Oxford Bulletin of Economics and Statistics

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Forecasting Climate Change Using a Multivariate Cointegrated System

Jennifer L. Castle[1] | Jurgen A. Doornik[2] | David F. Hendry[2] | Luke P. Jackson[3]

[1]Magdalen College and Climate Econometrics, University of Oxford, Oxford, UK | [2]Climate Econometrics, Nuffield College, University of Oxford, Oxford, UK | [3]Department of Geography, University of Durham, Durham, UK

Correspondence: Jennifer L. Castle (jennifer.castle@magd.ox.ac.uk)

Received: 20 December 2024 | **Revised:** 15 December 2025 | **Accepted:** 6 January 2026

Keywords: climate sensitivity | cointegration | energy-balance model | radiative forcing

ABSTRACT

A cointegrated vector equilibrium correction model of key climate variables including sea surface temperature, ocean heat content, Arctic sea-ice extent and sea-level change is built, driven by radiative forcing in which a stochastic trend arises due to anthropogenic emissions of greenhouse gases. A valid and congruent statistical model requires saturation estimation to model breaks in trends, while also conditioning on natural radiative forcings and El Niño–Southern Oscillation. The model is stable over 150 years, reflecting the slow adjustment of the deep oceans to increased greenhouse gas concentrations, and predicts an equilibrium climate sensitivity of 2.6°C. Projections out to 2100 highlight the many uncertainties over the coming decades.

JEL Classification: C32, C53, Q54

1 | Introduction

Forty years on from the groundbreaking publication of the Oxford Bulletin of Economics and Statistics on Cointegration, it is clear that cointegration is both theoretically and empirically important in modelling economic time series. 2026 will also be the centenary of Udny Yule's [1] path-breaking analysis of nonsense relations which he ascribed to unit roots in the generating process, and to Bradford Smith's linking levels and differences in econometric models [2]. Cointegration is a key route to utilising unit root information for a subset of variables linked by long-run stable relationships as conceived by Smith, so all three anniversaries should be celebrated together.

Cointegration as a concept has extended beyond economics. Any data characterised by non-stationary time series with long-run stable relationships between trending variables and short-run dynamic adjustments with feedback lends itself to modelling using cointegration techniques. One field in which cointegration has proven to be effective in modelling underlying physical relationships is climate. As Stern and Kaufmann [3] note, 'changes in radiative forcing might introduce a stochastic trend in temperature if the radiative forcing variables have a stochastic trend. This is likely because the concentrations of trace gases and sulphate aerosols are driven by anthropogenic emissions, which are determined by the stochastic trends that characterize many macroeconomic time series'. This paper explores the role of cointegration in modelling climate variables.

Pretis [4] shows that an energy-balance model (EBM) of global temperature, ocean heat content and radiative forcing is equivalent to a cointegrated system estimated from discrete time data. We use this mapping to build an extended model of global climate variables including air and sea temperatures, sea-ice extent and sea-level change. Cointegration techniques model the long-run stable physical relationships between radiative forcing and temperature outcomes, while allowing for short-run dynamics which can be non-linear. The model produces long-run forecasts of

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Oxford Bulletin of Economics and Statistics* published by Oxford University and John Wiley & Sons Ltd.

aggregate climate variables, first assuming endogenous anthropogenic greenhouse gas (GHG) and aerosol radiative forcing and then implementing scenario analysis by switching greenhouse gas and aerosol forcings to be exogenous and conditioning on future paths of greenhouse gases given by Shared Socio-Economic Pathways (SSPs; [5]) used in IPCC AR6 (2021).

The statistical empirical model is intended to complement physics-based general circulation models (GCMs) and falls within the category of simple climate/climate emulation modelling that has developed in the last 20 years [6, 7]. Many approximations are made by aggregation across time and space, as well as by focusing on a small number of climate variables. This suits the cointegrated vector autoregression (CVAR) analysis as it allows for all forms of non-stationarity, including unit roots, outliers and distributional shifts to be explicitly modelled, which is not typically feasible in GCMs. However, the complex physical relationships embedded in GCMs (see, e.g., [8]) are not fully captured. Earth system models of intermediate complexity (EMICs) are an attempt to bridge the gap between EBMs and GCMs, see Weber [9] for a summary, but do not explicitly model non-stationarity that CVARs are designed for.

The empirical analysis will extend current cointegrated climate models with sea-ice extent and sea level. This enables us to confirm the EBM relations established in the literature and evaluate the impacts on other climate variables within the system. We use saturation estimation to model distributional shifts and breaks within the system that would otherwise obscure the long-run relations. The CVAR methodology enables quantification of uncertainty, including model uncertainty, parameter estimation uncertainty and uncertainty over future paths of greenhouse gas emissions. The framework ensures a congruent model forms the basis of forecasting, where forecasts are driven by long-run stable cointegrating relationships proxying physical relationships.

The structure of the paper is as follows. Section 2 discusses the literature on cointegration in climate models, reviewing the early literature disputing the evidence of stochastic trends in climate data and addressing the debate over integrated of order 1, $I(1)$ versus $I(2)$. Section 3 outlines the data used in the empirical analysis and Section 4 develops a multivariate cointegrated model of climate variables. Section 5 assesses the robustness of the model, Section 6 produces scenario forecasts up to 2100, and Section 7 concludes, while details of the data are given in Appendix A. All empirical analysis is conducted in OxMetrics version 9, and PcGive version 16.[1]

2 | Literature Review

Cointegration methods have been applied to climate data since the early 2000s, with the seminal paper of Kaufmann and Stern [10] introducing cointegration to jointly model surface temperature and radiative forcing. They model Northern and Southern Hemisphere temperature as a function of radiative forcing of greenhouse gases, sulphur emissions, solar irradiance, and stratospheric sulphates. After finding two cointegrating relations they test for restrictions and find the just identified model implies that temperature in each hemisphere is determined by radiative forcing in that hemisphere and temperature in the opposite hemisphere. The components of radiative forcing do not cointegrate among themselves and the empirical results show that both natural variability and human activity drives the increase in surface temperature.

There followed numerous papers applying cointegration techniques to climate data, including Harvey and Mills [11], Liu and Rodríguez [12], and Kaufmann et al. [13]. Juselius [14] provides a non-technical summary of the methodology, along with a detailed analysis of the relationship between sea temperatures and land temperatures using the tools of multivariate cointegration. Looking at annual data over 1500–2000, she finds that sea temperatures drive land temperatures with the long-run trend in sea temperatures associated with solar radiation, greenhouse gasses and volcanic eruptions.

Pretis [4] first made the link between cointegrated VAR models and physical EBMs. He argues that the cointegration framework enables the quantification of parameter estimation uncertainty not feasible in the physical climate models, and it allows the bi-directional feedback between the climate and the economy to be modelled. A two component energy balance model explicitly models the heat transfer from the upper layer to the deep ocean which has a higher heat capacity so is slower to adjust. The heat transfer is assumed to be proportional to the difference in temperatures. Any stochastic trends present in radiative forcing (where the stochastic component is driven by GHG emissions) will be imparted onto the temperatures of the upper component, through which it transfers onto the lower component.

The use of cointegration analysis for climate variables has drawn some debate. Stern and Kaufmann [3] find evidence of stochastic trends in radiative forcing and temperatures. Gay-Garcia et al. [15], Estrada et al. [16], and Estrada et al. [17] argue that climate variables exhibit deterministic trends with breaks rather than stochastic trends, and temperature and radiative forcings are not integrated processes once breaks in trend are accounted for. Furthermore, Gadea-Rivas et al. [18] uses spatially disaggregated gridded data and conclude that global average temperatures are trend stationary with a structural break. But there is clear evidence of an integrated long-run earth system response to anthropogenic forcing, for example, the linear relation between cumulative carbon emissions and global temperature [19], and a substantial body of empirical evidence supports the view that climate variables contain stochastic trends, see, for example, Kaufmann et al. [20] and Kaufmann et al. [21]. Table A2 in Appendix A reports statistical tests for the climate data used in the analysis below, supporting the evidence for stochastic trends.

The next debate in the literature concerned whether the characteristic behaviour of the observed climate system is cointegrating $I(1)$ or multicointegrating $I(2)$ [25]. Bruns et al. [26] finds that the cumulation of the cointegration disequilibrium from surface temperature and radiative forcing (measuring system heat content) cointegrates with surface temperature. Benati [27] finds that radiative forcing, land and ocean temperatures are $I(2)$, and he uses a Bayesian VAR approach within an $I(2)$ cointegrated VECM model. Beenstock et al. [28] do not find evidence of a shared common stochastic trend using polynomial cointegration, but Pretis and Hendry [29] show that the time-series properties and degrees

of integrability of data can change over time. Hence, it is important to model regime shifts, which we allow for in the model below.

One of the benefits of the cointegration approach is that climate sensitivity can be calculated, which is policy relevant as it impacts the carbon budget for temperature targets. Climate sensitivity varies between observational and modelled estimates but it needs to be prescribed in any climate model to allow inter-comparison and context for results, see Sherwood et al. [30]. The transient climate response (TCR) is the temperature increase at the point associated with a doubling of atmospheric CO_2, whereas equilibrium climate sensitivity (ECS) refers to the temperature increase when the climate system fully adjusts to a doubling of CO_2. Both measures can be inferred from the CVAR as the model directly estimates the system equilibrium. Mills [31] finds a stable cointegrating relation between temperature and total radiative forcing over the period 1850–2000, and calculates a robust estimate of the temperature sensitivity to a doubling of radiative forcing to be in the range of 1°C – 3°C, with a point estimate of just over 2°C, whereas Bruns et al. [26] calculates climate sensitivity to be 2.8°C.

Other climate applications of the cointegration framework include Brock and Miller [32] who allow for polar amplification where the polar latitudes are warming faster. Schmith et al. [33] undertakes cointegration analysis between global sea-level change and land-ocean surface temperature, although they do not find sea-level rise depending on radiative forcing, probably due to the long adjustment process of the oceans. Jackson et al. [34] model the changing mass of polar ice sheets within a cointegrating framework and they find evidence that supports an $I(2)$ model with a bipolar relationship between Greenland and West Antarctica.

3 | Data

We estimate a six variable cointegrated VAR conditioning on two exogenous regressors, for annual data over the period 1880–2023. Table A1 in Appendix A summarises the data used in the analysis and their sources. The climate data (recorded in Figure 1, including (a) land-sea temperature anomaly *AT*; (b) ocean heat content *OHC*; (c) global mean sea-level change *GSL*; and (d) Arctic sea-ice extent *ICE*) are non-stationary, exhibiting stochastic trends and distributional shifts (although *OHC* is marginal for the [23] test of a unit root with trend break at unknown location).[2]

There are significant measurement changes in the data series over time. *OHC* is the volume integral of the product over the three-dimensional ocean of temperature, density and heat capacity at each measured point. *OHC* is observational data from 1957 (from three independent sources including the National Oceanic and Atmospheric Administration, the Meteorological Research Institute and the Institute of Atmospheric Physics) but we need a longer series to capture a stable relationship between *OHC*, *GSL*, and *ICE* given the slow (probably decadal) adjustment process. To extend the series to 1880 we use the independent global

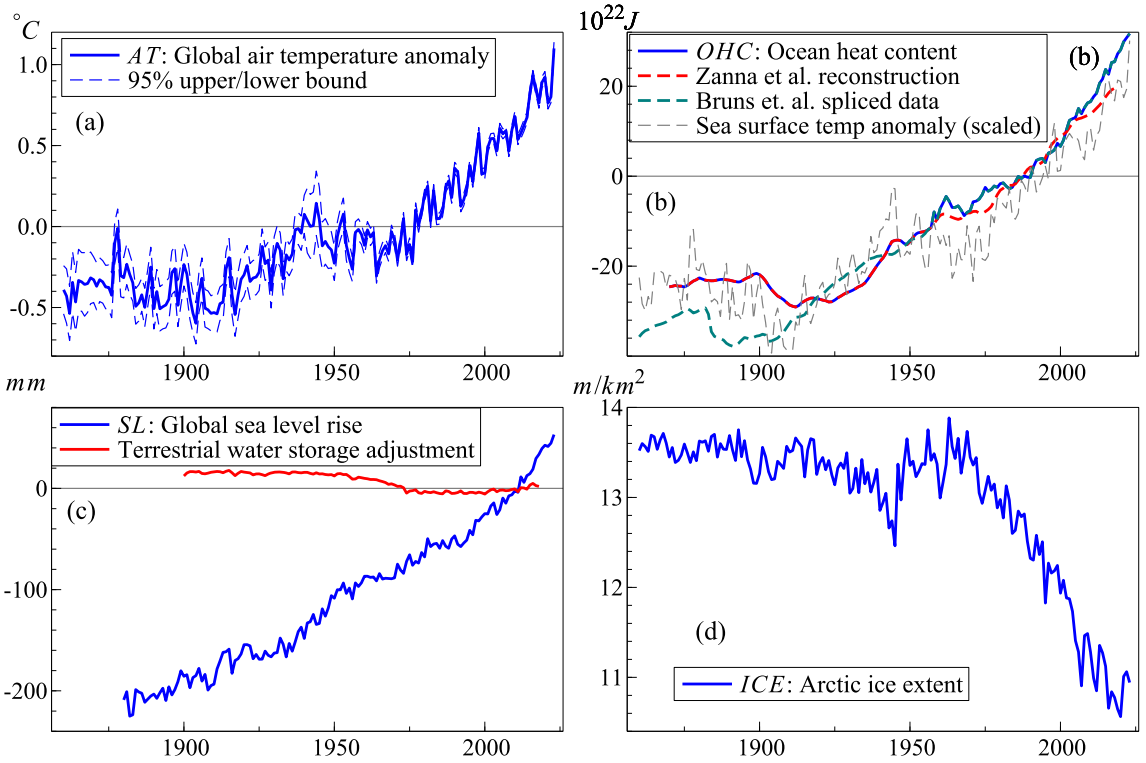

FIGURE 1 | (a) Mean land-sea temperature anomaly, *AT* (relative to the 1961–1990 average temperature) in °C, (b) ocean heat content down to 2000 m, *OHC* (measured relative to the 1971–2000 average) in 10^{22} J (with sea surface temperature anomaly in °C scaled to match the mean and variance of *OHC* for visual comparison), (c) global mean sea-level change, *GSL*, (measured relative to the 2002–2018 average sea level) in mm, adjusted for terrestrial water storage (in red), and (d) Arctic sea-ice extent (measuring the total area of the oceans with sea-ice concentrations of 15% or more, averaged across all months of the year) in m/km^2.

reconstruction of ocean heat content provided by Zanna et al. [35], who use a method to reconstruct global ocean temperatures that does not use any of the exogenous drivers in our model. The data are spliced in 1957, taking the average of the three observational series for the latter part of the sample. Figure 1b records the spliced data that we use, along with the full reconstruction from Zanna et al. [35] to compare with the observational data and the latent measure by Bruns et al. [26] based on an $I(2)$ model for comparison. We also plot sea surface temperature anomalies matched by means and ranges for context. Figure 1b shows the increase in variance for the observational data relative to the simulated data for the earlier part of the sample.

