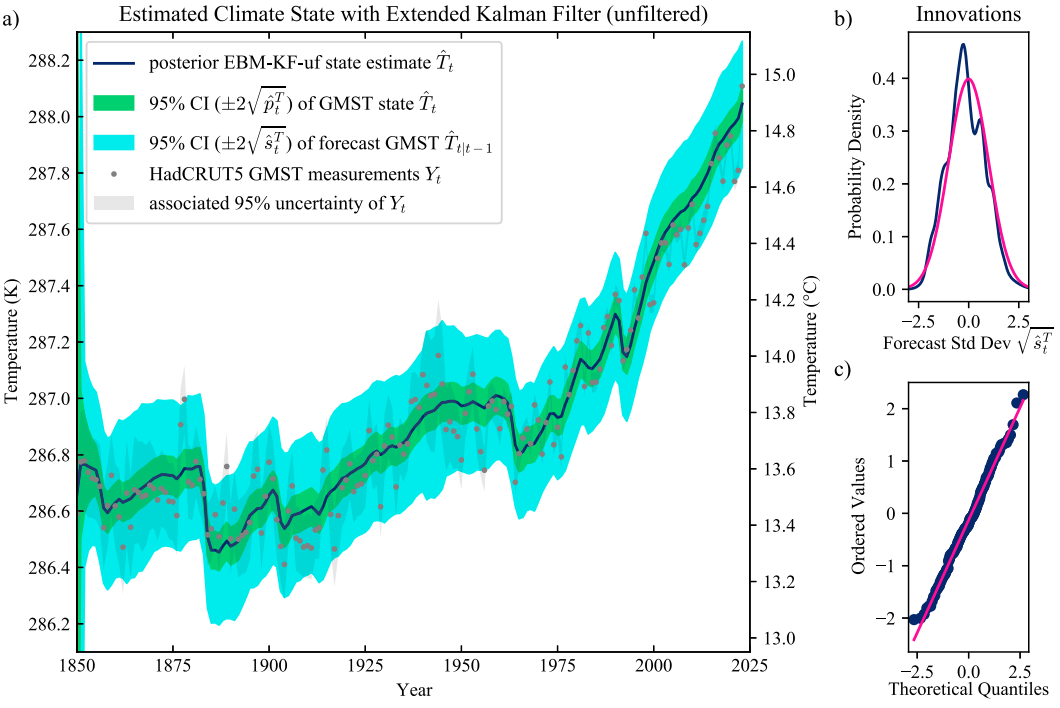

FIG. 4. EBM–KF–uf and associated uncertainties. (a) The EBM–KF–uf climate state estimate (navy blue line $\hat{T}_t$) is drawn with a 95% or extremely likely CI (light green area) of its posterior uncertainty $\pm 2\sqrt{\hat{p}_t^T}$. Annual-mean HadCRUT5 GMST measurements are assimilated (gray dots and gray area mostly within the light blue). A 95% CI in light blue indicates the forecast uncertainty $\pm 2\sqrt{\hat{s}_t^T}$ (b) The Gaussian mixture of innovations z_t^T (deviations between measurements minus Y_t the projected climate state $\hat{T}_{t|t-1}$) with each year's associated measurement uncertainty (navy blue), normalized onto a horizontal axis labeled with standard deviations $\sqrt{\hat{s}_t^T}$ of the ideal forecast covariance (pink). (c) Quantile–quantile plot of these normalized innovations ($z_t^T/\sqrt{\hat{s}_t^T}$) All panels demonstrate that the gray HadCRUT5 GMST observations appropriately fill out the 95% CI of the forecast uncertainty (light blue) around the EBM–KF–uf state estimate (navy blue, $\hat{T}_t$).

found to be the best fit to the data using the decomposed normalized maximum likelihood (Okada et al. 2020). See appendix B for further details.

3. Results

a. *Examination of the EBM–KF climate state (1850–the present)*

A primary product of this paper is the EBM–KF–uf climate state, spanning from 1850 to the present. Recall that the forward EBM uses published literature values: This is not an empirical fit to GMST and OHCA data, but rather the EBM–KF (in all variants) assimilates these data. We first examine the GMST component $\hat{T}_t$ of the Kalman-filtered climate state $\hat{\mathbf{x}}_t$. There are two distinct Gaussian distributions relevant to understanding our method: the uncertainty in the current GMST climate state $T_t \pm 2\sqrt{\hat{p}_t^T}$, as graphed in narrow green envelope in Fig. 4a, and the uncertainty window of possible next-year (forecast) GMST measurements $T_{t|t-1} \pm 2\sqrt{\hat{s}_t^T}$, as graphed in the light blue wider envelope in Fig. 4a.

Further examination of the "update" difference [(16)] between the posterior estimated states and forecast states $\hat{T}_t - \hat{T}_{t|t-1}$ reveals that in any individual year after 1855, assimilation of the GMST measurement only shifts the forecast GMST state projection $\hat{T}_{t|t-1}$ by -0.001 ± 0.017 K yr^{-1}, range [−0.020–0.022] K yr^{-1}. This update value is minuscule compared with the GMST adjustment in $\tilde{T}_t$ from the blind, forward EBM contribution of forced climate state change of $+0.025 \pm 0.053$ K yr^{-1} since 1975, and $+0.002 \pm 0.055$ K yr^{-1} from 1850 to 1975, while the forecast change can be as large as range [−0.191–0.053] K in a single year due to volcanic eruptions. However, as shown in Fig. 3, repeated small updates in the same direction (due to repeatedly lower or higher than expected GMST measurements) can drift $\hat{T}_t$ away from the blind model estimate $\tilde{T}_t$. This "accumulated correction" $\hat{T}_t - \tilde{T}_t$ is +0.004 K on average and as much as [−0.086–0.096 K]. Accumulated corrections are 3–4 times larger in extreme than the most extreme updates, indicating that these updates had accumulated over >4 years prior to 1886 and 2022 (5 and 8 years, respectively; see Fig. 3a). Note the mean accumulated correction is slightly positive, while the mean update is slightly negative because of the influence of OHCA corrections (see below and Fig. 3b).

The EBM–KF–uf state $\hat{T}_t$ is still very highly correlated with the blind, forward EBM $\tilde{T}_t$ ($r^2 = 0.992$). Measurements Y_t have nearly equal warming and cooling contributions to the underlying $\hat{T}_t$ climate state, forming the expected Gaussian distribution of normalized innovations ($z_t^T/\sqrt{\hat{s}_t^T}$) as demonstrated over the entire time series in Fig. 4b and in every full 50-yr period in supplemental Fig. 12. The GMST observations since 2000 slightly cool the EBM (supplemental Figs. 12d,h) indicating that the EBM may have oversized positive climate feedbacks, an issue which could be rectified with parameter adjustment (section 4c).

After an initial convergence period of about a decade, the green 95% CI of the GMST state uncertainty $\pm 2\sqrt{\hat{p}_t^T}$ slightly shrinks from ±0.067 K in the 1870s to ±0.062 K since 1980. The 95% CI of forecast uncertainty $\pm 2\sqrt{\hat{s}_t^T}$ is drawn in light blue around the forecast estimated GMST state projection $\hat{T}_{t|t-1}$, showing where the Kalman filter expects the subsequent year's temperature measurement to be. This forecast uncertainty converges from roughly ±0.26 K in the 1870s to ±0.223 K since 1980. Both reductions reflect the improvement in the GMST component of the time-varying measurement uncertainty $\mathbf{R}_t^{\text{var}}$ with modern observations, but these reductions are modest compared to the 76% reduction in time-varying HadCRUT5 measurement uncertainty over the same period because the EBM–KF is also assuming time-invariant levels of chaotic internal climate process uncertainty ($\mathbf{Q}$ and the associated $\mathbf{R}^{\text{const}}$).

The empirical projection probability distribution (a Gaussian mixture of all measurement uncertainties relative to the EBM–KF forecast distribution) and an ideal Gaussian distribution closely match (Fig. 3b), confirming that the annual measurements of GMST can be interpreted as Gaussian noise around an underlying climate state. The quantile–quantile plot (Fig. 3c) demonstrates this same finding, just using gray points of innovations z_t^T (the difference between EBM–KF forecasts $\hat{T}_{t|t-1}$ and measurements Y_t) rather than each innovation being a distribution (with variance from $\mathbf{R}_t^{\text{var}}$) as in Fig. 3b. Each innovation point is normalized to the forecast uncertainty ($z_t^T/\sqrt{\hat{s}_t^T}$), and then, these are sorted from lowest to highest and plotted on the vertical axis. Along the horizontal (theoretical quantiles) axis, the percentile of each innovation is plotted where it would lie on an ideal Gaussian distribution, showing the real GMST weather measurements from HadCRUT5 are distributed around the EBM–KF–uf GMST climate state in precisely the expected Gaussian distribution.

As we hoped, the EBM–KF–uf GMST climate state estimate over 1850 to the present is not substantively different from the 30-yr running average except for the impact of major volcanoes (see Fig. 10a; $r^2 = 0.923$); thus, $\hat{T}_t \approx \overline{{}_{30}Y_t}$ in nonvolcano years. The LENS2 hindcasts depart from both in the interval from 1940 to 2000 (see Fig. 10a) causing a lower $r^2 = 0.906$ over all 174 years between EBM–KF–uf and LENS2. The EBM–KF–uf with unfiltered volcanic forcing can thus be interpreted as a middle ground between the 30-yr running average and a LENS2 ensemble average (which are farther apart with $r^2 = 0.820$). The performance of the GMST and OHCA portions of EBM–KF–uf model do vary; the most noticeable biases (see Fig. 3) are that the blind OHCA is significantly corrected toward the Zanna et al. (2019) reconstruction of OHCA from 1875 to 2005 (assimilation of this data reconstruction continues through 2018), but these correction periods are not evident as persistent biases in the EBM–KF (Fig. 5). Forward model biases may be ameliorated by automated, optimized tuning of parameters. This is addressed in section 4c and is well studied in Kalman filter applications (Zhang and Atia 2020); the potential adoption of these tools to climate science is a key advantage of the EBM–KF hybrid.

Figure 5 shows the deep OHCA component of the EBM–KF and its associated uncertainties. While the OHCA measurements from the Zanna et al. (2019) hybrid product are more autocorrelated than the HadCRUT5 GMST (relatively less year-to-year variability), the innovations for OHCA are again approximately Gaussian (Figs. 5b,c). In the context of this empirical probability distribution, each member of the Gaussian mixture has a larger gray window given by the time-varying measurement uncertainties $\mathbf{R}_t^{\text{var}}$ from the OHCA measurements. In simpler language, the light blue forecast window is large because it must encapsulate the gray measurement uncertainty window, which moves around within it. To achieve the nearly Gaussian empirical probability distribution in Fig. 5b, it is unsurprising that most EBM–KF estimated states are pulled very close to the autocorrelated OHCA measurements in Figs. 3b and 5a. This is a situation dominated by measurement uncertainty $\mathbf{R}_t^{\text{var}}$, which is different than observable dynamic weather variability (innovations z_t^T) filling the full forecast distribution (light blue) in Fig. 4a. As a result, the OHCA component of the EBM–KF pays much more attention to these measurements ψ_t than relying mostly on the blind EBM (see Fig. 3b). This updates the OHCA state estimate ($\hat{H}_t - \hat{H}_{t|t-1}$) after 1855 by 0.05 ± 7.38 ZJ yr^{-1}, range [−8.16–9.78] ZJ yr^{-1}, comparable with the OHCA change in $\tilde{H}_t$ from the blind, forward EBM contribution 3.07 ± 10.60 ZJ yr^{-1}, range [−25.31–14.72] ZJ yr^{-1}. Unsurprisingly, the EBM–KF takes a substantially different track than the blind EBM, yielding an accumulated correction of up to +91.6 ZJ in 1998. Reflecting this improvement in measurement accuracy (as incorporated via $\mathbf{R}_t^{\text{var}}$), the OHCA components of both state uncertainty $\pm 2\sqrt{\hat{p}_t^H}$ and forecast uncertainty $\pm 2\sqrt{\hat{s}_t^H}$ shrink dramatically over the 174-yr run. The term $\pm 2\sqrt{\hat{p}_t^H}$, the envelope for the OHCA climate state estimate, has a very slow initial convergence that reaches ±45.1 ZJ by 1865 and then gradually falls to ±29.4 ZJ by 2000, a 35% decrease. The term $\pm 2\sqrt{\hat{s}_t^H}$, the 95% forecast envelope for OHCA, drops from ±115.0 ZJ by 1865 to ±66.2 ZJ by 1970 (42% decrease) and then remains near this value through the present, range ±[63.4–71.2] ZJ. This reduction in forecast uncertainty directly reflects a 48% decrease in the uncertainty from the Zanna et al. (2019) hybrid product over the equivalent time period.

b. Using the EBM–KF to determine policy threshold crossing

A single GMST measurement is not an accurate measurement of anthropogenic climate change due to the large

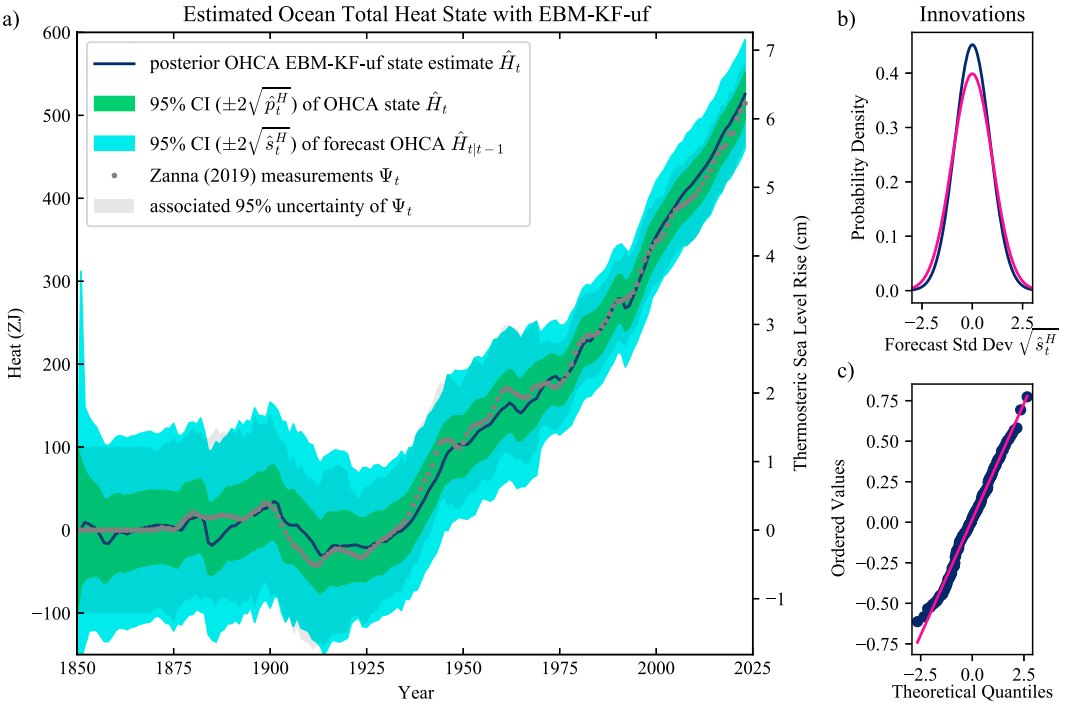

FIG. 5. EBM–KF state estimate (navy blue) for deep ocean OHCA in zettajoules and approximate thermosteric sea level from the same EBM–KF run as in Fig. 3. 95% CI of forecast estimate is drawn in light blue, and posterior 95% CI is drawn in green. Annual-mean (Zanna et al. 2019) reconstructions are assimilated (gray dots and gray area almost entirely within the light blue). Other panels and colors as in Fig. 4. All panels demonstrate that the uncertainty window of the assimilated OHCA data (gray) closely corresponds to the 95% CI of the forecast uncertainty (light blue) around the EBM–KF state estimate (navy blue).

internal variability of the system, and so a single annual temperature above a particular policy threshold is not a guarantee of the climate state crossing that threshold. One interpretation of "crossing" is when the climate state underlying GMST (e.g., the standard climate normal or 30-yr running mean of GMST) is determined with a given probability to have passed a policy threshold. This climate state above the threshold definition was used by Tebaldi and Knutti (2018) for regional thresholds and the IPCC AR6 (Lee et al. 2021) who state "the time of GSAT exceedance is determined as the first year at which 21-yr running averages of GSAT exceed the given policy threshold."[4] A second interpretation would be the chance that next year's annual-mean GMST will exceed the policy threshold, or "annual temperature forecast above" the threshold. The EBM–KF generates probability distributions for both the climate state above and the annual temperature forecast above interpretations of whether a policy threshold has been crossed.

For the first interpretation, the climate state threshold as in the IPCC definition is given in the EBM–KF by a Gaussian distribution (green in Fig. 6a) about the state $\hat{T}_t$ with a variance $\hat{p}_t^T$. The IPCC probability distribution is drawn from an ensemble of models over both the historical period and future projections (including those from LENS2 in Fig. 6b), so the fraction of the climate states (21-yr means in the IPCC definition) of each j of the ensemble members $(\overline{{}_{21}Y_t})_j$ found above a given policy threshold determines the overall probability that the climate policy threshold was crossed (see Fig. 6d). Within our notation, we not only reuse Y_t to represent a GMST time series but also add the j subscript to indicate the jth LENS2 hindcast (a simulation ensemble with 90 members), to distinguish an ensemble member from an observed historical record. This empirical approach assumes the ensemble spread is a good representation of the real-world GMST uncertainty. However, caution with this assumption is needed as recent IPCC reports discount the 90% ensemble spread to a 66% confidence range because coarse climate models underrepresent internal variability and model uncertainty (Collins et al. 2013; Lee et al. 2021). The EBM–KF (all variants) does not require a future projection to arrive at a present-day climate state, because it already provides an instantaneous and continual estimate of $\hat{T}_t$. The uncertainty $2\sqrt{\hat{p}_t^T}$ around the posterior climate state $\hat{T}_t$ is used to calculate the probability of threshold crossing (see Fig. 6a) as follows: The probability of the climate state exceeding the policy threshold q is the integral of the probability density of the GMST

[4] We use a 30-yr averaging window nearly everywhere, but for consistency with IPCC practices only in Figs. 6b and 12a–e, we use a 21-yr averaging window for raw ESM simulations. The EBM–KF climate state covariance is still chosen to reflect the uncertainty in the 30-yr average of real-world GMST (see section 2c) using $\mathbf{R}^{\text{const}}$ and $\mathbf{Q}$ matrices reflecting the 21-yr means to match the IPCC definition would be a trivial modification.

