

## A machine learning approach to rapidly project climate responses under a multitude of net-zero emission pathways

Vassili Kitsios <sup>1,2✉</sup>, Terence John O’Kane<sup>3</sup> & David Newth<sup>4</sup>

Navigating a path toward net-zero, requires the assessment of physical climate risks for a broad range of future economic scenarios, and their associated carbon concentration pathways. Climate models typically simulate a limited number of possible pathways, providing a small fraction of the data needed to quantify the physical risk. Here machine learning techniques are employed to rapidly and cheaply generate output mimicking these climate simulations. We refer to this approach as QuickClim, and use it here to reconstruct plausible climates for a multitude of concentration pathways. Higher mean temperatures are confirmed to coincide with higher end-of-century carbon concentrations. The climate variability uncertainty saturates earlier, in the mid-century, during the transition between current and future climates. For pathways converging to the same end-of-century concentration, the climate is sensitive to the choice of trajectory. In net-zero emission type pathways, this sensitivity is of comparable magnitude to the projected changes over the century.

<sup>1</sup>Environment, Commonwealth Scientific and Industrial Research Organisation, Aspendale, VIC, Australia. <sup>2</sup>Laboratory for Turbulence Research in Aerospace and Combustion, Department of Mechanical and Aerospace Engineering, Monash University, Melbourne, VIC, Australia. <sup>3</sup>Environment, Commonwealth Scientific and Industrial Research Organisation, Hobart, Tasmania, Australia. <sup>4</sup>Environment, Commonwealth Scientific and Industrial Research Organisation, Canberra, Australian Capital Territory, Australia. ✉email: [vassili.kitsios@csiro.au](mailto:vassili.kitsios@csiro.au)

The global community is currently facing the challenge of how to meet the needs and desires of humanity<sup>1,2</sup>, while not exceeding the planetary boundaries that the Earth can safely sustain<sup>3</sup>. Anthropogenic carbon emissions have been attributed to global warming<sup>4</sup> and modifications to climate variability<sup>5,6</sup>. Such changes to the climate influence the human experience in numerous ways, including: food production<sup>7</sup>, affordability<sup>8–10</sup> and trade<sup>11</sup>; water security<sup>12</sup>; bushfire risk<sup>13</sup>; spread of zoonotic diseases<sup>14</sup>; human health and mortality<sup>15–17</sup>; labour productivity<sup>18</sup>; and economic development<sup>19</sup>. Navigating the transition to a low carbon net-zero emission economy is key to mitigating these and other climate related impacts.

Efforts are currently being undertaken to assess the transitional and physical climate risks as a result of the uncertain path toward a net-zero emissions world<sup>20,21</sup>. Transition risks relate to how economic development may be influenced by changes in future policy, technology, and/or consumer preferences brought on by the constraining emissions pathway. These risks are often assessed via integrated assessment models<sup>22,23</sup> to determine the economic settings compatible with prescribed target emissions trajectories. A measure of the transition risk is the spread of various possible economic measures associated with a range of plausible emissions trajectories.

Physical climate risks are the damages imparted upon society as a result of changing climatic conditions (chronic) and extreme weather events (acute)<sup>24</sup>. To quantify the relevant future environmental hazards, one requires spatially resolved estimates of not only the mean global warming response<sup>25</sup>, but also the probabilities of a variety of natural phenomena<sup>26</sup>. Existing Coupled Model Inter-comparison Project (CMIP) climate projections<sup>27,28</sup> provide data for only a small subset of the universe of possible future emissions trajectories. These general circulation model simulations of the global climate are computationally expensive and complex<sup>29</sup>. Consequently, it is not feasible to run such simulations for each and every bespoke concentration pathway required by the various stakeholders in the global economy.

To undertake consistent transitional and physical risk assessments across a broad range of economic scenarios, a computationally fast means of determining how climate variability and change might respond to arbitrary concentration pathways is required. Existing approaches to estimating globally averaged temperatures, involve the simulation of prescribed emissions or concentration trajectories by reduced complexity models<sup>30</sup>. The globally averaged temperature can then be used as an input to estimate static spatially dependent maps of climate change related modifications to temperature (or rainfall), using pattern scaling techniques<sup>31</sup>. These methods provide no time varying information, nor any information pertaining to the variability nor extremity of these future climates. Systems that provide time varying gridded fields for a prescribed globally averaged temperature, do so only for monthly averaged surface temperature<sup>32</sup>.

Here we apply purpose-built model reduction approaches in conjunction with machine learning (ML) techniques to produce time varying gridded climate fields in response to prescribed concentration pathways. Our approach captures both climate variability and change. It is also multi-variate and can simultaneously reconstruct multiple user-defined climate variables in a dynamically consistent manner. We refer to this capability as QuickClim. QuickClim utilises existing climate model data and the associated atmospheric carbon dioxide equivalent ( $CO_{2e}$ ) concentrations, to learn the relationships between the two. The adopted class of ML methods<sup>33,34</sup> captures the dependence of the data at a given time, on the data at previous times, and on the applied external forcing (i.e.,  $CO_{2e}$ ). The ML model coefficients can also be regime dependent. This approach has previously been successfully applied to climate attribution problems<sup>5,35–37</sup>, and

finance applications<sup>38–40</sup>. The power and novelty of QuickClim, is that it can estimate the climate variability and change for any arbitrary concentration pathway, and for any selection of climate variables. At present there is no other method in the literature capable of doing so.

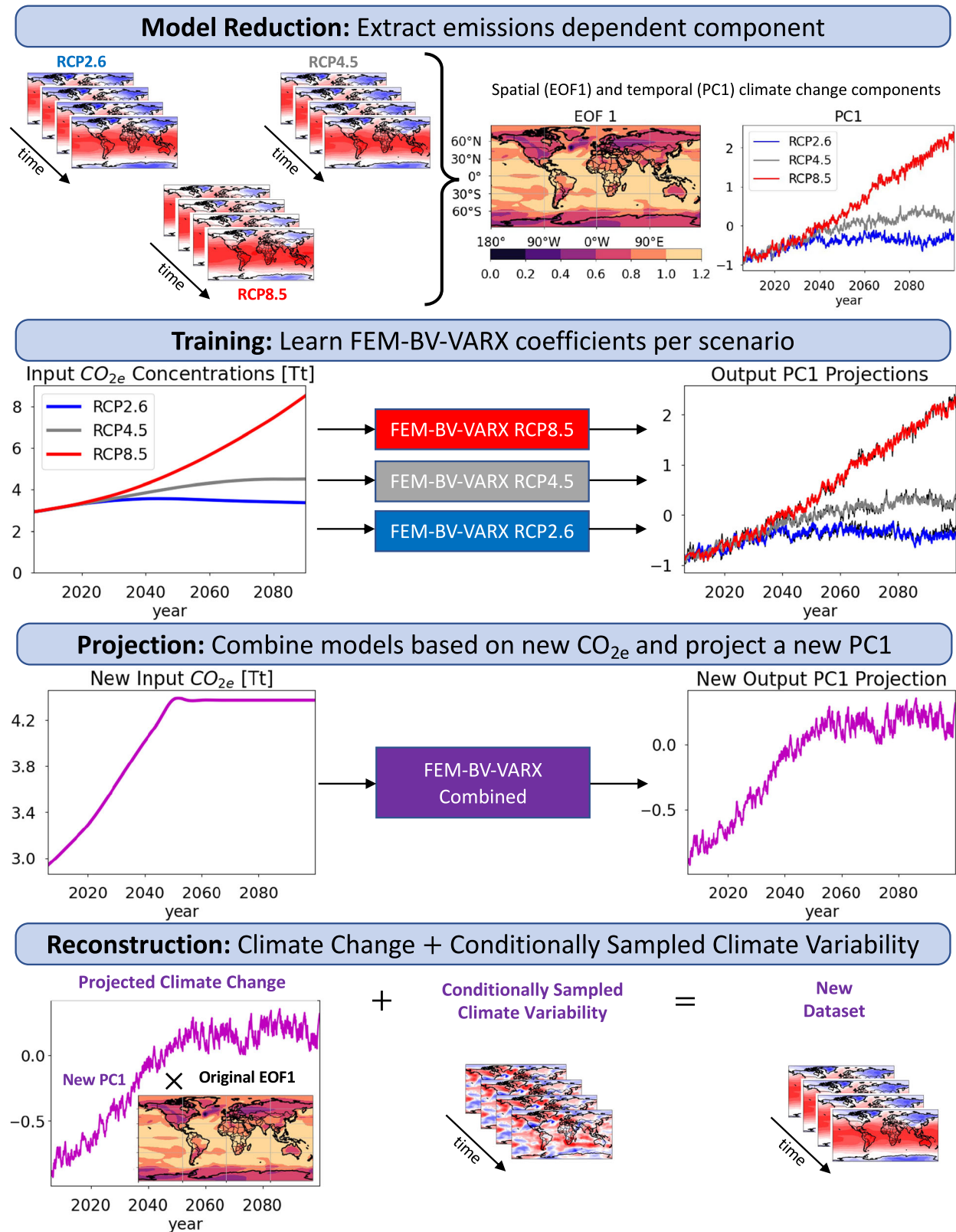
The majority of research on the climatic response to net-zero emissions pathways, has addressed the mean change to a small number of specific trajectories. Here, via multiple case studies, we use QuickClim to examine how a multitude of possible concentration pathways influence the climate. We first determine the influence of the average decarbonisation rate. Then for a fixed average decarbonisation rate and fixed end-of-century concentration, we infer the influence of the specific trajectory. Finally, we assess a family of concentration pathways approaching that of RCP2.6, to be more in line with a net-zero emissions type scenario. To understand these sensitivities, QuickClim readily facilitates the rapid reconstruction of thousands of climate responses approximating the behaviour of the available CMIP models to the explored pathways.

In each of the considered case studies, we not only characterise the climate response by changes to its mean state, but also by changes to measures of its variability, symmetry and extremity. We analyse statistical changes to surface temperature, and also to a broader set of climate variables required to more directly estimate the influence of the climate on human health. For simplicity, the concentration pathways explored herein have been designed to be functions of time. In the discussion section we also illustrate how QuickClim could equally be adopted to generate climate data for scenarios as defined by integrated assessment models, spanning various settings for population growth, affluence and technological development.

## Results

**Climate estimation approach.** As mentioned above, QuickClim learns the relationships between atmospheric  $CO_{2e}$  concentrations and associated climate model output. The climate model data ingested in this study comprises of the projections contributed to the fifth Coupled Model Inter-comparison Project (CMIP5)<sup>27</sup>. These projections were generated by multiple climate models from various research centres worldwide, for a predefined set of future scenarios, referred to as Representative Concentration Pathways (RCP). Each RCP represents a possible future climate, with a specific  $CO_{2e}$  trajectory. In order of increasing levels of warming, the three scenarios adopted herein are denoted by RCP2.6, RCP4.5, and RCP8.5. A set of 15 general circulation models supplied the climate variables required for this study, with simulations run up to the year 2100 for all of the three aforementioned RCPs. QuickClim simultaneously learns the relationships between the  $CO_{2e}$  concentrations and the desired climate variables. The process is repeated for each individual CMIP5 model. This allows one to produce an ensemble of climate model projections for a given arbitrary concentration pathway. The approach can reconstruct fields of multiple climate variables at once, so long as the surface air temperature (SAT) field is available.