Modern *GSL* reconstructions combine tide gauge measurements and, after 1992, satellite altimetry, so that the estimated precision varies through time. There are three main contributions to *GSL*, ocean thermal expansion which can be mapped to the energy balance model and will be modelled by *OHC*, land ice, including melting, calving and dynamics, and terrestrial water storage (TWS). TWS is not directly affected by global climate change though it has varied through time due to human activity (e.g., dam construction, ground-water extraction). We correct the GSL reconstruction by the GSL-equivalent TWS contribution in Frederikse et al. [36] for 1900–2018, and extrapolating to 2023 using an unobserved components model. It is reasonable to assume a negligible effect on GSL by TWS prior to 1900.

Figure 2 shows estimates of effective radiative forcing, *ERF*, including natural forcings (*NF*) due to solar irradiance and stratospheric aerosols from volcanic eruptions (with some large volcanoes named in the figure), and those from greenhouse gases (*GHG*) and aerosols (*AF*), combined to give forcings from anthropogenic sources. *GHG* and *AF* are $I(1)$ with breaks, whereas *NF* are $I(0)$. Miller et al. [37] report a smoothed version of the anthropogenic radiative forcing data using a lowess smoother with a 20-year window to address the variability in the simulation estimates (the solid lines in Figure 2) which we use in the empirical analysis below. All but *NF* are non-stationary, although the evidence for *GHG* allowing for a structural break is mixed due to the smoothing but we proceed assuming it is $I(1)$. We model the three forcings separately to allow for different short-run and long-run effects, including *GHG* and *AF* as endogenous variables with *NF* as an exogenous regressor. *GHG* and *AF* are included separately as *GHG* should respond to changes in sea ice and land and sea temperature as they affect photosynthesis, respiration, the solubility of carbon in ocean water, and oceanic circulation, but they should have little or no effect on *AF*.[3] We rely on simulated rather than observational data as we do not have long-run time series of CH_4 and N_2O, but the close correspondence between ERF of GHGs and CO_2 emissions, with a correlation coefficient of 0.98 (0.99 for the spline smoothed data) supports our use of the simulated data.

Not all natural changes in Earth's energy balance are forcings. There is also endogenous natural variability including quasi-periodic cycles caused by atmosphere–ocean interactions at multi-annual timescales such as El Niño Southern Oscillation (ENSO) and North Atlantic Oscillation (NAO). The magnitude of climate variability, regardless of its cause (e.g., [42, 43]), is often sufficient to overprint anthropogenic warming so that periods of enhanced or damped global temperatures are observed (e.g., the early 2000's hiatus). Thus, accounting for climate modes such as ENSO that reflect these atmosphere–ocean interactions are important to include when constructing the CVAR (Figure 3). Although NAO (measuring the pressure difference between Iceland and Azores) is more regional, it does partially modulate Arctic sea-ice extent so is included in the initial specification although is found not to be significant. ENSO is the main driver of natural variability at our timescale and is highly significant.

Pretis et al. [44] shows that the response of temperatures to volcanic eruptions is fairly standard, with emissions blocking solar radiation which reduces temperatures, but then gradually get removed from the atmosphere resulting in a 'v' shaped response. Including *NF* as an exogenous regressor, partialing it out from the anthropogenic components of effective radiative forcing, means

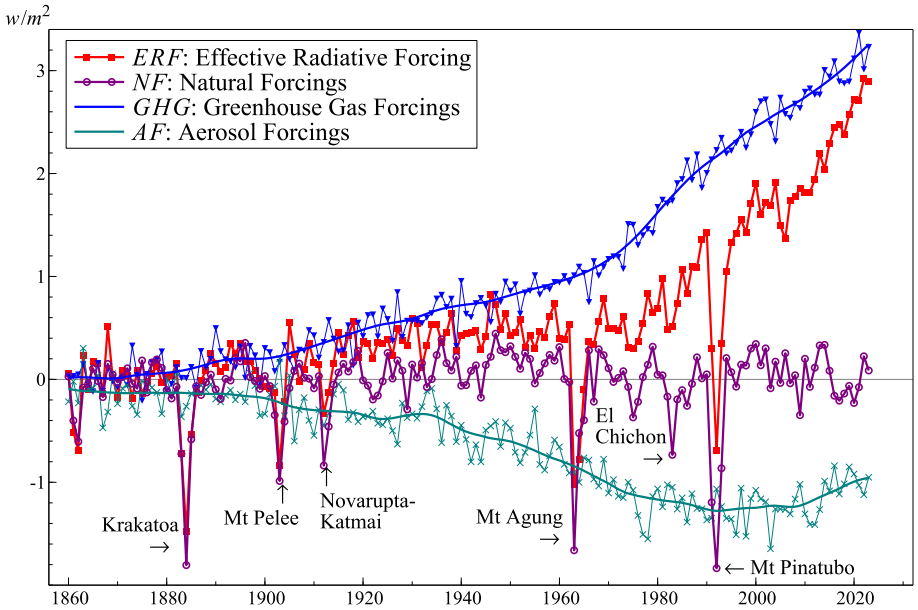

FIGURE 2 | Radiative forcings from IPCC CMIP6 Simulations using GISS-E2.1 (in W/m^2), see Miller et al. [37].

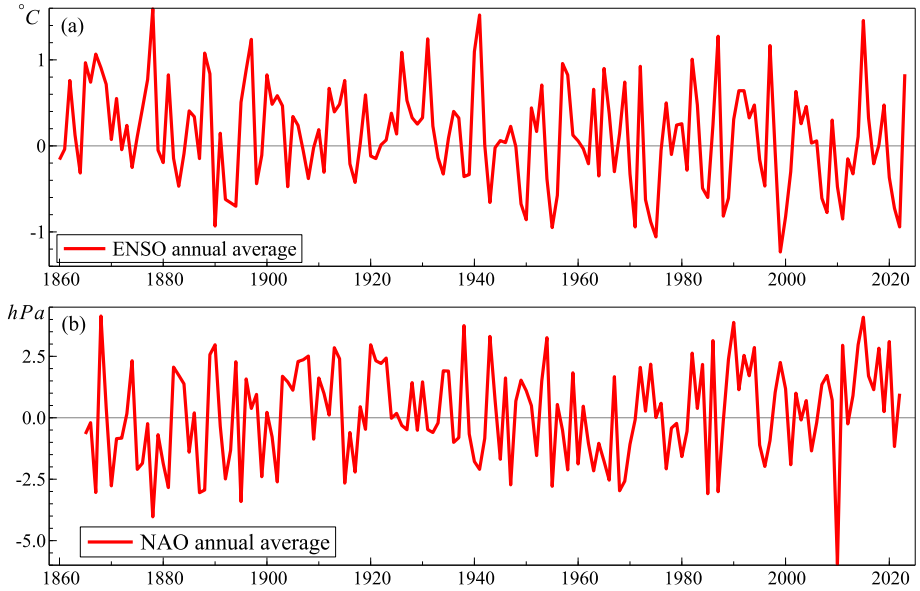

FIGURE 3 | Annually averaged climate series for (a) El Niño Southern Oscillation (ENSO) in °C; and (b) North Atlantic Oscillation (NAO) annual index in hectopascals (equivalent to a millibar of pressure).

that the volcanic eruptions do not enter into the long-run equilibrium of the system.

4 | A Multivariate Model of Climate Variables

4.1 | Econometric Methodology

We begin by estimating a six variable ($p = 6$) VAR in levels, commencing with four lags ($m = 4$) and condition on a set of restricted ($\mathbf{z}_{r,t}$) and unrestricted ($\mathbf{z}_{u,t}$) stationary regressors. The vectors of variables are denoted by $\mathbf{y}_t = (GHG, AF, AT, OHC, ICE, GSL)_t$; $\mathbf{z}_{r,t} = t$, where t is a full sample deterministic trend and $\mathbf{z}_{u,t} = (\mu, NF, ENSO, NAO)_t$, where μ is a full sample constant. The initial VAR is:

$$\mathbf{y}_t = \sum_{j=1}^{4} \mathbf{\Pi}_j \mathbf{y}_{t-j} + \mathbf{\Gamma}_u \mathbf{z}_{u,t} + \mathbf{\Gamma}_r \mathbf{z}_{r,t} + \sum_{k=2}^{T-1} \lambda_{s,k} S_{1882+k} \sum_{k=3}^{T-1} \lambda_{\tau,k} \tau_{1882+k} + \boldsymbol{\epsilon}_t \tag{1}$$

where $\boldsymbol{\epsilon}_t \sim \mathrm{IN}[\mathbf{0}, \Omega_\epsilon]$ and Ω_ϵ is a positive-definite symmetric covariance matrix of the error process.

Step indicator saturation (SIS) [45, 46] and trend indicator saturation (TIS) [24] is applied to the system (i.e., not equation by equation) at $\alpha = 0.001$, forcing all other regressors to be retained and selecting over just the indicators. Trend indicators are denoted τ_{year} and take the values $(-t, -t+1, -t+2, \ldots, 0, 0, 0 \ldots)$. Then selection is applied over lags and exogenous regressors at $\alpha = 0.01$. System and equation diagnostic tests are undertaken to ensure a congruent model specification, see Hendry and Doornik [47].

We then map to a vector equilibrium correction model (VECM):

$$\Delta \boldsymbol{y}_t = \left(\Pi, \Pi_l\right) \begin{pmatrix} y_{t-1} \\ t \end{pmatrix} + \sum_{i=1}^{m-1} \Gamma_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{d}_t + \boldsymbol{\epsilon}_t \tag{2}$$

where $\Pi = \boldsymbol{I}_p - \sum_{j=1}^{m} \Pi_j$ and $\boldsymbol{d}_t$ includes all the unrestricted retained regressors after selection including the intercept, saturation indicators and cyclical regressors.

When $\boldsymbol{y}_t \sim I(1)$, the cointegrating rank hypothesis for $r \leq p$ is:

$$\mathrm{rank}\left(\Pi, \Pi_l\right) \leq r, \quad \text{so} \quad \left(\Pi, \Pi_l\right) = \boldsymbol{\alpha}\left(\boldsymbol{\beta}', \gamma\right)$$

where $\boldsymbol{\alpha}$ is a $(p \times r)$ matrix of adjustment coefficients describing which equations adjust, and $\boldsymbol{\beta}$ is a $(p \times r)$ matrix of coefficients describing r long-run relations $\boldsymbol{\beta}' \boldsymbol{y}_t$ [48, 49].

If step or trend indicators are retained from the saturation estimation, (2) needs to be modified to allow for structural breaks, see Johansen et al. [50] and Kurita and Nielsen [51], where the model allows for different deterministic terms in sub-samples but maintains common dynamic parameters, so (2) holds for each sub-sample. Kurita and Nielsen [51] approximate the distribution of the cointegrating rank test statistic using a gamma-distribution and we use their response surface to calculate the quantiles of the cointegrating rank test below.

4.2 | Mapping From Physical Theory to Empirical Model

In the two-component EBM of Pretis [4], the stochastic trends present in radiative forcing driven by anthropogenic emission of GHGs from economic activity will be imparted onto the temperatures of the upper component, through which it transfers onto the lower component, captured by the postulated cointegrating relations:[4]

$$\boldsymbol{y}_t^{\dagger} = \begin{pmatrix} ERF \\ AT \\ OHC \end{pmatrix}; \qquad \beta' = \begin{pmatrix} 1 & -\lambda & 0 \\ 0 & 1 & -1 \end{pmatrix} \tag{3}$$

We extend this model by decomposing effective radiative forcing into two anthropogenic components with very different trends, as well as including sea level and sea-ice extent. There is no direct mapping of the physical relationships extending the two component energy balance model to sea-level change and sea-ice extent as the processes are multifaceted and complex.

For *GSL* the main contributions can be separated into three parts: ocean thermal expansion, land ice (both melting and changes around the fringes including calving and dynamics along with sublimation), and terrestrial water storage including extraction of water and dams. We correct the observed global mean sea-level data by the impact of terrestrial water storage (which is not directly induced by climate change) such that the global sea level modelled by the CVAR is purely climate induced. Additionally, changes in sea-ice (volume) does not affect sea level so we impose this constraint in the cointegrating relation. We do not include a separate variable for land-ice in the cointegrated system as the response of glaciers to radiative forcing is quite different to ice sheets (Greenland and Antarctica), which also have idiosyncratic dynamics, but allow for these changes to be modelled via the response of sea level to radiative forcing and surface air temperature.

Extending the energy balance model to include the postulated statistical relationships between climate variables results in the cointegrating relations as:

$$\boldsymbol{y}_t = \begin{pmatrix} GHG \\ AF \\ AT \\ OHC \\ ICE \\ GSL \end{pmatrix}; \qquad \boldsymbol{\beta}' = \begin{pmatrix} 1 & -\lambda_{1,a} & -\lambda_{1,b} & 0 & 0 & 0 \\ 0 & 0 & 1 & -\lambda_2 & 0 & 0 \\ \lambda_{3,a} & \lambda_{3,b} & 0 & \lambda_{3,c} & 1 & 0 \\ -\lambda_{4,a} & 0 & -\lambda_{4,b} & -\lambda_{4,c} & 0 & 1 \end{pmatrix} \tag{4}$$

where $\lambda_{1,b}$ determines the equilibrium climate sensitivity but we allow for different weights on the anthropogenic forcings driving temperature. The second cointegrating relation relates air temperature to ocean heat content ([4] relates the mixed layer temperature anomaly to the deep compartment temperature anomaly with a cointegrating relation $(1, -1)$, but we do not impose a unit cointegrating relation initially). Our third cointegrating relation captures the effect of greenhouse gas radiative forcing on the extent of Arctic sea ice, allowing for warming oceans to reinforce the rate of ice melt, and the fourth maps the relationship from greenhouse gas radiative forcing to sea level via ocean heat content (leading to thermal expansion) while also allowing for indirect effects of land-ice melt via either radiative forcing or temperature, but imposing the constraint that Arctic ice melt does not lead to sea-level rise.