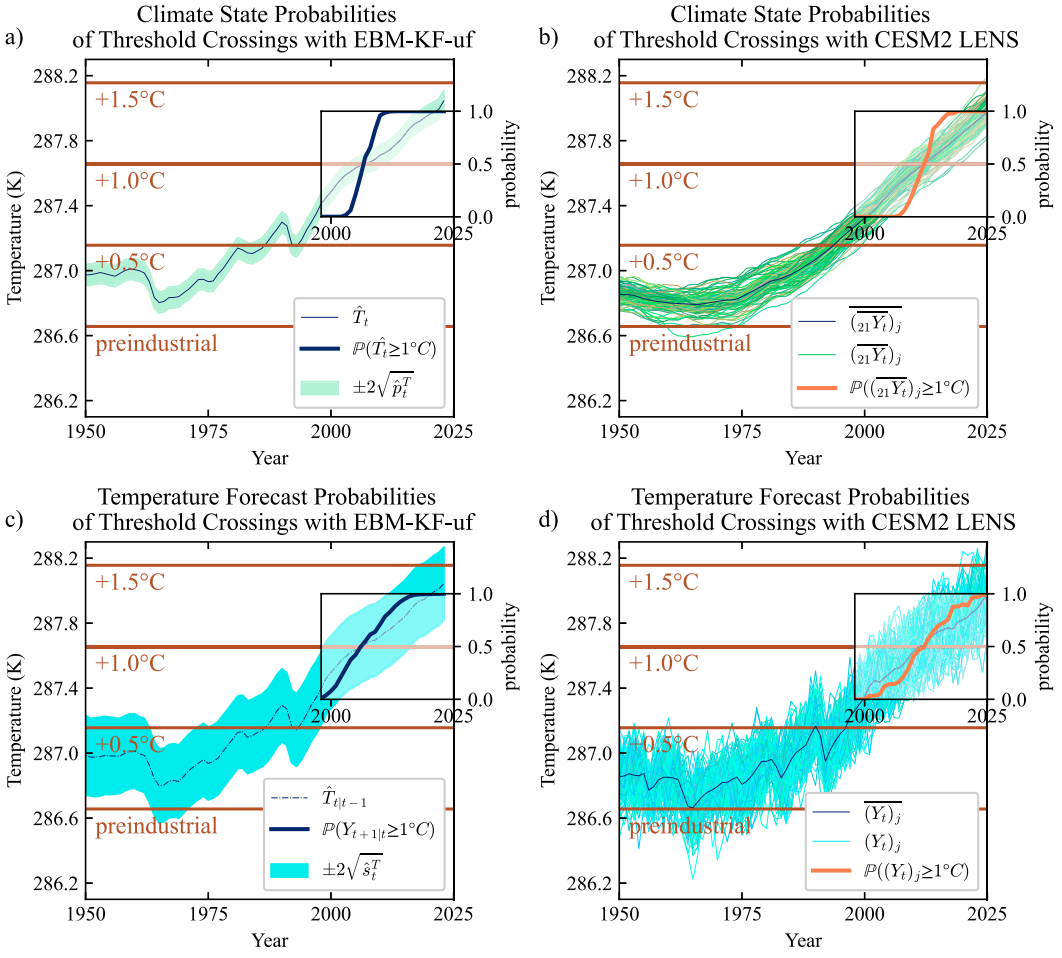

FIG. 6. (a) Climate state crossing policy thresholds: As in Fig. 4, the EBM–KF–uf GMST state estimate (navy blue line) $\hat{T}_t$ and 95% CI of this estimate (light green) $\pm 2\sqrt{\hat{p}_t^T}$ is shown. Policy thresholds (brown lines) are shown at +0.5°, +1.0°, and +1.5°C relative to the preindustrial baseline. The inset axis indicates the +1°C threshold crossing probability (thick navy blue; from 0 to 1). (b) 21-yr running mean of each LENS2 member is plotted in green $(\overline{_{21}Y_t})_j$, along with the ensemble average in black $\overline{(\overline{_{21}Y_t})_j}$. The inset axis shows the fraction of these running means above the +1°C policy threshold. (c) Temperature forecasts: The projected GMST weather 95% CI: $\pm 2\sqrt{\hat{s}_t^T}$ is shown in light blue around the forecast EBM–KF–uf GMST state estimate (navy blue dashed-dotted line) $\hat{T}_{t|t-1}$ The inset axis indicates the prior likelihood that a GMST measurement will be above 1°C (purple; from 0 to 1). (d) Each LENS2 ensemble member is plotted as a blue or green line $(Y_t)_j$ along with the ensemble average in dark blue $\overline{(Y_t)_j}$ The inset axis shows the fraction of these members with annual GMST above the +1°C policy threshold.

climate state above q, equivalently 1 minus the integrated probability below q. The Gaussian cumulative distribution function centered at $\hat{T}_t$ with variance set to $\hat{p}_t^T$, evaluated at q, is this cumulative probability below the threshold:

$$\mathbb{P}(\hat{T}_t \geq q) = 1 - \mathrm{CDF}_{\mathcal{N}(\hat{T}_t, \hat{p}_t^T)}(q) = \frac{1}{2}\left\{1 + \mathrm{erf}\left[(\hat{T}_t - q)/\sqrt{2\hat{p}_t^T}\right]\right\}. \quad (23)$$

For the second interpretation, temperature forecast above the policy threshold, the EBM–KF–uf predicts a relevant window (blue in Fig. 6c) of possible next-year GMST measurements. This EBM–KF window is a Gaussian distribution centered at the projected state $\hat{T}_{t|t-1}$ (dashed dark blue line) with a variance $\hat{s}_t^T$: in other words, a simulated draw from the forecast state. This uncertainty range reflects and encapsulates actual annual GMST measurements, not the uncertainty in the climate. For LENS2, an ensemble of ESMs, the analogous temperature forecast probability is the fraction of unfiltered individual ensemble members $(Y_t)_j$ at year t that are warmer than the policy threshold (blue lines in Fig. 6d).

There is additional ambiguity regarding whether "crossing a policy threshold" should specify an instant or a brief period. Here, we define (based on the 1σ confidence interval or the likely range in IPCC calibrated language) the likely "policy

threshold crossing period" to span from the earliest year when 15.9% of climate states or temperature forecasts exceed the policy threshold to the latest year when 84.1% of climate states or temperature forecasts exceed that policy threshold. A policy threshold crossing instant is the year when the probability of exceeding the policy threshold is nearest to 50% while continuing to increase (or as likely as not to have crossed the policy threshold in IPCC calibrated language).

Regardless of whether an ESM ensemble (see Figs. 6b,d) or EBM–KF–uf (see Figs. 6a,c) is used, the temperature forecast above threshold period (Figs. 6c,d) has a longer duration than the climate state above period (Figs. 6a,b) because the uncertainty/ensemble spread in the annual forecasts is wider than the uncertainty/ensemble spread of the time-averaged states. Both the ESM ensemble (LENS2) and EBM–KF–uf methods report similar policy threshold crossing instants (Fig. 12). Interestingly, the Mount Pinatubo eruption in 1991 resets the +0.5°C threshold crossing repeatedly in both the EBM–KF–uf and raw ESM ensemble (LENS2) by elevated volcanic emissions. As shown below in Fig. 10, the EBM–KF–ta only crosses this threshold once, much like the 21-yr running means of LENS2 (Fig. 6b).

Figure 6 quantifies the probability of crossing policy thresholds as a function of time (dark blue or orange), inset on top of the relevant GMST time series and spread. The EBM–KF climate state estimate in Fig. 6a and annual temperature forecast in Fig. 6c are aligned by year, although these two quantities are in entirely different probability domains. As the EBM–KF state estimate approaches any given policy threshold, the cumulative temperature policy threshold approaches 50% at a policy threshold crossing instant. The +1.0°C policy threshold's crossing instant was in 2010. For the annual temperature forecast in Fig. 6c, the likely crossing period was 2003–15 for +1.0°C. The likely crossing period for the climate state in Fig. 6a is briefer: 2008–12 for +1.0°C. For comparison using LENS2, the analogous climate state thresholds are plotted in Figs. 6b and 6d, although these do not precisely align temporally due to the cold bias of LENS2 during this decade. All threshold crossing periods and instants including future projections under SSP3-7.0 are compared directly in Fig. 12.

c. *The spread from one member: Using EBM–KF to generate an analog for an ESM large ensemble spread*

There are many more past and future climate scenarios that researchers wish to investigate than there are computational resources to run a full large ensemble for each scenario. Fortunately, the EBM–KF can project the climate state distribution when assimilating only one or a handful of ESM simulations, reducing the need for simulating an entire ensemble just to estimate its GSAT spread (similar to approaches for emulation of ensembles of ice sheet models in Edwards et al. 2021; Van Katwyk et al. 2023). Of course, there are interannual differences which persist between runs of the ensemble and skew some climate states persistently cooler and others warmer (supplemental Fig. 6), and an ESM ensemble provides regional information, and these effects are not captured by a Kalman filter framework.

Figure 7a shows the comparison between the EBM–KF–uf GMST climate state uncertainty distribution (light green) and the LENS2 Kalman-filtered ensemble members (orange). This Kalman filtering was performed using the same EBM–KF, momentarily assuming that one of the ensemble members' hindcast was the actual measured temperature and ocean heat content record. Each of the orange lines is a simulation-based estimate of the climate state that is comparable to the blue line of the real observation-based GMST climate state. Sometimes, the (observation-corrected) EBM–KF–uf climate state uncertainty distribution contains the Kalman-filtered LENS2 ensemble members, such as in 1900 and 1935, but at other times, it does not, such as in 1950. In corresponding panels within supplemental Fig. 14, we show the histogram (supplemental Fig. 14a) and quantile–quantile comparison (supplemental Fig. 14b) which both demonstrate a clear bias. This bias indicates that the LENS2 climate state disagrees with the observed climate state within the EBM–KF framework.

We could interpret the Kalman-filtered ensemble spread versus the climate state uncertainty distribution of one ensemble member in a similar fashion. This comparison has a different purpose, as now we are testing whether the EBM–KF can predict the spread of the Kalman-filtered LENS2 ensemble correctly, regardless of whether the LENS2 ensemble matches the observed temperatures. If so, that would indicate that from one ensemble simulation we could effectively predict the time series of the statistical distribution of the ensemble spread spanning all of the other ensemble members. As expected, there is a distribution of results, where some of the ensemble members are close to the center of the distribution and others are outliers.

We can statistically calculate the expected error in our predicted ensemble of Kalman-filtered LENS2 states from a single member versus the true ensemble of Kalman-filtered LENS2 states. Figure 7b shows the error in spread (standard deviation) and error in bias by repeatedly making this prediction of a distribution from single members of LENS2 and comparison to the whole Kalman-filtered LENS2 ensemble. Examining the centroid (cross symbol), this is an unbiased estimate of the distribution (centered left–right, as it should be). However, the ensemble of Kalman-filtered LENS2 is distributed with a standard deviation that is 1.22 times larger than the average prediction from one ensemble member. At worst, it is 1.54 times larger than any single ensemble member's estimate. Figure 7b labels two examples of where one ensemble member predicts the time series of the ensemble spread distribution. A good fit (best quartile) is shown as a circle, and the worst fit is shown as a square. The poorness of fit between two (time series of) distributions is measured via the Kullback–Leibler divergence. Supplemental Figs. 14c and 14d show these two comparisons in more detail. To provide context to this error in predicting the hindcast spread distribution, we repeat this procedure but now using Kalman-filtered observations to predict the Kalman-filtered LENS2 hindcast (green star). These Kalman-filtered observations yield a worse prediction of the Kalman-filtered LENS2 hindcast than the prediction from any ensemble member: larger Kullback–Leibler divergence, inappropriately large spread (large predicted

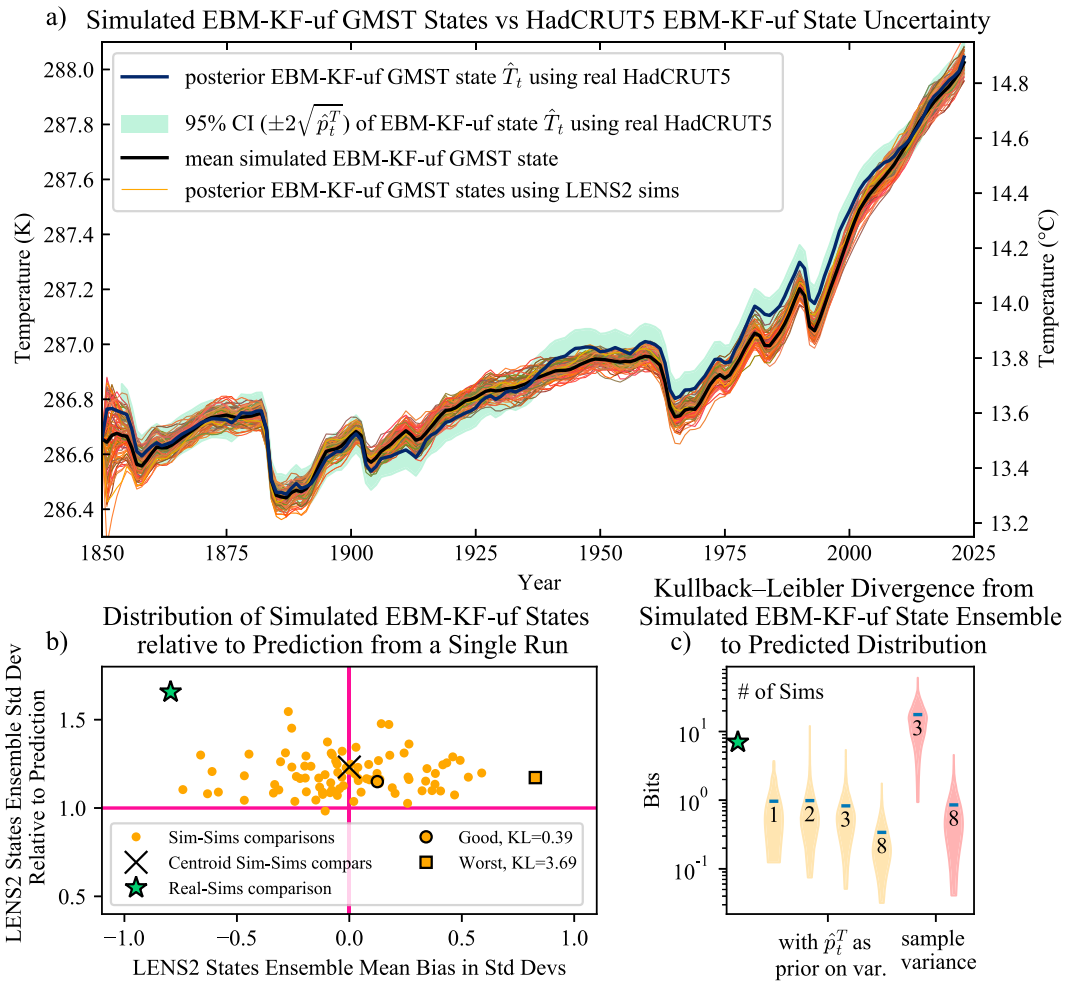

FIG. 7. Comparison of the GMST KF states across the LENS2 ensemble. (a) The EBM–KF–uf $\hat{T}_t$ from HadCRUT5 (thick blue) and its 95% CI (light green) $\pm 2\sqrt{\hat{p}_t^T}$, along with EBM–KF state estimates for each individual CESM2 ensemble member (orange lines) and their mean (thick black line). (b) Climate states and associated uncertainties arising from each of 90 LENS2 simulations and HadCRUT5 are compared to all other LENS2 climate states, and the relative bias and standard deviation of the resulting empirical distributions with respect to a particular ensemble member's $\sqrt{\hat{p}_t^T}$. (c) Violin plots compare the Kullback–Leibler divergence (on a log scale, smaller indicates a better match) for a variety of methods of predicting the LENS2 time-filtered ensemble spread. In yellow, $\hat{p}_t^T$ from 1, 2, 3, or 8 EBM–KF–uf LENS2 runs is averaged and used in combination with the time-varying sample variance. In red, 3 or 8 of these time-filtered ensemble members are used to predict an ensemble distribution from time-varying sample variance alone. Taking a single EBM–KF–uf LENS2 run with $\hat{p}_t^T$ approximates the time-filtered LENS2 ensemble with similar accuracy as taking the time-varying sample variance of 8 time-filtered ensemble members.

standard deviation), and more negatively biased (too cold).[5] From this, we conclude that the error in predicted distribution from one ensemble member is small in comparison with the distance between the model and reality. Thus, this approach is effective in making such comparisons, with a typical bias error in single ensemble member estimate of order ±0.007 K with a range [−0.0265–0.0268]K.

If the ensemble spread distribution can be estimated by one ensemble member ($N = 1$), it seems to be a reasonable hypothesis that adding a few more ensemble members (but still fewer than the original LENS2 ensemble $N \ll 90$) would improve the estimate. Within Fig. 7c, the Kullback–Leibler divergence is utilized to evaluate the utility of using the EBM–KF state uncertainty as a prior estimate of the spread between Kalman-filtered LENS2 ensemble members. At each year, this GMST state variance $\hat{p}_t^T$ is combined in a weighted mean with the variance of a small subset of LENS2 members as written in (24). This is shown in yellow violin plots of Fig. 7c, with a number indicating the number of members

[5] Just one "worst" fit ensemble member prediction (shown in supplemental Fig. 14c) was positively biased to a greater extent than the Kalman-filtered observations' prediction was negatively biased.

taken: $N = 1, 2, 3$, or 8. This mean adds the GMST state variance (averaged across all subset runs) to the sum of squared differences from the mean of the LENS2 subset and then divides by the size of the subset, essentially treating $\hat{p}_t^T$ as an extra sample and taking Bessel's correction:[6]

$$({}_{\text{ens}}\sigma_t^T)^2 = \underbrace{\frac{1}{90}\left\{\sum_{j=1}^{90}\left[(Y_t)_j - \overline{(Y_t)_j}\right]^2\right\}}_{\text{all of LENS2}}$$

$$\approx \frac{1}{N}\left\{\underbrace{\frac{\sum_{j=1}^{N}(\hat{p}_t^T)_j}{N}}_{\text{from KF}} + \underbrace{\sum_{j=1}^{N}\left[(Y_t)_j - \frac{1}{N}\sum_{j=1}^{N}(Y_t)_j\right]^2}_{\text{sample of LENS2}}\right\}. \quad (24)$$

This is a frequentist estimate of variance but the reader may notice that analogously to a Bayesian approach, it employs a growing data term from direct samples of LENS2 and uses the pooled EBM–KF state uncertainty as a prior estimate [note that the last term in (24) vanishes with $N = 1$].

Taking a subset of two members does not improve the predicted distribution of LENS2 using (24), as there is a significant chance that two members which are close together will be selected, incorrectly shrinking the predicted ensemble spread. With three LENS2 members, the predicted distribution slightly improves. The yellow violin plots show the advantage of using the first term in parentheses in (24) provided from the KF in addition to the raw ensemble variance [second term in parentheses in (24)] over just using a raw ensemble variance alone [using only the second term in parentheses in (24)] when the number of ensemble members is small ($N \ll 90$). Without using this prior estimate (and allowing the sample variance to change over time, red violin plots), at least eight LENS2 members are required to generate a predictive ensemble distribution that is comparable to using a single LENS2 member and the Kalman filter $\hat{p}_t^T$ as the ensemble's variance. Figure 7c demonstrates this with three and eight LENS2 members with a time-varying sample standard deviation (red: 3 or 8) again using Bessel's correction. Thus, Fig. 7 shows the power of the parametric Gaussian statistics generated by the EBM–KF over a raw ensemble member sample estimate.

LENS2 runs are more similar to each other than to the real Earth, especially regarding outputs such as OHCA (see supplemental Fig. 13) and Arctic or Antarctic sea ice extent (Horvat 2021; Roach et al. 2020; Rosenblum and Eisenman 2017). In comparison with the observation-assimilating EBM–KF, LENS2 has a profound cold bias from 1940 to 2000 (maximum separation of LENS2 ensemble average in 1983 of 0.262°C, average absolute separation 0.088°C, root-mean-square error 0.085°C, and $r^2 = 0.907$). Also, the current generation of ESMs tends to underestimate the appropriate full spread of climate variability (Lee et al. 2021; Box 4.1). For instance, some weather models use stochastic noise to push their distribution wider than dynamic variation alone (Buizza et al. 1999), and other numerical climate models perturb parameters to achieve the same distribution-widening effect (Duffy et al. 2023; Keil et al. 2021).