QuickClim is extensively described and validated in the Methods section, with a brief overview provided below. Figure 1 illustrates the four key stages: model reduction; training; projection; and reconstruction. In the model reduction phase, an empirical orthogonal function decomposition<sup>41,42</sup> is calculated across all time and scenarios to extract the climate change related spatial pattern (EOF1) and temporal evolution (PC1). During the training, separate ML models (red, grey, and blue boxes) are learnt linking the  $CO_{2e}$  concentrations with the associated climate change time series. In the projection phase these separate ML



**Fig. 1 QuickClim machine learning approach to generating space and time varying climate data for an arbitrary CO<sub>2e</sub> concentration pathway.** The QuickClim approach comprises of four phases: model reduction; ML training; projection; and reconstruction. The adopted ML approach is the Finite Element Method Bounded Variation Vector Auto-regressive model with exogenous factors (FEM-BV-VARX)<sup>33,34</sup>. While CO<sub>2e</sub> concentrations are presented here in terratonnes (Tt), the required ML input factor adopted in this study is the natural log of CO<sub>2e</sub> in gigatonnes. Note, any linear scaling of this exogenous input factor from the beginning of the process, due to for instance a change of units, would have no impact on the performance.

representations are combined in a manner weighted by the model relevance. This combined ML model (purple box) is then applied to calculate the temporal climate change response (PC1) to the prescribed concentration pathway of interest. In the reconstruction phase the new projected temporal response is multiplied by the original climate change related spatial pattern (EOF1). The climate variability fields are then sampled from the original CMIP scenario associated with the ML model most relevant in each month. The final dataset is the sum of the climate change and variability components.

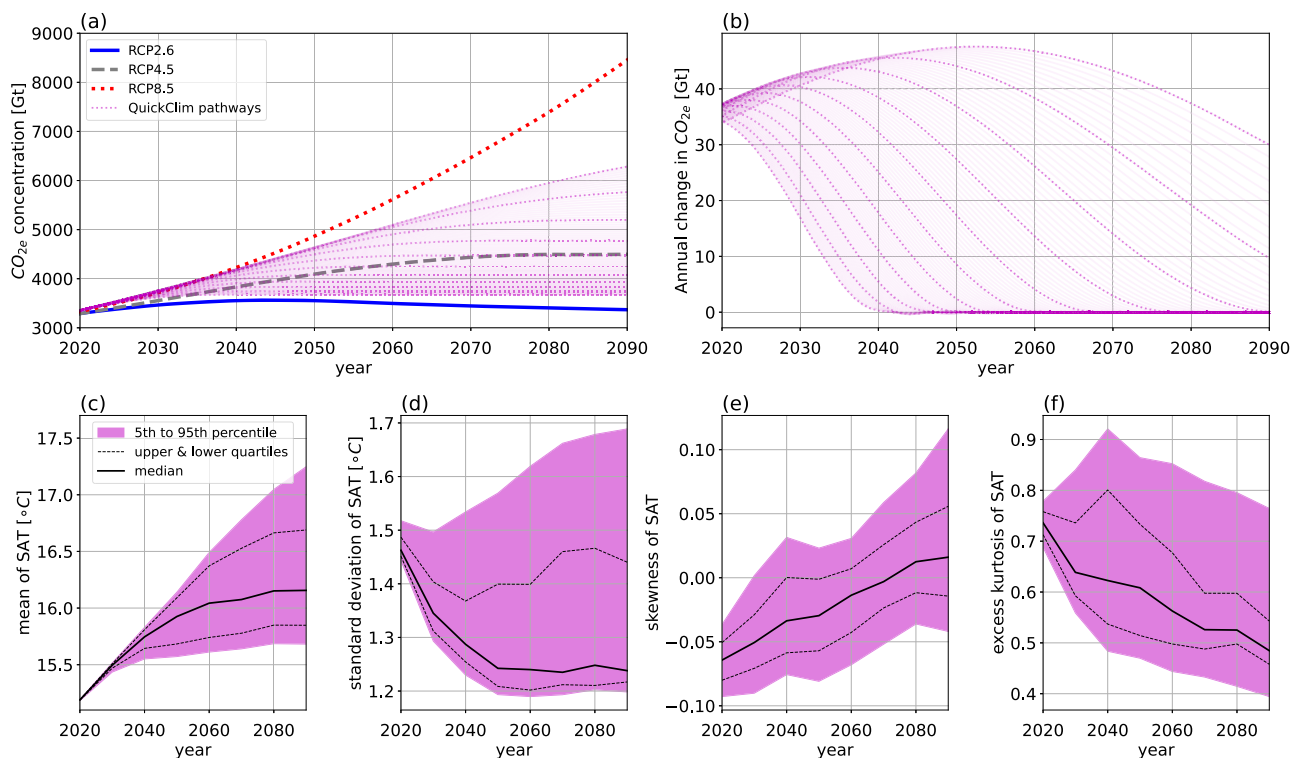
Multiple case studies are undertaken in the following sections. QuickClim is used to assess how carbon concentrations impact measures of the climate shown to be important for the human condition. We first explore the influence of the average decarbonisation rate, with the envelope of pathways engulfing the moderate decarbonisation scenario of RCP4.5. Next, we assess the sensitivity of the climate to a series of concentration pathways, all converging to the same end-of-century trajectory, akin to that of RCP4.5. Finally, we determine the climate sensitivity for an envelope of pathways converging to that of the net-zero emission type scenario of RCP2.6. In all cases we quantify the decadal statistics of SAT at a global scale. The response of wet-bulb-globe-temperature (WBGT) is also highlighted, as it is a determining factor in economic labour productivity<sup>19</sup> and human health<sup>17</sup>.

**Sensitivity to the average decarbonisation rate.** Here we explore the influence the decarbonisation rate has upon the climate. QuickClim reconstructs the climate response to 100 prescribed pathways with a range of decarbonisation rates, each attaining constant atmospheric concentrations in different future years. The longer the decarbonisation period, the larger the end-of-century  $CO_{2e}$  concentration. These pathways are illustrated by the magenta lines in Fig. 2a, with a subset of these pathways dotted

for clarity. The CMIP5 pathways of RCP2.6, RCP4.5, and RCP8.5 are also included in this figure for reference. The annual change in atmospheric  $CO_{2e}$  concentration of the reconstructed pathways, is illustrated in Fig. 2b. Pathways with the most aggressive decarbonisation attain a constant  $CO_{2e}$  concentration within the 2040s, while the least aggressive pathways continue to grow at the end of the century.

QuickClim produces spatio-temporally varying fields for all of these 100 individual concentration pathways. Multi-variate reconstructions are made of SAT, sea-level pressure and specific humidity, for each of the CMIP5 models that provide all of these variables. WBGT is then calculated from these reconstructed variables. For each model and concentration pathway, the time mean of SAT and WBGT is calculated over each decade. The standard deviation, skewness and excess kurtosis are similarly calculated, but on the climatological anomalies with a 5-year centred moving averaged subtracted. This is done so that these higher order statistics represent changes separate from the seasonal cycle and any warming trends. Percentile maps are calculated across the multiple concentration pathways for each of the decadal statistics. These percentile maps are then averaged over the output mimicking each of the CMIP models.

For each decade, Fig. 2c–f illustrates the globally averaged time mean, standard deviation, skewness and kurtosis of SAT. Within these plots, the solid black line is the median across all concentration pathways, the dashed lines are the upper and lower quartiles, and the magenta shaded region spans the 5th to 95th percentiles. For the decadal time mean, the median and inter-quantile ranges all continue to grow toward the end of the century, as the range across the carbon concentrations also grows in time. The median of the globally averaged standard deviation decreases toward the end of the century, as more of these trajectories attain a constant atmospheric  $CO_{2e}$  concentration.



**Fig. 2** Influence of decarbonisation rate on the decadal statistics of SAT, for case study one. **a** Atmospheric  $CO_{2e}$  concentration pathways (magenta), with subset of pathways dotted for clarity, and including RCP8.5 (red dotted), RCP4.5 (grey dashed), and RCP2.6 (blue solid) pathways for reference; and **b** annual change in atmospheric  $CO_{2e}$  concentration, with same subset of pathways dotted for clarity. For SAT, the global average and CMIP model ensemble average of the: **c** time mean; **d** anomalous standard deviation; **e** skewness; and **f** excess kurtosis; per decade. Legend in **c**, applicable to **d–f**.

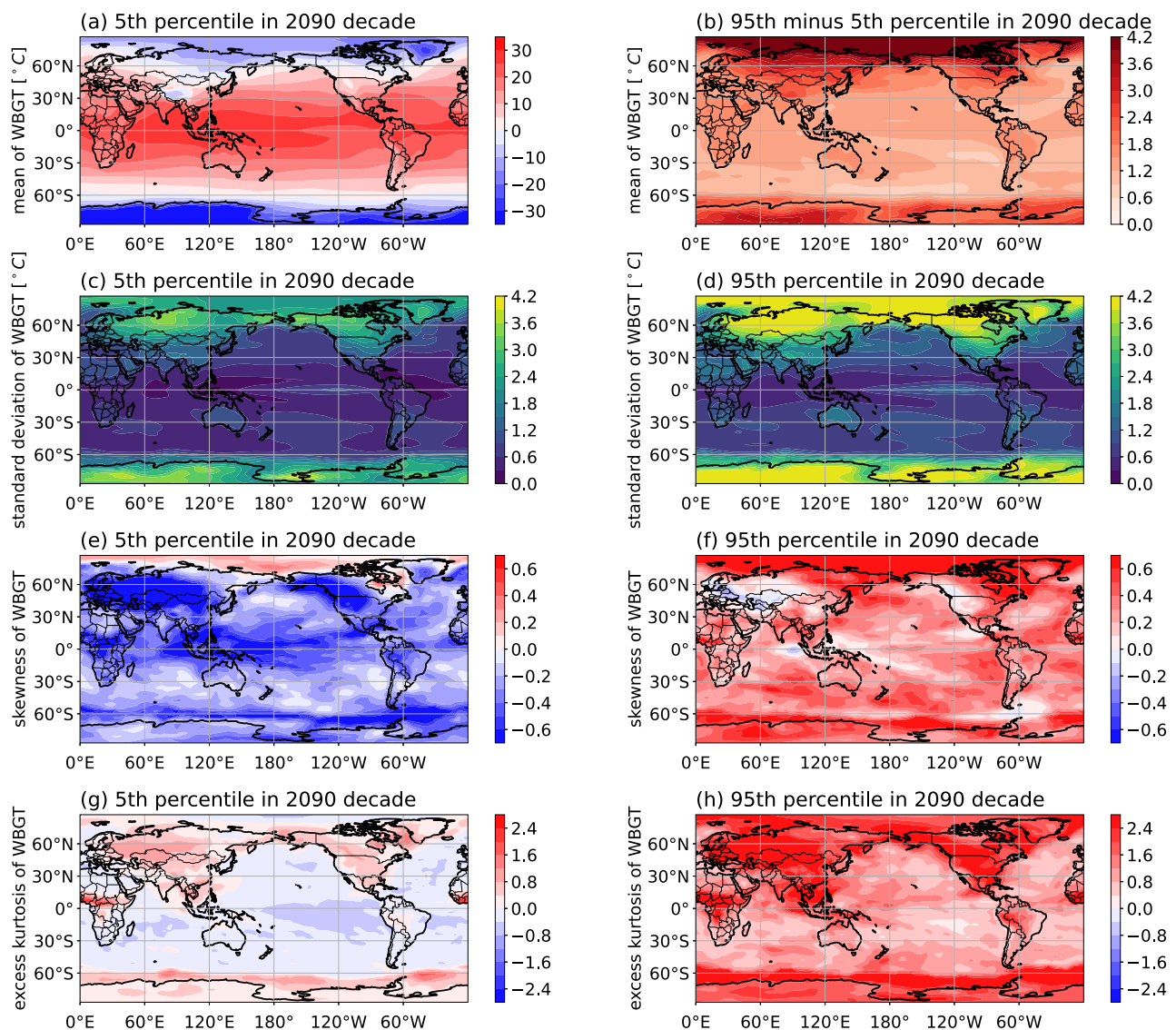


However, the uncertainty continues to grow in time, as quantified by the expansion of the inter-quantile ranges. At the beginning of the century the entire inter-percentile range of the globally averaged skewness is negative, and increases to being centred at zero by the end of the century. The globally averaged excess kurtosis (i.e., kurtosis minus three) is positive throughout the entire century, but is on a downward trajectory. The skewness and excess kurtosis both tend toward zero, indicating a return to Gaussian-like conditions. Their inter-quantile ranges also saturates by mid this century. In Supplementary Fig. 1, we also observe similar behaviour for SAT when averaged over: Africa; Asia; Oceania; Europe; North America; and South America.

