JGR | Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2025JH000698

Process-Based Machine Learning Observationally Constrains Future Regional Warming Projections

Sophie Wilkinson[1], **Peer Nowack**[2,3], **and Manoj Joshi**[1]

[1]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK, [2]Institute of Theoretical Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany, [3]Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe Institute of Technology, Karlsruhe, Germany

Key Points:

- We demonstrate a new approach to observationally constrain climate modeling uncertainty in future near-surface temperature projections
- The approach builds on climate-invariant relationships learned to reliably predict daily temperatures from a set of controlling factors
- The results of our observational constraint are in line with previous work suggesting warming in high sensitivity models may be too high

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. Wilkinson,
sophie.wilkinson@uea.ac.uk

Citation:

Wilkinson, S., Nowack, P., & Joshi, M. (2025). Process-based machine learning observationally constrains future regional warming projections. *Journal of Geophysical Research: Machine Learning and Computation*, *2*, e2025JH000698. https://doi.org/10.1029/2025JH000698

Received 20 MAR 2025
Accepted 27 MAY 2025

Abstract We present the results of a novel process-based machine learning method to constrain climate model uncertainty in future regional temperature projections. Ridge-ERA5—a ridge regression model—learns coefficients to represent observed relationships between daily near-surface temperature anomalies and predictor variables from ERA5 reanalysis in Northern Hemisphere land regions. Combining the historically constrained Ridge-ERA5 coefficients with inputs from CMIP6 future projections enables a derivation of observational constraints on regional warming. Although the multi-model mean falls within the constrained range of temperatures in all tested regions, a subset of models which predict the greatest degree of warming tend to be excluded and decomposition of the constraint into predictor variable contributions suggests error-cancellation of feedbacks in some models and regions.

Plain Language Summary Knowledge about future global and regional warming is essential for effective adaptation planning. Future temperature projections are based on the output of global climate models which simulate future warming under a range of possible future emissions pathways. Although climate models agree that there will be warming in the future, there are still significant discrepancies in the degree of warming projected under a given scenario. Here, we make use of a machine learning (ML)-based method which learns relationships from daily observations-based data to predict daily temperature anomalies in climate models. These observations-based relationships are applied to future climate model projections to produce a constrained range of future temperature predictions. For regions across the Northern Hemisphere, we find that, the mean of models broadly falls within the observationally constrained range whilst those few models which predict the most extreme future warming are most likely to be excluded by the constraint.

1. Introduction

Despite improvements in resolution and modeling of physical processes between subsequent generations of CMIP models, constraining uncertainty in future projections of climate change remains a challenge, even in key impact variables such as changes in surface temperature and precipitation. Changes in the uncertainty range of key climate variables between CMIP5 and CMIP6 are mixed with some variables showing little difference between generations of climate models or even a widening of the uncertainty range (Cos et al., 2022; Li et al., 2021; S. Zhang & Chen, 2021), potentially related to more processes being added to models which then contribute additional uncertainties. For instance, the equilibrium climate sensitivity of several CMIP6 models exceeds the range of the previous generation of CMIP5 models (Tokarska et al., 2020). Lack of certainty in end of century warming projections constitutes a major barrier to consensus on climate change adaptation and mitigation policy. Uncertainty in future projections of climate can be divided amongst three primary sources: scenario uncertainty, model uncertainty, and internal variability (Hawkins & Sutton, 2009). Here, we seek to constrain model uncertainty.

One of the simplest ways to assess and address model uncertainty in future climate change projections is to combine information from multiple climate models. Although equally weighted, multi-model means on average improve projection reliability for the expected value of change (Ghafarianzadeh & Monteleoni, 2013; Weigel et al., 2010), there are cases where performance can be further optimized by alternative weighting schemes (Brunner et al., 2019; Giorgi & Mearns, 2002; Knutti et al., 2017; Sanderson et al., 2015a, 2015b).