In summary, Fig. 7 shows that the EBM–KF climate state based on HadCRUT5 temperatures or based on any one of the LENS2 ensemble members shows the expected level of consistency and (potentially biased) Gaussian differences with the rest of the LENS2 ensemble. Thus, using the EBM–KF on any one of the ensemble members does a good job of estimating the GMST climate state (i.e., averaged over internal variability) and its uncertainty as simulated by the spread of the entire LENS2 ensemble. Further comparisons between the EBM–KF, such as comparing the unfiltered ensemble spread to the forecast prior distribution, would be revealing.

d. *Sampling future projections from a non-Gaussian volcanic distribution*

In standard climate assessments (e.g., Lee et al. 2021), future volcanism has long been singled out as an unknown aspect of projected climate change in any given future year, particularly regarding tropical eruptions' contribution to planetary albedo (Marshall et al. 2022). The forcing of historical period climate models includes the effects of known past volcanoes, while the forcing of future climate models includes only "background forcing from volcanoes," i.e., an expected average forcing value in future years. Applying an average forcing misses the potential impact of individual volcanic events on the global climate state (compare blue lines to black lines in Fig. 8) and underestimates nonlinearities in the climate system. Individual volcanoes can shift crossing thresholds (as section 3b and Fig. 12 show), and so they affect near-term decades (see Figs. 11a,b). However, running an ESM ensemble of sufficient size to explore the low probability of a volcanic eruption in any potential year is computationally challenging using traditional ESMs and has motivated specialized experiments and model intercomparison projects (Timmreck et al. 2018; Zanchettin et al. 2016; Bethke et al. 2017). By contrast, robust sampling of rare events is easily accomplished with the inexpensive EBM–KF. For simplicity, only the volcanic AOD effect is included; the added volcanic contribution of CO_2 and other greenhouse gases is not, as their annual greenhouse gas contribution is minuscule compared to anthropogenic emissions: 20 times smaller in 1900 and 130 times smaller in 2010 (Gerlach 2011). Slightly different climate responses have been modeled to occur when volcanic events occur at different phases of climate oscillation patterns, such as the Pacific decadal oscillation (PDO) and North Atlantic Oscillation (NAO) (Illing et al. 2018). Due to its low-dimensional state space and limited representation of variations about the climate state, the EBM–KF neglects such complexities.

Figures 8 and 9 show the future projections of GMST and OHCA using EBM–KF–uf, including sampling for future volcanoes for two emission scenarios. SSP1-2.6 shown in Figs. 8a and 9a has anthropogenic CO_2 emissions that sharply decline after 2020 to keep GMST rise below 2 K (van Vuuren et al. 2007, 2017). SSP3-7.0 shown in Figs. 8b and 9b is a higher

[6] Bessel's correction involves a prefactor of $1/(N-1)$ when taking a sample variance. We use N instead of $N+1$ in the prefactor denominator outside of the parentheses on the second line approximation of (24), even though the parentheses include $N+1$ estimates including the KF contribution.

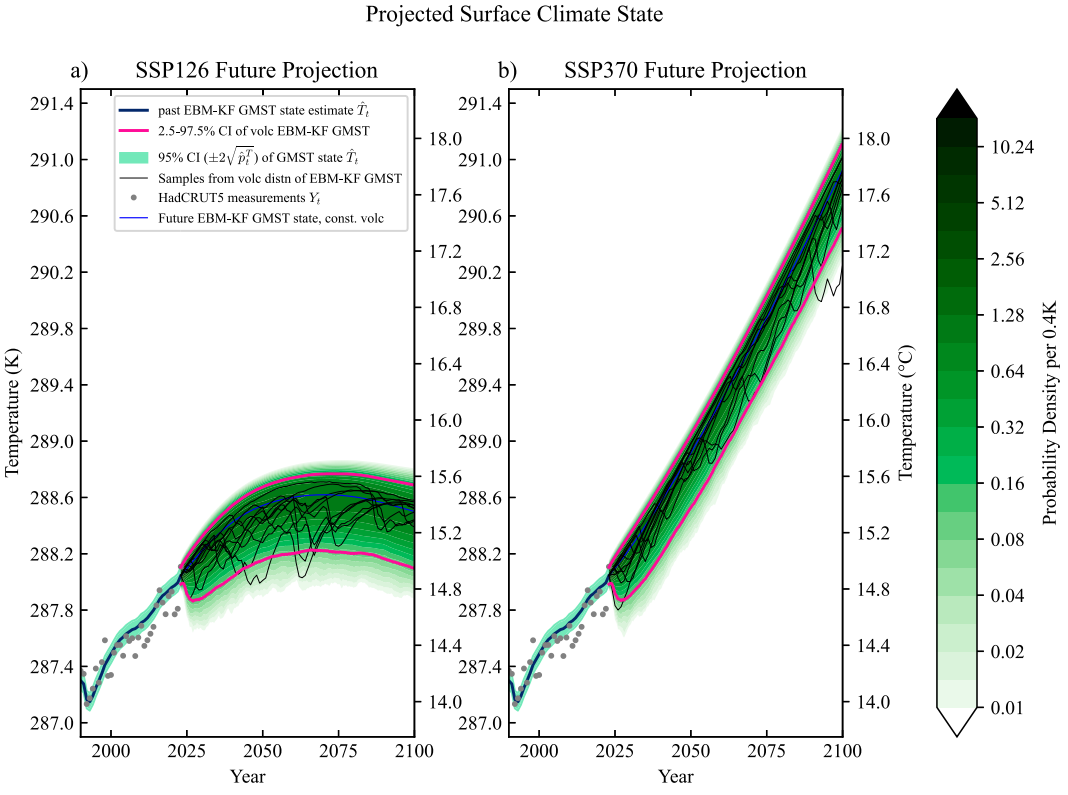

FIG. 8. Future GMST projections of (a) SSP1-2.6 and (b) SSP3-7.0 scenarios using sampled measures of volcanic activity and greenhouse gas concentrations calculated according to MAGICC7.0 (Meinshausen et al. 2020). The historical Mount Pinatubo eruption in 1991 is shown in the lower-left corner of both graphs for scale. About 10 of the sampled 6000 potential future climate states from the volcanic probability distribution are graphed (thin black), along with a future climate state projection that uses constant volcanism with average over the entire 1850–2014 record $\overline{\text{AOD}_t} = 0.0123$ (blue). The probability density function formed by taking the summation of all sampled Gaussian kernels at each time point is shaded in green on a logarithmic scale (note these probability densities are not probabilities so they can exceed 1). Pink lines show the 2.5%–97.5% CI of these probability density functions, which are very asymmetrical (negatively skewed) due to the sampled volcanic eruptions' impact on GMST.

anthropogenic emission scenario in which CO_2 emissions double by 2100 (Fujimori et al. 2017).

Figures 8 and 9 show that the volcanic ensemble probability density is not symmetrical for GMST—there is a much longer tail on the cooler side because of intermittent cooling by volcanic aerosols. In Fig. 8, the cooler side of the distribution takes a few years (2024–26) to fully expand out because large eruptions generally did not produce their maximal effect on AOD_t (and thus the GMST climate state) until 1–2 years after the eruption (and no major eruptions are ongoing at present). In any single future sampled scenario of volcanic eruptions, there is usually a significant gap between major volcanic eruptions (as our model indicates by the thin black lines in Figs. 8 and 9), representing an autocorrelation (see appendix B). These gaps are not reflected in the 95% CI (pink) which samples thousands of independent futures. Indeed, the volcanic eruptions dominate the future uncertainty over the slowly growing GMST climate state uncertainty and rival or exceed the scenario uncertainty up until about 2050 (assuming known model parameters; Fig. 11a). By contrast, the LENS2 using "constant background" future volcanism has a symmetrical distribution about the mean for future projections of the same SSPs (supplemental Fig. 6, right of dashed line). The effects of volcanism on OHCA (Fig. 9) are much smaller than on GMST (Fig. 8), but there is still a longer tail toward the cooler, low OHCA side.

Regarding future GMST policy threshold crossings, the volcanic eruptions widen the likely threshold crossing periods and lessen the difference between the climate state above and the annual temperature forecast above interpretation periods. Occasionally, major volcanic eruptions can cause a policy threshold to be "uncrossed." For example, if we were to examine one arbitrary policy threshold, 0.27°C above preindustrial, the 30-yr running average GMST uncrosses this global warming policy threshold (crossing first in 1944, then dipping back under the threshold to uncross in 1956, and crossing again in 1965) because the eruption of Mount Agung reduced GMST for about a decade after its eruption in 1963. Because the 30-yr running average incorporates future information, it anticipated the future eruption and started cooling in the late 1950s. The EBM–KF–uf, EBM–KF–ta, and LENS2 ensemble average similarly warm, cool, and then warm again in this period, although the cooling periods follow the causative Mount Agung

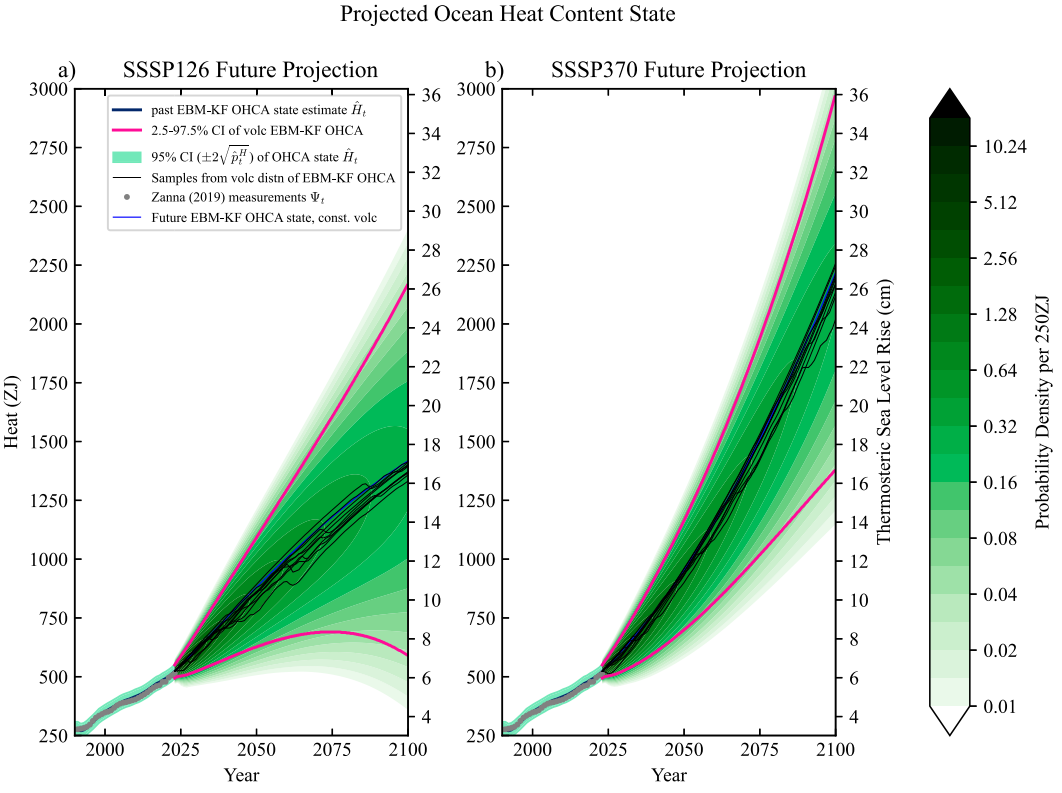

FIG. 9. Future OHCA projections of (a) SSP1-2.6 and (b) SSP3-7.0 scenarios using sampled measures of volcanic activity and greenhouse gas concentrations calculated according to MAGICC7.0 (Meinshausen et al. 2020). About 10 of the sampled 6000 potential future climate states from the volcanic probability distribution are graphed (thin black), along with a future climate state projection that uses constant volcanism with average over the entire 1850–2014 record $\overline{\text{AOD}}_t = 0.0123$ (blue). The probability density function formed by taking the summation of all sampled Gaussian kernels at each time point is shaded in green on a logarithmic scale (note these probability densities are not probabilities so they can exceed 1). Pink lines show the 2.5%–97.5% CI of these probability density functions, which are only slightly asymmetrical because the sampled volcanic eruptions have a much smaller impact on OHCA.

eruption (Fig. 10). In contrast, following the Mount Pinatubo eruption in 1991, the EBM–KF–ta and the 30-yr running average do not uncross the 0.5°C above preindustrial threshold, whereas the EBM–KF–uf and LENS2 ensemble average do. These distinctions are lost when using "background volcanic activity" or "human attributable warming" to estimate policy threshold crossings (Forster et al. 2023; Haustein et al. 2017).

Across many future simulations, the dynamic model Jacobian matrix $\mathbf{\Phi}_t$ happens to remain nearly constant at values of

$$\mathbf{\Phi}_t \approx \begin{bmatrix} 0.893 & 0.000\,253\ \text{K}\Big/\dfrac{\text{W yr}}{\text{m}^2} \\ 11.1\dfrac{\text{W yr}}{\text{m}^2}\Big/\text{K} & 0.999 \end{bmatrix},$$

which are nearly unit triangular. Due to this Jacobian matrix shape and the 0.893 factor, $\hat{p}_t^T$ grows sublinearly, with yearly growth less than the upper-left (GMST-exclusive) component of $\mathbf{Q}$: $0.010\,99/30\ \text{K}^2 = 0.000\,37\ \text{K}^2$ [see (22)]. Over a 78-yr future projection (2023–2100), the GMST state 95% confidence interval $\pm 2\sqrt{\hat{p}_t^T}$ only grows from ±0.0620 to ±0.1619 K. This 2.6-fold increase is small over the twenty-first century compared to the GMST dips that occur under volcanic eruptions (see Figs. 8 and 10). The effect of volcanoes on historical state (Figs. 3 and 4) and future projections (Fig. 8) is therefore worthy of specialized treatment in addition to measurement uncertainty and internal chaotic variability (see Fig. 11 in the discussion below). In contrast, the OHCA component of the state uncertainty 95% confidence interval $\pm 2\sqrt{\hat{p}_t^H}$ grows exponentially due to the 11.1 value in the lower-left entry of $\mathbf{\Phi}_t$, and volcanoes have a negligible effect on projected OHCA trajectories (see Fig. 9). The ocean state uncertainty 95% CI = $\pm 2\sqrt{\hat{p}_t^H}$, initially at ±29.4 ZJ in 2023, balloons to ±801.4 ZJ by 2100. These uncertanties grow similarly in all SSP projections, for instance under SSP5-85 (see supplemental Figs. 16 and 17) the final nonvolcanic uncertainties are very slightly larger: $\pm 2\sqrt{\hat{p}_{2100}^T} = \pm 0.1650$ K and $\pm 2\sqrt{\hat{p}_{2100}^H} = \pm 811.6$ ZJ.

4. Discussion

The EBM–KF climate state estimate resembles other standard estimates of climate state, but it has advantages they do

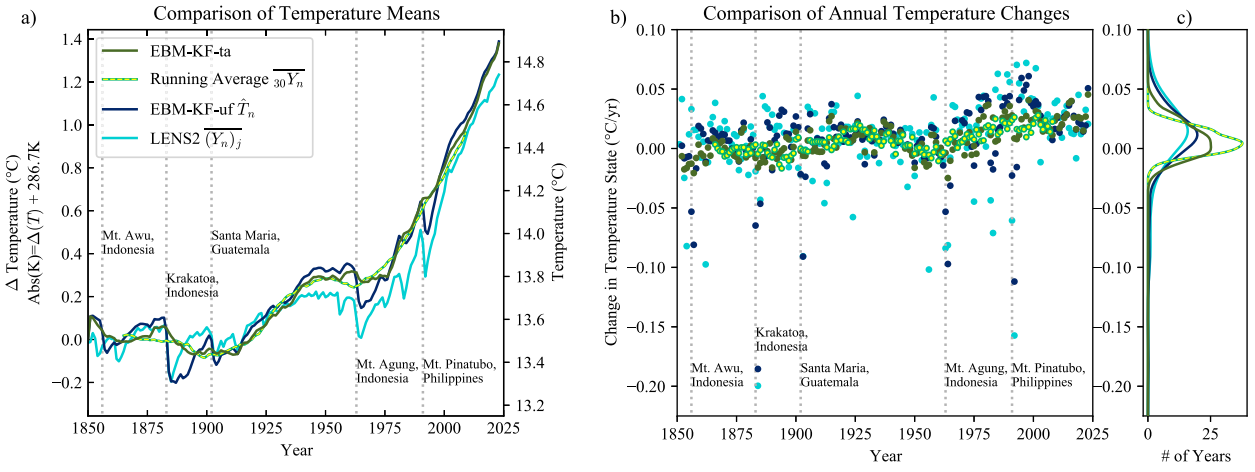

FIG. 10. GMST climate state comparisons. Major volcanic eruptions are labeled in both (a) and (b) with dotted vertical light gray lines. In all panels, the 30-yr averaged GMST (yellow-green dashed) is close to the EBM–KF–ta state (dark green), whereas the EBM–KF–uf state (navy blue) resembles the ensemble mean of GSAT in the LENS2 simulations (sky blue) in responses to volcanic eruptions. (a) Direct GMST temperatures of the 4 climate states. (b) The innovations (derivatives) are plotted against time. Colors are the same as in (a). (c) A smoothed empirical density with respect to yearly change in temperature is linked to (b). (This empirical density is simply an approximation of a histogram, and the kernel densities are not provided by elements of the KF as in Figs. 4b and 5b).

not share. The EBM–KF algorithm, because of its relationship with a forward or blind EBM, can be projected forward in time without temperature observations and thus can be used in many situations. Unlike an ESM, the EBM–KF benefits from data assimilation due to its Kalman filter nature and thus remains close to observations or synthetic data (e.g., the examples in section 4 of reproducing the LENS2 from a few ensemble members). The OHCA component is particularly sensitive to assimilated observations (see Fig. 3b), largely because of the reduced understanding of the ocean dynamics that drive deep ocean heat uptake compared to atmospheric radiative feedbacks. The EBM has a correspondingly simpler model of ocean physics. Unlike an ensemble Kalman filter approach that can reweight a full-physics ESM ensemble toward observations, the EBM–KF has negligible computational cost and can thus examine rare, long-tailed events such as volcanoes. Additionally, tuning of the EBM parameters and uncertainty quantification of these results can benefit from the Kalman filter literature and algorithms to optimize our Kalman filter parameters.

a. *Comparison to previous estimation methods of the climate state*

Although they are different types of average, a direct comparison (Fig. 10) of the state estimated from the EBM–KF (Fig. 4) and that estimated by the 30-yr running mean (Fig. 1) and the LENS2 ensemble mean (supplemental Fig. 6), the EBM–KF has slightly more year-to-year variation than the 30-yr mean and less than the LENS2 ensemble mean. Departures from the main Gaussian cloud in all methods represent volcanoes. The five largest eruptions which caused the largest dip in the EBM–KF state are labeled in Fig. 10, corresponding to the five peaks in $\mathrm{AOD}_t \geq 0.06$ plotted in Fig. B1a in appendix B. The climate effects of these major tropical volcanic eruptions have been studied extensively (Jones and Kelly 1996; McCormick et al. 1995). Note for the eruptions listed, plus many others, the dips in the EBM–KF mean state correspond with dips in the sample mean of the LENS2 simulations.[7]

Based on this interpretation of Fig. 10, we now see that the LENS2 ensemble average (light blue) is closer to the EBM–KF–uf (navy blue, with uf abbreviating unfiltered AOD forcing) regarding sensitivity to volcanoes than the 30-yr running mean (yellow green). In response to this, we will distinguish two variants of AOD_t forcing: one that directly uses the annual measured values of AOD_t (EBM–KF–uf, navy blue, as above) and one that takes a 15-yr trailing average combined in equal weight with the overall time series AOD_t mean (EBM–KF–ta in dark green, ta abbreviating "trailing average"). This trailing average is the best real-time estimate for the centered 30-yr average of AOD_t that can be made before observing the 15-yr of future AOD_t and is displayed as a green line in Fig. B1a. Figure 10 shows that this trailing average preparation of the AOD_t forcing brings the EBM–KF–ta (dark green) close to the 30-yr running mean (yellow green) regarding sensitivity to volcanoes. Their maximum separation was in 1962 with the 30-yr running average −0.073°C cooler; otherwise, their average absolute separation was 0.025°C, root-mean-square error 0.030°C, and $r^2 = 0.987$. We experimented with a true centered average rather than a trailing average, and the results did not improve (see supplemental Fig. 20).