4.3 | Econometric Modelling Results

The VAR(4) was reduced to VAR(3), selected by a system test at 1%. Saturation on the VAR(3) over 1882–2023 at $\alpha = 0.001$ results in two trend breaks, (τ_{1963} and τ_{1992}). Both breaks are likely precipitated by Mt. Agung (1963) and Pinatubo (1991) which are known to have caused temporary (1–2 year), measurable falls in *AT* and a brief slowing of *GSL*. Also see Ivanov and Evtimov [52]

TABLE 1 | System diagnostic tests.

$F_{ar}(72,544) = 1.2$	Second-order vector residual autocorrelation [54, 55]
$\chi^2_{nd}(12) = 14.9$	Vector residual non-normality [22, 56, 57]
$F_{het}(270,538) = 1.2$	Vector residual heteroscedasticity [58, 59]
$F_{reset}(72,544) = 1.7^{**}$	RESET functional form [60]
$F_{nl}(378,296) = 1.3^{*}$	Non-linearity index test [61]

Note: "*" indicates test significance at 5%, and "**" at 1%.

TABLE 2 | Correlations between residuals (standard deviations on diagonal).

	GHG	***AF***	***AT***	***OHC***	***ICE***	***GSL***
GHG	0.00					
AF	0.03	0.002				
AT	0.09	0.06	0.08			
OHC	0.01	−0.08	0.08	0.69		
ICE	0.12	0.04	−0.23	0.11	0.18	
GSL	0.02	0.06	0.16	0.11	−0.07	6.96

in which 1963 has been widely identified as a break in Northern Hemisphere annual surface temperature. Impulse indicator saturation (IIS) [53] identified outliers in *OHC*; $D_{1958} = 1$ in 1958 and -1 in 1959 which captures the point at which the *OHC* data is spliced between the latent and observational data, and $I_{2016} = 1$ for 2016 and 0 otherwise. A dummy is included for 1943–1945 when Arctic sea-ice extent and sea level measurements are not accurately recorded during the second world war. *NAO* was not significant but *ENSO* is retained, along with lag 0 and 1 of *NF*. The model passes most system diagnostic tests, see Table 1, but there is evidence of functional form misspecification and non-linearity, likely due to the change in measurement of *OHC* mid-sample. The diagnostic tests are conducted conditional on the selection of outliers and breaks which affect the properties of the tests, discussed in the references given.

Table 2 reports the correlations between the residuals with the equation standard deviations on the diagonal. There are large residual correlations between air temperature with Arctic sea-ice extent and sea level which lend themselves to being modelled.

Figure 4 shows the recursive residuals for each equation in the top row, the one-step Chow [62] tests to evaluate one-step ahead forecasts testing for constancy of each additional observation in the second row [63], and the breakpoint Chow test where each point is the value of the Chow *F*-test for that date against the final period, scaled by its 1% critical value, with the forecast horizon decreasing from left to right in the bottom row [64]. While the vast majority of the one-step residuals lie within their anticipated 95% confidence intervals, so we do not reject constancy overall, there is a clear break in 1958 corresponding to the point at which the latent simulated data on *OHC* is spliced to the observational data which is first recorded in 1957. This is also reflected in an increase

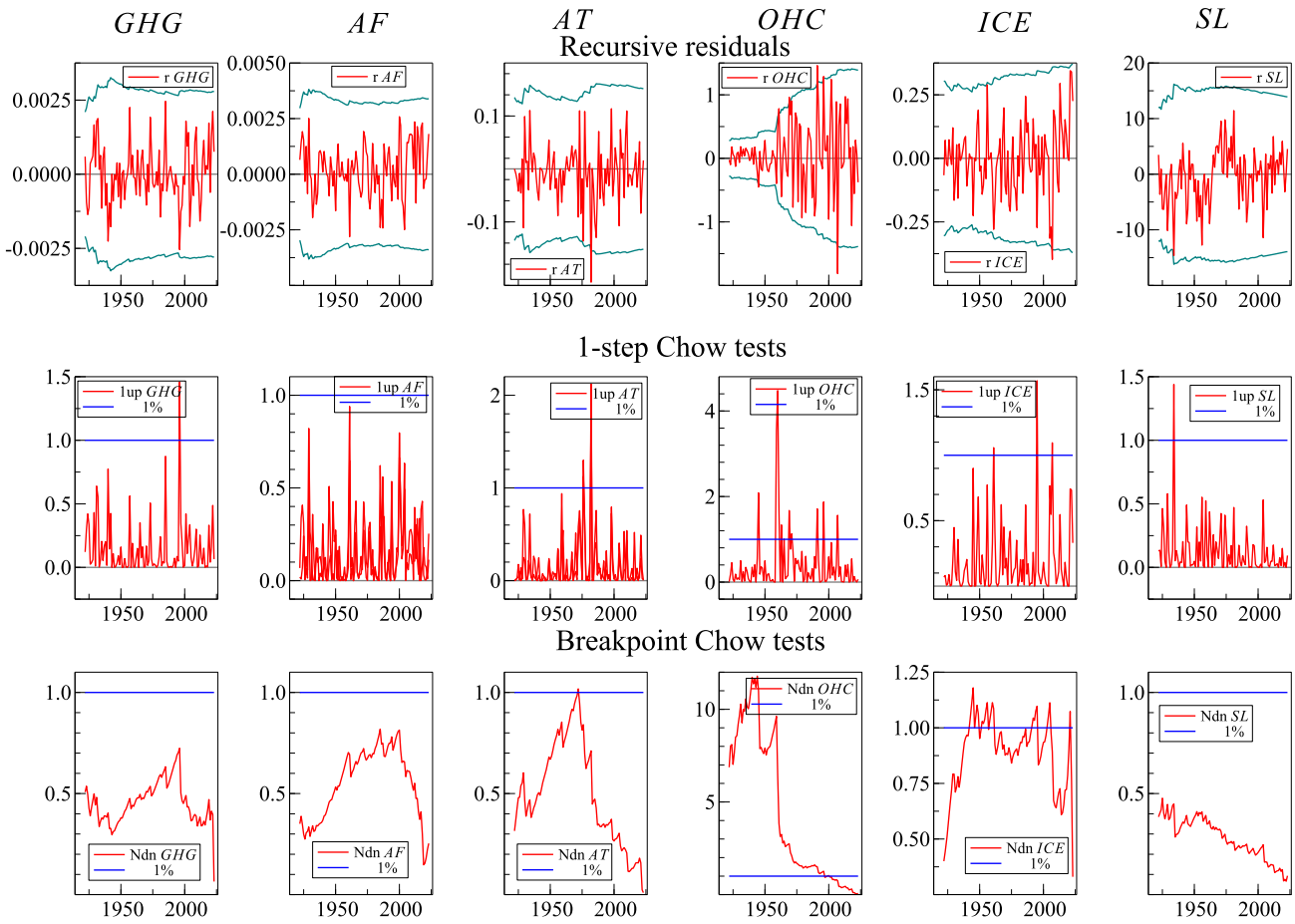

FIGURE 4 | Recursive residuals estimates and constancy tests for the system.

in variance over the latter period. As this is a construct of the data measurement and not an aspect of the underlying physical process, we proceed by including the 'blip' dummy variable D_{1958} to account for the measurement change.

Figure 5 records the roots of the companion matrix. We undertake cointegration analysis on a partial system conditioning on weakly exogenous variables (*NF* and *ENSO*) with broken trends, using Kurita and Nielsen [51] to obtain the relevant quantiles for determining the cointegrating rank.

The two trend breaks result in three regimes to calculate the pseudo-likelihood ratio test (PLR) critical values.[5] The resulting quantiles and p values are recorded in Table 3, along with the PLR test statistics for the determination of cointegrating rank.

The rank is greater than zero which points to cointegration between the variables, and there are not more roots close to unity than the dimension of the system which rules out an $I(2)$ representation. The trace test is marginal for a rank of 3 with stronger evidence for a rank of 4 which aligns with our postulated model. The data does not accept the restriction that *GHG* does not cointegrate with *ICE* or *GSL*, so they are driven directly by the anthropogenic stochastic trend as well as indirectly through *OHC*, and *AT* for *GSL*.

We initially impose the restrictions given in (4) which are accepted with the likelihood ratio test given by $\chi^2(1) = 1.91$ [$p = 0.17$]. However, this does not identify $\boldsymbol{\beta}$ so we impose a further restriction that aerosol forcings do not enter the *ICE* cointegrating relation. Although aerosols affect aerosol optical depth, which affects the fraction of solar insolation that reaches the ground and hence, the rate at which ice melts, we find that restriction is accepted with $\chi^2(1) = 2.02[0.15]$, and results in an identified $\boldsymbol{\beta}$ matrix. As such, (5) differs to (4) by one element. We undertake further restrictions on the $\boldsymbol{\alpha}$ matrix based on tests of significance, and the joint set of restrictions is comfortably accepted ($\chi^2(9) = 12.14[0.21]$) such that the $\boldsymbol{\Pi}$ matrix is identified.

The cointegrating relations are recorded in Figure 6, along with the sub-sample estimates from 1958 onwards (to coincide with the observational data on *OHC*). All cointegration relations are stationary based on Augmented Dickey Fuller tests although the ocean heat content and sea level cointegrating relations do exhibit a small but not statistically significant trend in the second half of the sample. This is not as evident in the sub-sample estimates and could be due to the switch from latent to observational data, or omitted Antarctic Ice data.[6] Equation (5) reports the cointegrating relations, excluding trend indicators, along with the $\boldsymbol{\alpha}$ matrix:

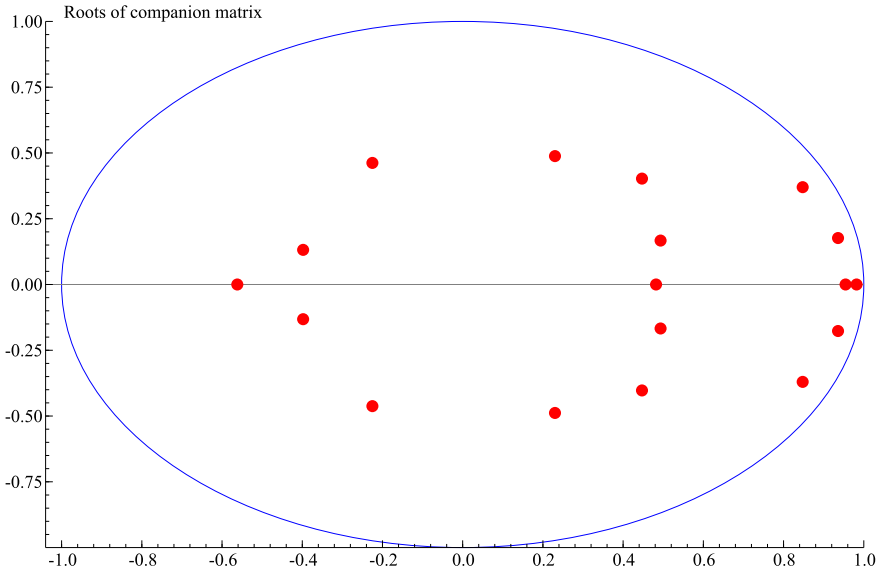

FIGURE 5 | Eigenvalues of the dynamic system.

TABLE 3 | Testing for cointegrating rank.

	$r = 0$	$r \leq 1$	$r \leq 2$	$r \leq 3$	$r \leq 4$	$r \leq 5$
$PLR\{H_l(r)\|H_l(6)\}$	276.9 [0.00][a]	195.1 [0.00][a]	132.0 [0.008][a]	80.9 [0.07][a]	44.6 [0.31]	14.1 [0.74]
95% limit quantiles	199.9	158.2	120.2	86.2	56.0	29.0

Note: Figures in square brackets are *p* values.
[a]Significance at the 10% level.

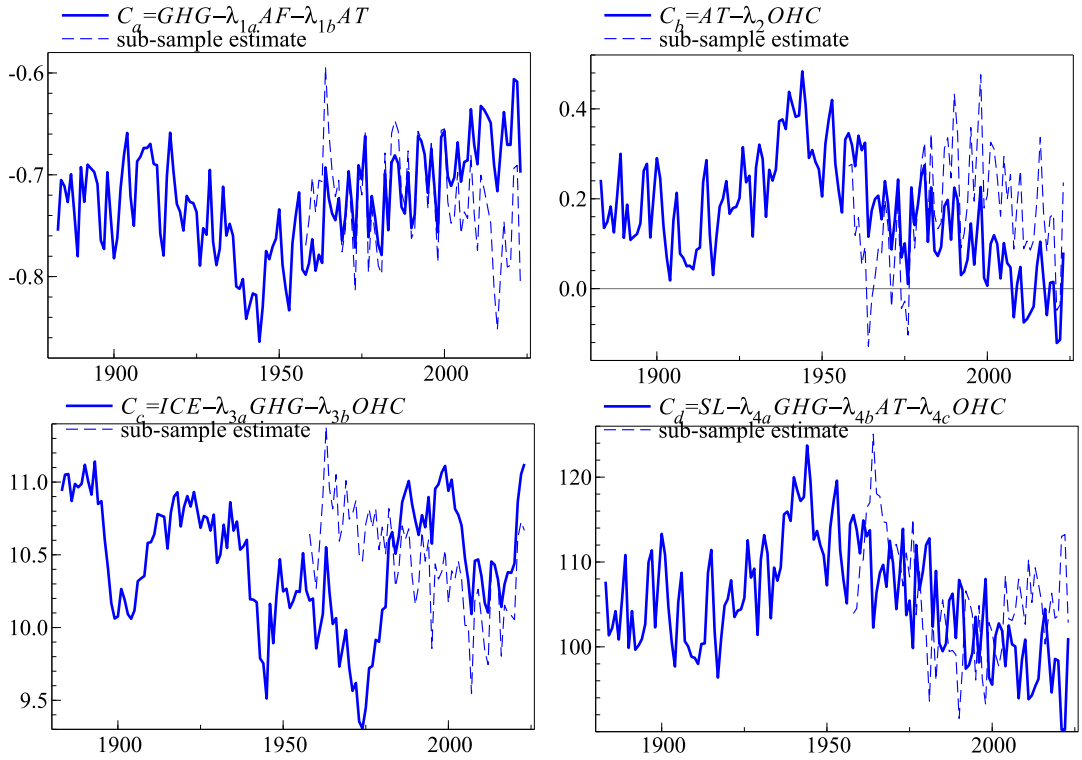

FIGURE 6 | Cointegrating relations.