There are approaches to constraining model uncertainty that are based on detection and attribution methodologies using optimal fingerprinting (Allen & Tett, 1999; Hasselmann, 1997; Hegerl et al., 1997; Kettleborough

et al., 2007; Stott et al., 2000; Stott & Kettleborough, 2002). Bayesian statistical principles are also applied to adjust the probability distribution of future changes on the basis of observational evidence (Douville, 2023; Murphy et al., 2009, 2018; Qasmi & Ribes, 2022; Ribes et al., 2021, 2022; Sexton et al., 2012). An alternative approach to model weighting is the emergent constraint framework. Derived from physically explainable relationships between model simulations of a variable X in the current climate and projections of a different variable Y in the future climate across large ensembles of climate models (e.g., CMIP6), observational constraints on historical variable X are translated into a constraint on future projections of Y (Hall et al., 2019; Hall & Qu, 2006).

Machine learning (ML) methods have also been applied for uncertainty constraints. For example, training ML models using climate model data to learn a relationship between historical and future climate in climate models have been applied to observations to constrain future values of a climate variable (Schlund, Eyring, et al., 2020) or to predict when climate thresholds will be reached (Diffenbaugh & Barnes, 2023; Yu et al., 2022). Here, a novel approach is taken: applying ML algorithms to learn approximately climate-invariant controlling factor relationships to produce constraints on model uncertainty in regional near-surface temperature projections. This has advantages over methods such as emergent constraint frameworks which may not be transferable between subsequent generations of climate models (Nowack & Watson-Parris, 2025; Schlund, Lauer, et al., 2020) as our underlying constraint is derived from observational rather than model relationships. We also find our constraint to be weakest where ML model skill is lowest reducing the risk of overconfident constraints (Sanderson et al., 2021).

Our new observational constraint approach to constrain near-surface temperature projections builds on the idea of controlling factor analyses for climate science. The approach is to formulate regression functions that predict a variable to be constrained (e.g., near-surface temperature) on the basis of predictor variables, or controlling factors, that represent the processes that drive anomalies in this target variable. Such approaches have already been used to constrain uncertainties in specific climate feedback processes such as changes in clouds (Andersen et al., 2023; Ceppi & Nowack, 2021; Klein et al., 2017; Wilson Kemsley et al., 2024) and in stratospheric water vapor (Nowack et al., 2023). In a well-performing controlling factor framework, differences in the regression coefficients (or, sensitivities) when learned from observational data as compared with those learned from GCM output can then be used to constrain responses in the target variables (Nowack & Watson-Parris, 2025). The novelty in our approach comes from the climate invariance property in the relationships learned between controlling factors and the target temperature variable which is achieved in the context of high-dimensionality thanks to the Ridge regression regularization term. This climate invariance allows us to reproduce model consistent temperature predictions under end-of-century warming conditions and to derive an observational constraint by substituting the model consistent responses with responses learned from observations.

2. Methods

The first step is to train the underlying ML model to make accurate physically consistent predictions of the target variable: the daily near-surface temperature anomaly. These ground truth relationships will form the basis of the observational constraint and therefore need to be learned from an observationally based gridded data source. We use the ERA5 (Hersbach et al., 2020) reanalysis product as a proxy for direct observations. While ERA5 is known to be of high quality and we base our initial constraints on this particular reanalysis, we underline that future work could investigate further observational uncertainties, for example, by extending our approach to other reanalyzes as done in (Ceppi & Nowack, 2021; Nowack et al., 2023).

To characterize climate model uncertainty, we use data from the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016) at a consistent temporal and spatial resolution. This climate model data are used to separately train ML models for performance testing of our approach in a perfect model framework and is later combined with ML models trained on reanalysis data to form our observational constraint.

Our analysis will focus on daily June–July–August (JJA) data to train functions for the observational constraint approach. All data are interpolated bilinearly to a common $3° \times 3°$ longitude-latitude grid, which encapsulates the coarsest model resolution of the CMIP6 models used here. For further details on the CMIP6 models and data pre-processing steps, we refer to Supporting Information S1.

Whist the model presented here is trained to predict Northern Hemisphere summer temperature anomalies over land, the methodology in principle may be applied to other geographic regions, seasons, and climate variables by tailoring the selection of an appropriate controlling factor framework, in combination with the corresponding

training data. In this instance, the controlling factors or predictor variables are selected with physical links to the target variable (the daily temperature anomaly) in mind. The model could be used for other seasons, but other variables, for example, snow cover would be needed. We seek to select controlling factors that both increase the predictive skill of the model and aid in interpretability of the models' predictions when these predictions are decomposed into contributions from each predictor variable.