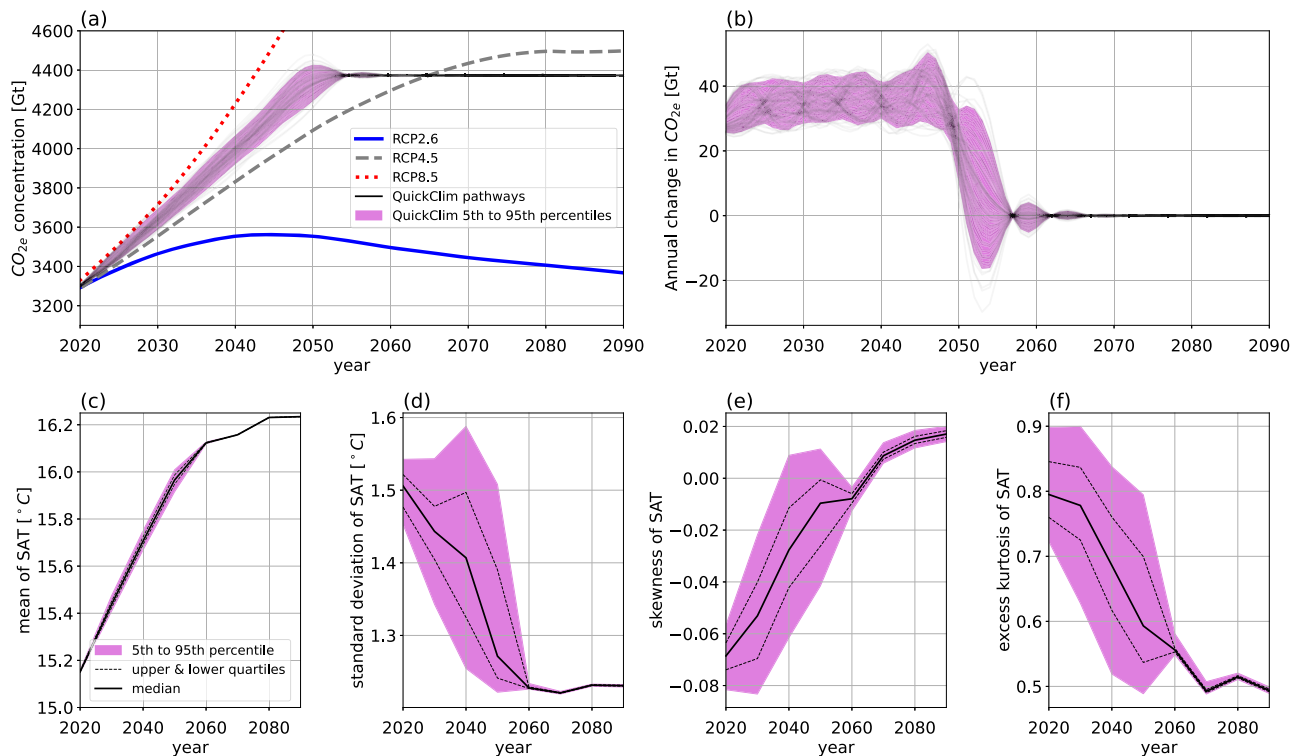
We now examine the spatial structure of the statistical range in potential climate responses due to the various reconstructed concentration pathways considered. For the decade centred at 2090, Fig. 3a illustrates the 5th percentile across the concentration pathways, of the time mean WBGT. To accentuate the spread in potential outcomes, Fig. 3b illustrates the difference between the 95th and 5th percentiles. This indicates enhanced warming in the high latitude regions. Figure 3c, d present the standard deviation, Fig. 3e, f the skewness and Fig. 3g, h the kurtosis. In each of these

plots the 5th and 95th percentiles are illustrated in the left and right columns, respectively. All plots demonstrate large differences across the percentile range. The standard deviation has potential for enhanced variability at the high latitudes. The 5th percentile of the skewness is predominantly negative over the entire globe, while the 95th percentile is almost exclusively positive. For excess kurtosis the lower percentile is consistently near zero, while the upper percentile is strongly positive over the entire globe. Hence, depending on the decarbonisation rate, not only is there potential for increased warming (via the mean), but also increased variability (standard deviation), tendency toward larger positive WBGT (skewness), and longer tailed probability distributions (kurtosis). Observations made on the basis of WBGT are also valid for SAT. The unknown future decarbonisation rate, adds to the total uncertainty of climate projections. Other sources of uncertainty include climate variability and choices made in the numerical modelling.

**Sensitivity to the specific trajectory.** As evidenced by the historical record, the atmospheric  $\text{CO}_2e$  concentration over the past decades has not evolved smoothly in time. The carbon



**Fig. 3** Wet-bulb-globe-temperature decadal statistics centred at 2090, for case study one. Time mean: **a** 5th percentile across all the concentration pathways; and **b** 95th minus 5th percentile. The following are calculated from the climatological anomalies. Standard deviation: **c** 5th; and **d** 95th percentile. Skewness: **e** 5th; and **f** 95th percentile. Excess kurtosis: **g** 5th; and **h** 95th percentile. All maps are averaged over the CMIP model ensemble.



**Fig. 4 Sensitivity of the decadal climate statistics to concentration pathways with fixed average decarbonisation rate, and fixed end-of-century atmospheric  $CO_{2e}$  concentration, for case study two.** **a** 5th to 95th percentile range of atmospheric  $CO_{2e}$  concentration pathways (shaded magenta), with individual pathways indicated (translucent black), including RCP8.5 (red dotted), RCP4.5 (grey dashed) and RCP2.6 (blue solid) pathways for reference; and **b** 5th to 95th percentile range of annual change in atmospheric  $CO_{2e}$  concentration (shaded magenta), with individual pathways indicated (translucent black). For SAT, the global average and CMIP model ensemble average of the: **c** time mean; **d** anomalous standard deviation; **e** skewness; and **f** excess kurtosis; per decade. Legend in **c**, applicable to **d-f**.

concentration is rather influenced by a manner of external factors, including but not limited to: global pandemics; economic expansions; and recessions. During the decarbonisation of the future global economy, the eventual concentration pathway, could quite likely be more volatile than what has previously been observed.

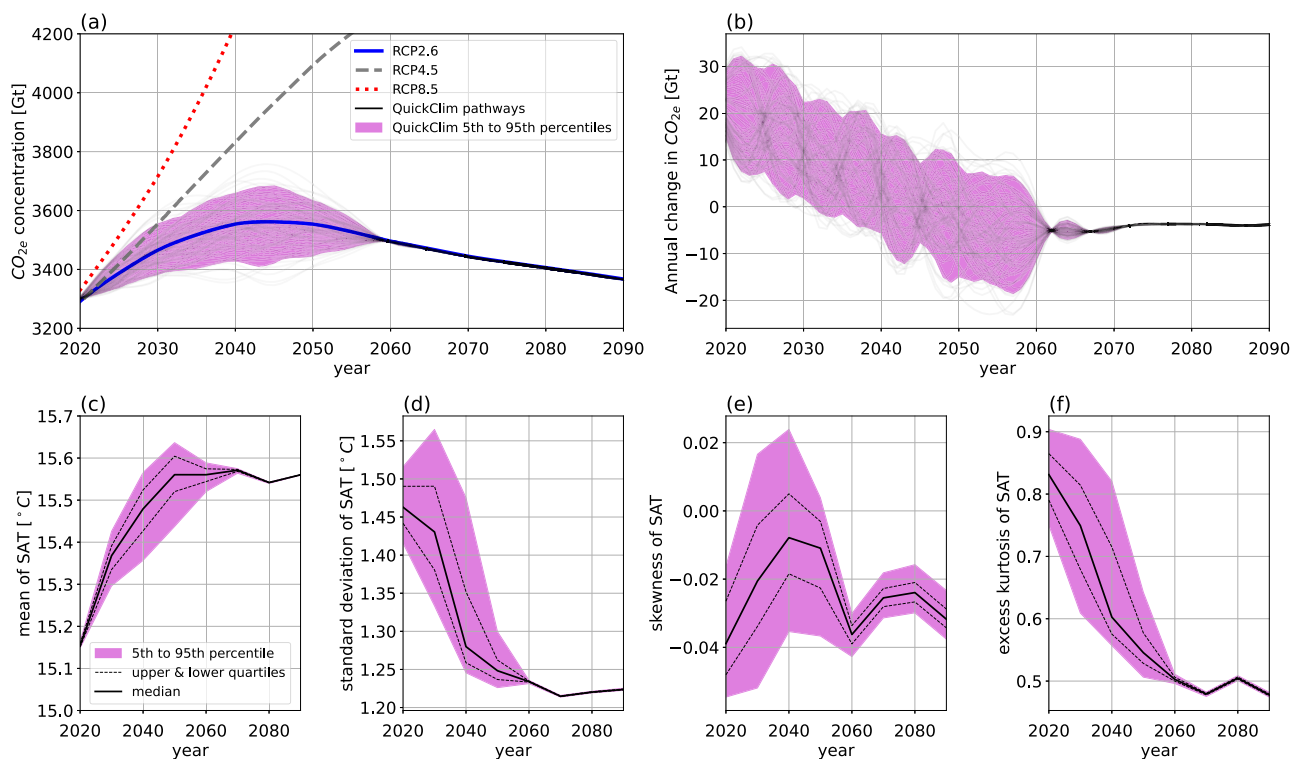
Recognising this, we assess another 100 concentration pathways, all achieving constant concentrations mid this century. They also all have the same fixed end-of-century  $CO_{2e}$  concentration, but with unique trajectories. The magenta shaded region in Fig. 4a spans the 5th and 95th percentiles of the concentration pathways, with the black lines illustrating all 100 individual trajectories. The RCP2.6, RCP4.5, and RCP8.5 concentration pathways are again included for reference. Figure 4b illustrates the annual changes in the QuickClim concentrations in terms of the 5th to 95th percentile range, and individual trajectories. In some of these trajectories the  $CO_{2e}$  concentration gradually increases year on year as they approach the final end-of-century conditions. Certain trajectories, however, overshoot and then decrease to the final  $CO_{2e}$  concentration. The 5th percentile of the annual reduction in  $CO_{2e}$  concentration across all of the scenarios is less than 20 Gt of  $CO_{2e}$  per year. This is approximately the level of carbon dioxide removal assumed at certain stages of this century in various integrated assessment model economic studies<sup>20,21</sup>.

The globally averaged SAT decadal time mean, standard deviation, skewness, and excess kurtosis, are illustrated in Fig. 4c–f. The median across the concentration pathways of all the decadal statistics follow trajectories consistent with the previous case study. The decadal mean increases, standard deviation decreases, skewness increases from a negative value to approach zero, and the excess kurtosis decreases while remaining

positive. However, for the global mean SAT, its inter-quantile range is small in comparison to the change in its median throughout the reconstructed time period. This is to be expected, since the envelope of concentration pathways is much smaller in this case study. On the other hand, for the standard deviation, skewness and excess kurtosis, their respective uncertainties peak in the 2040s, and are of the same magnitude as the change in their medians over the century. Additionally, in 2040 the magnitude of the inter-quantile range of the standard deviation and excess kurtosis, is similar in magnitude to that observed in the previous case study. This is despite the former case study spanning a much wider range of  $CO_{2e}$  concentrations. The above observations are also true for the statistics averaged over the majority of continental regions in Supplementary Fig. 2.