$$\alpha\beta' = \begin{bmatrix} 0.014 & -0.002 & 0 & -0.069 \\ (0.006) & (0.010) & & (0.014) \\ -0.052 & -0.052 & -0.003 & 0.017 \\ (0.009) & (0.014) & (0.001) & (0.004) \\ 0.867 & 1.956 & 0 & 0 \\ (0.172) & (0.465) & & \\ -0.089 & 0 & 0 & 0 \\ (0.227) & & & \\ -3.173 & 6.673 & -1.027 & 0.003 \\ (0.915) & (1.453) & (0.118) & (0.000) \\ 0 & -15.24 & 0 & -0.002 \\ & (23.49) & & (0.004) \end{bmatrix} \begin{bmatrix} 1 & 0.308 & -3.072 & 0 & 0 & 0 \\ & (0.071) & (0.055) & & & \\ 0 & 0 & 1 & 0.014 & 0 & 0 \\ & & & (0.006) & & \\ 11.46 & 0 & 0 & -0.054 & 1 & 0 \\ (1.304) & & & (0.016) & & \\ 56.41 & 0 & -63.83 & -0.208 & 0 & 1 \\ (4.13) & & (1.074) & (0.0.047) & & \end{bmatrix} \tag{5}$$

To calculate the equilibrium climate sensitivity (ECS) we estimate the cointegrating relations excluding the broken trends, implementing the same cointegrating relations.[7] The ECS measures the equilibrium temperature response to a doubling of CO_2 concentrations, via radiative forcing. Using the IPCC estimate that a doubling of CO_2 increases radiative forcing by $3.7\,\mathrm{Wm}^2$ (see [65], 357) with $\tilde{\lambda}_{1b} = -1.418$ we obtain $ECS = 2.6^{\circ}C$, which is similar to Bruns et al. [26] and in line with IPCC AR6 (2021) which states there is a 90% or more chance that the ECS is between 2°C and 5°C.

Next, we map to a cointegrated system and undertake model reduction by eliminating regressors in each equation that were insignificant at 5% [66]. We augment the system with non-linear functions of the regressors to check for unmodelled non-linearity but none were retained. The series of reductions results in 120 restrictions which are accepted with $\chi^2(120) = 142[0.08]$. The resulting model estimated over 1883–2023 is:

$$\begin{aligned}
\widehat{\Delta GHG}_t &= \underset{(0.006)}{-0.033} - \underset{(0.002)}{0.014} C_{a,t-1} - \underset{(0.001)}{0.065} C_{d,t-1} + \underset{(0.076)}{1.28}\,\Delta GHG_{t-1} - \underset{(0.074)}{0.32}\,\Delta GHG_{t-2} \\
&\quad + \underset{(0.041)}{0.14}\,\Delta AF_{t-1} - \underset{(0.042)}{0.10}\,\Delta AF_{t-2} \\
\widehat{\Delta AF}_t &= \underset{(0.015)}{-0.001} - \underset{(0.006)}{0.039} C_{a,t-1} - \underset{(0.010)}{0.046} C_{b,t-1} - \underset{(0.0006)}{0.002} C_{c,t-1} - \underset{(0.0002)}{0.001} C_{d,t-1} \\
&\quad + \underset{(0.054)}{1.58}\,\Delta AF_{t-1} - \underset{(0.053)}{0.71}\,\Delta AF_{t-2} \\
\widehat{\Delta AT}_t &= \underset{(0.474)}{1.62} + \underset{(0.141)}{0.94} C_{a,t-1} + \underset{(0.393)}{2.24} C_{b,t-1} - \underset{(0.578)}{1.30}\,\Delta GHG_{t-1} + \underset{(0.010)}{0.09}\,ENSO_t \\
&\quad + \underset{(0.017)}{0.10}\,NF_{t-1} + \underset{(0.047)}{0.19}\,D_{WWII} + \underset{(0.001)}{0.002}\Delta GSL_{t-1} + \underset{(0.001)}{0.002}\Delta GSL_{t-2} \\
\widehat{\Delta OHC}_t &= \underset{(5.98)}{-16.24} - \underset{(3.21)}{8.60} C_{b,t-1} + \underset{(0.235)}{0.65}\, C_{c,t-1} - \underset{(0.0005)}{0.001} C_{d,t-1} + \underset{(0.157)}{0.32}\,NF_{t-1} \\
&\quad + \underset{(0.070)}{0.18}\,\Delta OHC_{t-1} + \underset{(0.065)}{0.25}\,\Delta OHC_{t-2} - \underset{(0.266)}{1.43}\,\Delta ICE_{t-1} - \underset{(0.264)}{0.84}\,\Delta ICE_{t-2} \\
&\quad + \underset{(0.481)}{2.86}\,D_{1958} - \underset{(0.680)}{2.40}\,I_{2007} - \underset{(0.686)}{2.55}\,I_{2016} \\
\widehat{\Delta ICE}_t &= \underset{(1.912)}{17.5} - \underset{(0.648)}{3.67} C_{a,t-1} + \underset{(1.180)}{5.07} C_{b,t-1} - \underset{(0.076)}{0.91} C_{c,t-1} + \underset{(0.0003)}{0.003} C_{d,t-1} \\
&\quad + \underset{(1.365)}{4.48}\,\Delta GHG_{t-2} + \underset{(5.307)}{12.0}\,\Delta AF_{t-2} - \underset{(0.110)}{0.50}\,D_{WWII} - \underset{(0.171)}{55}\,I_{2007} \\
\widehat{\Delta GSL}_t &= \underset{(6.246)}{-3.44} - \underset{(0.0006)}{0.060} C_{d,t-1} - \underset{(0.077)}{0.366}\Delta GSL_{t-1}
\end{aligned} \tag{6}$$

Figure 7 records the model fit, residuals, residual QQ plot (see [67], who recommend QQ plots over the Kernel density of residuals) and correlogram, and Table 4 reports the model diagnostics, noting that misspecification tests are conducted after model selection.

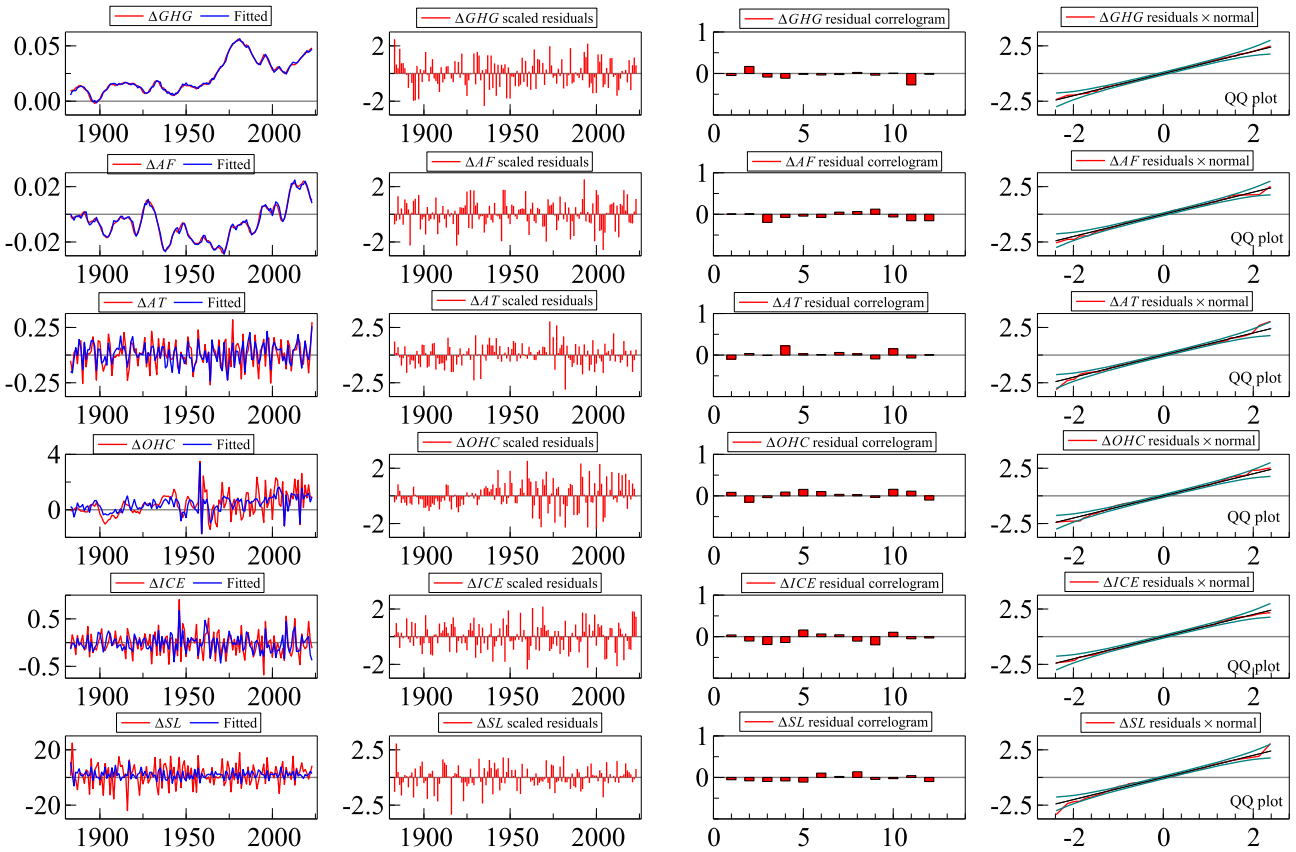

FIGURE 7 | Model fit, residuals, residual correlograms and residual QQ-plots for (6).

TABLE 4 | Equation standard errors and diagnostic tests for (6).

	ΔGHG	ΔAF	ΔAT	ΔOHC	ΔICE	ΔGSL
Single-equation tests						
F_{ar}	2.31 [0.10]	0.20 [0.82]	1.76 [0.18]	7.09 [0.002]**	1.69 [0.19]	1.87 [0.16]
F_{arch}	1.12 [0.29]	0.91 [0.34]	0.11 [0.74]	1.19 [0.28]	0.31 [0.58]	2.22 [0.14]
χ^2_{nd}	1.02 [0.60]	0.79 [0.67]	3.19 [0.20]	0.57 [0.75]	0.89 [0.64]	5.53 [0.06]
F_{het}	0.40 [0.96]	1.09 [0.37]	1.92 [0.03]*	1.40 [0.14]	1.03 [0.43]	1.53 [0.20]
$\hat{\sigma}$	0.001	0.002	0.078	0.669	0.175	7.47
$s.d.(\Delta y)$	0.015	0.012	0.119	0.893	0.242	8.09
System tests						
$F_{ar}(72,636) = 1.06[0.34]$						
$F_{arch}(36,569) = 1.23[0.17]$						
$\chi^2_{nd}(12) = 10.8[0.55]$						
$F_{het}(282,533) = 1.29[0.006]^{**}$						
LR test of over-identifying restrictions: $\chi^2(120) = 142[0.08]$						

Note: $\hat{\sigma}$ is the residual standard deviation with $s.d.(\Delta y)$ the unconditional standard deviation of Δy for comparison. p values in brackets. F_{ar} tests residual autocorrelation (see [54], where [55], allows for ADL and VAR models with stationary, unit and explosive roots and polynomial trends) F_{arch} tests autoregressive conditional heteroscedasticity [68], F_{het} tests residual heteroscedasticity (see [58], and [59], who show that residual heteroscedasticity tests are robust to outlier removal), $\chi^2_{nd}(2)$ tests non-Normality (see [22], and [57], who show normality testing is fragile after selection with a fixed cut off). "*" denotes significance at 5% and "**" denotes significance at 1%.

The model is well-specified apart from some evidence of autocorrelation in ΔOHC and system wide heteroscedasticity due to the measurement change in 1957 (observed in the residual variance of ΔOHC in Figure 7). The equation standard deviations are close to those of the unrestricted VAR in (2).

The first cointegrating relation which captures the stable relation between anthropogenic forcings and air temperature is highly significant in the AT equation, highlighting the feedthrough effects of the stochastic trend from anthropogenic sources onto temperatures, along with cyclical factors including natural radiative forcing and $ENSO$. ΔGSL_t and ΔGSL_{t-1} are retained to satisfy tests of reduction but are marginal in significance. Ocean heat content is primarily driven by the second, third and fourth cointegrating vectors but there are significant short-run dynamics from natural radiative forcing and ICE. Arctic ice melt is driven by all four long-run relations as well as short-run anthropogenic effects. Finally, sea-level rise has long-run trends picked up by the relationship between GHG forcing and sea-level rise, along with some inertia. The trend indicators retained in the VAR(3) enter the long-run cointegrating relations and only two additional impulse indicators for ocean heat content (2007 and 2016) are needed in the VECM to obtain a constant parameter stable model over a century and a half.

5 | Robustness

To evaluate model stability we undertake a dynamic simulation of the model conditioning on initial conditions up to 1904. We use actual values of the non-modelled variables (NF and $ENSO$, along with deterministic terms) but simulated values of the endogenous variables (similar to conditional dynamic forecasting within the estimation period). Figure 8 records the simulated outcomes for $\Delta \boldsymbol{y}$ in the top row and $\boldsymbol{y}$ in the bottom row. Although the variation in the differences is hard to model, the system captures the long term trends closely. The estimates of the turning points such as aerosol forcings around the turn of the century and the reduction in ice extent in the mid-1960s tend to lag the actual turning points but the overall trends in levels are fairly accurate. However, see the critiques of dynamic simulation as a method of evaluating econometric systems in Hendry and Richard [69] and Chong and Hendry [70]. As the trends are crucial for long-term forecasting this gives us confidence that absent future breaks due to anthropogenic changes, the system could extrapolate reliable trends.