The steps for derivation of the observational constraint using our controlling factor framework to constrain surface temperature projections are outlined in Figure 1.

The algorithm for deriving the observational constraint using our controlling factor framework is:

1. Train Ridge-ERA5 models using historical reanalysis data and separately train Ridge-CMIP6 models for each individual climate model using climate model data (one function per model), then validate performance;
2. Assess *climate-invariance* of relationships represented in Ridge models using a perfect model framework—do the relationships learned by Ridge-CMIP models from historical climate model data only still hold predictive power under strong future climate change scenarios?;
3. Merge observations-based relationships learned by Ridge-ERA5 with future projections to produce observationally calibrated realizations of future temperature projections—the basis of the observational constraint;
4. Incorporate prediction error/uncertainty—error in the relationship between Ridge-CMIP model-consistent predictions and actual CMIP projections is accounted for as to produce the final constraint.

Each of these four steps is now explained in detail, starting with the model training process:

Through a systematic testing of variable combinations and domain sizes at a random sample of Northern Hemisphere land grid cells, a suitable predictor variable setup was determined. The motivation behind the selection of these controlling factor variables is not only a statistical correlation with the target near-surface temperature anomaly and an increase in model skill but also the presence of physical processes in the climate system that can explain how they influence temperature. Here, these controlling factors are selected with the Northern Hemisphere mid-latitudes summer in mind, aiming to capture key information about physical processes associated with heat extremes in these regions. To take advantage of spatial information about each input variable, predictor variables are provided in square $n \times n$ grid cell domains centered on the target grid cell for prediction resulting in n^2 inputs to the model for each predictor variable. This means that each controlling factor provides multiple inputs to the regression model covering the local and surrounding grid cells. The fact that these adjacent input variables are likely to be highly correlated motivates our use of Ridge regression as a method which deals well with co-linearity as well as having the ability to extrapolate (being a linear method), which is crucial if we want to train only on data under 1979–2020 forcing conditions and predict on SSP5-8.5 conditions, that is, well beyond the training climate regime. In our problem formulation, vanilla neural networks would fail in this extreme extrapolation challenge. We show in our paper that the linear relationships learned are indeed approximately climate-invariant, see also detailed discussion in Nowack and Watson-Parris (2025). Additionally, Ridge is a relatively simple, interpretable method that is computationally cheap and demonstrated competitive performance scores against non-linear methods such as Random Forest Regression and Gaussian Process Regression during model development. For further detail on Ridge regression, please refer to the Supporting Information S1. A Ridge-ERA5 model is trained at each Northern Hemisphere land grid cell using historical data from ERA5 during JJA. During training, a 5-fold cross-validation procedure is applied to determine the optimum value of the Ridge regularization parameter. Each daily temperature anomaly prediction $\left(y_{ridge,era5}\right)$ can be expressed as a linear combination of controlling factor anomalies (x_p):

$$y_{ridge,era5} = \theta_0 + \sum_{i=1}^{P} \theta_p x_p. \quad (1)$$

The final controlling factor setup is listed in Table 1 alongside the domain size in grid cells centered on the target location over which the information is provided to the ridge model.

As the relationships learned by the Ridge-ERA5 models will be applied for future projection, we want to validate that these relationships are climate invariant—that they will still hold predictive power under significant climate change (Nowack & Watson-Parris, 2025). We test this climate invariance property using a perfect model framework, turning to climate model data where we have consistent realizations of both historical and future