The statistical WBGW maps presented for the previous case study, are calculated for this new ensemble of trajectories and illustrated in Supplementary Fig. 3. These maps are for the decade centred in 2040. All of the statistical responses have similar spatial structure to those of the previous example. The inter-percentile range of WBGW in Supplementary Fig. 3b illustrates a smaller difference in the decadal mean response across the assessed concentration pathways as compared to the previous case study. This is also consistent with the smaller range in the globally averaged decadal mean in Fig. 4c. However, the standard deviation, skewness and excess kurtosis of WBGW indicates that there are large differences between the 5th and 95th percentiles. As such the choice of specific trajectory, can have a measurable influence on the higher order statistics, even for a given fixed average decarbonisation rate and fixed final  $CO_{2e}$  concentration.

Finally, to explore concentration pathways perhaps more compatible with a net-zero emissions scenario, we reconstruct a



**Fig. 5 Sensitivity of the decadal climate statistics to concentration trajectories converging to RCP2.6 toward the end of the century, for case study three.** **a** 5th to 95th percentile range of atmospheric  $\text{CO}_{2e}$  concentration pathways (shaded magenta), with individual pathways indicated (translucent black), including RCP8.5 (red dotted), RCP4.5 (grey dashed), and RCP2.6 (blue solid) pathways for reference; and **b** 5th to 95th percentile range of annual change in atmospheric  $\text{CO}_{2e}$  concentration (shaded magenta), with individual pathways indicated (translucent black). For SAT, the global average and CMIP model ensemble average of the: **c** time mean; **d** anomalous standard deviation; **e** skewness; and **f** excess kurtosis; per decade. Legend in **c**, applicable to **d-f**.

family of different trajectories all converging to the RCP2.6 pathway by the end of the century. The carbon dioxide concentration in RCP2.6 is similar to those of the first shared socio-economic pathways in the sixth CMIP, which encompasses net-zero emissions like scenarios<sup>43,44</sup>. The QuickClim pathways, and associated globally averaged statistics are illustrated in Fig. 5. All of the decadal statistics in Fig. 5c–f exhibit similar inter-quantile ranges as compared to case study two. However, the inter-quantile range of the time average mean is no longer negligible, as compared to the milder increase in the median SAT over the century. These observations are consistent with the continental averages illustrated in Supplementary Fig. 4. Finally, Supplementary Fig. 5 illustrates the same WBGW statistics as in Fig. 3, but with statistics centred in 2040. These maps exhibit similar spatial structure and magnitude to those of the second case study. The choice of specific trajectory toward this net-zero emissions type scenario, is again shown to have a measurable influence on the decadal statistics.

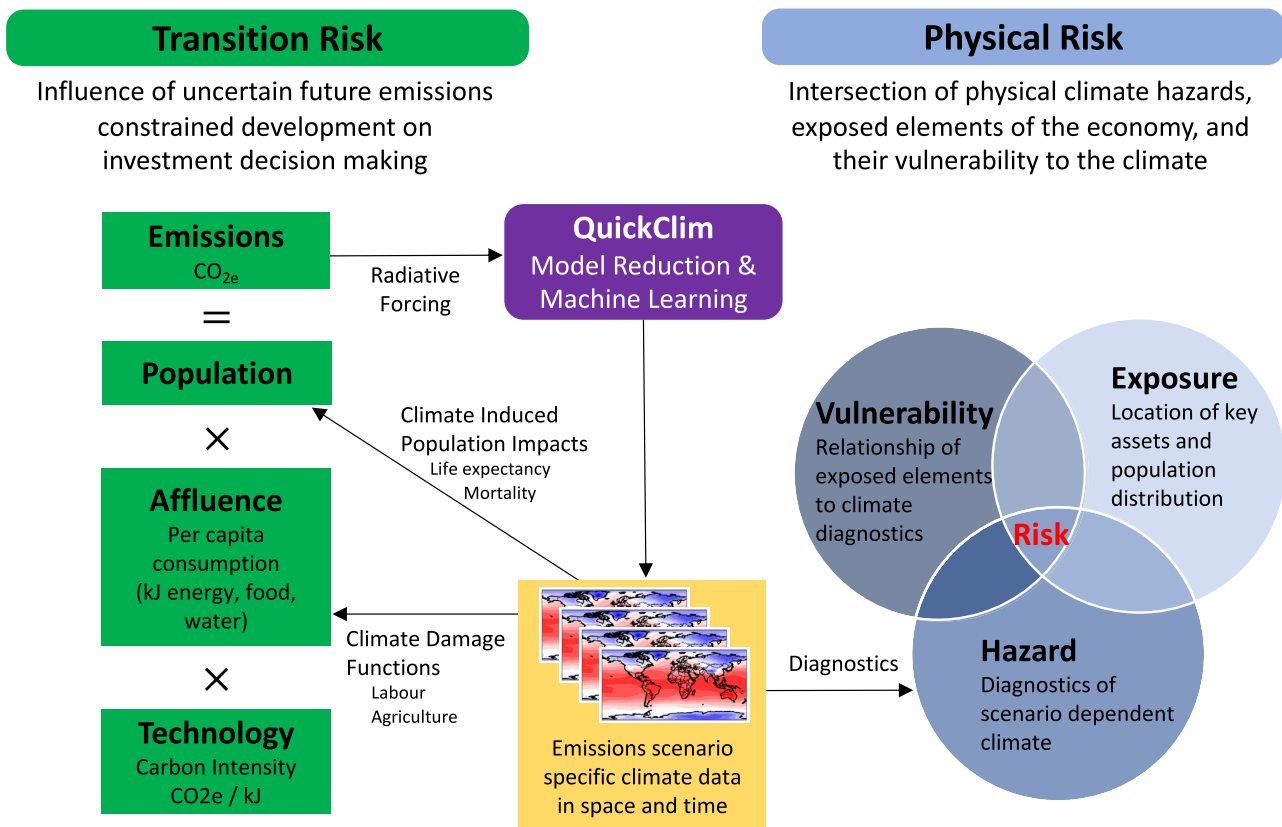
## Discussion

The pre-industrial era was a statistically stationary, stable, and cooler world than the one we currently live in. The end-of-century climate will almost certainly be warmer, but potentially also statistically stationary. Our results show that the most unstable time of high risk may in fact be now, as we adjust from a pre-industrial climate to some new future stable one. This current transition period corresponds to one where disruptions and even tipping points might manifest in changes to rainfall patterns and variations in major climate teleconnections. There may well be relatively little time to adapt to such rapidly changing features of the Earth system.

To summarise, QuickClim enabled the rapid and computationally cheap estimation of the physical climate response across an extensive range of concentration pathways. This would not be possible using numerical simulations of the climate alone. Three case studies were performed. In the first case study, slower average rates of decarbonisation, leading to higher end-of-century atmospheric  $\text{CO}_{2e}$  concentrations, were shown to produce higher temperatures. Skewness and kurtosis were also shown to steadily approach Gaussian-like conditions, with their uncertainties saturating mid this century. In the second case study, for a fixed average decarbonisation rate and fixed final  $\text{CO}_{2e}$  concentration, the climate variability was shown to be sensitive to the specific concentration pathway. The inter-percentile ranges of the standard deviation, skewness and kurtosis, were of the same order as the change of their medians over the century. This was also true for the net-zero emissions type family of concentration pathways in the third case study. Here, the inter-percentile range of the mean was also of similar magnitude to its change over the century due to the milder warming.

On the basis of these case studies, and the physical representativeness of the climate model output, the take home messages are as follows. Climate variability and change are both strongly dependent upon the average decarbonisation rate. The specific trajectory for a given fixed decarbonisation rate, can also have a quantifiable effect upon measures of the climate variability, symmetry and extremity. This effect may also be present in the mean climate change response if the warming is sufficiently mild, as in net-zero emissions scenarios.

Determining a fair, equitable, and efficient pathway forward toward net-zero emissions requires a holistic and consistent assessment of both the climate and economy. Figure 6 illustrates



**Fig. 6 QuickClim provides emissions trajectory specific climate data sets, enabling consistent transitional and physical risk assessments.** Carbon emissions are defined as the product of population, per capita energy consumption, and the carbon intensity of the energy sources. These emissions influence the radiative forcing, which QuickClim can use to generate gridded fields of associated climate variables. These variables can then be used to quantify environmental hazards, which are an essential element of physical climate risk.

the role that QuickClim could play in enabling such an assessment of transitional and physical climate risk. In assessing transition risk, the emissions act as a constraint on the economic system. The annual emissions are the product of the population, embodied energy consumed per person (proportional to affluence), and the carbon intensity of the energy sources<sup>45,46</sup>—see green boxes in Fig. 6. Associated concentrations are readily ingested by QuickClim to generate compatible spatio-temporally varying climate data. These QuickClim outputs can then be used to assess: the climate induced population impacts<sup>15,17</sup>; affluence reducing climate damage functions<sup>7,18,19</sup>; and the environmental hazards. Physical climate risks are the intersection of the environmental hazards, location of the exposed assets and people, and their vulnerabilities to these hazards<sup>47</sup>—see blue circles in Fig. 6.

While QuickClim can rapidly assess the climate response, we in no way see it as a replacement for the general circulation models used in CMIP. In fact, we envisage that this technology would serve as a complementary tool, to quickly explore the large concentration pathway parameter space. Pathways with particularly risky and/or interesting QuickClim output, could be verified by one or more general circulation models. Any new simulated datasets could then be additionally ingested by QuickClim, which would further improve its representation of the concentration pathway parameter space.

Finally, the case studies presented within are representative of potential outcomes, but certainly do not encompass all potential future risks. The motivation for developing QuickClim, is to enable researchers, policymakers, and other stakeholders to investigate future climate risks for economic scenarios to which they might be particularly exposed. We propose that a greater understanding of the potential risks of climate variability and

change can only help the global community in the decarbonisation effort over the coming decades.

**Methods**

As illustrated in Fig. 1, QuickClim comprises of four key stages: model reduction; training; projection; and reconstruction. This methods section includes a description and illustration of each of these four stages for one given CMIP model, with additional validation provided in Supplementary Note 2. An assessment of the reconstruction process across all of the available CMIP models is presented in Supplementary Note 3.

**Model reduction.** The first step is to reduce the dimension of the problem via some form of model reduction. For a given climate model, each surface field is a function of time  $t$ , and space  $\mathbf{y} = (\lambda, \phi)$  of longitude  $\lambda$ , and latitude  $\phi$ . Additionally, the data is dependent upon the scenario  $\mathbf{s}$ , which here is specifically  $\mathbf{s} \equiv (s_1, s_2, s_3) = (RCP2.6, RCP4.5, RCP8.5)$  of length  $N_s = 3$ . To facilitate the discussion we define the state vector  $\mathbf{q}(\mathbf{y}, t, \mathbf{s})$ , containing at a minimum the SAT, as denoted by  $T(\mathbf{x}, t, \mathbf{s})$ , followed by any other field variables required to be reconstructed. The number of variables contained in vector  $\mathbf{q}$  is denoted by  $N_q$ . It is essential that SAT is included, since it has a strong relationship between the concentrations and the response of the climate system. The data sets are standardised according to

$$\mathbf{q}'(\mathbf{y}, t, \mathbf{s}) = \frac{\mathbf{q}(\mathbf{y}, t, \mathbf{s}) - \tilde{\mathbf{q}}(\mathbf{y}, o)}{\sigma(\mathbf{y}, o)}, \tag{1}$$

where  $o$  is the calendar month of the year, with  $\tilde{\mathbf{q}}(\mathbf{y}, o)$  and  $\sigma(\mathbf{y}, o)$  the climatological average and climatological standard deviation



calculated across all time and scenarios. The resulting standardised variables are all now non-dimensional. A empirical orthogonal function (EOF) decomposition is then performed on these standardised fields such that

$$\mathbf{q}'(\mathbf{y}, t) = \sum_{n=1}^{N_{\text{modes}}} a^{[n]}(t) \Phi^{[n]}(\mathbf{y}), \quad (2)$$

where  $N_{\text{modes}}$  is the number of modes used in the reconstruction, with  $a^{[n]}(t)$  and  $\Phi^{[n]}(\mathbf{y})$  the  $n$ -th principal component and  $n$ -th EOF mode, respectively. Each EOF mode is a vector quantity containing the SAT, followed by any other included field variables. This means that one set of EOF modes is used to compress the data across all of the three scenarios. This is an essential property, since we will be combining model representations and output across the various scenarios. It is, therefore, important that the reduction is applied consistently across these scenarios.