5.1 | In-Sample Forecast Evaluation

To assess the role of cointegration and saturation in the empirical model we undertake an in-sample forecasting exercise, evaluating 30 years of dynamic forecasts over 1994–2023, but note that forecast performance is not a measure of the validity of the model, see Castle and Hendry [71]. We condition on known exogenous regressors (NF and $ENSO$) and in-sample structural breaks, holding the model specification as fixed but the parameters are estimated over 1883–1993. We compare model (6), denoted VECM(ISE) to indicate it is the vector equilibrium correction model with indicator saturation estimation derived from the CVAR specification, with a VAR in levels (VAR) and differences (DVAR) both with and without indicator saturation (ISE), and some benchmarks including an AR(1), random walk (RW) and trend indicator saturation (TIS). Levels forecasts are obtained from the differenced forecasts for VECM and DVAR. Table 5 reports the root mean square forecast errors (RMSFE) over the 30 dynamic forecasts for each model, along with tests of forecast accuracy comparison using Harvey et al. [72] which is a Diebold and Mariano [73] test with a bias correction for small samples (30

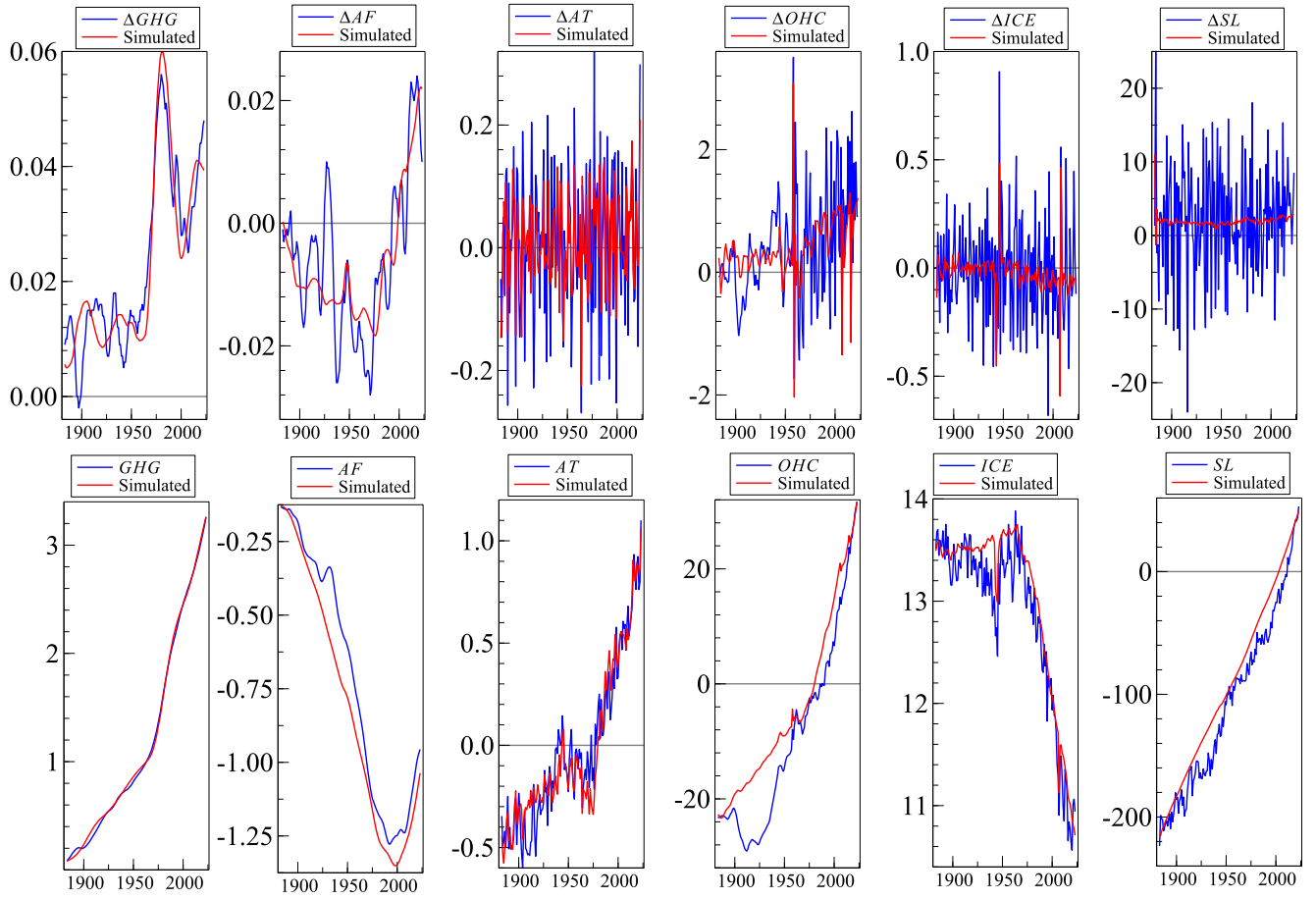

FIGURE 8 | Simulated outcomes from (6) for Δ*GHG*, Δ*AF*, Δ*AT*, Δ*OHC*, Δ*ICE* and Δ*GSL* in the top panels and *GHG*, *AF*, *AT*, *OHC*, *ICE* and *GSL* in the bottom panels, conditioning on known values for *NF*, *ENSO* and impulses, commencing in 1904.

TABLE 5 | RMSFEs for dynamic forecasts over 1994–2023.

	VECM		VAR		DVAR				
	ISE		ISE		ISE		AR(1)	TIS	RW
GHG	**0.02**	0.20**	0.61**	0.47**	0.19**	0.05*	0.57**	0.03	*0.59***
AF	**0.02**	0.05**	0.12**	0.35**	0.24**	0.23**	*0.47***	0.33**	0.15**
AT	**0.07**	0.08	0.31**	0.17*	0.16*	0.15*	*0.74***	0.41**	0.49**
OHC	3.92	**3.30***	8.47**	8.41**	10.1**	10.6**	11.0**	3.86	*15.3***
ICE	**0.26**	0.36**	0.45**	0.97**	0.30**	0.86**	*1.84***	0.54**	1.10**
GSL	**6.61**	11.7**	55.5**	22.7**	10.8**	37.8**	55.2**	23.0**	*59.7***

Note: Smallest RMSFE in bold, largest in italics. Tests of significant forecast differences using Harvey et al. [72] against VECM(ISE) as the benchmark, where ** and * denotes significance at 1% and 5%, respectively.

forecasts in our evaluation). The benchmark forecasts are given by VECM(ISE) (first column) and pairwise comparisons are evaluated across all other forecasts, recording significant differences at 1% and 5%.

Model (6) produces the smallest RMSFE forecasts for most variables, almost all of which are significant at 1%. However, the VECM without saturation (used to calculate ECS) dominates for *OHC* (significant at 5%). Explicitly modelling the long-run relations using cointegration improves forecasts relative to the VAR and DVAR suggesting the physical relations embodied in the VECM are critical in extrapolating future trends. The evidence on modelling breaks is more mixed. ISE mostly improves the VECM forecasts, but for misspecified models the saturation estimators are likely to be capturing misspecification and hence the wrong trends can be extrapolated. (6) outperforms all naive benchmarks, indicating that there is additional information in the system rather than modelling the climate variables equation by equation. Hence, we conclude that modelling long-run relationships within a system via multivariate cointegration is important for climate modelling.

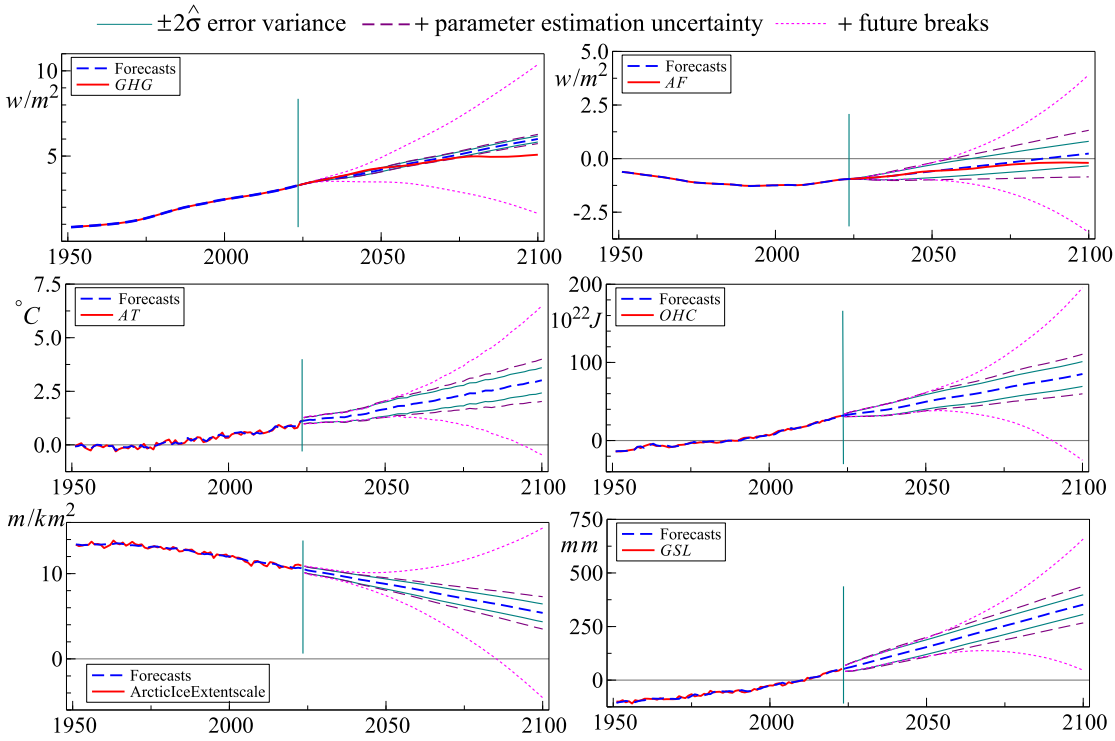

FIGURE 9 | Model forecast of each climate variable from 2024 to 2100 for system (6) and their associated uncertainties.

6 | Forecasting Climate Variables

The VECM (6) is a statistical approximation to the underlying physical model but is not a facsimile so is not informative about theory. It can, however, be useful as a forecasting device, see Clements and Hendry [74] who consider the implications of imposing cointegration for forecasting. Other than *ENSO* and *NF* the model is closed, that is, all variables are modelled and so absent future shifts, for example driven by policy or technological change, we can project infinitely far into future. Hendry and Mizon [75] provide a forecast error taxonomy for open models which highlights the many additional forecasting errors that can arise from open systems. To keep the system closed we forecast the exogenous variables using Cardt (see Doornik et al. [76]), but as both are $I(0)$ regressors which only have temporary effects on the system they will not affect the long-horizon forecasts.

Figure 9 records projections for the system in levels up to 2100. As we have projections for *GHG* and *AF* from Miller et al. [37] up to 2100 (using the E2.1-G NINT f2 model with SSP2-4.5) we plot our forecasts against those projections. The IPCC embed current and future climate mitigation strategies showing greenhouse gas forcing is anticipated to dampen over the second half of the century. As our model is agnostic to policy interventions we do not forecast this slow down, that is, based on the current trajectory for radiative forcing we forecast global temperatures to increase by 3.4°C above the 1850–1900 average temperature, far exceeding the Paris accord. Ocean heat content is projected to increase to 85×10^{22}J above the 1971–2000 average, within the IPCC uncertainty bands [77] of $80 - 310 \times 10^{22}$ J. The sea level is projected to rise by 352 mm above the 1993–2008 average, and Arctic sea-ice extent is projected to fall to 5,400,000 m/km^2.

The $\pm 2\hat{\sigma}$ uncertainty bands in Figure 9 reflect the degree of uncertainty in the system. The narrowest bands show the error variance only. The slightly larger dashed bands reflect the additional parameter estimation uncertainty, all close to the error variance bands reflecting a low degree of uncertainty over the parameters. The widest bands are computed for the VECM with no saturation, denoted '+ future breaks'. Conditioning on past breaks suggests greater confidence in future forecasts as it assumes no breaks over the forecast horizon. If the same frequency and magnitude of breaks were to occur over the forecast horizon as during the in-sample period, then we would expect much larger uncertainty bands. The uncertainties increase dramatically if we look more than 50 years out, as expected.

The forecasts in Figure 9 are unconditional, but as the future path of GHGs is highly uncertain, dependent on human use of fossil fuels and agriculture, we look at projections conditional on possible future scenarios. Meinshausen et al. [78] provide annual observations for well-mixed GHG surface air mole fraction concentrations under 9 alternative SSP scenarios [5] which we convert to *GHG* radiative forcing, see Appendix A for details. The range of scenario *GHG* forcing paths under the various SSPs is given in Figure 10.

To implement the forecasts assuming a projected path for *GHG* forcing we change *GHG* and *AF* from endogenous to exogenous and estimate the four variable system conditioning on *GHG* and *AF*. The model is unchanged in-sample, but by setting the radiative forcing variables to exogenous we condition on their 'known' future values. This allows us to implement the scenario paths for *GHG* forcings and the IPCC central projections for *AF* (shown in Figure 9b). The cointegrating relations are identities so are

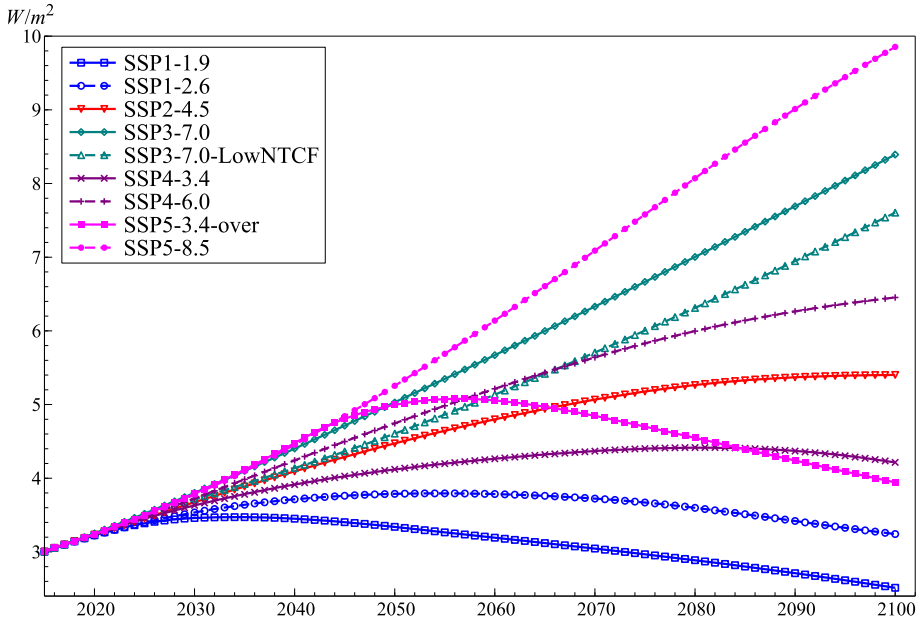

FIGURE 10 | Trajectories of radiative forcing induced by GHG emissions for SSPs (see Appendix A for conversion details).