10.1029/2025JH000698

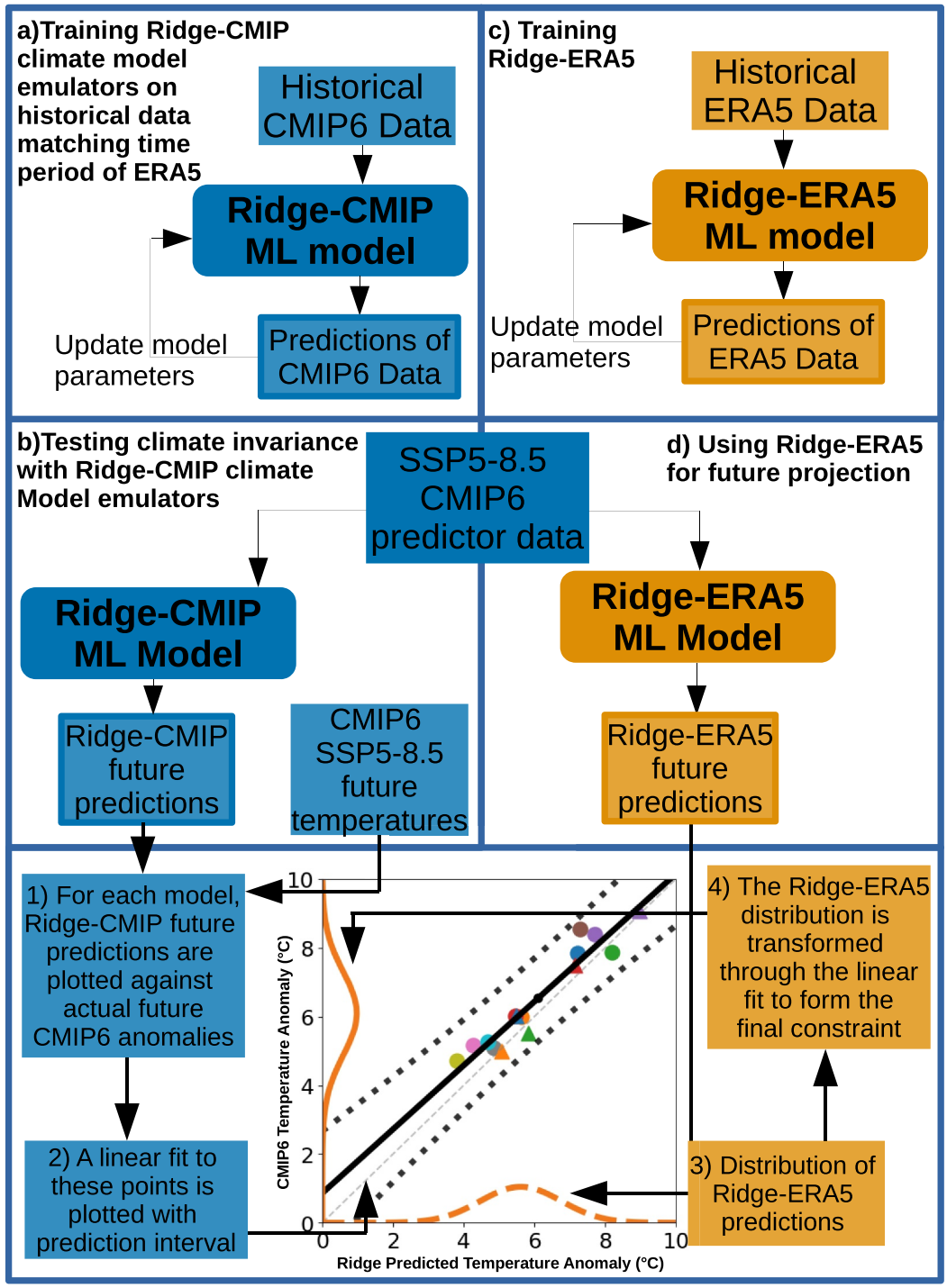

Figure 1. Illustration of the Ridge-ERA5 observational constraint framework. We first train Ridge-CMIP (a) and Ridge-ERA5 (c) models on historical data using the predictor variables identified during the controlling factor analysis. Climate invariance of the learned relationships is validated through a perfect model framework using input data from the most extreme future warming scenario: SSP5-8.5—as input to the functions trained on historical data (matching the time period from ERA5 for the observational function). Ridge-CMIP predicted temperatures are plotted against actual CMIP6 temperature projections (points) alongside a least squares fit to points (black line) & prediction interval (black dashed line) calculated according to Wilks (2006), refer to Supporting Information S1 for further details (b). Climate model inputs from this future warming scenario are then fed into the observations-based Ridge-ERA5 model to produce observationally moderated realizations of the future warming projections (orange dashed line) which form the basis of the final observational constraint (solid orange line).