It is beyond the scope of this paper to detail the characteristics of the large and growing variety of "mean state" definitions, but a summary is useful. For all methods we have

[7] However, the earliest AOD values provided by Sato et al. (1993) also demonstrate a major spike at 1856, which is not reflected in the LENS2 simulations. This may correspond to either the 1856 eruptions of Koma-ga-take, Japan, or Mount Awu, Indonesia, and we labeled this with the latter eruption because tropical volcanic eruptions typically have a much larger climate impact (Marshall et al. 2022).

examined regarding the GMST (30-yr mean—Figs. 1 and 10; EBM–KF–uf—Figs. 4 and 10; LENS2 model ensemble mean—Fig. 10, supplemental Fig. 6; purely statistical methods—supplemental Figs. 4c, 4d and 5; EBM–KF–ta—Fig. 10), the differences in the estimated climate state are relatively small in available years (on the order of 0.1 K—see supplemental Fig. 7, column 1). The largest differences seen between these methods lie in the spread of the changes from year to year (see supplemental Fig. 7, column 2) and persistent mean anomalies relative to observations, particularly concerning volcanism and biases in the forward, blind LENS2 ensemble (see supplemental Fig. 7, column 4).

The different variants of the EBM–KF forcing preparations apply to different intended applications. Preparation refers to providing something for use; in this case, aerosol forcing is filtered in time. When we are trying to directly match the behavior of ensembles such as LENS2 (light blue), the EBM–KF–uf (navy blue) is the correct choice. As noted in section 3a, LENS2 versus EBM–KF–uf is useful for examining biases in LENS2, and in Fig. 10b, we see their responses to volcanic events are very similar. When we are trying to emulate 30-yr climate normal (up to the present), then the EBM–KF–ta is the best estimator based only on information available before the present. When we are trying to project both the weather and climate states without bias, for next-year predictions and beyond (so AOD_t will be unavailable), the optimal method is to run many predictions using the EBM–KF–uf and sampling a volcanic probability distribution, as in section 4d.

For policy thresholds, it is important to actually sample across volcanic probability distributions rather than use background volcanism, as all climate state estimates are capable of "uncrossing" a threshold directly because of a volcanic eruption. The EBM–KF–uf and LENS2 are just more sensitive to such eruptions than the EBM–KF–ta and 30-yr running mean. Section 4d notes that Mount Agung caused all four climate state estimates to uncross the 0.27°C threshold, while Mount Pinatubo caused only the EBM–KF–uf and LENS2 climate state estimates to uncross the 0.5°C threshold. Similarly, future eruptions may cause a policy threshold to be uncrossed, and only sampling the volcanic probabilities anticipates the odds of such uncrossings.

The primary distinction of our EBM–KF method and all existing alternative definitions is the integrated quantification of uncertainty. While many methods exhibit a relationship between the mean state and "sample" that varies in time, the EBM–KF quickly converges to a stable GMST state uncertainty of $\pm 2\sqrt{\hat{p}_t^T} = \pm 0.065$ K. The RTS filter (supplemental Fig. 2) has a narrower ± 0.044 K uncertainty but involves past and future information in a given year as does the 30-yr climate normal. Our choice of method was motivated by the mathematical compatibility between the governing equation for a Kalman filter and that of an EBM, which is not true of many alternatives, e.g., a Butterworth filter or Bayesian changepoint analysis or a more complex dynamical model such as an ESM. We also emphasize again that our EBM–KF infers the climate state directly via yearly signal processing, which is faster and less complex than simulating future weather over the next 15 years by calculating many 30-yr means. In the next section, we discuss how the EBM–KF uncertainties compare to those of ESM ensembles.

b. *Comparison to Earth system models (CESM2 large ensemble and CMIP5)*

The chief advantage of EBM–KF over an ensemble of ESMs is that it replicates many statistical features while being trivial to compute. Figure 7 suggested that any of the ensemble members or the observed temperature record could be used together with EBM–KF to recreate the climate state, but now we examine whether we can anticipate or improve on the ensemble statistics without a single ensemble member.

First, we examine the basic statistical character of LENS2. The distribution of annual differences in all ESM trajectories from the ensemble mean is remarkably close to Gaussian (see supplemental Fig. 10a). Therefore, again due to the central limit theorem, this fundamental assumption of the EBM–KF is also met by GSAT as simulated by the CESM2. The standard deviation rises insignificantly with time in LENS2 over the entire simulation duration ($p = 0.168$). Before 2065, this rise is significant ($p = 1.2 \times 10^{-6}$; see supplemental Fig. 10b) while relatively small (linear trend $r^2 = 0.105$ and only 8.9% rise in ${}_{\mathrm{ens}}\sigma_t^T$ from 1850 to 2065). The time-averaged standard deviation (note 1σ is used within this section to describe the bulk of the ensemble) of 0.127 K was close to both chosen total GMST-exclusive (top-left) measurement noise from $\mathbf{R}_t$ {range [0.107–0.136] K; see section 2c, (21)} and half the converged values in the EBM–KF of $\sqrt{\hat{s}_t^T}$: 0.13 K in 1865 and later 0.112 K in 2000. Examining skewness and kurtosis, the uncertainty in climate as indicated by the distribution of simulations about the LENS2 GSAT ensemble mean is not meaningfully altered as the climate warms (see supplemental Figs. 10c,d).

Next, we evaluate how well the LENS2 captures the overall shape of the observed HadCRUT5 temperatures, given that it is not constrained directly by these observations. The absolute temperature of the LENS2 runs had to be revised down by a full 1.75 K to match its ensemble 1850–1949 100-yr average GMST to HadCRUT5. Other authors have also noted this high absolute temperature as well as the high climate sensitivity of CESM2, the model used in LENS2 (Feng et al. 2020; Gettelman et al. 2019; Zhu et al. 2022). Recall HadCRUT5 was recalibrated to a 1960–89 30-yr climate normal (Jones and Harpham 2013) of 13.85°C (287.00 K), and the LENS2 average has a slightly lower temperature during this 30-yr climate normal of 13.71°C (286.86 K).

Comparing the EBM–KF projections (Fig. 8) with LENS2 projections (supplemental Fig. 6), both Fig. 8b and supplemental Fig. 6 trace out roughly the same shape, as both are forced by warming SSP projections. The largely symmetric uncertainty in the LENS2 about the ensemble mean is driven by dynamical instability. This is fundamentally different from the EBM–KF, which in addition to scaled, data-driven weather dynamical uncertainty also samples a noisy distribution of volcanic eruptions, yielding an asymmetrical distribution. LENS2 projections based on SSP3-7.0 achieve a slightly higher mean temperature in 2100 (291.3, +4.6 K warming) than the equivalent EBM–KF

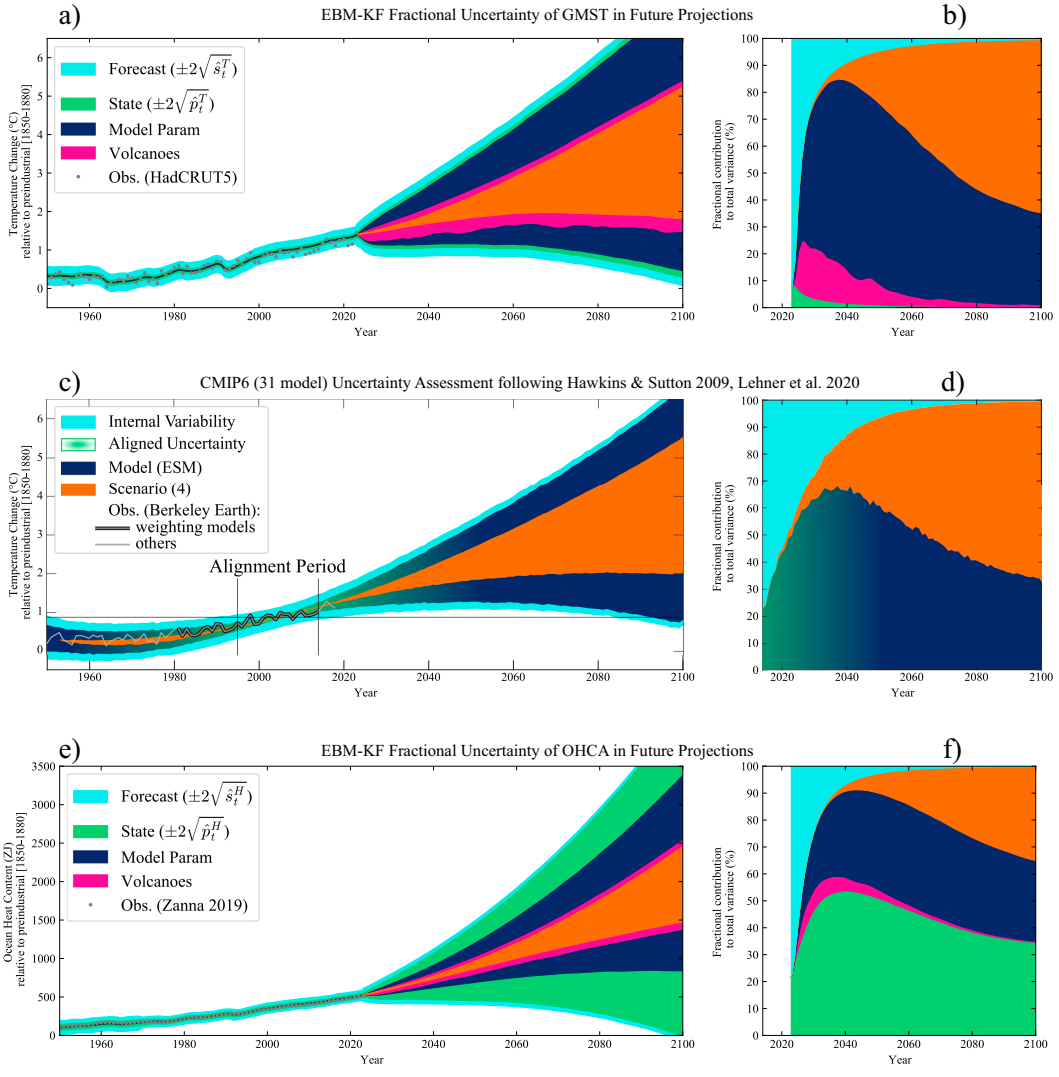

FIG. 11. "Hawkins plots" (Hawkins and Sutton 2009) of the sources of uncertainty (95% CI on the left and fractional variance on the right) in future projections, with (a),(b) the GMST state projections from the EBM–KF, (c),(d) global mean surface air temperature GSAT from CMIP6 (following Lehner et al. 2020), and (e),(f) the OHCA projections from the EBM–KF. Internal dynamical or forecast variability is colored light blue in all figures, and while initially dominant (≥80%), it quickly falls off within the first decade, to eventually be replaced with emissions scenario uncertainty in orange. The smoothed CMIP6 ESMs have been both calibrated to the same baseline over an alignment period (1995–2014) and weighted according to their correlation with a longer trend (1981–2014). The spread of smoothed ESMs after both alignment and weighting is colored in light green in the alignment window, as it is roughly analogous to the climate state covariance within the EBM–KF. This aligned uncertainty melds into model uncertainty (navy blue) as the ESMs diverge in (c) and (d). Future uncertainty related to volcanoes (in magenta) is negatively skewed and very important in the first 3–25 years of the EBM–KF's projections of GMST.

projection (290.9, +4.2 K warming; see Fig. 8b), despite the LENS2 simulations being cooler throughout most of the twentieth century and early twenty-first century (see Fig. 10a). Across all CMIP6 models (Lee et al. 2021; Tebaldi et al. 2021), the projected warming under this scenario is 3.9 K with 5%–95% range [2.8–5.5] K, closer to the EBM–KF projection.

Continuing beyond LENS2 to compare against the multimodel CMIP6 ensemble, a projected uncertainty decomposition is created following Hawkins and Sutton (2009) and Lehner et al. (2020) in Fig. 11. In their adopted uncertainty decomposition method, ESMs are smoothed with fourth-degree polynomials, but here we show 95% CI and annual (rather than decadal) internal variability. By our methods in section 4c, the EBM–KF adds the new entry of volcanic emissions uncertainty into this picture (Figs. 11a,c, pink). A second advantage is that the climate state uncertainty (due to the cumulative reliability of measurements with respect to a model, green) and the model uncertainty (due to the confidence in

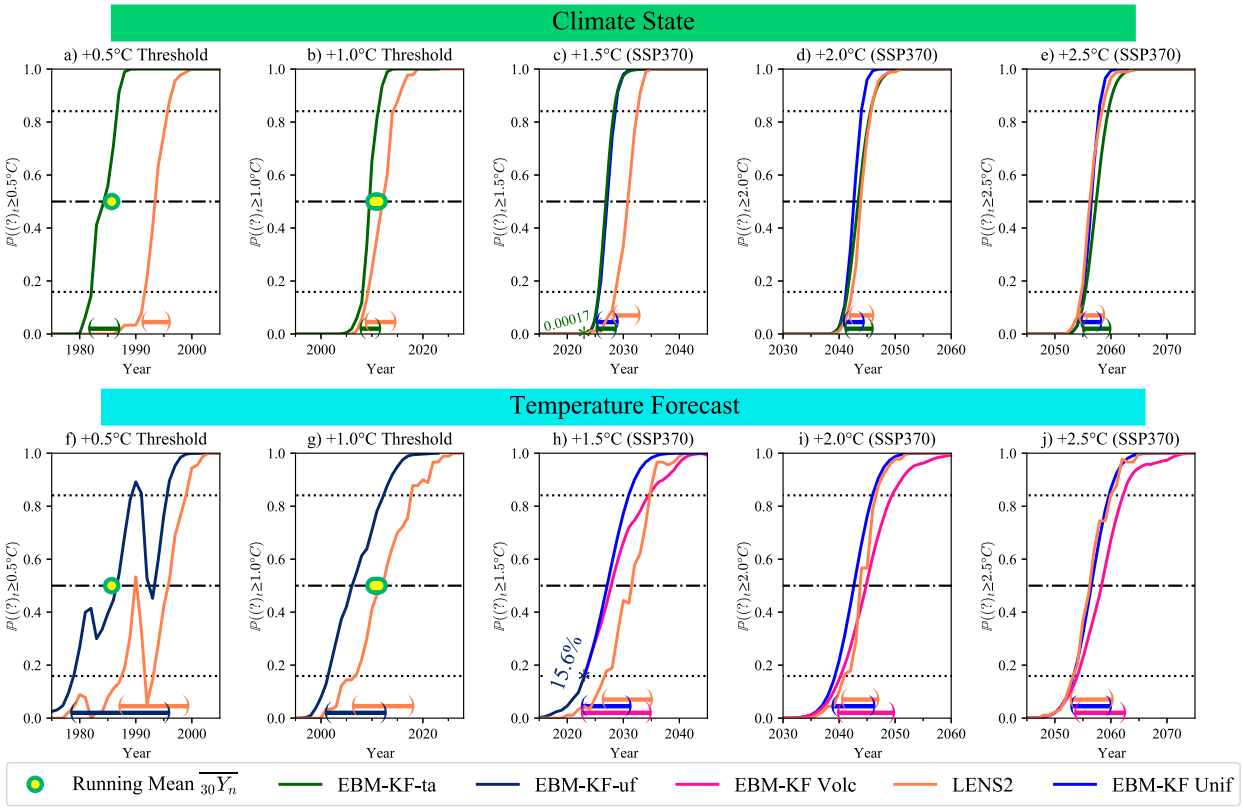

FIG. 12. Comparison of 0.5–2.5 K GMST policy threshold crossing probabilities for various relevant preparations of the EBM–KF and CESM2 LENS simulations (orange). (a)–(e) Comparison of climate states in the EBM–KF with 21-yr averages of the LENS2 simulations. (f)–(j) Comparison of next-year temperature forecasts from the EBM–KF directly with the LENS2 simulations. Recall from section 3b that these are the integrated probabilities of the GMST climate states or temperature forecasts below that policy threshold, with policy crossing instants when this probability is at 0.5. Historical EBM–KF–uf estimates of temperature forecasts are in dark blue in (f) and (g) (see Fig. 6c). EBM–KF–ta states (climate state thresholds) are shown in green in (a) and (b). These EBM–KF–ta state estimates come the closest to matching the instants (yellow-green dots) when the 30-yr running average crossed the 0.5°C threshold in 1985 (or very likely from a linear trend will have crossed the 1.0°C threshold in 2010 or 2011). Two versions of future EBM–KF state estimates are shown: in (h)–(j) an amalgamation of samples in pink from the volcanic distribution shown in Fig. 8 and in (c)–(e) and (h)–(j) a single run in bright blue with uniform $\mathrm{AOD}_t = \overline{\mathrm{AOD}}_{1850-2024} = 0.0123$ mirroring how LENS2 treats volcanism. In future climate state projections in (c)–(e), samples of future volcanism are preprocessed according to EBM–KF–ta. Policy threshold crossing windows (thick bracketed lines at the bottom) are also shown.

the model structure and parameters, blue) can be distinguished, whereas in CMIP6, they are combined because they are calculated together from the multimodel ensemble spread (green-blue shaded region). For simplicity, we estimated the model and parameter uncertainty of the EBM–KF by just varying the cloud feedback parameter [samples from $\mathcal{N}(0.42,\ 0.36^2)$, based on Fig. 7.10 and Table 7.10 of AR6 (Forster et al. 2021)] and the ocean heat conductivity [samples from $\mathcal{N}(0.67,\ 0.15^2)$, based on Geoffroy et al. (2013b)]. Incomplete understanding of cloud feedback is a primary source of uncertainty within ESMs, leading to diverging predictions within CMIP6 (Ceppi and Nowack 2021; Zelinka et al. 2017). Even though the cloud and OHCA dynamics of EBM–KF are oversimplified (Cheng et al. 2022; Newsom et al. 2023) and sparse long-term records yield disparate OHCA reconstructions before 2005 (Gulev et al. 2021, Fig. 2.26), the GMST and OHCA uncertainty ranges from the EBM–KF can help quantify beyond what is known how to estimate in CMIP6. Were we to go further and assimilate the CMIP6 temperature and OHCA records into the EBM–KF (as done for the LENS2 in section 3c), these additional quantifications of uncertainty could be brought to bear on the CMIP6 ensemble, akin to fair-calibrate (Smith et al. 2024).