The decomposition is calculated by solving the eigenvalue problem

$$\mathbf{C}\mathbf{a}^{[n]} = \Lambda^{[n]}\mathbf{a}^{[n]}, \quad (3)$$

where  $\mathbf{C}$  is a covariance matrix, with  $\mathbf{a}^{[n]}$  and  $\Lambda^{[n]}$  its  $n$ -th eigenvector and eigenvalue, respectively<sup>42</sup>. The eigenvectors are scaled such that  $\Lambda^{[n]} = \mathbf{a}^{[n]} \cdot \mathbf{a}^{[n]} / N_s / N_T$ . The elements of  $\mathbf{C}$  are given by

$$C_{k+N_T j, i+N_T l} = \frac{1}{N_T} \left\langle \mathbf{q}'(\mathbf{y}, t_k, s_j), \mathbf{q}'(\mathbf{y}, t_i, s_l) \right\rangle, \quad (4)$$

for  $i, k \in [1, N_T]$  and  $j, l \in [1, N_s]$ , with  $N_T$  the number of samples. The inner product in (4) is defined as

$$\begin{aligned} \left\langle \mathbf{q}'(\mathbf{y}, t_k, s_j), \mathbf{q}'(\mathbf{y}, t_i, s_l) \right\rangle = \\ \int_V \left[ T'(\mathbf{y}, t_k, s_j) T'(\mathbf{y}, t_i, s_l) + \frac{1}{N_q - 1} \sum_{v=2}^{N_q} q'_v(\mathbf{y}, t_k, s_j) q'_v(\mathbf{y}, t_i, s_l) \right] d\mathbf{y}, \end{aligned} \quad (5)$$

with  $V$  the integration volume. This inner product is designed such that in instances where fields other than SAT are included, for a given element in the covariance matrix, the SAT contributes half, with the other half coming from the remaining variables. This ensures that the model reduction process produces a first principle component (PC) time series predominantly associated with the temperature field, for which there is a strong association with the concentrations.

By construction of the inner product, each eigenvector is a concatenation of the temporal evolution of each scenario where

$$\mathbf{a}^{[n]} = [\mathbf{a}^{[n]}(s_1), \mathbf{a}^{[n]}(s_2), \mathbf{a}^{[n]}(s_3)], \quad (6)$$

with each vector  $\mathbf{a}^{[n]}(s_j)$ , containing the instances in the time series  $a^{[n]}(t, s_j)$ , representing the evolution of the EOF mode  $n$  for scenario  $j$ . The EOF patterns are recovered according to

$$\Phi^{[n]}(\mathbf{y}) = \frac{1}{N_T N_s \Lambda^{[n]}} \sum_{k=1}^{N_T} \sum_{j=1}^{N_s} a^{[n]}(t_k, s_j) \mathbf{q}'(\mathbf{y}, t_k, s_j), \quad (7)$$

which exploits the orthogonality property  $\langle \Phi^{[n]}(\mathbf{y}), \Phi^{[r]}(\mathbf{y}) \rangle = \delta_{nr}$ , where  $\delta_{nr}$  is the Kronecker delta function.

To illustrate the process a multi-variate EOF decomposition was undertaken for the climate model BCC-CSM1-1, with key properties consistent across all of the assessed CMIP5 models. The surface variables included are SAT, zonal wind, meridional wind, relative humidity, sea-level pressure, and precipitation flux. SAT is the key variable in terms of determining the relationship with the concentrations, and the other variables are included in the EOF decomposition to ensure they are all reconstructed consistently. The time series associated with each scenario per PC is illustrated for modes 1, 2, and 3 in Fig. 7a–c, respectively. The titles of these plots list the percentage of explained variance, with the first mode containing 26.87%, followed by 2.21% and 1.83%

for the second and third most energetic modes. Clearly, mode  $n=1$  captures the climate change signal with clear distinctions between each of the scenarios. The RCP8.5 time series increases most strongly, followed by that of the intermediate emission pathway RCP4.5, and then the lowest emission pathway RCP2.6. In contrast, for PCs of  $n=2$  and  $n=3$ , a given mode has negligible differences across each of the scenarios in terms of the time mean and variance. This is also true for the higher order PCs with  $n > 3$ .

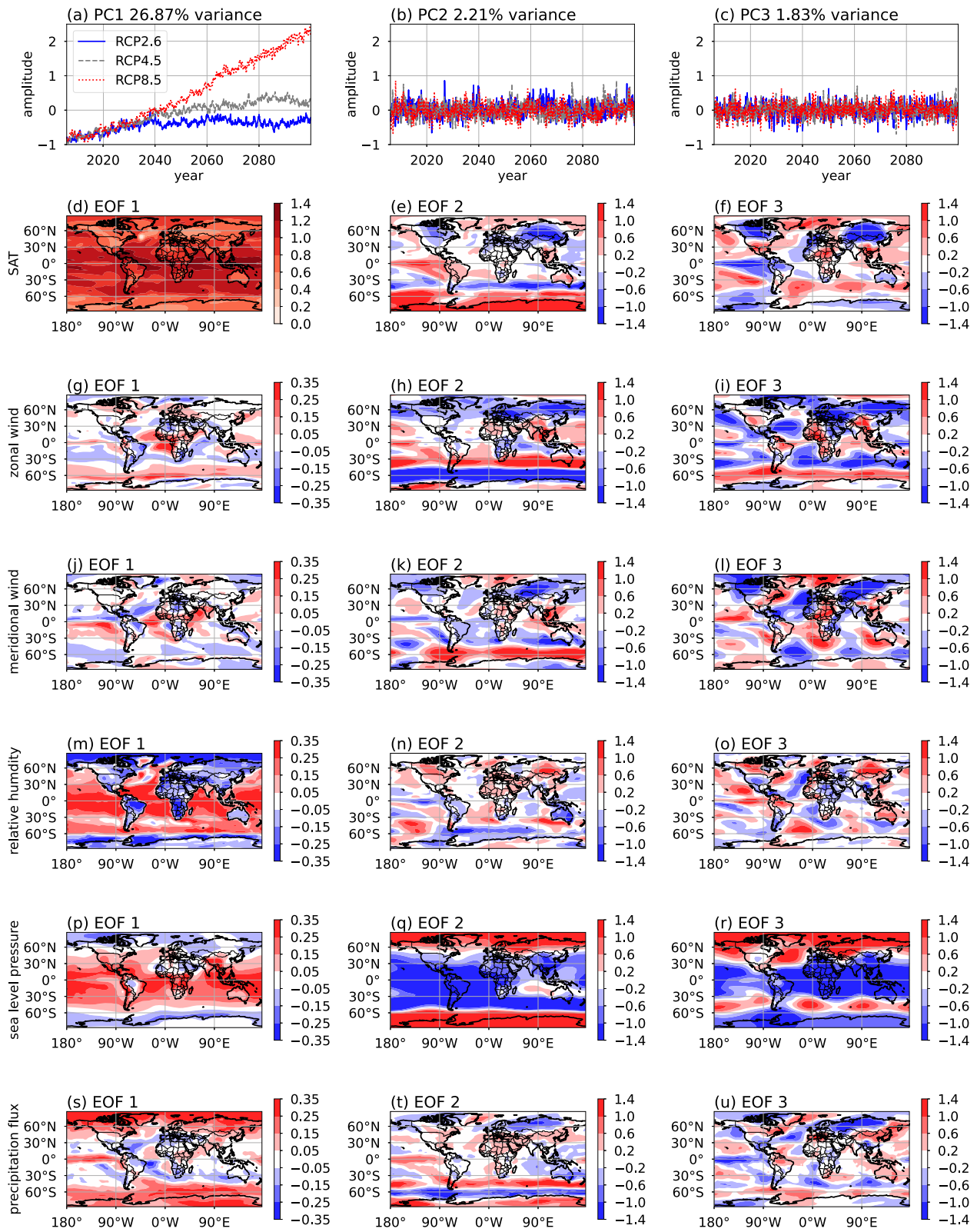
The multi-variate spatial patterns are illustrated from the second row downwards in Fig. 7. Modes 1, 2, and 3 are illustrated in the left, middle, and right columns, respectively. The second row contains the patterns associated with the surface temperature. Mode 1 is clearly a climate change associated mode, exhibiting a predominantly zonal structure, which is warming at all locations. Mode 2 and 3 exhibit El Niño Southern Oscillation type structures over the Pacific. The higher order modes ( $n > 3$ ) have progressively smaller scale spatial structures with shorter timescales. The remaining rows in descending order illustrate the associated patterns per mode for the zonal wind, meridional wind, relative humidity, sea-level pressure and the precipitation flux.

In summary, the first PC is the only mode to have time series exhibiting a clear statistical distinction between each of the scenarios. Note, one might expect a climate change mode to have larger warming at the poles and extra-tropics as compared to the tropics. Recall, however, that these patterns have all been normalised by the standard deviation of the field. The magnitude of the variability of the surface temperature is less in the tropical regions, which is why this nondimensionalised pattern has increased magnitude in this region. The QuickClim reconstruction process in its entirety, however, does produce the expected global warming patterns of increased warming at the poles, as evidenced in Fig. 3b, Supplementary Figs. 3b and 5b.

**Training.** For each scenario, we now build ML representations for the first PC time series dependent upon its associated  $\text{CO}_2e$  concentrations. We adopt the finite element method (FEM) clustering approach, with bounded variation (BV) regularisation, based on a distance metric governed by a vector auto-regressive model with exogenous factors (VARX). This technique is referred to as FEM-BV-VARX<sup>33,34</sup>, with the mathematical details provided in Supplementary Note 1. By combining FEM-BV-VARX representations of each scenario in a judicious way, we can produce climate projections for any user defined exogenous carbon concentration pathway. Note, that we also tested the generation of FEM-BV-VARX representations for the higher order PCs, and found the model coefficients associated with the exogenous concentrations were negligible. We, therefore, require a concentration dependent projection of only the first PC.