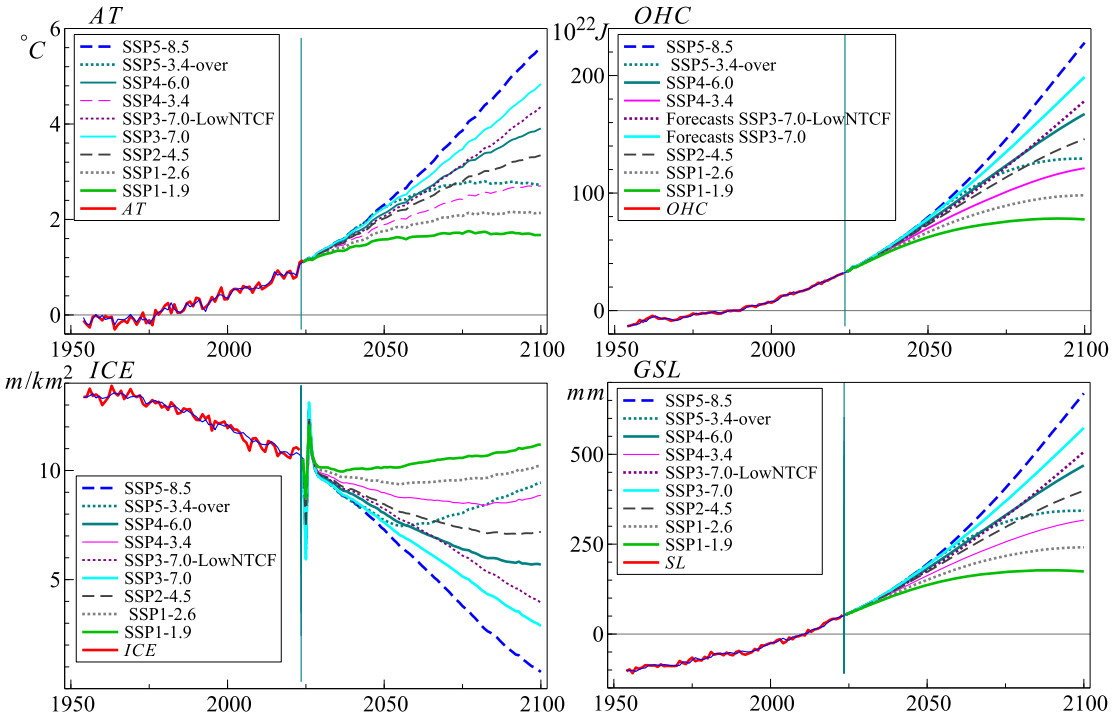

FIGURE 11 | Projections of climate variables from 2024 to 2100 for SSPs. Uncertainties in 2050 and 2100 detailed in Figure 12.

forecast within the system using the scenario projections for *GHG* and *AF* central projections.

Figure 11 records the range of forecasts for the climate variables based on the scenarios for *GHG* given by Figure 10.[8] The figure reflects an alternative measure of uncertainty to that in Figure 9, namely uncertainty over future mitigation policy. The range of projections is recorded in Figure 12 for comparison. The largest uncertainty comes from future breaks while the scenarios also demonstrate the wide degree of uncertainty depending on our ability to mitigate *GHG* emissions.

We can compare the scenario projections with those from the IPCC [79], notwithstanding some caveats. The IPCC [79] projections commence in 2015 so our forecasts condition on 8 more years of observed data over which the climate conditions have significantly worsened; the IPCC scenarios use different benchmarks for the anomalies so we adjust our forecasts for a mean shift; and the IPCC Arctic sea-ice projections are for September so we adjust our annual average forecasts for the difference between the annual average and September for a comparison, also see Diebold and Rudebusch [80]. Table 6 records the range from the IPCC along with our point estimates for comparison.

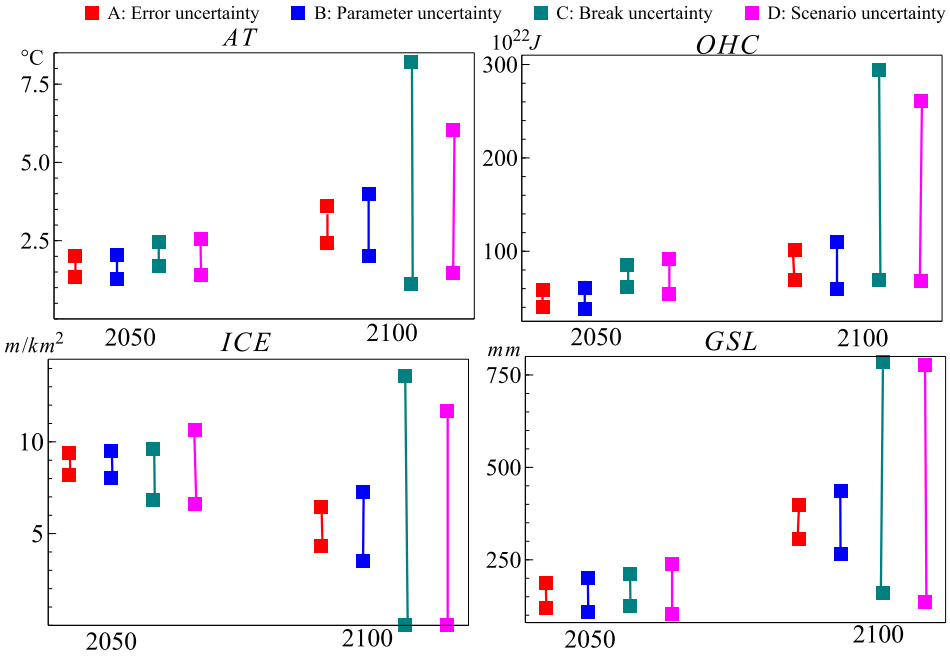

FIGURE 12 | Ninety-five percent uncertainty bands for system model Equation (6) with (A) just error variance; (B) error variance and parameter estimation uncertainty; (C) excluding all breaks (step and trend indicators from the cointegrating relations and impulse indicators from the equations); (D) range of forecasts from the SSP1-5 projections for *GHG*.

TABLE 6 | Comparison of long-run scenario projections.

	AT (°C)		*OHC* (10^{24} *J*)		*ICE* (10^6 km^2)		*GSL* (*m*)	
	IPCC	**(6)**	**IPCC**	**(6)**	**IPCC**	**(6)**	**IPCC**	**(6)**
2050								
SSP1-2.6	1.39–2.16	**2.12**	30.2–74.3	78.9	0.20–6.18	**3.72**	0.17–0.26	**0.17**
SSP2-4.5	1.61–2.40	**2.39**	39.3–71.0	85.3	0.03–5.84	**2.64**	0.18–0.27	**0.19**
SSP3-7.0	1.74–2.56	2.60	42.1–76.0	90.7	0.04–5.79	**1.76**	0.19–0.28	**0.21**
SSP5-8.5	1.90–2.87	**2.68**	45.2–82.5	91.8	0.03–4.82	**1.40**	0.22–0.31	0.21
2100								
SSP1-2.6	1.32–2.32	2.50	62.4–177.4	**110.4**	0.03–6.17	**4.48**	0.40–0.69	0.26
SSP2-4.5	2.14–3.57	3.71	92.6–204.8	**158.2**	0–4.78	**1.45**	0.50–0.81	0.41
SSP3-7.0	3.07–5.03	5.20	124.6–244.4	**211.1**	0–2.93	**0**	0.58–0.92	**0.59**
SSP5-8.5	3.56–6.22	**5.98**	144.0–306.0	**240.3**	0–2.11	**0**	0.69–1.05	**0.69**

Note: Comparison of IPCC [79] projections (90% ranges) and this study (point estimates using system model (6)) for climate variables in 2050 and 2100 under different SSPs. IPCC projections commence in 2015, (6) projections condition on information up to 2023. Bold indicates (6) point projection within IPCC bands. *AT*: Adjusted by −0.365°C for 1850–1900 benchmark. *OHC*: adjusted by 12.3×10^{22}J to correspond to Yu et al. [81] with bands in 10^{24} Joules relative to 1995–2014.
GSL: Adjustment of −0.016 m for IPCC 1995–2014 baseline; *ICE*: Adjustment of $5.73 \times 10^6 km^2$ to approximate September estimate from annual average for comparison.

Our projections for *ICE* all lie within the IPCC ranges, but our projections for *AT* are on the high-side relative to IPCC, likely due to the increased observed temperature over the 2015–2023 period relative to the IPCC scenario projections.

Our estimates for *OHC* are higher than the IPCC bands for 2050 but lie within the bands for 2100. Differences are likely due to transient response of the ocean to radiative forcing with depth. While our model is sensitive to changes up to 2000 m water depth (constrained by the data) but not below it has been shown that warming at this depth was (a) delayed by Holocene climate in the 20th century and (b) is projected to contribute later in the 21st century [82].

Our estimates for sea-level change are on the low side relative to the IPCC, likely due to the treatment of thermal expansion and land-ice as one variable despite their different temporal response to radiative forcing [83]. Thermal expansion and glacier contributions dominate most of the observed sea level record, with Antarctica and Greenland only recently increased their share [36]. Thus, our model mostly represents the thermal expansion and glacier components. This hypothesis is supported by comparing the sum of these projected components in Fox-Kemper et al. [77] giving a 2100 sea-level range of 15 – 570 mm with our 170 – 690 mm range across SSPs, which shows strong overlap.

7 | Conclusions

The two component EBM is statistically well-represented by a cointegrated vector equilibrium correction model. We extended the model of surface temperature and ocean heat content being driven by radiative forcing to include Arctic sea-ice extent and global mean sea-level change. Four cointegrating relations led to a stable system in which the stochastic trend in radiative forcing due to anthropogenic activity explained the long-run evolution of global temperature rise, sea-ice decline and sea-level rise. Indicator saturation estimation to model breaks in mean and trend was crucial to obtaining stable cointegrating relations, which allowed for long-run dynamic forecasts to be made.

The cointegrated VAR methodology produced more accurate ex post forecasts than the equivalent statistical models that ignored either the long-run (the VAR and DVAR) or system effects (univariate forecasting models). It also enabled the quantification of various measures of uncertainty within the system, including model and parameter estimation uncertainty, uncertainty as to future mitigation paths, and uncertainty over future structural breaks or shifts. The approach to measuring uncertainty is complementary to physical GCMs which focus on uncertainty over initial conditions.

A number of authors have shown the relevance of the cointegrated VAR approach to climate data. The CVAR framework provides sufficient structure to capture the slow adjustments to long-run physical relations as well as endogeneity between anthropogenic influences and the physical climate. Given the 40th anniversary of the Oxford Bulletin of Economics and Statistics Special Issue on Cointegration, it is an opportune moment to reflect on the benefits of the methodology outside of the economic sphere. The scope for extensions including spatial disaggregation to allow for polar amplification and ocean circulation and measures of salinity, glaciers, and permafrost, and so on are all left for future research, but we anticipate the cointegration framework will prove useful in all of these contexts.

Acknowledgements

The paper is for inclusion in the Oxford Bulletin of Economics and Statistics Special Issue Celebrating 40 Years of Unit Roots and Cointegration. We wish to thank the special issue editors for inviting the paper and participants of the conference celebrating 40 years of Unit Roots and Cointegration at University of Oxford, April 2025, for their invaluable feedback. Our thanks also to two anonymous referees, as well as Jonas Kurle and participants of the H.O. Stekler Research Program on Forecasting at George Washington University, Stockholm Business School, and the Korean Society of Economics and Trade for helpful comments and suggestions. We are pleased to acknowledge financial support from the Research Council of Norway, project 324,472, on 'Model invariance and constancy in the face of large shocks to the Norwegian macroeconomic system', and Nuffield College.

Funding

This work was supported by Nuffield College, University of Oxford and Norges Forskningsråd (324472).

Endnotes

[1] https://www.doornik.com/.

[2] See Table A2 in Appendix A for descriptive statistics of the data.

[3] Results for a model in which all radiative forcing is aggregated is available in the Supporting Information material. $I(0)$ and $I(1)$ regressors can be aggregated without changing the statistical properties of the data as the $I(1)$ trend dominates, allowing for a system driven by endogenous total effective radiative forcing.

[4] The notation and ordering used here differs from Pretis [4] to correspond to the variable labels used in this paper.

[5] Sub-sample 1 spans 1883–1962 (80 observations), sub-sample 2 spans 1963–1991 (108 observations over the two sub-samples) and the final sub-sample extends from 1992 to 2023. These break dates imply $a = 0.21$ and $b = 0.23$ in Kurita and Nielsen [51]. To calculate the quantiles we define $p = 8$ (endogenous and exogenous regressors), $m = 6$ (endogenous regressors), $k = 3$ (lag length) and $T = 141$ (number of observations).

[6] ADF t-statistics for a constant, with lag length chosen by AIC: $C_a = -5.11^{**}$; $C_b = -4.38^{**}$; $C_c = -2.65^{*}$; $C_d = -4.71^{**}$, where $^{*} = 10\%$ and $^{**} = 1\%$.

[7] The trend breaks are excluded to avoid collinearity with the anthropogenic forcings. The estimated cointegrating relations excluding restricted trends are:

$$\beta' = \begin{bmatrix} 1 & \underset{(0.125)}{0.483} & \underset{(0.000)}{-1.418} & 0 & 0 & 0 \\ 0 & \underset{(0.088)}{-0.340} & 1 & \underset{(0.002)}{-0.011} & 0 & 0 \\ \underset{(0.000)}{-0.325} & 0 & 0 & \underset{(0.101)}{0.505} & 1 & 0 \\ \underset{(11.07)}{12.55} & 0 & \underset{(1.59)}{-9.29} & \underset{(0.6004)}{-0.865} & 0 & 1 \end{bmatrix}.$$

[8] The blip in projections for *ICE* at the start of the forecast horizon is due to the forecast initialization and does not affect the forecast trajectory.

References

1. G. U. Yule, "Why Do We Sometimes Get Nonsense-Correlations Between Time-Series? A Study in Sampling and the Nature of Time Series (With Discussion)," *Journal of the Royal Statistical Society* 89 (1926): 1–64, https://www.jstor.org/stable/2341482.