Regarding the various types of climate policy thresholds, the LENS2 can be used to generate very similar results to the EBM–KF (Figs. 6 and 12). Differences in absolute probability and policy threshold crossing instants reflect differences in the modeled climate states: particularly that the LENS2 ensemble was slightly cooler than the EBM–KF model after correcting to the same preindustrial temperature, so policy thresholds were crossed 3–5 years later (Fig. 12). The eruption of Mount Pinatubo caused the policy threshold of +0.5 K to be crossed in three instants within the EBM–KF model, because this eruption temporarily cooled the climate state back below the

threshold temperature. The first of these EBM–KF crossings coincides very closely with the (single) policy threshold crossing instant of the 30-yr running mean (indicated by large yellow-green dots). The 21-yr running averages of the LENS2 simulations only crossed the 0.5°C threshold once. Future threshold crossings (1.5°, 2.0°, and 2.5°C) under the SSP3-7.0 projection scenario show close temporal alignment in the threshold instants between LENS2 and the EBM–KF estimates that sample for volcanic uncertainty. Although shifted, the overall shapes of these cumulative distribution functions and spans of the threshold crossing windows are more similar between LENS2 and a single EBM–KF future estimate that like LENS2 keeps AOD constant (see Fig. 12). In contrast, the EBM–KF–uf sampling over potential volcanic futures has a long tail (pink lines, lower row) regarding temperature forecast thresholds, extending the later bound of the crossing period by about 5 years, because there remains a modest chance that a large volcano will erupt and tip the temperature forecast below that threshold.

c. *Potential issues with the EBM–KF and future extensions*

This first climate Kalman filter does not generate regional temperatures nor other essential climate variables, such as precipitation. These variables are often highly non-Gaussian and may require an understanding of regional "dynamical tipping points" or other important nonlinear process aspects of climate change. Additionally, this two-component EBM–KF lacks a "memory ENSO state" to allow for the prediction of 2–7-yr quasi-periodic El Niño events (Hu and Fedorov 2017), and without such a state, our EBM–KF wrongly assumes that weather innovations z_t^T have no autocorrelation. Therefore, this first EBM–KF is far from generating the information required to replace many aspects of large ensembles. An expanded global climate state vector, including precipitation, seasonal temperature, or eigenvalues of spatially decomposed principal components (e.g., ENSO modes), might be appended into this statistical framework with appropriate physical forward modeling (Yang et al. 2018).

Astute readers may note the estimated climate state and covariance within the EBM–KF are influenced by the choice of reconstructed HadCRUT5 GMST and Zanna et al. (2019) OHCA. With only minor modifications, the EBM–KF method could be used with multiple annual reconstructions at the same time, e.g., GISTEMP GMST (Lenssen et al. 2019) or other OHCA reconstructions (Cheng et al. 2017; Ishii et al. 2017), considering each as only an estimate of the true GMST or OHCA (Willner et al. 1976). Reconstructions of sea level rise could be used from different sources as further constraints on OHCA (Fox-Kemper et al. 2021; Palmer et al. 2021, 2018b).

Here, we use preselected, constant parameters at their published values in the EBM–KF. However, methods for tuning parameters, including time-dependent parameters, within Kalman filters are much more extensively studied mathematically (Chen et al. 2021, 2018; Zhang and Atia 2020) than the methods thus far applied in climate sciences to diagnose parameter variations within EBMs or ESMs [e.g., the regional effects diagnosed in Armour et al. (2013) and the global effects found by Gregory and Andrews (2016)]. Our EBM–KF hybrid presents an opportunity to adopt KF parameter optimization methods for the GMST and OHCA projection optimization problem. In a preliminary experiment with a Bayesian parameter search to give better estimates of the coefficients in the blind EBM, the prior distributions of these coefficients (rather than point estimates) were extracted from climate science literature, followed by a Metropolis–Hastings search. Several parameters required further care or tuning to achieve desired constraints (e.g., balanced energy transfer in the preindustrial climate), such as the main longwave radiation coefficient and the temperature exponent. However, identifiability and overfitting are challenges of this approach and deserve more attention than the scope of this paper allows. In this first illustration of the system, opportune imperfections in the point estimates given by literature sources allow demonstration of the course-corrective properties of the EBM–KF (Fig. 4).

d. *Policy utility*

Has the climate already passed the 1.5°C policy threshold? Real-time, accurate knowledge of policy threshold crossing will allow for more prudent planning and more comprehensible communication of climate science to the public. For instance, while the "Climate Clock" (https://climateclock.world) intends to communicate the urgency of the climate crisis with a countdown to the estimated expenditure of our remaining carbon budget, only a static date informs it. In contrast, an EBM–KF threshold reflects the most recently measured state of the Earth system and up-to-date emissions and presents limits on future emissions. As can be seen in Fig. 12h, there was a substantial (15.5%) chance that the 2023 GMST measurement could have exceeded the 1.5°C threshold. Indeed, the HadCRUT5 2023 number came close at 1.45°C, and others with slightly different methodologies reported 1.52°C above preindustrial (Burgess 2024). Rather than relying on sponge proxy data to posit that the climate state has exceeded this threshold (McCulloch et al. 2024), or using an overlap window between ESM projections and smoothed observations that is challenging to translate into probability (Hausfather 2024a,b), the EBM–KF–ta can simply give a p value (subject to our chosen point parameters and their quantifiable uncertainties in Fig. 11). By this method, we have not yet crossed the 1.5°C climate policy threshold: The EBM–KF–ta states (Fig. 12c) that there is $p = 0.000\,17$ that the climate state exceeded 1.5°C in 2023.

Climate modeling with ESMs is slow, computationally expensive, and typically performed with blind models that do not respond to the latest observations. The relatively simple question "How did the COVID-19 lockdowns and the 8% reduction in CO_2 emissions impact the near-term climate?" required hundreds of ESM simulations to yield a statistically insignificant answer (Jones et al. 2021). That sort of modeling effort, arriving months or years after the question was posed, is an unsatisfactory prize for many aspects of communication and decision-making for the annual profit or election term. The EBM–KF

can produce the result that an 8% emissions reduction over 2 years cools the climate state by ≈0.0017 K and pushes back subsequent threshold crossing time by 1.2 months—an insufficient reduction in climate change, but at least precisely and rapidly quantified. The EBM–KF is sufficiently fast that, once fully calibrated, it could be easily embedded as an interactive web tool for such exploration. This demonstrates that like human attributable warming the EBM–KF is an "antifragile index" and therefore of greater use to planning climate mitigation strategies (Otto et al. 2015; Haustein et al. 2017; Forster et al. 2023).

Additionally, Kalman filters are often used for process control (Lee and Ricker 1994; Myers and Luecke 1991), and in this case, an EBM–KF could be used to optimize climate change mitigation or intervention strategies (Filar et al. 1996; Kravitz et al. 2016; MacMartin et al. 2014). For instance, within carbon offset, carbon sequestration, and geoengineering accreditation markets, credits could be assigned based on the projected delay in crossing policy thresholds. Once a space of potential climate solutions has been defined, the EBM–KF can work seamlessly with a variety of optimizers to find the maximum climate benefit at the lowest societal cost.

5. Conclusions

The EBM–KF–ta presented in this paper takes the best features from a 30-yr running mean of GMST (the historical definition of climate) and state-of-the-art ESM large ensembles such as CESM2 LENS. The EBM–KF–ta GMST climate state, which also tracks OHCA, is constructed to be very close to that of a running 30-yr mean but generates this climate state 15 years sooner: It has no lag in reporting after annual observations are collected. We did not empirically fit this emulator to the climate record: 12 of the 17 parameters within the energy-balance equation were directly obtained from literature estimates, whereas the remaining five parameters are inferred indirectly from assumed preindustrial climate equilibrium and literature estimates of climate sensitivities. Thus, while some of these parameters were calibrated to the historical climate record independently by other researchers, they were not adjusted to suit this novel EBM combination. This filtered climate state captures the overall shape of the 30-yr means of measured GMST (r^2 = 0.922) and OHCA (r^2 = 0.989). In comparison with the ensemble spread of a hindcast ensemble of an ESM (LENS2), which is the state-of-the-art method for quantifying internal variability and probabilistic futures, the EBM–KF–uf provides a similar Gaussian distribution. Using this distribution, EBM–KF–uf can annually assess the likelihood of whether a policy threshold, e.g., 1.5° or 2°C over preindustrial, has been crossed. The EBM–KF–uf is also accurate at inferring the behavior of an entire climate model large ensemble using only one or a few ensemble members and can be used to distinguish novel sources of uncertainty in future projections, such as rare but significant future volcanic eruptions.

The EBM–KF approach has transparent, clean physical parameters of the EBM that can be directly measured or taken from estimates in modeling literature, leading to trivial uncertainty quantification by the Kalman filter machinery under fixed parameters. This uncertainty quantification revealed important aspects of GMST and OHCA uncertainty, in both hindcast and future projection contexts, with and without volcanoes. While the EBM–KF does not predict all climate variables of interest, it is a powerful, transparent, and inexpensive tool that may be readily combined with other approaches.

Acknowledgments. BFK's contributions were funded by ONR N00014-17-1-2393, NOAA NA19OAR4310366, and NSF OIA 2316271. JMN was funded by a Brown University Fellowship, a Brown 2023 OVPR Seed Award, and the 3CRS Project (NSF OIA 2316271). Conversations with Elizabeth Yankovsky, Anna Lo Piccolo, Joel Feske, Jochem Marotzke, Piers Forster, Lorraine E. Lisiecki, James (Jamie) Pringle, Zebedee Nicholls, Laure Zanna, Larissa Nazarenko, and Jung-Eun Lee helped to deepen and focus this work and locate data and other resources.

Data availability statement. This study performed reanalysis of existing datasets openly available at locations provided in appendix A regarding historical CO_2 and AOD_t, for SSP projections at https://greenhousegases.science.unimelb.edu.au/ and for LENS2 at https://www.earthsystemgrid.org/dataset/ucar.cgd.cesm2le.atm.proc.monthly_ave.TS.html. For critical measurements of the climate state, GMST via HadCRUT5 is at https://www.metoffice.gov.uk/hadobs/hadcrut5/data/current/download.html and OHCA from Zanna et al. (2019) is at https://zenodo.org/records/4603700. Further documentation about data processing, copies of the utilized datasets, and EBM–KF Python code is available through Harvard Dataverse at http://doi.org/10.7910/DVN/XLY8C2.

APPENDIX A

Derivation of the Blind Energy-Balance Model

a. Overall structure of the model

In the schematic diagram (Fig. 2), one stream of incoming solar shortwave energy $^1/_4 G_{SC}$ is successively fractionated by three reflective layers until a portion warms the ground and surface ocean. Then, this surface layer radiates longwave infrared energy back to space $\sigma_{sf} T^4$, again with greenhouse "reflection" in two layers. The surface ocean warms the deep ocean with fixed thermal insulation between them.

Temperature-dependent feedbacks are shown as cyclical arrows, with positive and negative feedback indicated relative to the overall energy balance. Positive feedbacks increase the energy flowing to the surface at higher surface temperatures T either by decreasing the fraction of shortwave reflection or by increasing the greenhouse reflection. Prescribed forcings are indicated by gear (cog) symbols. Unknown coefficients β_0, β_1, β_2, and β_3 exist, respectively, within the terms, $\tilde{g}(t), f_{H_2O}(T), f_{\alpha A}(T, t), f_{\alpha S}(T)$, in addition to the unknown exponent η. All these symbols are defined below.

Reiterating the overall structure in the model with discrete difference equations, T_t is the temperature of the surface in calendar year t (e.g., 2000), θ_t is the Conservative

Temperature of the deep ocean in that same year, and H_t is the total ocean heat content combining the heat in the surface ocean and deep ocean. The time step k (in Kalman filter literature) is 1 year. The overall energy flow into the Earth system and surface is

$$\Delta\text{energy in total} = \mathscr{F}_{\text{SW}} - \phi_{\text{LW}}, \tag{A1}$$

$$\Delta\text{energy at surface} = \mathscr{F}_{\text{SW}} - \phi_{\text{LW}} - Q_{\text{surf}\to\text{deep}}. \tag{A2}$$

Equations (1) and (2) within section 2a are describing the surface and deep ocean layers:

$$\begin{aligned}(T_{t+1} - T_t) \times C_{\text{surf}} = &\underbrace{(\tfrac{1}{4}G_{\text{SC}})_t \times \tilde{d}_t \times f_{\alpha A}(T_t) \times f_{\alpha S}(T_t)}_{\mathscr{F}_{\text{SW}}} \\ &- \underbrace{\sigma_{\text{sf}} T_t^4 \times \tilde{g}_t \times f_{\text{H}_2\text{O}}(T_t)}_{\phi_{\text{LW}}} \\ &- \underbrace{\gamma \times (T_t - \theta_t - \zeta_0)}_{Q_{\text{surf}\to\text{deep}}},\end{aligned} \tag{A3}$$

$$(\theta_{t+1} - \theta_t) \times C_{\text{deepO}} = \gamma \times (T_t - \theta_t - \zeta_0). \tag{A4}$$

Equation (3) to calculate OHCA is repeated below, along with its inverse transformation to obtain the deep ocean temperature:

$$H_t = (T_t - T_{1850}) \times C_{\text{upperO}} + (\theta_t - \theta_{1850}) \times C_{\text{deepO}}, \tag{A5}$$

$$\theta_t = [H_t - (T_t - T_{1850}) \times C_{\text{upperO}}]/C_{\text{deepO}} + \theta_{1850}. \tag{A6}$$

Equation (A5) also applies to the subsequent time step (substituting $t \to t + 1$), and then, (A4)–(A6) are substituted:

$$\begin{aligned}H_{t+1} = &(T_{t+1} - T_{1850}) \times C_{\text{upperO}} + \gamma \times (T_t - \theta_t - \zeta_0) \\ &+ (\theta_t - \theta_{1850}) \times C_{\text{deepO}}.\end{aligned} \tag{A7}$$

Equation (A5) is again substituted into (A7):

$$H_{t+1} - H_t = (T_{t+1} - T_t) \times C_{\text{upperO}} + \gamma \times (T_t - \theta_t - \zeta_0). \tag{A8}$$

Derivatives of θ_t from (A6) are as follows:

$$\frac{\partial \theta_t}{\partial H_t} = 1/C_{\text{deepO}}, \tag{A9a}$$

$$\frac{\partial \theta_t}{\partial T_t} = C_{\text{upperO}}/C_{\text{deepO}}. \tag{A9b}$$

On the right-hand side of (A3), both the incoming shortwave radiative flux $\mathscr{F}_{\text{SW}}$ and the outgoing longwave radiative flux ϕ_{LW} take the same form: ($\frac{1}{4}G_{\text{SC}}$ or $\sigma_{\text{sf}}T^4$) × [prescribed attenuation from forcing: $\tilde{d}(t)$ or $\tilde{g}(t)$] × [attenuation functions with feedback: $f_{\alpha A}(T, t) \times f_{\alpha S}(T)$ or $f_{\text{H}_2\text{O}}(T)$]. The term C_{surf}, the heat capacity of the surface (including the atmosphere, thermally active soil, and an 86 m upper layer of the ocean), was known least precisely of all coefficients: 17 ± 7 W yr m^{-2} K^{-1} (Schwartz 2007). The deep ocean layer (technically the zone where most of the ocean warming occurs) was chosen for the purpose of heat capacity estimation to be an additional 1141 m (Geoffroy et al. 2013a) within the 71% of the area covered by ocean based on the previous work of this heat transfer process. Converting this depth to a heat capacity, C_{deepO} = 1141 m × 0.71 × 1030 kg m^{-3} × 4180 W s^{-1} kg^{-1} K^{-1} × 1 yr (3.154 × 10^7 s)$^{-1}$ = 155.7 W yr m^{-2} K^{-1}. Constants γ, ζ_0 form a linear heat flux $Q_{\text{surf}\to\text{deep}}$ into the deep ocean, as discussed below.

b. Functional forms of components

For brevity, derivations and detailed explanations of each of these components have been moved to supplemental (A1) and (A2). Here, the functional form of each component is provided. We begin with the heat flowing from the surface layer into the deep ocean:

$$\begin{aligned}Q_{\text{surf}\to\text{deep}} &= \gamma \times (\Delta T_t - \Delta\theta_t) \\ &= \gamma \times [T_t - \theta_t - \underbrace{(T_{1850} - \theta_{1850})}_{\zeta_0}].\end{aligned} \tag{A10}$$

The fraction of shortwave (incoming) light reflected by AOD$_t$ is from Harshvardhan and King (1993):

$$\tilde{d}(t) \approx \frac{9.07}{\text{AOD}_t + 9.73}. \tag{A11}$$

The fraction of longwave radiation absorbed by greenhouse gases is

$$\tilde{g}(t) = 1 - \beta_0 \log_{10}([\text{eCO}_2]_t) < 1. \tag{A12}$$

Blackbody radiation, the source of longwave outgoing radiation, is the term $\sigma_{\text{sf}}T^4$, and the whole outgoing longwave radiation flux could be alternatively described in two ways:

$$\begin{aligned}\phi_{\text{LW}}(\text{outgoing}) &= \sigma_{\text{sf}} T_t^4 - \frac{\phi_{\text{LW(absorbed)}}}{2} \\ &= \sigma_{\text{sf}} T_t^4 \times \tilde{g}(t) \times f_{\text{H}_2\text{O}}(T_t).\end{aligned} \tag{A13}$$

In this paper, we use the form of (A13) in the second line which relates CO_2 to a fraction absorbed (similarly to albedo). Other authors favor the expression in the center of (A13), as it relates the absorption of a greenhouse gas to a power (W m^{-2}) rather than an expression.