The projection of the first PC time series,  $a^{[1]}(t, s_j)$ , must be performed independently for each scenario  $s_j$ . For clarity of notation we define the scalar time series  $x(t) = a^{[1]}(t, s_j)$ . The goal is to build a mathematical model for the time series  $x(t)$ , subject to an exogenous forcings  $u(t)$ , characterised by a time-dependent set of model parameters  $\theta(t)$ . Both  $x(t)$  and  $u(t)$  have  $T+1$  snapshots over the time interval  $[0, T]$ . In dealing with time series consisting of trends, the main idea is to break the problem down into a series of local time chunks (i.e., finite elements) that are treated individually as being stationary. Within a given local chunk of time the data is represented by one of potentially multiple clusters governed by the auto-regressive model

$$x(t) = \mu^{(i)} + \sum_{q=1}^m A_q^{(i)} x(t - q\tau) + B^{(i)} u(t), \quad (8)$$



**Fig. 7** Empirical orthogonal function decomposition for a given climate model. Modes 1 **a, d, g, j, m, p, s, 2 b, e, h, k, n, q, t**, and 3 **c, f, i, l, o, r, u**, illustrated by their principle components time series **a-c**, and spatial patterns of the nondimensionalised surface variables: SAT **d-f**; zonal wind **g-i**; meridional wind **j-l**; relative humidity **m-o**; sea-level pressure **p-r**; and precipitation flux **s-u**. Legend in **a** applicable to **b, c**.

where  $m$  is the number of endogenous lags,  $\mu^{(i)}$  is a constant coefficient,  $A_q^{(i)}$  the endogenous coefficients, and  $B^{(i)}$  the exogenous coefficients of cluster  $i$ . Time-dependent probabilities (or model affiliations), denoted by  $\gamma^{(i)}(t)$ , determine which of the multiple clusters best represents that chunk of time. The  $\gamma^{(i)}(t)$  affiliations are non-negative at all times  $t$  and for all clusters  $i$ . At all times the sum of  $\gamma^{(i)}(t)$  over all clusters is equal to 1. The bounded variation constraint incorporates additional information pertaining to the temporal persistency, where the parameter  $p$  sets a lower bound on the average duration between cluster transitions. The cluster dependent parameter vector is then defined as

$$\theta^{(i)} = (\mu^{(i)}, A_1^{(i)}, A_2^{(i)}, \dots, A_m^{(i)}, B^{(i)}), \text{ with} \quad (9)$$

$$\theta(t) = \sum_{i=1}^K \gamma^{(i)}(t) \theta^{(i)}, \quad (10)$$

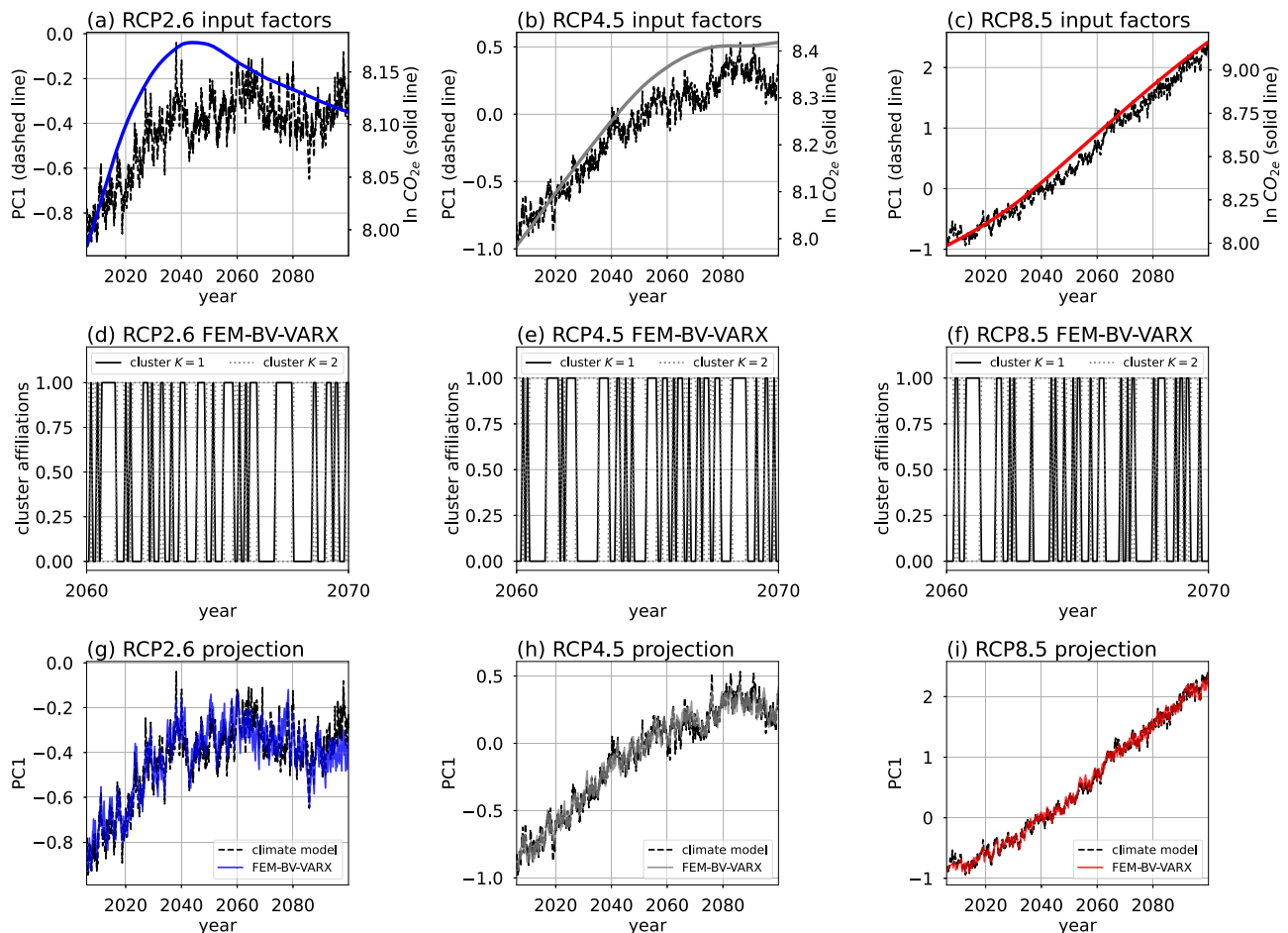
the time varying global parameter set. In the following section, we refer to this global parameter set for a given scenario  $s$  as  $\theta_s(t)$ .

In summary, for a given  $x(t)$  and  $u(t)$ , the number of clusters  $K$ , VARX order  $m$  (i.e., memory length), and state persistence duration  $p$ , constitute the set of hyper-parameters that must be chosen beforehand when applying the above procedure. For simplicity, we assume the same lag order  $m$  for all  $K$  clusters. The FEM-BV-VARX inputs and outputs are illustrated for the scenarios: RCP2.6 in Fig. 8a, d, g; RCP4.5 in Fig. 8b, e, h; and RCP8.5 in Fig. 8c, f, i. These plots are for the climate model BCC-CSM1-1, with  $K=2$ ,  $m=3$  months and  $p=3$  months. For each of these scenarios, the

top row in Fig. 8 illustrates the endogenous climate model PC by the dashed black line. The solid coloured lines are the exogenous time series, which is given by the natural logarithm ( $\ln$ ) of the associated  $\text{CO}_{2e}$  concentrations. The  $\ln$  of  $\text{CO}_{2e}$  is proportional to the radiative forcing acting on the atmosphere.

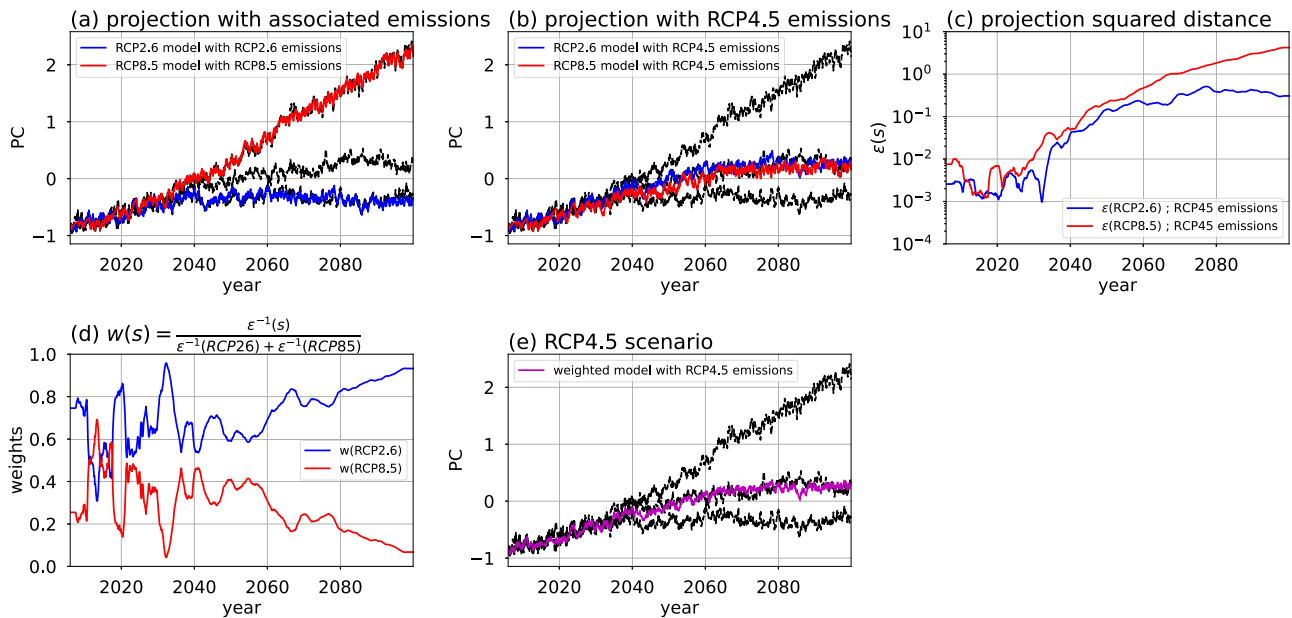
The relationship between this first PC and  $\ln \text{CO}_{2e}$  appears to become more direct, as the scenarios become more carbon intensive from left to right. For each scenario these time series are the inputs in the FEM-BV-VARX optimisation problem. The solution to this optimisation problem generates the cluster dependent FEM-BV-VARX model coefficients listed in supplementary table 1, and the time varying cluster affiliation sequences illustrated in the middle row of Fig. 8. For clarity we have zoomed in to a period between 2060 and 2070, to more clearly illustrate the switching between clusters states. The time varying parameters for a given scenario are calculated following (10). That is, the cluster dependent coefficients in Supplementary Table 1 are multiplied by the time varying affiliations of the associated cluster, and then summed over all clusters. The bottom row illustrates the climate model PC (dashed black line) compared with the FEM-BV-VARX projection (solid coloured line) of the associated scenario. The FEM-BV-VARX projection adopts the time varying parameters along with the associated exogenous  $\ln \text{CO}_{2e}$  time series.

**Projection.** Now that we have the ability to generate FEM-BV-VARX representations of existing climate data sets, the question



**Fig. 8** FEM-BV-VARX representation of the first principle component (PC1). For: RCP2.6 **a, d, g**; RCP4.5 **b, e, h**; and RCP8.5 **c, f, i** scenarios. **a-c** illustrates the endogenous climate model PC1 (black dashed left axis) and the exogenous  $\text{CO}_{2e}$  concentrations (colour solid right axis). **d-f** illustrates the cluster affiliation sequences. **g-i** illustrates the climate model PC1 (black dashed) compared with the FEM-BV-VARX projection (colour solid).