2. B. B. Smith, "Combining the Advantages of First-Difference and Deviation-From-Trend Methods of Correlating Time Series," *Journal of the American Statistical Association* 21 (1926): 55–59, https://doi.org/10.1080/01621459.1926.10502158.

3. D. I. Stern and R. K. Kaufmann, "Detecting a Global Warming Signal in Hemispheric Temperature Series: A Structural Time Series Analysis," *Climatic Change* 47 (2000): 411–438, https://doi.org/10.1023/A:1005672231474.

4. F. Pretis, "Econometric Models of Climate Systems: The Equivalence of Two-Component Energy Balance Models and Cointegrated Vector Autoregressions," *Journal of Econometrics* 214 (2020): 256–273, https://doi.org/10.1016/j.jeconom.2019.05.013.

5. K. Riahi, D. P. van Vuuren, E. Kriegler, et al., "The Shared Socioeconomic Pathways and Their Energy, Land Use, and Greenhouse Gas Emissions Implications: An Overview," *Global Environmental Change* 42 (2017): 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009.

6. Z. R. J. Nicholls, M. Meinshausen, J. Lewis, et al., "Reduced Complexity Model Intercomparison Project Phase 1: Introduction and Evaluation of Global-Mean Temperature Response," *Geoscientific Model Development* 13, no. 11 (2020): 5175–5190, https://doi.org/10.1029/2020EF001900.

7. C. Tebaldi, N. Selin, R. Ferrari, and G. Flierl, "Emulators of Climate Model Output," *Annual Review of Environment and Resources* 50 (2025): 709–737, https://doi.org/10.1146/annurev-environ-012125-085838.

8. M. Stute, A. Clement, and G. Lohmann, "Global Climate Models: Past, Present, and Future," *PNAS* 98 (2001): 10529–10530, https://doi.org/10.1073/pnas.191366098.

9. S. L. Weber, "The Utility of Earth System Models of Intermediate Complexity (Emics)," *WIREs Climate Change* 1 (2010): 243–252, https://doi.org/10.1002/wcc.24.

10. R. K. Kaufmann and D. I. Stern, "Cointegration Analysis of Hemispheric Temperature Relations," *Journal of Geophysical Research* 107, no. 2 (2002): ACL 8-1–ACL 8-10, https://doi.org/10.1029/2000JD000174.

11. D. I. Harvey and T. C. Mills, "Modelling Global Temperature Trends Using Cointegration and Smooth Transitions," *Statistical Modelling* 1 (2001): 143–159, https://doi.org/10.1177/1471082X0100100204.

12. H. Liu and G. Rodríguez, "Human Activities and Global Warming: A Cointegration Analysis," *Environmental Modelling & Software* 20 (2005): 761–773, https://doi.org/10.1016/j.envsoft.2004.03.017.

13. R. K. Kaufmann, H. Kauppi, and J. H. Stock, "Emissions, Concentrations, and Temperature: A Time Series Analysis," *Climatic Change* 77 (2006): 249–278, https://doi.org/10.1007/s10584-006-9062-1.

14. K. Juselius, "Cointegration Analysis of Climate Change: An Exposition," Working Paper, 2007, https://api.semanticscholar.org/CorpusID:18027870.

15. C. Gay-Garcia, F. Estrada, and A. Sánchez, "Global and Hemispheric Temperatures Revisited," *Climatic Change* 94 (2009): 333–349, https://doi.org/10.1007/s10584-008-9524-8.

16. F. Estrada, C. Gay, and A. Sãnchez, "A Reply to: 'Does Temperature Contain a Stochastic Trend? Evaluating Conflicting Statistical Results' by RK Kaufmann et al.," *Climatic Change* 101 (2010): 407–414, https://doi.org/10.1007/s10584-010-9928-0.

17. F. Estrada, P. Perron, and B. Martínez-López, "Statistically Derived Contributions of Diverse Human Influences to Twentieth-Century Temperature Changes," *Nature Geoscience* 6 (2013): 1050–1055, https://doi.org/10.1038/ngeo1999.

18. M. D. Gadea-Rivas, J. Gonzalo, and A. Ramos, "Trends in Temperature Data: Micro-Foundations of Their Nature," *Economics Letters* 244 (2024): 111992, https://doi.org/10.1016/j.econlet.2024.111992.

19. H. D. Matthews, N. P. Gillett, P. A. Stott, and K. Zickfeld, "The Proportionality of Global Warming to Cumulative Carbon Emissions," *Nature* 459, no. 7248 (2009): 829–832, https://doi.org/10.1038/nature08047.

20. R. K. Kaufmann, H. Kauppi, and J. H. Stock, "Does Temperature Contain a Stochastic Trend? Evaluating Conflicting Statistical Results," *Climatic Change* 101 (2010): 395–405, https://doi.org/10.1007/s10584-009-9711-2.

21. R. K. Kaufmann, H. Kauppi, M. L. Mann, and J. H. Stock, "Does Temperature Contain a Stochastic Trend: Linking Statistical Results to Physical Mechanisms," *Climatic Change* 118 (2013): 729–743, https://doi.org/10.1007/s10584-012-0683-2.

22. J. A. Doornik and H. Hansen, "An Omnibus Test for Univariate and Multivariate Normality," *Oxford Bulletin of Economics and Statistics* 70 (2008): 927–939, https://doi.org/10.1111/j.1468-0084.2008.00537.x.

23. E. Zivot and D. W. K. Andrews, "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis," *Journal of Business & Economic Statistics* 10 (1992): 251–270, https://doi.org/10.2307/1391541.

24. J. L. Castle, J. A. Doornik, D. F. Hendry, and F. Pretis, "Trend-Indicator Saturation," 2019 Working paper, Nuffield College, Oxford University.

25. C. W. J. Granger and T. Lee, "Multicointegration," in *Advances in Econometrics, Volume 8*, ed. G. F. Rhodes and T. B. Fomby (JAI Press, 1990).

26. S. B. Bruns, Z. Csereklyei, and D. I. Stern, "A Multicointegration Model of Global Climate Change," *Journal of Econometrics* 214, no. 1 (2020): 175–197, https://doi.org/10.1016/j.jeconom.2019.05.010.

27. L. Benati, "Forecasting Global Temperatures by Exploiting Cointegration With Radiative Forcing," Discussion Paper 23-08, University of Bern, Department of Economics, Bern, 2023, https://www.econstor.eu/handle/10419/278545.

28. M. Beenstock, Y. Reingewertz, and M. Paldor, "Polynomial Cointegration Tests of Anthropogenic Impact on Global Warming," *Earth System Dynamics* 3 (2012): 173–188, https://doi.org/10.5194/esd-3-173-2012.

29. F. Pretis and D. F. Hendry, "Comment on "Polynomial Cointegration Tests of Anthropogenic Impact on Global Warming" by Beenstock Et al. (2012)—Some Hazards in Econometric Modelling of Climate Change," *Earth System Dynamics* 4, no. 2 (2013): 375–384.

30. S. C. Sherwood, M. J. Webb, J. D. Annan, et al., "An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence," *Reviews of Geophysics* 58, no. 4 (2020): e2019RG000678, https://doi.org/10.1029/2019RG000678.

31. T. C. Mills, "How Robust Is the Long-Run Relationship Between Temperature and Radiative Forcing?," *Climatic Change* 94 (2009): 351–361, https://doi.org/10.1007/s10584-008-9525-7.

32. W. A. Brock and J. I. Miller, "Polar Amplification in a Moist Energy Balance Model: A Structural Econometric Approach to Estimation and Testing," SSRN Working Paper, SSRN, 2023, https://doi.org/10.2139/ssrn.4450492.

33. T. S. Schmith, S. J. Johansen, and P. Thejll, "Statistical Analysis of Global Surface Temperature and Sea Level Using Cointegration Methods," *Journal of Climate* 25 (2012): 7822–7833, https://doi.org/10.1175/JCLI-D-11-00598.1.

34. L. P. Jackson, K. Juselius, A. B. Martinez, and F. Pretis, "Modelling the Dependence Between Recent Changes in Polar Ice Sheets: Implications for Global Sea-Level Projections," *Advances in Econometrics: Econometrics of Climate, Energy and Green Transmission* 47B (2026): Chapter 3 (forthcoming), https://ssrn.com/abstract=3912725.

35. L. Zanna, S. Khatiwala, J. M. Gregory, J. Ison, and P. Heimbach, "Global Reconstruction of Historical Ocean Heat Storage and Transport," *Proceedings of the National Academy of Sciences of the United States of America* 116 (2019): 1126–1131, https://doi.org/10.1073/pnas.1808838115.

36. T. Frederikse, F. Landerer, L. Caron, et al., "The Causes of Sea-Level Rise Since 1900," *Nature* 584 (2020): 393–397, https://doi.org/10.1038/s41586-020-2591-3.

37. R. L. Miller, G. A. Schmidt, L. Nazarenko, et al., "CMIP6 Historical Simulations (1850–2014) With GISS-E2.1," *Journal of Advances in Modeling Earth Systems* 13, no. 1 (2021): e2019MS002034, https://doi.org/10.1029/2019MS002034.

38. R. J. Hodrick and E. C. Prescott, "Postwar US Business Cycles: An Empirical Investigation," *Journal of Money, Credit and Banking* 29, no. 1 (1997): 1–16. First published in 1980 as Carnegie Mellon University Discussion Paper No. 451.

39. J. E. Walsh, W. L. Chapman, F. Fetterer, and S. Stewart, *Gridded Monthly Sea Ice Extent and Concentration, 1850 Onward, Version 2* (NSIDC: National Snow and Ice Data Center, 2019), https://doi.org/10.7265/jj4s-tq79.

40. J. A. Church and N. J. White, "Sea-Level Rise From the Late 19th to the Early 21st Century," *Surveys in Geophysics* 32 (2011): 585–602, https://doi.org/10.1007/s10712-011-9119-1.

41. J. Hurrell and A. Phillips, "The Climate Data Guide Hurrell North Atlantic Oscillation (Nao) Index (Station-Based)," Technical report, National Center for Atmospheric Research Staff, 2024, https://climatedataguide.ucar.edu/climate-data/hurrell-north-atlantic-oscillation-nao-index-station-based.

42. Y. Kosaka and S.-P. Xie, "Recent Global-Warming Hiatus Tied to Equatorial Pacific Surface Cooling," *Nature* 501 (2013): 403–407, https://doi.org/10.1038/nature12534.

43. F. Pretis, M. L. Mann, and R. K. Kaufmann, "Testing Competing Models of the Temperature Hiatus: Assessing the Effects of Conditioning Variables and Temporal Uncertainties Through Sample-Wide Break Detection," *Climatic Change* 131, no. 4 (2015): 705–718, https://doi.org/10.1007/s10584-015-1391-5.

44. F. Pretis, L. Schneider, J. E. Smerdon, and D. F. Hendry, "Detecting Volcanic Eruptions in Temperature Reconstructions by Designed Break-Indicator Saturation," *Journal of Economic Surveys* 30 (2016): 403–429, https://doi.org/10.1111/joes.12148.

45. J. L. Castle, J. A. Doornik, D. F. Hendry, and F. Pretis, "Detecting Location Shifts During Model Selection by Step-Indicator Saturation," *Econometrics* 3, no. 2 (2015): 240–264, https://doi.org/10.3390/econometrics3020240.

46. H. B. Nielsen and M. Qian, "Asymptotic Properties of the Gauge and Power of Step-Indicator Saturation," *Econometric Theory* (2025): 1–38, https://doi.org/10.1017/S0266466625100145.

47. D. F. Hendry and J. A. Doornik, *Empirical Model Discovery and Theory Evaluation* (MIT Press, 2014), https://doi.org/10.7551/mitpress/9780262028356.001.0001.

48. S. Johansen, "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics & Control* 12 (1988): 231–254, https://doi.org/10.1016/0165-1889(88)90041-3.

49. S. Johansen, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (Oxford University Press, 1995).

50. S. Johansen, R. Mosconi, and B. Nielsen, "Cointegration Analysis in the Presence of Structural Breaks in the Deterministic Trend," *Econometrics Journal* 3 (2000): 216–249, https://doi.org/10.1111/1368-423X.00047.

51. T. Kurita and B. Nielsen, "Partial Cointegrated Vector Autoregressive Models With Structural Breaks in Deterministic Terms," *Econometrics* 7 (2019): 42, https://doi.org/10.3390/econometrics7040042.

52. M. A. Ivanov and S. N. Evtimov, "1963: The Break Point of the Northern Hemisphere Temperature Trend During the Twentieth Century," *International Journal of Climatology* 30, no. 11 (2010): 1738–1746, https://doi.org/10.1002/joc.2002.

53. D. F. Hendry, S. Johansen, and C. Santos, "Automatic Selection of Indicators in a Fully Saturated Regression," *Computational Statistics* 23 (2008): 317–335, https://doi.org/10.1007/s00180-007-0054-z.

54. L. G. Godfrey, "Testing for Higher Order Serial Correlation in Regression Equations When the Regressors Include Lagged Dependent Variables," *Econometrica* 46 (1978): 1303–1313, https://doi.org/10.2307/1913830.

55. B. Nielsen, "Order Determination in General Vector Autoregressions," in *Time Series and Related Topics: In Memory of Ching-Zong Wei*, ed. H.-C. Ho, C.-K. Ing, and T. Lai (Institute of Mathematical Statistics, 2006), 93–112, https://users.ox.ac.uk/~nuff0078/Discuss/LNMS5206.pdf.

56. L. Kilian and U. Demiroglu, "Residual-Based Tests for Normality in Autoregressions: Asymptotic Theory and Simulation Evidence," *Journal of Business & Economic Statistics* 18 (2000): 40–50, https://doi.org/10.1080/07350015.2000.10524846.

57. V. Berenguer-Rico and B. Nielsen, "Normality Testing After Outlier Removal," *Econometrics and Statistics* (forthcoming), https://doi.org/10.1016/j.ecosta.2023.06.001.

58. H. White, "A Heteroskedastic-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48 (1980): 817–838, https://doi.org/10.2307/1912934.

59. V. Berenguer-Rico and I. Wilms, "Heteroscedasticity Testing After Outlier Removal," *Econometric Reviews* 40 (2020): 51–85, https://doi.org/10.1080/07474938.2020.1735749.