The expression reported by Forster et al. (2023) for the blocked outgoing longwave radiation follows the center form and is written below in (A14). To be used within our EBM, this expression must be converted into a fraction to solve for β_0 in (A12):

$$\phi_{\text{LW}}^{\text{CO}_2} = 12.74 \log_{10}([\text{eCO}_2]_t) - 31.15. \tag{A14}$$

Functions proposed by the authors for the water feedback (on absorbing fraction of longwave radiation), atmospheric

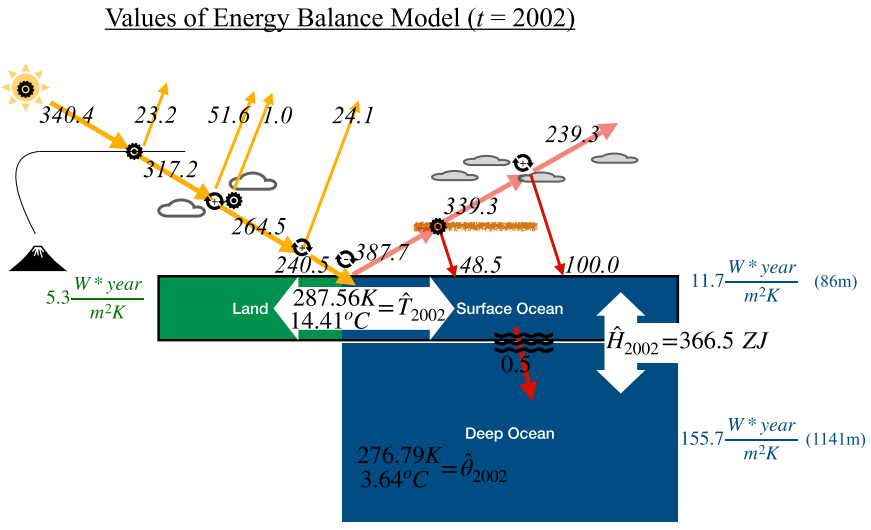

FIG. A1. Diagram with energy fluxes, temperatures, and total ocean heat content for the blind run of EBM in 2002 (when many of the reflectivity values were first measured by the CERES satellite). All numbers without units are in watts per square meter (W m^{-2}). Rounding to the nearest 0.1 W m^{-2} was performed after calculations.

albedo feedback, and surface albedo feedback are as follows:

$$f_{H_2O}(T_t) := \beta_1\left(\frac{1}{T_t}\right)^{\eta} \tag{A15}$$

$$f_{\alpha A}(T_t, t) := 0.834[1 + \beta_2(T_t - Y_{2002})] + \frac{AC_t - AC_{2002}}{(1/4 G_{SC})_t \times d_{2002}}, \tag{A16}$$

$$f_{\alpha S}(T_t) := 0.909[1 + \beta_3(T_t - Y_{2002})]. \tag{A17}$$

Note that the values of 0.834 and 0.909 came from the CERES satellite in the early 2000s (Loeb et al. 2009; Wielicki et al. 1996). Solving for all the coefficients by differentiating [see supplemental (SA17)–(SA20)], we find from feedbacks assessed in ESM (CMIP6 and AR6)

$$\eta = 1.615, \tag{A18a}$$

$$\beta_2 = 0.001\,36\,K^{-1}, \tag{A18b}$$

$$\beta_3 = 0.001\,63\,K^{-1}. \tag{A18c}$$

By assuming the climate was at equilibrium before 1850 and assimilating the longwave anthropogenic greenhouse gas and anthropogenic aerosol energy fluxes published by Forster (Forster et al. 2023) at https://github.com/ClimateIndicator/forcing-timeseries/tree/main/output, we find

$$\beta_1 = 6592.35, \tag{A19a}$$

$$\beta_0 = 0.046\,585. \tag{A19b}$$

This yields the following energy fluxes in 2002 displayed in Fig. A1, which are comparable to those obtained by Wild et al. (2015, 2019).

c. *Differentiating to find the Jacobian matrix*

The above derivation yielded a blind energy-balance model with good skill at predicting the GMST (orange dashed line in Fig. 2), $r^2 = 0.908$ blind versus HadCRUT5. Rewriting the overall model with β coefficients,

$$T_{t+1} = T_t + \frac{(1/4 G_{SC})_t \times 0.758 \times 9.068}{C_{surf} \times (AOD_t + 9.731)}\left[1 + \beta_2(T_t - Y_{2002}) + \frac{AC_t - AC_{2002}}{\left(1/4\overline{G_{SC}}\right) \times \tilde{d}_{2002} \times 0.834}\right][1 + \beta_3(T_t - Y_{2002})]$$

$$- \frac{\sigma_{sf}\beta_1}{C_{surf}}(T_t)^{2.385}[1 - \beta_0 \log_{10}([eCO_2]_t)] - \frac{\gamma}{C_{surf}}(T_t - \theta_t - \zeta_0). \tag{A20}$$

Partial derivatives of this updated equation are taken below, using the partial derivatives of θ_t written above in (A9) and also substituting $(1/4 G_{SC})_t \approx \left(1/4\overline{G_{SC}}\right) = 340.2$:

$$\frac{\partial T_{t+1}}{\partial T_t} = 1 + \frac{137.6}{(\mathrm{AOD}_t + 9.731)}\left[\beta_2 + \beta_3 + 2\beta_2\beta_3(T_t - Y_{2002}) + \beta_3\frac{\mathrm{AC}_t - \mathrm{AC}_{2002}}{\left(\frac{1}{4}\overline{G_{\mathrm{SC}}}\right) \times \tilde{d}_{2002} \times 0.834}\right]$$

$$- \frac{2.385\sigma_{\mathrm{sf}}\beta_1}{C_{\mathrm{surf}}}(T_t)^{1.385}[1 - \beta_0 \log_{10}([\mathrm{eCO_2}]_t)] - \frac{\gamma}{C_{\mathrm{surf}}}(1 - \underbrace{C_{\mathrm{upperO}}/C_{\mathrm{deepO}}}_{\frac{\partial\theta_t}{\partial T_t}}). \tag{A21}$$

$$\frac{\partial T_{t+1}}{\partial H_t} = \frac{\gamma}{C_{\mathrm{surf}}} \times \frac{\partial\theta_t}{\partial H_t} = \frac{\gamma}{C_{\mathrm{surf}}C_{\mathrm{deepO}}}. \tag{A22}$$

The ocean heat content update equation is written in (A7) with $r^2 = 0.910$ blind OHCA vs Zanna et al. (2019), and related partial derivatives are

$$\frac{\partial H_{t+1}}{\partial H_t} = C_{\mathrm{upperO}}\frac{\partial T_{t+1}}{\partial H_t} + \gamma \times \left(0 - \frac{\partial\theta_t}{\partial H_t}\right) + C_{\mathrm{deepO}}\frac{\partial\theta_t}{\partial H_t}$$

$$= \frac{\gamma}{C_{\mathrm{deepO}}}\left(\frac{C_{\mathrm{upperO}}}{C_{\mathrm{surf}}} - 1\right) + 1, \tag{A23}$$

$$\frac{\partial H_{t+1}}{\partial T_t} = C_{\mathrm{upperO}}\frac{\partial T_{t+1}}{\partial H_t} + \gamma \times \left(1 - \frac{C_{\mathrm{upperO}}}{C_{\mathrm{deepO}}}\right) + C_{\mathrm{upperO}}. \tag{A24}$$

The Jacobian matrix is thus complete, as

$$\mathbf{\Phi}_t = \begin{bmatrix} \frac{\partial T_{t+1}}{\partial T_t} & \frac{\partial T_{t+1}}{\partial H_t} \\ \frac{\partial H_{t+1}}{\partial T_t} & \frac{\partial H_{t+1}}{\partial H_t} \end{bmatrix}.$$

APPENDIX B

Generation of Volcanic Eruption Samples

As can be appreciated in Fig. B1a, long periods of no major volcanic eruptions (for instance, 1935–60) alternated with periods of many eruptions occurring in rapid succession (1883–1914 and 1960–94). This observed pattern prevented one Poisson distribution from describing the data well, but an exponential mixture did much better.

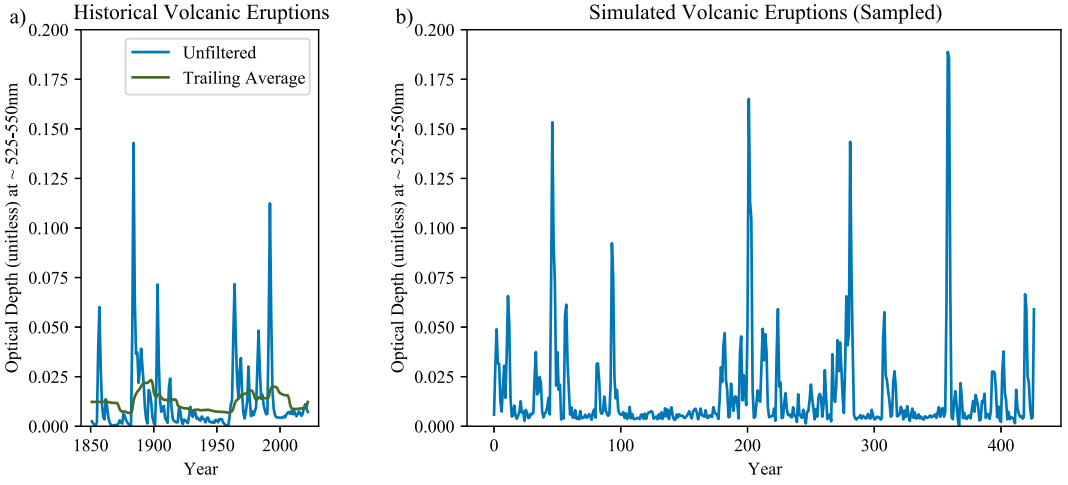

FIG. B1. Comparison of (a) historical volcanic eruptions with (b) simulated volcanic eruptions, generated from a combination of several probability distributions. In (a), the unfiltered aerosol optical depths AOD_t are plotted in blue, whereas the trailing average filter is plotted in green (it combines 15 years of a trailing average and 15 years of future projections at the mean AOD).

TABLE B1. Exponential parameters of volcano generating function. This generating function starts with a list of zero values for all AOD_t, and first samples several of these t years to be major volcanic eruptions. "Interval between" refers to the interval in years between the peaks of two successive major volcanic eruptions.

Exponential distributions	Random variable	Scale (units)	ℙ(used)
Interval between: round $(i_{t,0} + 2.6)$	$i_{t,0} \sim$ exp	2.263 (years)	88.9%
Interval between: round $(i_{t,1} + 2.6)$	$i_{t,1} \sim$ exp	24.2 (years)	11.1%
Peak size: $AOD_t = h_t + 0.0082$	$h_t \sim$ exp	0.0339 (unitless ∅)	Only "eruption" years

TABLE B2. Gaussian parameters of volcano generating function. These distributions are sampled after the major eruptions have already been filled in by the exponential distributions in Table B1.

Truncated Gaussian distributions	Random variable	Mean μ (unitless ∅)	Std dev σ (unitless ∅)
Prepeak −1: $AOD_{t-1} = a_{-1} \times AOD_t$	$a_{-1} \sim (\mathcal{N} > 0)$	0.51	0.25
Postpeak 1: $AOD_{t+1} = a_1 \times AOD_t$	$a_1 \sim (\mathcal{N} > 0)$	0.61	0.16
Postpeak 2: $AOD_{t+2} = a_2 \times AOD_t$	$a_2 \sim (\mathcal{N} > 0)$	0.32	0.16
Other years: $AOD_t = a_0$	$a_0 \sim (\mathcal{N} > 0)$	0.003 71	0.002 86

Eruptions that occurred within 3 years were indistinguishable in the historical dataset, so the minimum time interval between simulated volcanic eruptions was 2.6 years plus a sample (Table B1) from the exponential mixture model i_t (Okada et al. 2020). These intervals were rounded to integers. Similarly, the size of each volcanic eruption h_t was approximated using another shifted exponential distribution. The preceding year and 2 years following the eruption peak were positive fractions of the maximum aerosol optical depth, with Gaussian blur. Similarly, nonvolcanic years were positive Gaussian noise (Table B2). Figure B1b shows a sample from this combined generating function.

The overall procedure was to first create a series of spaced volcanic eruptions using Table B1 and then infill all the adjacent and nonvolcano years using Table B2. It is beyond the scope of this paper to investigate whether this statistical pattern has some relation to magma or tectonic dynamics or is merely an artifact of phasing.

APPENDIX C

Mathematical Notation and Symbols

Table C1 is provided for the user's reference to disambiguate how various bracket notations are used throughout the paper. These specific meanings are also specified within the text main body near the location of their first use. Table C2 is similarly provided as a comprehensive glossary of all mathematical symbols used within this paper.

TABLE C1. Notation involving brackets/fencing and subscripts/superscripts.

Example	Section used	Meaning
{[()]}	All	General AMS hierarchy for nesting fencing both within the text and equations
$\mathbf{F}(\mathbf{x}_{t-1}; u_t)$	2a	Parentheses for function inputs with two classes: primary input $\mathbf{x}_{t-1}$ and time-varying control parameters u_t
T_t, AC_t	2a	Trailing subscript may indicate a constructed time series or raw time series input at time t
$\tilde{g}(t)$	2a	Constructed function that is time varying
ζ_0, c_3, $i_{t,0}$	2a, appendix B	Trailing subscript may also indicate multiple parameters (baselines or coefficients) or random variables
$\hat{p}_t^T$	2b	Superscript T indicates temperature (GMST) component of the state variance (not usual AMS notation)
$\mathbf{\Phi}_t^*$	2b	Asterisk superscript indicates matrix transposition (not usual AMS notation)
$(\mathbf{S}_t)^{-1}$	2b	The −1 superscript indicates matrix inversion
$\overline{{}_{30}Y_t}$	1	Preceding subscript indicates a series of 30-yr subsets (running mean) are taken from the GMST time series
$[eCO_2]_t$	2a	Square brackets here indicate this is an effective CO_2 concentration in ppm
$\mu(a–b)$	1	95% confidence interval (following IPCC guidelines, extremely likely interval)
$[a–b]$	2a, 3a	Finite or closed range from a to b (following IPCC guidelines); does not apply to year ranges
$[a, b]$	2b	Defines a 1×2 matrix (a vector with two indices)
$\{a, b, c, d, \ldots, z\}$	2b	Constructing a specific time series, similar to set notation

TABLE C2. Glossary of mathematical symbols.

Symbol		Units
	Meaning within statistics	
p, $\mathbb{P}$(event)	Probability of observed result for a particular hypothesis test (e.g., slope is positive)	[0–1]
r^2	Coefficient of determination: fraction of variance explained by a model	[0–1]
μ	Mean of a set or distribution	Any
σ	Standard deviation ($\sqrt{\text{variance}}$)	Any
$\mathcal{N}(\mu, \sigma^2)$	Gaussian (normal) distribution	Any
$\mu \pm 2\sigma$ = 95% CI	95% confidence interval (extremely likely) under Gaussian distribution	Any
Cov()	Covariance of a random vector (here $\mathbf{y}_t$ has length 2, so its covariance is 2×2)	Square matrix
	Meaning within energy-balance model	
t, k	Time index, time step	Year
T_t	GMST surface temperature climate state, idealized	K (°C)
θ_t	Deep ocean conservative temperature state, idealized	K (°C)
H_t	Ocean heat content anomaly, idealized	W yr m^{-2} (ZJ)
u_t	Set of time-varying forcing inputs to the atmosphere (four items below)	—
$[\text{eCO}_2]_t$	Amount of total greenhouse gas in the atmosphere, in effective concentration of CO_2	ppm
AOD_t	Aerosol optical depth (from top of the atmosphere), affected by volcanoes	∅ (AOD)
AC_t	Cloud radiative forcing due to change in reflectivity by anthropogenic aerosols	W m^{-2}
$(1/4 G_{\text{SC}})_t$	Top of the atmosphere total solar irradiance	W m^{-2}
$\mathscr{F}_{\text{SW}}$, ϕ_{LW}	Net radiative fluxes (shortwave, longwave) at the top of the atmosphere	W m^{-2}
Δenergy at surface	Net heat flow into the surface layer	W m^{-2}
$Q_{\text{surf}\to\text{deep}}$	Heat flow into the deep ocean layer	W m^{-2}
C_{surf}, C_{upperO}, C_{deepO}	Heat capacities of the surface, surface ocean, and deep ocean	W yr m^{-2} K^{-1}
$\sigma_{\text{sf}} T^4$	Source of outgoing longwave radiation (blackbody or Planck feedback)	W yr m^{-2}
σ_{sf}	Stefan–Boltzmann constant = 5.670×10^{-8}	W m^{-2} K^{-4}
$\tilde{d}_t$, $\tilde{g}_t$	Prescribed, time-varying attenuations from AOD_t and $[\text{eCO}_2]_t$, respectively	∅
$f_{\alpha A}(T, t) \times f_{\alpha S}(T)$	Attenuations of incoming shortwave radiation due to albedo of the atmosphere and land surface, respectively (feedback from T_t)	∅
$f_{\text{H}_2\text{O}}(T)$	Attenuation of outgoing longwave radiation by water vapor (feedback from T_t)	∅
ζ_0	Equilibrium temperature difference between the surface and deep ocean	K (°C)
HCa_t	The HadCRUT5 anomaly record (Morice et al. 2021)	K (°C)
ζ_1	Baseline temperature for HadCRUT5 to achieve the appropriate 1960–89 climate normal (Jones and Harpham 2013)	K (°C)
β_0	Solved coefficient on $\log_{10}([\text{eCO}_2]_t)$ within a sequential filter atmosphere approximately	∅
β_1, η	Solved coefficient and exponent for the $f_{\text{H}_2\text{O}}(T)$ water vapor longwave feedback	∅
β_2, β_3	Solved coefficients for $f_{\alpha A}(T, t)$, $f_{\alpha S}(T)$ atmosphere and surface albedo feedbacks	∅
c_1, c_2, c_3, c_4	Simplifications of constants within the EBM for (4)–(6)	See Table 1
$[\tilde{T}_{t+1}, \tilde{H}_{t+1}] = \mathbf{F}(\tilde{T}_t, \tilde{H}_t; u_t)$	Blind energy-balance model, deterministic from prior climate state, no data assimilation	K, W yr m^{-2}
	Meaning within (extended) Kalman filter	
$\mathbf{x}_t = [T_t, H_t]$	Idealized true climate estate, with dynamic model noise	K, W yr m^{-2}
$\mathbf{y}_t = [Y_t, \psi_t]$	Measurements with noise of the climate state, GMST from HadCRUT5 (Jones and Harpham 2013) and OHCA from Zanna et al. (2019)	K, W yr m^{-2}
$\mathbf{Q} = \text{Cov}(\mathbf{w}_t)$	Assumed dynamic model error and model covariance matrix	$\begin{bmatrix} \text{K}^2 & \text{K}\frac{\text{W yr}}{\text{m}^2} \\ \text{K}\frac{\text{W yr}}{\text{m}^2} & \left(\frac{\text{W yr}}{\text{m}^2}\right)^2 \end{bmatrix}$
$\mathbf{R} = \text{Cov}(\mathbf{v}_t)$	Assumed measurement error and measurement covariance matrix	As $\mathbf{Q}$ above
$\mathbf{R}_t = \mathbf{R}_t^{\text{var}} + \mathbf{R}^{\text{const}}$	Actual model and measurement covariance matrices used in the EBM–KF,	As $\mathbf{Q}$ above
$\mathbf{Q} = \mathbf{R}^{\text{const}}/30$	defined to mimic the statistics of the 30-yr running mean	
$\hat{\mathbf{x}}_t = [\hat{T}_t, \hat{H}_t]$	Posterior estimated state (after measurement assimilation)	K, W yr m^{-2}
$\mathbf{P}_t$	Posterior estimated state covariance (after measurement assimilation)	As $\mathbf{Q}$ above
$[\hat{p}_t^T, \hat{p}_t^H] = \text{diag}(\mathbf{P}_t)$	Elements of state variance exclusive to GMST and OHCA	[K^2, (W yr m^{-2})^2]

TABLE C2. (*Continued*)

Symbol		Units
$\mathbf{\Phi}_t = \left.\frac{\partial \mathbf{F}(\mathbf{x}; u_t)}{\partial \mathbf{x}}\right\|_{\mathbf{x}=\hat{\mathbf{x}}_{t-1}}$	Linearized Jacobian tensor derivative of the (blind) EBM	$\begin{bmatrix} \varnothing & \mathrm{K}/\frac{\mathrm{W\ yr}}{\mathrm{m}^2} \\ \frac{\mathrm{W\ yr}}{\mathrm{m}^2}/\mathrm{K} & \varnothing \end{bmatrix}$
$\hat{\mathbf{x}}_{t\|t-1} = [\hat{T}_{t\|t-1}, \hat{H}_{t\|t-1}]$	Forecast state projection (before new measurement)	K, W yr m^{-2}
$\mathbf{P}_{t\|t-1}$	Forecast covariance projection (before new measurement)	As **Q** above
$\mathbf{z}_t = [z_t^T, z_t^H]$	Innovation residual	K, W yr m^{-2}
$\mathbf{S}_t$	Innovation covariance	As **Q** above
$[\hat{s}_t^T, \hat{s}_t^H] = \mathrm{diag}(\mathbf{S}_t)$	Elements of innovation variance exclusive to GMST and OHCA	$[\mathrm{K}^2, (\mathrm{W\ yr\ m}^{-2})^2]$
$\mathbf{K}_t$	Kalman gain	$\begin{bmatrix} \varnothing & \varnothing \\ \varnothing & \varnothing \end{bmatrix}$
	Meaning within ESM ensembles (LENS2)	
$(Y_t)_j$	The jth ensemble member's annual mean at time t of near-surface air temperature	K
$(\psi_t)_j$	The jth ensemble member's annual mean at time t of total ocean heat content	W yr m^{-2}
$\overline{(Y_t)_j}$	Ensemble average (across all members e.g., 90) at year t	K
$\overline{({}_{21}Y_t)_j}$	The 21-yr running mean of ensemble member j	K
$\overline{(\overline{{}_{21}Y_t})_j}$	The cross-ensemble average of all 21-yr running means	K
${}_{\mathrm{ens}}\sigma_t^T$	The cross-ensemble standard deviation of GMST; see (24)	K
N	Number of ensemble members within a subset of the larger ensemble; see (24)	$\varnothing$
	Meaning within volcanic eruption distribution	
$i_{t,0}$, $i_{t,1}$	Exponential mixture random variables to determine intervals between major eruptions	years
h_t	Exponential random variable to determine size of a particular major eruption	$\varnothing$ (AOD)
a_{-1}, a_1, a_2, a_0	Truncated Gaussian distributions to determine the atmospheric optical depth in eruption-adjacent and noneruption years	$\varnothing$ (AOD)
	Meanings (miscellaneous contexts)	
q	Location of a climate policy threshold; see (23)	K (°C)
$\hat{\mathbf{x}}_t$, $\hat{\mathbf{P}}_t$, $\hat{\mathbf{K}}_t$	Rauch–Tung–Striebel (RTS) smoother reestimated state estimate, state covariance, and Kalman gain following backward sweep; see supplemental A3	As above for KF
$\overline{{}_{30}Y_t}$, ${}_{30}\hat{\mathbf{y}}_t$	Standard climate normal, a 30-yr running mean of GMST or (GMST, OHCA) measurements, undefined before 1865 or after 2008 (as of this publication in 2024)	K, (K, W yr m^{-2})

REFERENCES

Abraham, J. P., and Coauthors, 2013: A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Rev. Geophys.*, **51**, 450–483, https://doi.org/10.1002/rog.20022.