**Fig. 9 Reconstruction of RCP4.5 principle component using FEM-BV-VARX representations of RCP2.6 and RCP8.5 climate model output.** **a** Projection of RCP2.6 and RCP8.5 FEM-BV-VARX models using their associated exogenous concentrations, compared to PCs of all three scenarios (black dashed). **b** Projection of RCP2.6 and RCP8.5 FEM-BV-VARX models using RCP4.5 exogenous concentrations. **c** Centred 3-year moving average of the squared distance between the FEM-BV-VARX projections using RCP4.5 exogenous concentrations and the PCs of the associated scenarios (RCP2.6, RCP8.5) from which the FEM-BV-VARX models were learnt. **d** Weights for combining time varying FEM-BV-VARX parameters. **e** FEM-BV-VARX projection of RCP4.5 (magenta) using the  $w_s(t)$  weighted average of the RCP2.6 and RCP8.5 model parameters, compared to the first PC of each scenario (black dashed).

is how we can use this to generate climate model output for an arbitrary user-defined pathway. To validate the approach we use climate model output from only two of the three scenarios (RCP2.6, RCP4.5, RCP8.5) in order to reconstruct the excluded dataset. This will enable us to compare the FEM-BV-VARX based reconstruction with the original excluded dataset. Below we illustrate the process using the RCP2.6 and RCP8.5 scenarios to reconstruct RCP4.5 for one particular climate model. Note, excluding RCP4.5 from the initial model reduction process produces exceedingly similar spatial EOF patterns, as those calculated when all of the scenarios are included. The inclusion of RCP4.5 in the model reduction stage, therefore, has negligible impact on the following validation metrics.

Figure 9a illustrates the reconstruction of the FEM-BV-VARX projections of RCP2.6 (blue line, as in Fig. 8g) and RCP8.5 (red line, as in Fig. 8i) along with the PCs of the climate model output (associated dashed black lines). These projections were undertaken using the parameters listed in Supplementary Table 1, the cluster affiliations illustrated in Fig. 8, along with their associated exogenous  $CO_{2e}$  concentrations. The target RCP4.5 PC (centre dashed black line) is also included in this plot as a reference between these two bounding scenarios. Figure 9b illustrates the projection of these models, using the same model parameters, but instead with the exogenous carbon concentrations of the RCP4.5 scenario. We refer to these as out-of-sample projections. The projection using the RCP2.6 model (blue) is lifted to approach the target RCP4.5 PC, and the projection using the RCP8.5 model (red) has fallen to also approach the RCP4.5 PC. We now require an approach of combining these two estimates of the RCP4.5 scenario.

We assume that the less a model projection diverts from its original trajectory, as a result of the new RCP4.5 exogenous input, the more representative it might be. As a measure of such a diversion, the blue line in Fig. 9c illustrates a centred 3-year moving averaging of the squared difference between the RCP2.6 PC and the model projection using instead the RCP4.5 exogenous

time series. The same calculation is performed for the model built using the RCP8.5 data, illustrated by the red line. This distance metric is denoted by  $\epsilon_s(t)$  for scenario  $s$ . As the simulation progresses and the exogenous factors begin to take effect. The projections using the RCP4.5 concentrations become more distinct from the PC of the scenario the model coefficients were learnt from, and hence these distance measures increase. Toward the latter period, in particular, the FEM-BV-VARX model built using the RCP2.6 data has deviated to a lesser degree, and is presumably the more representative model.

We then calculate a time series of weights,  $w_s(t)$  for each scenario  $s \in (RCP2.6, RCP8.5)$ , by normalising the distance metric inverse at each instant in time, by its sum across the scenarios, such that

$$w_s(t) = \epsilon_s^{-1}(t) / \sum_{s \in (RCP2.6, RCP8.5)} \epsilon_s^{-1}(t). \quad (11)$$

By construction these weights are non-negative and sum to one at all instances in time, and are illustrated in Fig. 9d. The time varying global parameter sets of each scenario are combined using these weights to produce the final set of parameters

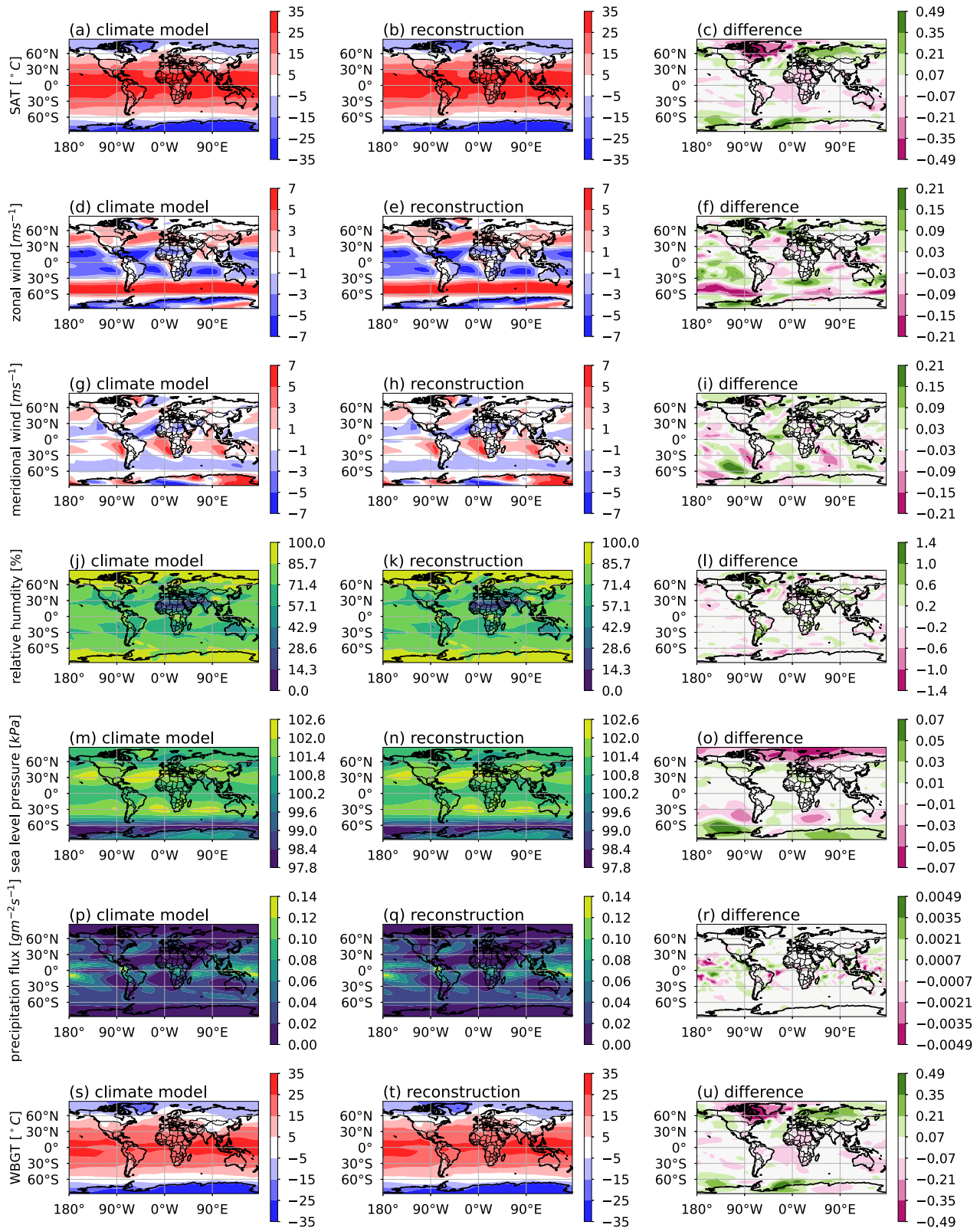
$$\hat{\theta}(t) = \sum_{s \in (RCP2.6, RCP8.5)} \theta_s(t) w_s(t). \quad (12)$$

This time varying global parameter set along with the exogenous RCP4.5 concentrations is then used to produce the projection of the RCP4.5 scenario, denoted by  $\hat{x}(t)$ , and illustrated in Fig. 9e by the magenta line.

**Reconstruction and validation.** Given the above projection of the first PC, we can now create gridded output for all of the climate variables involved in the initial EOF decomposition. Gridded fields of the climate change component are then given by

$$\mathbf{q}'_{cc}(\mathbf{y}, t) = \hat{x}(t) \Phi^{[1]}(\mathbf{y}), \quad (13)$$





**Fig. 10 Comparison of original climate model output and reconstructed surface fields averaged over the simulation period.** Original RCP4.5 data **a, d, g, j, m, p, s**, reconstruction of the RCP4.5 scenario **b, e, h, k, n, q, t**, and original minus the reconstruction **c, f, i, l, o, r, u**, for the: **a-c** SAT; **d-f** zonal wind; **g-i** meridional wind; **j-l** relative humidity; **m-o** sea level pressure; **p-r** precipitation flux; and **s-u** wet bulb globe temperature.

where  $\hat{x}(t)$  is the FEM-BV-VARX projected first PC, and  $\Phi^{[1]}(\mathbf{y})$  are the EOF patterns of mode 1 as illustrated by the maps in the first column of Fig. 7. To determine the climate variability component, we return to the standardised data instances. At each instant in time  $t$ , the scenario with the maximum value of  $w_s(t)$  is identified according to

$$s^*(t) = \operatorname{argmax}_s w_s(t). \quad (14)$$

The climate variability component of the new dataset is then defined as the data instance of scenario  $s^*(t)$  minus the first PC contribution, such that

$$\mathbf{q}'_{cv}(\mathbf{y}, t) = \mathbf{q}'(\mathbf{y}, t, s^*(t)) - a^{[1]}(t, s^*(t)) \Phi^{[1]}(\mathbf{y}). \quad (15)$$

The complete data set is then given by the sum of the time varying climate change and climate variability fields

$$\mathbf{q}'_{new}(\mathbf{y}, t) = \mathbf{q}'_{cc}(\mathbf{y}, t) + \mathbf{q}'_{cv}(\mathbf{y}, t). \quad (16)$$

Finally, we dimensionalise the fields by the associated climatological average and standard deviation according to

$$\mathbf{q}_{new}(\mathbf{y}, t) = \mathbf{q}'_{new}(\mathbf{y}, t) \odot \sigma(\mathbf{y}, o(t)) + \tilde{\mathbf{q}}(\mathbf{y}, o(t)), \quad (17)$$

where  $o(t)$  is the calendar month at time  $t$ , and  $\odot$  denotes element-wise multiplication between the vector components.

The time average of this reconstruction over the entire simulation period, is illustrated in the middle column of Fig. 10. The original RCP4.5 data is illustrated in the left column, again time averaged over the same period. The original data minus the reconstruction is illustrated in the right column. The associated SAT fields are illustrated in the top row, and demonstrate excellent agreement. As illustrated in the lower rows, the same is also true for the other variables in the EOF decomposition, namely the zonal wind, meridional wind, relative humidity, sea-level pressure, and precipitation flux. We additionally assessed the WBGT, which is a non-linear function of the SAT, relatively humidity and sea-level pressure. This diagnostic was calculated from the original data and also the reconstructed fields. It again produces excellent agreement.

In Supplementary Note 2, the temporal properties and higher order statistics are also shown to be well reproduced. FEM-BV-VARX projection tracks the climate model output on the basis of the globally averaged annual SAT in Supplementary Fig. 6a. The probability distribution function and auto-correlation function of the Niño4 index of the reconstruction matches that of the original data, as respectively shown in Supplementary Fig 6b, c. This indicates the dynamics associated with the El Niño Southern Oscillation are adequately represented. Gridded fields of various statistical moments of the SAT are also well reconstructed as illustrated in Supplementary Fig. 7. Pattern correlation between the reconstructed and original field is 1.00 for the time mean, and 0.98 for the standard deviation. Maps of skewness and kurtosis also exhibit very similar spatial patterns.

In the method section thus far, the process using the RCP2.6 and RCP8.5 scenarios has been used to undertake a multi-variate reconstruction of RCP4.5 for one particular climate model. We consider this experiment as an interpolation of the climate in the  $\text{CO}_2e$  concentration space, since the RCP4.5 scenario is bounded by RCP2.6 on the lower side and RCP8.5 on the higher side. Supplementary Note 3 presents summary statistics for how well this reconstruction approach also performs in an extrapolation sense. We additionally attempt to reconstruct either RCP2.6 or RCP8.5, using the other two non-bounding scenarios. We also assess the performance of the interpolated and extrapolated reconstructions for all of climate models simulating these three scenarios. In these tests, only SAT was included in the model reduction phase. Supplementary Fig. 8 shows that after accounting for the variability across the climate models, there is a high

degree of similarity between the reconstruction method, and the original climate data for all scenarios.