60. J. B. Ramsey, "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 31, no. 2 (1969): 350–371.

61. J. L. Castle and D. F. Hendry, "A Low-Dimension Portmanteau Test for Non-Linearity," *Journal of Econometrics* 158 (2010): 231–245, https://doi.org/10.1016/j.jeconom.2010.01.006.

62. G. C. Chow, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica* 28 (1960): 591–605, https://doi.org/10.2307/1910133.

63. B. Nielsen and A. Whitby, "A Joint Chow Test for Structural Instability," *Econometrics* 3, no. 1 (2015): 156–186.

64. R. L. Brown, J. Durbin, and J. M. Evans, "Techniques for Testing the Constancy of Regression Relationships Over Time (With Discussion)," *Journal of the Royal Statistical Society B* 37 (1975): 149–192, https://doi.org/10.1111/j.2517-6161.1975.tb01532.x.

65. O. Boucher, J. Haigh, D. Hauglustaine, et al., "Radiative Forcing of Climate Change," *IPCC TAR Climate Change 2001: The Scientific Basis* 6 (2001): 351–416, https://www.ipcc.ch/site/assets/uploads/2018/03/TAR-06.pdf.

66. J. A. Doornik and D. F. Hendry, *Modelling Dynamic Systems Using PcGive: Volume II*, 9th ed. (Timberlake Consultants Press, 2022).

67. E. Engler and B. Nielsen, "The Empirical Process of Autoregressive Residuals," *Econometrics Journal* 12 (2009): 367–381, https://doi.org/10.1111/j.1368-423X.2009.00282.x.

68. R. F. Engle, "Autoregressive Conditional Heteroscedasticity, With Estimates of the Variance of United Kingdom Inflation," *Econometrica* 50 (1982): 987–1007, https://doi.org/10.2307/1912773.

69. D. F. Hendry and J.-F. Richard, "On the Formulation of Empirical Models in Dynamic Econometrics," *Journal of Econometrics* 20 (1982): 3–33, https://doi.org/10.1016/0304-4076(82)90101-4.

70. Y. Y. Chong and D. F. Hendry, "Econometric Evaluation of Linear Macro-Economic Models," *Review of Economic Studies* 53 (1986): 671–690, https://doi.org/10.2307/2297611.

71. J. L. Castle and D. F. Hendry, "On Not Evaluating Economic Models by Forecast Outcomes," *Istanbul Business Research* 40, no. 1 (2011): 1–14, https://dergipark.org.tr/en/pub/iuisletme/issue/9251/115734.

72. D. Harvey, S. Leybourne, and P. Newbold, "Testing the Equality of Prediction Mean Squared Errors," *International Journal of Forecasting* 13, no. 2 (1997): 281–291.

73. F. X. Diebold and R. S. Mariano, "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics* 13 (1995): 253–263, https://doi.org/10.1198/073500102753410444.

74. M. P. Clements and D. F. Hendry, "Forecasting in Cointegrated Systems," *Journal of Applied Econometrics* 10 (1995): 127–146, https://doi.org/10.1002/jae.3950100204.

75. D. F. Hendry and G. E. Mizon, "Open-Model Forecast-Error Taxonomies," in *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, ed. X. Chen and N. R. Swanson (Springer, 2012), 219–240, https://doi.org/10.1007/978-1-4614-1653-1_9.

76. J. A. Doornik, J. L. Castle, and D. F. Hendry, "Short-Term Forecasting of the Coronavirus Pandemic," *International Journal of Forecasting* 38 (2022): 453–466, https://doi.org/10.1016/j.ijforecast.2020.09.003.

77. B. Fox-Kemper, H. T. Hewitt, C. Xiao, et al., "Ocean, Cryosphere and Sea Level Change," in *Climate Change 2021: The Physical Science Basis, Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, ed. IPCC (Cambridge University Press, 2021), 1211–1362, https://doi.org/10.1017/9781009157896.011.

78. M. Meinshausen, Z. R. J. Nicholls, J. Lewis, et al., "The Shared Socio-Economic Pathway (SSP) Greenhouse Gas Concentrations and

Their Extensions to 2500," *Geoscientific Model Development* 13, no. 8 (2020): 3571–3605.

79. IPCC, "IPCC, 2021: Summary for Policymakers," in *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, ed. V. Masson-Delmotte, P. Zhai, A. Pirani, et al. (Cambridge University Press, 2021), 3–32, https://doi.org/10.1017/9781009157896.001.

80. F. X. Diebold and G. D. Rudebusch, "Probability Assessments of an Ice-Free Arctic: Comparing Statistical and Climate Model Projections," *Journal of Econometrics* 231 (2022): 520–534, https://www.sciencedirect.com/science/article/pii/S0304407620304012.

81. Y. Yu, L. Hua, A. Mix, and B. Pearson, "Chapter 9 of the working group I Contribution to the IPCC Sixth Assessment Report—Data for figure 9.6 (v20220721). 15 may 2023, NERC EDS Centre for Environmental Data Analysis," 2023, https://doi.org/10.5285/439ccb0b0eb04c17b5c6897fb9cb550b.

82. L. Cheng, K. von Schuckmann, J. P. Abraham, et al., "Past and Future Ocean Warming," *Nature Reviews Earth & Environment* 3, no. 11 (2022): 776–794, https://doi.org/10.1038/s43017-022-00345-1.

83. A. Grinsted and J. H. Christensen, "The Transient Sensitivity of Sea Level Rise," *Ocean Science* 17, no. 1 (2021): 181–186, https://doi.org/10.5194/os-17-181-2021.

84. B. C. O'Neill, E. Kriegler, K. L. Ebi, et al., "The Roads Ahead: Narratives for Shared Socioeconomic Pathways Describing World Futures in the 21st Century," *Global Environmental Change* 42 (2017): 169–180, https://doi.org/10.1016/j.gloenvcha.2015.01.004.

Appendix A

TABLE A1 | Data sources.

Label	Units	Description	Source
AT	˚C	Global mean surface temperature anomaly[a]	OWID
T_{NH}	˚C	Northern hemisphere mean surface temperature anomaly[a]	OWID
T_{SH}	˚C	Southern hemisphere mean surface temperature anomaly[a]	OWID
SST	˚C	Global sea surface temperature anomaly[b]	OWID
SST_{NH}	˚C	Northern hemisphere sea surface temperature anomaly[b]	OWID
SST_{SH}	˚C	Southern hemisphere sea surface temperature anomaly[b]	OWID
OHC	10^{22} J	Ocean heat content in top 2000 m[c]	OWID & Zanna et al. [35]
OH_{700}	10^{22} J	Ocean heat content in top 700 m[c]	OWID
ERF	W/m^2	Total effective radiative forcing[d]	NASA
ERF_{GHG}	W/m^2	Effective radiative forcing for GHGs[d]	NASA
$ERF_{aerosols}$	W/m^2	Effective radiative forcing for aerosols[d]	NASA
$ERF_{natural}$	W/m^2	Natural effective radiative forcing[d]	NASA
ICE	m/km^2	Arctic sea-ice extent[e]	NSIDC & Copernicus
GSL	mm	Global mean sea-level change[f]	Frederikse et al. [36] with OWID and NOAA
$ENSO$	˚C	El Niño Southern Oscillation index[g]	NOAA JIASO
NAO	hPa	North Atlantic Oscillation index[i]	NCAR

[a] Average land-sea temperature anomaly relative to the 1961–1990 average temperature. Data available from Met Office Hadley Centre: CRUTEM5 surface air temperature dataset.

[b] Sea surface temperature anomaly relative to the 1961–1990 average temperature. Data available from Met Office Hadley Centre: HadSST4 sea-surface temperature dataset.

[c] Ocean heat content is measured relative to the 1971–2000 average, which is set at zero for reference. Original source: NOAA National Centers for Environmental Information. Three alternative measures include Institute of Atmospheric Physics (IAP); National Oceanic and Atmospheric Administration (NOAA); and Meteorological Research Institute (MRI). Annual data available from 1955 (700 m) and 1957 (2000 m). To extend the series to 1860 we use the latent measure of OHC derived from Zanna et al. [35], average the three observational series (with an adjustment for the observations for 2021, 2022 and 2023 as the MRI and IAP measures end in 2020) and splice the two together in 1957 so the latter part of the sample is observational.

[d] Effective radiative forcing at the top-of-the-atmosphere in the E2.1-G (f2) model, from CMIP6 Simulations, see Miller et al. [37]. Time-series are provided for all forcings together, just natural forcings, and greenhouse gas and aerosols separately, encompassing the historical times series (1850–2014) and SSP2-4.5 (2015–2100). Smoothed data obtained from Hodrick Prescott filter [38] with bandwidth 10.

[e] Sea-ice extent records total area of the oceans with sea-ice concentrations of 15% or more. Monthly gridded data from 1850 to 2017 from Walsh et al. [39] for northern hemisphere taken on 15th of each month. From 2018 onwards we use monthly data measuring sea-ice extent anomalies relative to 1991–2020 in the Arctic from Copernicus. We average the data over 12 months and infer the level from the anomaly data using NSIDC data over 1991–2020.

[f] Global mean sea-level change is measured relative to the 2002–2018 average sea level from Frederikse et al. [36] for the period 1900–2018 and corrected for groundwater depletion, water impoundment in artificial reservoirs and natural terrestrial water storage. The data are extrapolated backwards from 1900 to 1880 using the growth rate of sea-level change from Church and White [40]. Observations for 2019–2023 are updated from NOAA and spliced.

[g] ENSO data are spliced data from two sources. Data from 1950 to 2023 are from the NOAA National Weather Service Climate Prediction Center recording a 3 month running mean of ERSST.v5 SST anomalies in the Niño 3.4 region (5˚N–5˚S, 120˚–170˚W), based on centered 30-year base periods updated every 5 years. Annual data are averages across the 12 3-month averages for each year. Data for 1860–2010 are available from JIASO as monthly data on the average sea surface temperature anomaly with respect to the period 1950–1979, equatorward of 20˚ latitude (north and south) minus the average sea surface temperature poleward of 20˚. The different spatial areas of the two measures mean the standard deviations of the two series differ. The series are spliced using estimated standard deviations over the samples available: $ENSO = ENSOs \times (0.618/0.175)$.

[i] Hurrell Station-Based Annual NAO Index from Hurrell and Phillips [41]. Also see NOAA.

Descriptive Statistics for the Climate Data

TABLE A2 | Descriptive statistics.

	$\widehat{\mu}$	$\widehat{\sigma}$	$\widehat{\kappa}_1$	$\widehat{\kappa}_2$	Norm	ADF	ZA	TIS
ERF	0.66	0.75	1.28	0.98	88.5**	−3.11	−3.19	1963
GHG	1.18	0.95	0.70	−0.88	59.2**	−0.63	−5.32**	1966
AF	−0.70	0.42	−0.05	−1.60	39.7**	−2.96	−2.82	1964;1997
NF	−0.04	0.37	−2.65	8.91	242**	−6.57**	−6.35**	None
AT	−0.02	0.39	0.76	−0.21	32.7**	−2.52	−3.88	1940;1964
OHC	−11.84	19.0	0.40	−0.74	13.7**	−2.46	−4.38*	1910;1995
ICE	12.88	0.85	−1.31	0.55	137**	−1.41	−3.75	1967;1995;2019
GSL	−108.1	73.2	0.31	−0.95	15.0**	−2.01	−3.28	1965;2012

Note: Statistics reported for $T = 1880 - 2023$. GHG and AF statistics are for the smoothed data used in the empirical modelling. $\widehat{\mu}$ = sample mean; $\widehat{\sigma}$ = sample standard deviation; $\widehat{\kappa}_1$ = sample skewness; $\widehat{\kappa}_2$ = sample excess kurtosis. Norm = $\chi^2(2)$ test of normality (see Doornik and Hansen [22]). ADF is Augmented Dickey Fuller test with constant and trend for all variables apart from NF which was conducted with just constant. Lag length chosen by AIC. ZA is Zivot and Andrews [23] test for stationarity with an unknown structural break, where we allow for a break in trend. Critical values are −4.93 (1%); −4.42 (5%); and −4.11 (10%). TIS denotes the break dates selected by trend indicator saturation (Castle et al. [24]) with a forced constant and linear trend applying selection at $\alpha = 0.0001$. For all tests, ** and * denotes significance at 1% and 5%, respectively.

Calculation of *GHG* Forcings Under Alternative SSPs

O'Neill et al. [84] give five SSPs that embody different narratives on population, GDP and urbanisation trajectories, as well as assumptions on energy and land use sectors, that are used by the IPCC. Meinshausen et al. [78] provide annual observations for well-mixed GHG surface air mole fraction concentrations under 9 alternative SSP scenarios. To use these scenarios in our model we convert the concentrations to radiative forcing using Boucher et al. [65] (Table 6.2) taking the five largest components including CO_2, CH_4, N_2O, CFC-11 and CFC-12. First, defining:

$$C = CO_2 \text{ in ppm}$$
$$M = CH_4 \text{ in ppb}$$
$$N = N_2O \text{ in ppb}$$
$$X = CFC \text{ in ppb}$$

the conversions are given by:

$$\Delta F_{CO_2} = \alpha_1\big(g(C) - g\big(C_0\big)\big); \quad g(C) = \ln\big(1 + 1.2C + 0.005C^2 + 1.4 \times 10^{-6}C^3\big)$$
$$\Delta F_{CH_4} = \alpha_2\Big(\sqrt{M} - \sqrt{M_0}\Big) - \big(f\big(M, N_0\big) - f(M_0, N_0)\big)$$
$$\Delta F_{N_2O} = \alpha_3\Big(\sqrt{N} - \sqrt{N_0}\Big) - \big(f\big(M_0, N\big) - f(M_0, N_0)\big)$$
$$\Delta F_{CFC-11} = \alpha_4\big(X - X_0\big)$$
$$\Delta F_{CFC-12} = \alpha_5\big(X - X_0\big)$$

where $f(M, N) = 0.47\ln[1 + 2.01 \times 10^{-5}(MN)^{0.75} + 5.31 \times 10^{-15}M(MN)^{1.52}]$ and $\alpha_1 = 3.35$, $\alpha_2 = 0.036$; $\alpha_3 = 0.12$; $\alpha_4 = 0.25$; $\alpha_5 = 0.32$. Subscript 0 is taken to be the concentration level in year 1, also reported in Meinshausen et al. [78]. The radiative forcings are then summed to give *GHG* under the alternative concentration paths based on SSPs, recorded in Figure 10.