Armour, K. C., C. M. Bitz, and G. H. Roe, 2013: Time-varying climate sensitivity from regional feedbacks. *J. Climate*, **26**, 4518–4534, https://doi.org/10.1175/JCLI-D-12-00544.1.

Benhamou, E., 2018: Kalman filter demystified: From intuition to probabilistic graphical model to real case in financial markets. arXiv, 1811.11618v2, https://doi.org/10.48550/arXiv.1811.11618.

Bethke, I., S. Outten, O. D. Otterå, E. Hawkins, S. Wagner, M. Sigl, and P. Thorne, 2017: Potential volcanic impacts on future climate variability. *Nat. Climate Change*, **7**, 799–805, https://doi.org/10.1038/nclimate3394.

Betts, R. A., and Coauthors, 2023: Approaching 1.5°C: How will we know we've reached this crucial warming mark? *Nature*, **624**, 33–35, https://doi.org/10.1038/d41586-023-03775-z.

Bouttier, F., 1996: Application of Kalman filtering to Numerical Weather Prediction. *Workshop on Non-Linear Aspects of Data Assimilation*, Shinfield Park, Reading, ECMWF, 61–90, https://www.ecmwf.int/en/elibrary/73789-application-kalman-filtering-numerical-weather-prediction.

Budyko, M. I., 1969: The effect of solar radiation variations on the climate of the earth. *Tellus*, **21A**, 611–619, https://doi.org/10.3402/tellusa.v21i5.10109.

Buizza, R., M. Milleer, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908, https://doi.org/10.1002/qj.49712556006.

Burgess, S., 2024: Copernicus: In 2024, the world experienced the warmest January on record. ECMWF, https://climate.copernicus.eu/copernicus-2024-world-experienced-warmest-january-record.

Carré, M., J. P. Sachs, J. M. Wallace, and C. Favier, 2012: Exploring errors in paleoclimate proxy reconstructions using Monte Carlo simulations: Paleotemperature from mollusk and coral geochemistry. *Climate Past*, **8**, 433–450, https://doi.org/10.5194/cp-8-433-2012.

Ceppi, P., and P. Nowack, 2021: Observational evidence that cloud feedback amplifies global warming. *Proc. Natl. Acad.*

Sci. USA, **118**, e2026290118, https://doi.org/10.1073/pnas.2026290118.

Chan, D., and P. Huybers, 2021: Correcting observational biases in sea surface temperature observations removes anomalous warmth during World War II. *J. Climate*, **34**, 4585–4602, https://doi.org/10.1175/JCLI-D-20-0907.1.

Chen, X., and K.-K. Tung, 2018: Global-mean surface temperature variability: Space–time perspective from rotated EOFs. *Climate Dyn.*, **51**, 1719–1732, https://doi.org/10.1007/s00382-017-3979-0.

Chen, Z., C. Heckman, S. Julier, and N. Ahmed, 2018: Weak in the NEES?: Auto-tuning Kalman filters with Bayesian optimization. *2018 21st Int. Conf. on Information Fusion (FUSION)*, Cambridge, United Kingdom, Institute of Electrical and Electronics Engineers, 1072–1079, https://doi.org/10.23919/ICIF.2018.8454982.

——, ——, S. J. Julier, and N. R. Ahmed, 2021: Time dependence in Kalman filter tuning. *2021 IEEE 24th Int. Conf. on Information Fusion (FUSION)*, Sun City, South Africa, Institute of Electrical and Electronics Engineers, 1–8, https://doi.org/10.23919/FUSION49465.2021.9626864.

Cheng, L., K. E. Trenberth, J. Fasullo, T. Boyer, J. Abraham, and J. Zhu, 2017: Improved estimates of ocean heat content from 1960 to 2015. *Sci. Adv.*, **3**, e1601545, https://doi.org/10.1126/sciadv.1601545.

——, and Coauthors, 2022: Past and future ocean warming. *Nat. Rev. Earth Environ.*, **3**, 776–794, https://doi.org/10.1038/s43017-022-00345-1.

Coddington, O., J. L. Lean, D. Lindholm, P. Pilewskie, M. Snow, and NOAA CDR Program, 2017: NOAA Climate Data Record (CDR) of Total Solar Irradiance (TSI), NRLTSI version 2.1. NOAA National Centers for Environmental Information, accessed 10 July 2023, https://doi.org/10.7289/V56W985W.

Collins, M., and Coauthors, 2013: Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1029–1136, https://doi.org/10.1017/CBO9781107415324.024.

Compo, G. P., and Coauthors, 2011: The Twentieth Century Reanalysis project. *Quart. J. Roy. Meteor. Soc.*, **137** (654), 1–28, https://doi.org/10.1002/qj.776.

Duffy, M. L., B. Medeiros, A. Gettelman, and T. Eidhammer, 2023: Perturbing parameters to understand cloud contributions to climate change. *J. Climate*, **37**, 213–227, https://doi.org/10.1175/JCLI-D-23-0250.1.

Edwards, T. L., and Coauthors, 2021: Projected land ice contributions to twenty-first-century sea level rise. *Nature*, **593**, 74–82, https://doi.org/10.1038/s41586-021-03302-y.

Feng, R., B. L. Otto-Bliesner, E. C. Brady, and N. Rosenbloom, 2020: Increased climate response and earth system sensitivity from CCSM4 to CESM2 in mid-Pliocene simulations. *J. Adv. Model. Earth Syst.*, **12**, e2019MS002033, https://doi.org/10.1029/2019MS002033.

Filar, J. A., P. S. Gaertner, and M. A. Janssen, 1996: An application of optimization to the problem of climate change. *State of the Art in Global Optimization*, C. A. Floudas and P. M. Pardalos, Eds., Springer, 475–498, https://doi.org/10.1007/978-1-4613-3437-8_29.

Forster, P., and Coauthors, 2021: The Earth's energy budget, climate feedbacks and climate sensitivity. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 923–1054, https://doi.org/10.1017/9781009157896.009.

Forster, P. M., and Coauthors, 2023: Indicators of Global Climate Change 2022: Annual update of large-scale indicators of the state of the climate system and human influence. *Earth Syst. Sci. Data*, **15**, 2295–2327, https://doi.org/10.5194/essd-15-2295-2023.

Foster, G., and S. Rahmstorf, 2011: Global temperature evolution 1979–2010. *Environ. Res. Lett.*, **6**, 044022, https://doi.org/10.1088/1748-9326/6/4/044022.

Fox-Kemper, B., and Coauthors, 2021: Ocean, cryosphere and sea level change. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 1211–1362, https://doi.org/10.1017/9781009157896.011.

Friedrich, T., A. Timmermann, M. Tigchelaar, O. Elison Timm, and A. Ganopolski, 2016: Nonlinear climate sensitivity and its implications for future greenhouse warming. *Sci. Adv.*, **2**, e1501923, https://doi.org/10.1126/sciadv.1501923.

Fujimori, S., T. Hasegawa, T. Masui, K. Takahashi, D. S. Herran, H. Dai, Y. Hijioka, and M. Kainuma, 2017: SSP3: Aim implementation of Shared Socioeconomic Pathways. *Global Environ. Change*, **42**, 268–283, https://doi.org/10.1016/j.gloenvcha.2016.06.009.

García-Pintado, J., and A. Paul, 2018: Evaluation of iterative Kalman smoother schemes for multi-decadal past climate analysis with comprehensive earth system models. *Geosci. Model Dev.*, **11**, 5051–5084, https://doi.org/10.5194/gmd-11-5051-2018.

Geoffroy, O., D. Saint-Martin, D. J. L. Olivié, A. Voldoire, G. Bellon, and S. Tytéca, 2013a: Transient climate response in a two-layer energy-balance model. Part I: Analytical solution and parameter calibration using CMIP5 AOGCM experiments. *J. Climate*, **26**, 1841–1857, https://doi.org/10.1175/JCLI-D-12-00195.1.

——, ——, G. Bellon, A. Voldoire, D. J. L. Olivié, and S. Tytéca, 2013b: Transient climate response in a two-layer energy-balance model. Part II: Representation of the efficacy of deep-ocean heat uptake and validation for CMIP5 AOGCMs. *J. Climate*, **26**, 1859–1876, https://doi.org/10.1175/JCLI-D-12-00196.1.

Gerlach, T., 2011: Volcanic versus anthropogenic carbon dioxide. *Eos*, **92**, https://doi.org/10.1029/2011EO240001.

Gettelman, A., and Coauthors, 2019: High climate sensitivity in the Community Earth System Model Version 2 (CESM2). *Geophys. Res. Lett.*, **46**, 8329–8337, https://doi.org/10.1029/2019GL083978.

Gregory, J. M., 2000: Vertical heat transports in the ocean and their effect on time-dependent climate change. *Climate Dyn.*, **16**, 501–515, https://doi.org/10.1007/s003820000059.

——, and T. Andrews, 2016: Variation in climate sensitivity and feedback parameters during the historical period. *Geophys. Res. Lett.*, **43**, 3911–3920, https://doi.org/10.1002/2016GL068406.

Grewal, M. S., and A. P. Andrews, 2001: *Kalman Filtering: Theory and Practice Using MATLAB*. John Wiley and Sons, 416 pp., https://books.google.com/books?id=9UoZAQAAIAAJ.

Gulev, S. K., and Coauthors, 2021: Changing state of the climate system. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 287–422, https://doi.org/10.1017/9781009157896.004.

Guttman, N. B., 1989: Statistical descriptors of climate. *Bull. Amer. Meteor. Soc.*, **70**, 602–607, https://doi.org/10.1175/1520-0477(1989)070<0602:SDOC>2.0.CO;2.

Hakim, G. J., J. Emile-Geay, E. J. Steig, D. Noone, D. M. Anderson, R. Tardif, N. Steiger, and W. A. Perkins, 2016: The last millennium climate reanalysis project: Framework and first

results. *J. Geophys. Res. Atmos.*, **121**, 6745–6764, https://doi.org/10.1002/2016JD024751.

Harshvardhan, and M. D. King, 1993: Comparative accuracy of diffuse radiative properties computed using selected multiple scattering approximations. *J. Atmos. Sci.*, **50**, 247–259, https://doi.org/10.1175/1520-0469(1993)050%3C0247:CAODRP%3E2.0.CO;2.

Hasselmann, K., 1997: Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dyn.*, **13**, 601–611, https://doi.org/10.1007/s003820050185.

Hausfather, Z., 2024a: Analysis: What record global heat means reaching the 1.5C warming limit. https://www.carbonbrief.org/analysis-what-record-global-heat-means-for-breaching-the-1-5c-warming-limit/.

——, 2024b: When will the world really pass 1.5c? Twitter, https://x.com/hausfath/status/1757916875806392657?s=20.

Haustein, K., M. R. Allen, P. M. Forster, F. E. L. Otto, D. M. Mitchell, H. D. Matthews, and D. J. Frame, 2017: A real-time Global Warming Index. *Sci. Rep.*, **7**, 15417, https://doi.org/10.1038/s41598-017-14828-5.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1108, https://doi.org/10.1175/2009BAMS2607.1.

Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis, 2010: Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *J. Climate*, **23**, 2418–2427, https://doi.org/10.1175/2009JCLI3466.1.

Horvat, C., 2021: Marginal ice zone fraction benchmarks sea ice and climate model skill. *Nat. Commun.*, **12**, 2221, https://doi.org/10.1038/s41467-021-22004-7.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

Hu, S., and A. V. Fedorov, 2017: The extreme El Niño of 2015–2016 and the end of global warming hiatus. *Geophys. Res. Lett.*, **44**, 3816–3824, https://doi.org/10.1002/2017GL072908.

Huguenin, M. F., R. M. Holmes, and M. H. England, 2022: Drivers and distribution of global ocean heat uptake over the last half century. *Nat. Commun.*, **13**, 4921, https://doi.org/10.1038/s41467-022-32540-5.

Hummels, R., M. Dengler, and B. Bourlès, 2013: Seasonal and regional variability of upper ocean diapycnal heat flux in the Atlantic cold tongue. *Prog. Oceanogr.*, **111**, 52–74, https://doi.org/10.1016/j.pocean.2012.11.001.

Illing, S., C. Kadow, H. Pohlmann, and C. Timmreck, 2018: Assessing the impact of a future volcanic eruption on decadal predictions. *Earth Syst. Dyn.*, **9**, 701–715, https://doi.org/10.5194/esd-9-701-2018.

Ishii, M., Y. Fukuda, S. Hirahara, S. Yasui, T. Suzuki, and K. Sato, 2017: Accuracy of global upper ocean heat content estimation expected from present observational data sets. *SOLA*, **13**, 163–167, https://doi.org/10.2151/sola.2017-030.

Jones, C. D., and Coauthors, 2021: The climate response to emissions reductions due to COVID-19: Initial results from CovidMIP. *Geophys. Res. Lett.*, **48**, e2020GL091883, https://doi.org/10.1029/2020GL091883.

Jones, P. D., and P. M. Kelly, 1996: The effect of tropical explosive volcanic eruptions on surface air temperature. *The Mount Pinatubo Eruption*, G. Fiocco, D. Fuà, and G. Visconti, Eds., Springer, 95–111, https://doi.org/10.1007/978-3-642-61173-5_10.

——, and C. Harpham, 2013: Estimation of the absolute surface air temperature of the Earth. *J. Geophys. Res. Atmos.*, **118**, 3213–3217, https://doi.org/10.1002/jgrd.50359.

Julier, S. J., and J. K. Uhlmann, 1997: New extension of the Kalman filter to nonlinear systems. *Proc. SPIE*, **3068**, 182–193, https://doi.org/10.1117/12.280797.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Basic Eng.*, **82**, 35–45, https://doi.org/10.1115/1.3662552.

——, and R. S. Bucy, 1961: New results in linear filtering and prediction theory. *J. Basic Eng.*, **83**, 95–108, https://doi.org/10.1115/1.3658902.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp., https://doi.org/10.1017/CBO9780511802270.

Kaufman, D., and Coauthors, 2020: Publisher correction: A global database of Holocene paleotemperature records. *Sci. Data*, **7**, 271, https://doi.org/10.1038/s41597-020-00611-1.

Keil, P., H. Schmidt, B. Stevens, and J. Bao, 2021: Variations of tropical lapse rates in climate models and their implications for upper-tropospheric warming. *J. Climate*, **34**, 9747–9761, https://doi.org/10.1175/JCLI-D-21-0196.1.

Kim, Y., and H. Bang, 2018: Introduction to Kalman filter and its applications. *Introduction and Implementations of the Kalman Filter*, F. Govaers, Ed., IntechOpen, 7–22, https://doi.org/10.5772/intechopen.80600.

Kirtman, B., and Coauthors, 2013: Near-term climate change: Projections and predictability. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 953–1028, https://doi.org/10.1017/CBO9781107415324.023.

Kravitz, B., D. G. MacMartin, H. Wang, and P. J. Rasch, 2016: Geoengineering as a design problem. *Earth Syst. Dyn.*, **7**, 469–497, https://doi.org/10.5194/esd-7-469-2016.

——, and Coauthors, 2018: The climate effects of increasing ocean albedo: An idealized representation of solar geoengineering. *Atmos. Chem. Phys.*, **18**, 13 097–13 113, https://doi.org/10.5194/acp-18-13097-2018.

Lauritzen, S. L., 1981: Time series analysis in 1880: A discussion of contributions made by T.N. Thiele. *Int. Stat. Rev.*, **49**, 319–331, https://doi.org/10.2307/1402616.

——, 2002: *Thiele: Pioneer in Statistics*. Oxford University Press, 276 pp.

Lee, J. H., and N. L. Ricker, 1994: Extended Kalman filter based nonlinear model predictive control. *Ind. Eng. Chem. Res.*, **33**, 1530–1541, https://doi.org/10.1021/ie00030a013.

Lee, J.-Y., and Coauthors, 2021: Future global climate: Scenario-based projections and near-term information. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 553–672, https://doi.org/10.1017/9781009157896.006.

Lehner, F., C. Deser, N. Maher, J. Marotzke, E. M. Fischer, L. Brunner, R. Knutti, and E. Hawkins, 2020: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth Syst. Dyn.*, **11**, 491–508, https://doi.org/10.5194/esd-11-491-2020.

Lenssen, N. J. L., G. A. Schmidt, J. E. Hansen, M. J. Menne, A. Persin, R. Ruedy, and D. Zyss, 2019: Improvements in the GISTEMP uncertainty model. *J. Geophys. Res. Atmos.*, **124**, 6307–6326, https://doi.org/10.1029/2018JD029522.

Levitus, S., and Coauthors, 2017: NCEI ocean heat content, temperature anomalies, salinity anomalies, thermosteric sea level anomalies, halosteric sea level anomalies, and total steric sea

level anomalies from 1955 to present calculated from in situ oceanographic subsurface profile data. NOAA National Oceanic and Atmospheric Administration (NCEI Accession 0164586), accessed 13 February 2024, https://doi.org/10.7289/v53f4mvp.

Loeb, N. G., B. A. Wielicki, D. R. Doelling, G. L. Smith, D. F. Keyes, S. Kato, N. Manalo-Smith, and T. Wong, 2009: Toward optimal closure of the Earth's top-of-atmosphere radiation budget. *J. Climate*, **22**, 748–766, https://doi.org/10.1175/2008JCLI2637.1.

Loikith, P. C., and J. D. Neelin, 2019: Non-Gaussian cold-side temperature distribution tails and associated synoptic meteorology. *J. Climate*, **32**, 8399–8414, https://doi.org/10.1175/JCLI-D-19-0344.1.