In Supplementary Note 4, the sensitivity of the reconstruction performance to the hyper-parameter selection is assessed. Recall the hyper-parameters are the number of clusters ( $K$ ), the memory depth ( $m$ ), and the persistence parameter ( $p$ ). Supplementary Fig. 9 illustrates that the class of models that produced the best agreement between the climate data and the FEM-BV-VARX reconstructions across the CMIP5 ensemble has  $K=2$  clusters,  $m=3$  months and  $p=3$  months. This is the hyper-parameter set used in the body of the manuscript and the methods section. Note, reconstructions made in the case studies use of all three scenarios to generate the arbitrary concentration pathways. As such, one could consider the errors quantified in the validation sections to be an overestimate of the projection errors expected in the body of the manuscript.

### Data availability

All of the CMIP5 climate data adopted within<sup>27</sup> is available from <https://pcmdi.llnl.gov/mips/cmip5/data-portal.html>. A compact pre-curated version of the CMIP5 data is required to generate the figures in the manuscript and supplementary information has been made publicly available<sup>48</sup>.

### Code availability

The code developed to generate the data and produce all of the figures in the manuscript and supplementary information has been made publicly available<sup>49</sup>.

Received: 25 March 2023; Accepted: 18 September 2023;

Published online: 06 October 2023

### References

1. Raworth, K. *Doughnut economics: seven ways to think like a 21st-century economist* (Chelsea Green Publishing, White River Junction, Vermont, 2017).
2. UN-DESA. The sustainable development goals report 2022. Tech. Rep. [https://unstats.un.org/sdgs/report/2022/SDG2022\\_Flipbook\\_final.pdf](https://unstats.un.org/sdgs/report/2022/SDG2022_Flipbook_final.pdf), UN-DESA, New York, USA (2022).
3. Rockström, J. A safe operating space for humanity. *Nature* **461**, 472–475 (2009).
4. IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, United Kingdom and New York, 2013).
5. O’Kane, T. J., Risbey, J., Franzke, C. J. E., Horenko, I. & Monselesan, D. Changes in the metastability of the midlatitude southern hemisphere circulation and the utility of nonstationary cluster analysis and split-flow blocking indices as diagnostic tools. *J. Atmos. Sci.* **70**, 824–842 (2013).
6. Ghil, M. & Lucarini, V. The physics of climate variability and climate change. *Rev. Mod. Phys.* **92**, 1–77 (2020).
7. Cai, Y., Bandara, J. S. & Newth, D. A framework for integrated assessment of food production economics in south asia under climate change. *Environ. Model. Softw.* **75**, 459–497 (2016).
8. Gutierrez, L. Impacts of El Niño Southern Oscillation on the wheat market: a global dynamic analysis. *PLoS ONE* **12**, e0179086 (2017).
9. Ubilava, D. The role of El Niño southern oscillation in commodity price movement and predictability. *Amer. J. Agr. Econ.* **100**, 239–263 (2018).
10. Kitsios, V., De Mello, L. & Matear, R. Forecasting commodity returns by exploiting climate model forecasts of the El Niño Southern Oscillation. *Environ. Data Sci.* **1**, E7 (2022).
11. Porfiri, L., Newth, D., Finnigan, J. & Cai, Y. Economic shifts in agricultural production and trade due to climate change. *Palgrave Commun.* **4**, 1–9 (2018).
12. Allan, C., Xia, J. & Pahl-Wostl, C. Climate change and water security: challenges for adaptive water management. *Curr. Opin. Environ. Sustain.* **5**, 625–632 (2013).
13. Squire, D. T. et al. Unprecedented compound climate extremes and Australia’s 2019/2020 megafires. *J. Clim. Atmosph. Sci.* **4**, 64 (2021).
14. Gibb, R., Franklins, L., Redding, D. & Jones, K. E. Ecosystem perspectives are needed to manage zoonotic risks in a changing climate. *BMJ* **371**, m3389 (2020).

15. Bekkar, B., Pacheco, S., Basu, R. & De Nicola, N. Association of air pollution and heat exposure with preterm birth, low birth weight, and stillbirth in the us: a systematic review. *JAMA Netw Open*. **3**, e208243 (2020).
16. Peters, A. & Schneider, A. Cardiovascular risks of climate change. *Nat. Rev. Cardiol.* **18**, 1–2 (2021).
17. Poon, E., Kitsios, V., Pilcher, D., Bellomo, R. & Raman, J. Projecting future climate impact on national Australian respiratory-related intensive care unit demand, heart lung and circulation. *Heart Lung and Circulation* **32**, 95–104 (2023).
18. Dunne, J. P., Stouffer, R. J. & John, J. G. Reductions in labour capacity from heat stress under climate warming. *Nat. Clim. Change* **3**, 563–566 (2013).
19. Burke, M., Hsiang, S. & Miguel, E. Global non-linear effect of temperature on economic production. *Nature* **527**, 235–239 (2015).
20. Bertram, C. et al. NGFS Climate Scenarios Database: Technical Documentation V2.2. Tech. Rep. [https://www.ngfs.net/sites/default/files/ngfs\\_climate\\_scenarios\\_technical\\_documentation\\_\\_phase2\\_june2021.pdf](https://www.ngfs.net/sites/default/files/ngfs_climate_scenarios_technical_documentation__phase2_june2021.pdf), Network for Greening the Financial System (2021).
21. Whitten, S. et al. Exploring climate risk in Australia the economic implications of a delayed transition to net zero emissions. Tech. Rep., Commonwealth Scientific and Industrial Research Organisation, [https://ecos.csiro.au/wp-content/uploads/2022/02/Technical-Report\\_\\_Exploring-Climate-Risk-in-Australia\\_\\_20220204.pdf](https://ecos.csiro.au/wp-content/uploads/2022/02/Technical-Report__Exploring-Climate-Risk-in-Australia__20220204.pdf) (2022).
22. Nordhaus, W. Geography and macroeconomics: new data and new findings. *Proc. Natl Acad. Sci. USA* **103**, 3510–3517 (2006).
23. Cai, Y., Newth, D., Finnigan, J. & Gunasekera, D. A hybrid energy-economy model for global integrated assessment of climate change, carbon mitigation and energy transformation. *Appl. Energy* **148**, 381–395 (2015).
24. Earth Systems and Climate Change Hub. Scenario analysis of climate-related physical risk for buildings and infrastructure: climate science guidance. Tech. Rep. ESCC Hub Report No.21, National Environmental Science Program (NESP) Earth Systems and Climate Change Science (ESCC) Hub for the Climate Measurement Standards Initiative, <https://www.cmsi.org.au/reports> (2020).
25. Hsiang, S. et al. Estimating economic damage from climate change in the United States. *Science* **356**, 1362–1369 (2017).
26. Dietz, S., Bowen, A., Dixon, C. & Gradwell, P. 'Climate value at risk' of global financial assets. *Nat. Clim. Change* **6**, 676–679 (2016).
27. Taylor, K., Stouffer, R. J. & Meehl, G. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
28. Eyring, V. et al. Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).
29. McGuffie, K. & Henderson-Sellers, A. *A climate modelling primer* (John Wiley and Sons, Oxford, UK, 2005).
30. Meinshausen, M., Raper, S. C. B. & Wigley, T. M. L. Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, magicc6 – part 1: Model description and calibration. *Atmos. Chem. Phys.* **11**, 1417–1456 (2011).
31. Mitchell, T. D. Pattern scaling: an examination of the accuracy of the technique for describing future climates. *Clim. Change* **60**, 217–242 (2003).
32. Nath, S., Lejeune, Q., Beusch, L., Seneviratne, S. I. & Schleussner, C.-F. MESMER-M: an Earth system model emulator for spatially resolved monthly temperature. *Earth Syst. Dyn.* **13**, 851–877 (2022).
33. Horenko, I. On the robust estimatio of low-frequency variability trends in discrete Markovian sequences of atmospheric circulation patterns. *J. Atmos. Sci.* **66**, 1941–1954 (2009).
34. Horenko, I. On the identification of nonstationary factor models and their application to atmospheric data analysis. *J. Atmos. Sci.* **67**, 1559–1574 (2010).
35. O’Kane, T. J., Monselesan, D., Risbey, J., Horenko, I. & Franzke, C. J. E. On memory, dimension, and atmospheric teleconnections. *Math. Clim. Weather Forecast.* **3**, 1–27 (2017).
36. Quinn, C., O’Kane, T. J. & Kitsios, V. Application of local attractor dimension to reduced space strongly coupled data assimilation for chaotic multiscale systems. *Nonlinear Process. Geophys.* **27**, 51–74 (2020).
37. Quinn, C., Harries, D. & O’Kane, T. J. Dynamical analysis of a reduced model for the north Atlantic oscillation. *J. Atmos. Sci.* **78**, 1647–1671 (2021).
38. Horenko, I. Finite element approach to clustering of multidimensional time series. *SIAM J. Sci. Comp.* **32**, 62–83 (2010).
39. Metzner, P., Putzig, L. & Horenko, I. Analysis of persistent nonstationary time series and applications. *Comm. App. Math. Comp. Sci.* **7**, 175–229 (2012).
40. Marchenko, G., Gagliardini, P. & Horenko, I. Towards a computationally tractable maximum entropy principle for nonstationary financial time series. *SIAM J. Financial Math.* **9**, 1249–1285 (2018).
41. Lumley, J. L. *The structure of inhomogeneous turbulence*, 166–178 (Atmosphere Turbulence and Wave Propagation, Moscow: Nauka, 1967).
42. Cordier, L. & Bergmann, M. Proper Orthogonal Decomposition: an overview. In *Lecture series 2008 on post-processing of experimental and numerical data* (Von Karman Institute for Fluid Dynamics, 2008).
43. O’Neill, B. C. et al. The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.* **9**, 3461–3482 (2016).
44. Riahi, K. et al. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Glob. Environ. Change* **42**, 153–168 (2017).
45. Matsuhashi, Y. & Kaya, N. A study on economic measures for CO2 reduction in Japan. *Energy Policy* **21**, 123–132 (1993).
46. Raupach, M. R. et al. Global and regional drivers of accelerating CO2 emissions. *Proc. Natl Acad. Sci. USA* **104**, 10288–10293 (2007).
47. IPCC. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, chap. Technical Summary, 37–70 (IPCC, online, 2019).
48. Kitsios, V. Quickclim demonstration input data. *figshare Dataset*, <https://doi.org/10.6084/m9.figshare.24069741.v2> (2023).
49. Kitsios, V. vassilikitsios/quickclim\_fembvvarx: First release (v1.0.0). *Zenodo*, <https://doi.org/10.5281/zenodo.8307303> (2023).

## Acknowledgements

This research was supported by the Artificial Intelligence for Missions initiative at the Commonwealth Scientific and Industrial Research Organisation as funded by the federal government of Australia.

## Author contributions

V.Kitsios' contributions include: conceptualisation, methodology development, data curation, visualisation, investigation, writing, reviewing, and editing. T.J.O’Kane’s contributions include: conceptualisation, methodology development, reviewing, and editing. D.Newth’s contributions include: conceptualisation, reviewing, and editing.

## Competing interests

The authors declare no competing interests.

## Additional information


**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43247-023-01011-0>.

**Correspondence** and requests for materials should be addressed to Vassili Kitsios.

**Peer review information** *Communications Earth & Environment* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Kyung-Sook Yun and Clare Davis. A peer review file is available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© Crown 2023