MacMartin, D. G., B. Kravitz, and D. W. Keith, 2014: Geoengineering: The world's largest control problem. *2014 American Control Conf.*, Portland, OR, Institute of Electrical and Electronics Engineers, 2401–2406, https://doi.org/10.1109/ACC.2014.6858658.

Marotzke, J., and P. M. Forster, 2015: Forcing, feedback and internal variability in global temperature trends. *Nature*, **517**, 565–570, https://doi.org/10.1038/nature14117.

Marshall, L. R., E. C. Maters, A. Schmidt, C. Timmreck, A. Robock, and M. Toohey, 2022: Volcanic effects on climate: Recent advances and future avenues. *Bull. Volcanol.*, **84**, 54, https://doi.org/10.1007/s00445-022-01559-3.

McClelland, H. L. O., I. Halevy, D. A. Wolf-Gladrow, D. Evans, and A. S. Bradley, 2021: Statistical uncertainty in paleoclimate proxy reconstructions. *Geophys. Res. Lett.*, **48**, e2021GL092773, https://doi.org/10.1029/2021GL092773.

McCormick, M. P., L. W. Thomason, and C. R. Trepte, 1995: Atmospheric effects of the Mt Pinatubo eruption. *Nature*, **373**, 399–404, https://doi.org/10.1038/373399a0.

McCulloch, M. T., A. Winter, C. E. Sherman, and J. A. Trotter, 2024: 300 years of sclerosponge thermometry shows global warming has exceeded 1.5°C. *Nat. Climate Change*, **14**, 171–177, https://doi.org/10.1038/s41558-023-01919-7.

McDougall, T. J., P. M. Barker, R. M. Holmes, R. Pawlowicz, S. M. Griffies, and P. J. Durack, 2021: The interpretation of temperature and salinity variables in numerical ocean model output and the calculation of heat fluxes and heat content. *Geosci. Model Dev.*, **14**, 6445–6466, https://doi.org/10.5194/gmd-14-6445-2021.

McKinnon, K. A., A. Poppick, E. Dunn-Sigouin, and C. Deser, 2017: An "Observational Large Ensemble" to compare observed and modeled temperature trend uncertainty due to internal variability. *J. Climate*, **30**, 7585–7598, https://doi.org/10.1175/JCLI-D-16-0905.1.

Meehl, G. A., R. Moss, K. E. Taylor, V. Eyring, R. J. Stouffer, S. Bony, and B. Stevens, 2014: Climate model intercomparisons: Preparing for the next phase. *Eos*, **95**, https://doi.org/10.1002/2014EO090001.

Meinshausen, M., and Coauthors, 2020: The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geosci. Model Dev.*, **13**, 3571–3605, https://doi.org/10.5194/gmd-13-3571-2020.

Merchant, C. J., and Coauthors, 2019: Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Sci. Data*, **6**, 223, https://doi.org/10.1038/s41597-019-0236-x.

Millar, R. J., Z. R. Nicholls, P. Friedlingstein, and M. R. Allen, 2017: A modified impulse-response representation of the global near-surface air temperature and atmospheric concentration response to carbon dioxide emissions. *Atmos. Chem. Phys.*, **17**, 7213–7228, https://doi.org/10.5194/acp-17-7213-2017.

Miller, R. L., and Coauthors, 2014: CMIP5 historical simulations (1850–2012) with GISS ModelE2. *J. Adv. Model. Earth Syst.*, **6**, 441–478, https://doi.org/10.1002/2013MS000266.

Miller, R. N., 1996: Introduction to the Kalman filter. ECMWF Tech. Paper, 13 pp., https://www.ecmwf.int/sites/default/files/elibrary/1996/75692-introduction-kalman-filter_0.pdf.

Montgomery, D. C., and G. C. Runger, 2016: *Applied Statistics and Probability for Engineers*. 6th ed. John Wiley and Sons, 792 pp.

Morice, C. P., and Coauthors, 2021: An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *J. Geophys. Res. Atmos.*, **126**, e2019JD032361, https://doi.org/10.1029/2019JD032361.

Myers, M. A., and R. H. Luecke, 1991: Process control applications of an extended Kalman filter algorithm. *Comput. Chem. Eng.*, **15**, 853–857, https://doi.org/10.1016/0098-1354(91)80030-Y.

NASA/LARC/SD/ASDC, 2018: Global space-based stratospheric aerosol climatology version 2.0. NASA, accessed 13 February 2024, https://doi.org/10.5067/GLOSSAC-L3-V2.0.

Nazarenko, L. S., and Coauthors, 2022: Future climate change under SSP emission scenarios with GISS-E2.1. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002871, https://doi.org/10.1029/2021MS002871.

Newsom, E., L. Zanna, and J. Gregory, 2023: Background pycnocline depth constrains future ocean heat uptake efficiency. *Geophys. Res. Lett.*, **50**, e2023GL105673, https://doi.org/10.1029/2023GL105673.

Nieto-Reyes, A., 2022: On the non-Gaussianity of sea surface elevations. *J. Mar. Sci. Eng.*, **10**, 1303, https://doi.org/10.3390/jmse10091303.

Ogorek, B., 2019: Yet another Kalman filter explanation article. Towards Data Science. https://towardsdatascience.com/yet-another-kalman-filter-explanation-article-be0264d99937.

Okada, M., K. Yamanishi, and N. Masuda, 2020: Long-tailed distributions of inter-event times as mixtures of exponential distributions. *Roy. Soc. Open Sci.*, **7**, 191643, https://doi.org/10.1098/rsos.191643.

Otto, F. E. L., D. J. Frame, A. Otto, and M. R. Allen, 2015: Embracing uncertainty in climate change policy. *Nat. Climate Change*, **5**, 917–920, https://doi.org/10.1038/nclimate2716.

PAGES2k Consortium, 2017: A global multiproxy database for temperature reconstructions of the Common Era. *Sci. Data*, **4**, 170088, https://doi.org/10.1038/sdata.2017.88.

Palmer, M., and Coauthors, 2018b: UKCP18 marine report. Met Office Rep., 133 pp., http://nora.nerc.ac.uk/id/eprint/522257/.

Palmer, M. D., G. R. Harris, and J. M. Gregory, 2018a: Extending CMIP5 projections of global mean temperature change and sea level rise due to thermal expansion using a physically-based emulator. *Environ. Res. Lett.*, **13**, 084003, https://doi.org/10.1088/1748-9326/aad2e4.

——, C. M. Domingues, A. B. A. Slangen, and F. Boeira Dias, 2021: An ensemble approach to quantify global mean sea-level rise over the 20th century from tide gauge reconstructions. *Environ. Res. Lett.*, **16**, 044043, https://doi.org/10.1088/1748-9326/abdaec.

Papale, P., 2018: Global time-size distribution of volcanic eruptions on Earth. *Sci. Rep.*, **8**, 6838, https://doi.org/10.1038/s41598-018-25286-y.

Perron, M., and P. Sura, 2013: Climatology of non-Gaussian atmospheric statistics. *J. Climate*, **26**, 1063–1083, https://doi.org/10.1175/JCLI-D-11-00504.1.

Pielke, R., Jr., M. G. Burgess, and J. Ritchie, 2022: Plausible 2005–2050 emissions scenarios project between 2°C and 3°C of warming by 2100. *Environ. Res. Lett.*, **17**, 024027, https://doi.org/10.1088/1748-9326/ac4ebf.

Rauch, H. E., F. Tung, and C. T. Striebel, 1965: Maximum likelihood estimates of linear dynamic systems. *AIAA J.*, **3**, 1445–1450, https://doi.org/10.2514/3.3166.

Roach, L. A., and Coauthors, 2020: Antarctic sea ice area in CMIP6. *Geophys. Res. Lett.*, **47**, e2019GL086729, https://doi.org/10.1029/2019GL086729.

Robinson, A., and H. Stommel, 1959: The oceanic thermocline and the associated thermohaline circulation. *Tellus*, **11A**, 295–308, https://doi.org/10.3402/tellusa.v11i3.9317.

Rodgers, K. B., and Coauthors, 2021: Ubiquity of human-induced changes in climate variability. *Earth Syst. Dyn.*, **12**, 1393–1411, https://doi.org/10.5194/esd-12-1393-2021.

Rosenblum, E., and I. Eisenman, 2017: Sea ice trends in climate models only accurate in runs with biased global warming. *J. Climate*, **30**, 6265–6278, https://doi.org/10.1175/JCLI-D-16-0455.1.

Sætrom, J., and H. Omre, 2013: Uncertainty quantification in the ensemble Kalman filter. *Scand. J. Stat.*, **40**, 868–885, https://doi.org/10.1111/sjos.12039.

Samset, B. H., C. Zhou, J. S. Fuglestvedt, M. T. Lund, J. Marotzke, and M. D. Zelinka, 2022: Earlier emergence of a temperature response to mitigation by filtering annual variability. *Nat. Commun.*, **13**, 1578, https://doi.org/10.1038/s41467-022-29247-y.

Särkkä, S., 2013: *Bayesian Filtering and Smoothing*. Cambridge University Press, 232 pp.

Sato, M., J. E. Hansen, M. P. McCormick, and J. B. Pollack, 1993: Stratospheric aerosol optical depths, 1850–1990. *J. Geophys. Res.*, **98**, 22 987–22 994, https://doi.org/10.1029/93JD02553.

Schmidt, S. F., 1981: The Kalman filter—Its recognition and development for aerospace applications. *J. Guid. Control*, **4**, 4–7, https://doi.org/10.2514/3.19713.

Schwartz, S. E., 2007: Heat capacity, time constant, and sensitivity of Earth's climate system. *J. Geophys. Res.*, **112**, D24S05, https://doi.org/10.1029/2007JD008746.

——, Harshvardhan, and C. M. Benkovitz, 2002: Influence of anthropogenic aerosol on cloud optical depth and albedo shown by satellite measurements and chemical transport modeling. *Proc. Natl. Acad. Sci.*, **99**, 1784–1789, https://doi.org/10.1073/pnas.261712099.

Sellers, W. D., 1969: A global climatic model based on the energy balance of the earth-atmosphere system. *J. Appl. Meteor.*, **8**, 392–400, https://doi.org/10.1175/1520-0450(1969)008%3C0392:AGCMBO%3E2.0.CO;2.

Sherwood, S. C., and Coauthors, 2020: An assessment of Earth's climate sensitivity using multiple lines of evidence. *Rev. Geophys.*, **58**, e2019RG000678, https://doi.org/10.1029/2019RG000678.

Sippel, S., N. Meinshausen, A. Merrifield, F. Lehner, A. G. Pendergrass, E. Fischer, and R. Knutti, 2019: Uncovering the forced climate response from a single ensemble member using statistical learning. *J. Climate*, **32**, 5677–5699, https://doi.org/10.1175/JCLI-D-18-0882.1.

Smith, C., and Coauthors, 2021: PCC WGI AR6 chapter 7. Zenodo, accessed 3 July 2024, https://doi.org/10.5281/zenodo.5211357.

——, and Coauthors, 2024: fair-calibrate v1.4.1: Calibration, constraining, and validation of the FaIR simple climate model for reliable future climate projections. *Geosci. Model Dev.*, **17**, 8569–8592, https://doi.org/10.5194/gmd-17-8569-2024.

Smith, S. G. L., and S. T. Gille, 1998: Probability density functions of large-scale turbulence in the ocean. *Phys. Rev. Lett.*, **81**, 5249, https://doi.org/10.1103/PhysRevLett.81.5249.

Smith, S. W., 2003: Moving average filters. *Digital Signal Processing: A Practical Guide for Engineers and Scientists*, Newnes, 277–284.

Soden, B. J., R. T. Wetherald, G. L. Stenchikov, and A. Robock, 2002: Global cooling after the eruption of Mount Pinatubo: A test of climate feedback by water vapor. *Science*, **296**, 727–730, https://doi.org/10.1126/science.296.5568.727.

Stratonovich, R. L., 1959: Optimum nonlinear systems which bring about a separation of a signal with constant parameters from noise. *Radiofizika*, **2**, 892–901.

——, 1960: Application of the Markov processes theory to optimal filtering. *Radio Eng. Electron. Phys.*, **5**, 1–19.

Sura, P., and P. D. Sardeshmukh, 2008: A global view of non-Gaussian SST variability. *J. Phys. Oceanogr.*, **38**, 639–647, https://doi.org/10.1175/2007JPO3761.1.

——, and S. T. Gille, 2010: Stochastic dynamics of sea surface height variability. *J. Phys. Oceanogr.*, **40**, 1582–1596, https://doi.org/10.1175/2010JPO4331.1.

Susskind, J., G. A. Schmidt, J. N. Lee, and L. Iredell, 2019: Recent global warming as confirmed by AIRS. *Environ. Res. Lett.*, **14**, 044030, https://doi.org/10.1088/1748-9326/aafd4e.

Swerling, P., 1959: First-order error propagation in a stagewise smoothing procedure for satellite observations. RAND Corporation Research Memo. RM-2329, 31 pp., https://www.rand.org/pubs/research_memoranda/RM2329.html.

Szopa, S., and Coauthors, 2021: Short-lived climate forcers. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 817–922, https://doi.org/10.1017/9781009157896.008.

Tebaldi, C., and R. Knutti, 2018: Evaluating the accuracy of climate change pattern emulation for low warming targets. *Environ. Res. Lett.*, **13**, 055006, https://doi.org/10.1088/1748-9326/aabef2.

——, and Coauthors, 2021: Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6. *Earth Syst. Dyn.*, **12**, 253–293, https://doi.org/10.5194/esd-12-253-2021.

Thacker, N. A., and A. J. Lacey, 1998: Tutorial: The Kalman filter. Citeseer Tech. Memo. 1996-002, 10 pp., https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c631533d3095f0385dee2774675300a02c6cf6b7.

Timmreck, C., and Coauthors, 2018: The Interactive Stratospheric Aerosol Model Intercomparison Project (ISA-MIP): Motivation and experimental design. *Geosci. Model Dev.*, **11**, 2581–2608, https://doi.org/10.5194/gmd-11-2581-2018.

Van Katwyk, P., B. Fox-Kemper, H. Seroussi, S. Nowicki, and K. J. Bergen, 2023: A variational LSTM emulator of sea level contribution from the Antarctic ice sheet. *J. Adv. Model. Earth Syst.*, **15**, e2023MS003899, https://doi.org/10.1029/2023MS003899.

van Vuuren, D. P., M. G. J. den Elzen, P. L. Lucas, B. Eickhout, B. J. Strengers, B. van Ruijven, S. Wonink, and R. van Houdt, 2007: Stabilizing greenhouse gas concentrations at low levels: An assessment of reduction strategies and costs. *Climatic Change*, **81**, 119–159, https://doi.org/10.1007/s10584-006-9172-9.

——, and Coauthors, 2017: Energy, land-use and greenhouse gas emissions trajectories under a green growth paradigm. *Global Environ. Change*, **42**, 237–250, https://doi.org/10.1016/j.gloenvcha.2016.05.008.

Vernier, J.-P., and Coauthors, 2011: Major influence of tropical volcanic eruptions on the stratospheric aerosol layer during the last decade. *Geophys. Res. Lett.*, **38**, L12807, https://doi.org/10.1029/2011GL047563.

Wan, E. A., and R. Van Der Merwe, 2000: The unscented Kalman filter for nonlinear estimation. *Proc. IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symp. (Cat. No. 00EX373)*, Lake Louise, Alberta, Canada, Institute of Electrical and Electronics Engineers, 153–158, https://doi.org/10.1109/ASSPCC.2000.882463.

Wielicki, B. A., B. R. Barkstrom, E. F. Harrison, R. B. Lee III, G. L. Smith, and J. E. Cooper, 1996: Clouds and the Earth's Radiant Energy System (CERES): An earth observing system experiment. *Bull. Amer. Meteor. Soc.*, **77**, 853–868, https://doi.org/10.1175/1520-0477(1996)077<0853:CATERE>2.0.CO;2.

Wild, M., and Coauthors, 2015: The energy balance over land and oceans: An assessment based on direct observations and CMIP5 climate models. *Climate Dyn.*, **44**, 3393–3429, https://doi.org/10.1007/s00382-014-2430-z.

——, M. Z. Hakuba, D. Folini, P. Dörig-Ott, C. Schär, S. Kato, and C. N. Long, 2019: The cloud-free global energy balance and inferred cloud radiative effects: An assessment based on direct observations and climate models. *Climate Dyn.*, **52**, 4787–4812, https://doi.org/10.1007/s00382-018-4413-y.

Willner, D., C. B. Chang, and K. P. Dunn, 1976: Kalman filter algorithms for a multi-sensor system. *1976 IEEE Conf. on Decision and Control including the 15th Symp. on Adaptive Processes*, Clearwater, FL, Institute of Electrical and Electronics Engineers, 570–574, https://doi.org/10.1109/CDC.1976.267794.

Wills, R. C. J., D. S. Battisti, K. C. Armour, T. Schneider, and C. Deser, 2020: Pattern recognition methods to separate forced responses from internal variability in climate model ensembles and observations. *J. Climate*, **33**, 8693–8719, https://doi.org/10.1175/JCLI-D-19-0855.1.

Winton, M., K. Takahashi, and I. M. Held, 2010: Importance of ocean heat uptake efficacy to transient climate change. *J. Climate*, **23**, 2333–2344, https://doi.org/10.1175/2009JCLI3139.1.

Wunsch, C., 2020: Is the ocean speeding up? Ocean surface energy trends. *J. Phys. Oceanogr.*, **50**, 3205–3217, https://doi.org/10.1175/JPO-D-20-0082.1.

——, and P. Heimbach, 2007: Practical global oceanic state estimation. *Physica D*, **230**, 197–208, https://doi.org/10.1016/j.physd.2006.09.040.

Yang, S., Z. Li, J.-Y. Yu, X. Hu, W. Dong, and S. He, 2018: El Niño–Southern Oscillation and its impact in the changing climate. *Natl. Sci. Rev.*, **5**, 840–857, https://doi.org/10.1093/nsr/nwy046.

Yeager, S. G., and Coauthors, 2022: The Seasonal-to-Multiyear Large Ensemble (SMYLE) prediction system using the Community Earth System Model version 2. *Geosci. Model Dev.*, **15**, 6451–6493, https://doi.org/10.5194/gmd-15-6451-2022.

Zanchettin, D., and Coauthors, 2016: The Model Intercomparison Project on the climatic response to Volcanic forcing (VolMIP): Experimental design and forcing input data for CMIP6. *Geosci. Model Dev.*, **9**, 2701–2719, https://doi.org/10.5194/gmd-9-2701-2016.

Zanna, L., S. Khatiwala, J. M. Gregory, J. Ison, and P. Heimbach, 2019: Global reconstruction of historical ocean heat storage and transport. *Proc. Natl. Acad. Sci. USA*, **116**, 1126–1131, https://doi.org/10.1073/pnas.1808838115.

Zelinka, M. D., D. A. Randall, M. J. Webb, and S. A. Klein, 2017: Clearing clouds of uncertainty. *Nat. Climate Change*, **7**, 674–678, https://doi.org/10.1038/nclimate3402.

Zhang, A., and M. M. Atia, 2020: An efficient tuning framework for Kalman filter parameter optimization using design of experiments and genetic algorithms. *Navigation*, **67**, 775–793, https://doi.org/10.1002/navi.399.

Zhu, J., and Coauthors, 2022: LGM paleoclimate constraints inform cloud parameterizations and equilibrium climate sensitivity in CESM2. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002776, https://doi.org/10.1029/2021MS002776.