10.1029/2025JH000698

Table 1
Controlling Factors for Near-Surface Temperature Anomaly

Variable	Domain size
Soil moisture	1 × 1
Near-surface relative humidity	5 × 5
Monthly near-surface relative humidity	5 × 5
Cloud cover fraction	5 × 5
Precipitation rate	5 × 5
u Component of the 850 hPa wind	5 × 5
v Component of the 850 hPa wind	5 × 5
Magnitude of the 850 hPa wind	5 × 5
Monthly 850 hPa specific humidity	5 × 5

climate. Taking data from CMIP6, we train a set of 23 climate model emulators (Ridge-CMIP models) using historical climate model data (1979–2022) to predict daily mean temperature anomalies, shown in Figure 1a. These Ridge-CMIP models use the same controlling factor setup that was described for Ridge-ERA5 but learn from historical climate model data rather than reanalysis. The performance of these models for future prediction under extrapolation conditions (beyond the range of temperatures seen during training) can easily be tested using future scenario data. This testing is performed by providing each Ridge-CMIP emulator trained on historical climate model data with predictor variable inputs from future warming scenario SSP5-8.5 for the same climate model. The resulting Ridge-CMIP future temperature predictions are compared with the actual future projections from the corresponding CMIP6 model, as outlined in Figure 1b. By plotting the Ridge-CMIP prediction versus the CMIP6 projection for each climate model (Figure 3) and comparing with the one-to-one line we can identify any systematic under- or over-prediction by the Ridge-CMIP emulators:

$$y_{cmip} = a + b y_{ridge,cmip}. \tag{2}$$

Thirdly, to place a constraint on the future temperature anomalies projected by existing climate models we want to merge the observations-based coefficients learned by Ridge-ERA5 from historical reanalysis data with future projections of controlling factors from CMIP6. Considering uncertainty in future projections (Δy_{cmip}) to be a function of uncertainty in predictor variables (Δx_{cmip}) and uncertainty in the relationships between them $(\Delta \theta_{cmip})$:

$$\Delta y_{cmip} = f(\Delta x_{cmip}, \Delta \theta_{cmip}). \tag{3}$$

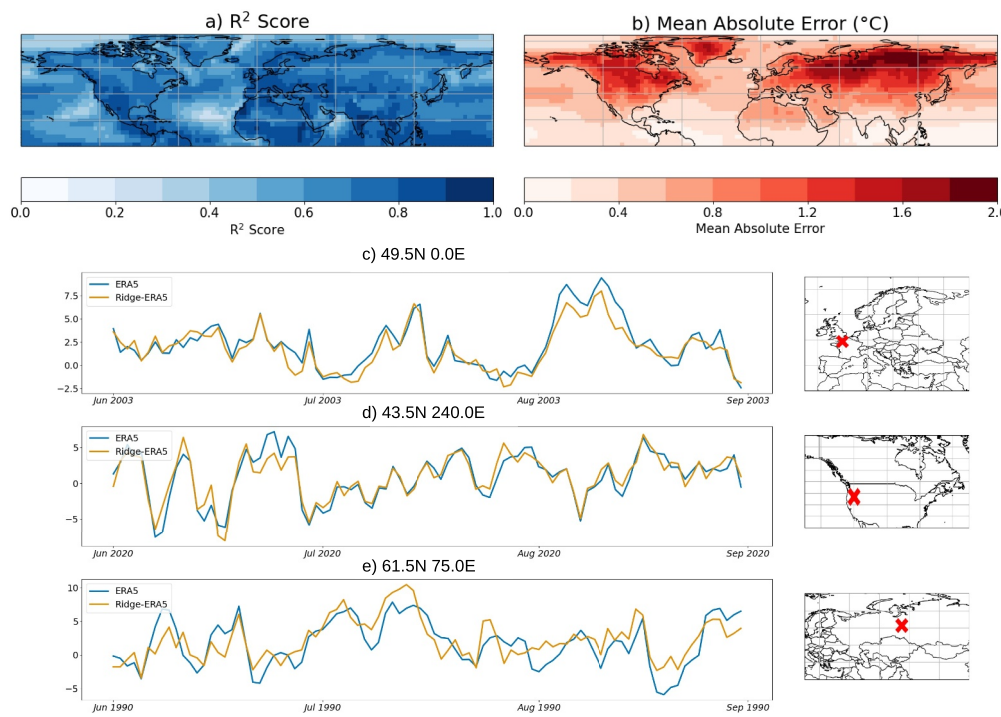

Figure 2. Overview of Ridge-ERA5 performance on held-out test data for daily JJA temperature anomalies. Northern Hemisphere maps of R^2 scores (a) and mean absolute error (b) for test predictions covering the period 1979–2022. Time series of Ridge-ERA5 predictions versus actual ERA5 temperature anomalies at three test locations: 49.5°N 0.0°E (c), 43.5°N 240.0°E (d), and 61.5°N 75.0°E (e) indicated by the red crosses in the inset maps.

We seek to constrain the $\Delta\theta_{cmip}$ by introducing the coefficients learned by Ridge-ERA5 to represent process-based relationships derived from observations (θ_{era5}). We make future predictions using Ridge-ERA5 by combining Ridge-ERA5 coefficients (θ_{era5}) with predictor variables from future scenario SSP5-8.5 in the CMIP6 archive (x_{cmip}) for each of the 23 models. This produces a distribution of observationally moderated predictions for end of century warming on a regional basis ($y_{ridge,era5}$).

As a final step, we additionally convolve this distribution with the prediction uncertainty in our ML framework to obtain the final constrained distribution. This first involves making use of the straight-line relationship between the Ridge-CMIP end-of-century temperature predictions and actual CMIP6 projections (refer to Equation 2). The distribution of observationally moderated predictions, $y_{ridge,era5}$, is transformed through this line:

$$\mathbf{z} = a + by_{ridge,era5}. \tag{4}$$

To fit a distribution to the translated temperatures, $\mathbf{z}$, 10^7 equidistant points are sampled from a normal distribution fitted to $\mathbf{z}$ using the prediction interval (Wilks (2006), refer to Supporting Information S1 for further detail) as the standard deviation. Finally, a Gaussian kernel density estimator with bandwidth 0.01 is fitted to these points to obtain the final distribution of constrained values.

Following calculation of the Ridge-ERA5 observational constraint we apply Shapley Additive ExPlanation (SHAP) value analysis (Lundberg & Lee, 2017; Shapley, 1953), to identify the source of the constraint in each region. SHAP values quantify the contribution of each input variable to a given model prediction, allowing a temperature prediction on a given day to be decomposed into contributions from each of the controlling factors, for example, specific humidity anomalies. SHAP values are additive, meaning that the sum of SHAP values for all variables adds up to the difference between the prediction of the model and a baseline value. In the case of linear models like Ridge, the calculation is simplified by the form of the model function with the baseline SHAP value equating to the linear regression intercept, whilst the SHAP value of a given variable, ϕ_j, is simply the product of the corresponding coefficient, θ_j, and the observed value for that case, x_j. Explainable AI methods are sensitive to the baseline relative to which the roles of features are quantified (Mamalakis et al., 2023). Since our controlling factor framework focuses on explaining and modeling temperature anomalies relative to the climatological mean, we consider the intercept of the ridge regression model as a straightforward baseline and aim to use SHAP values to explain the contributions of predictor variables relative to this expected value.

3. Predictive Skill of the Model

We first demonstrate the predictive skill of the Ridge-ERA5 ML model on held-out historical reanalysis data. Corresponding Ridge-CMIP emulators trained on historical climate model data are then applied to climate model projections to test climate invariance using model-consistent changes in the controlling factors, and then by extension to produce an observations-based constraint on future temperature anomalies as described in Section 2. Finally, the results of this constraint are broken down into SHAP contributions from individual predictor variables to identify the sources of the constraint in terms of controlling factors.

A test prediction time series covering the entire time period of the ERA5 data (1979–2022) is achieved by taking a *leave-one-year-out* approach that is rotating the year that is held-out for testing. So, for each year of historical data (e.g., 2000), we train a Ridge-ERA5 model using the remaining years of data (1979–1999 and 2001–2019), including a 5-fold, time-ordered cross-validation split to determine hyperparameter values. By repeating this process for each year of historical data, we produce a time series of Ridge-ERA5 test predictions (on data unseen during training) which covers the entire historical period 1979–2022. Figure S1 in Supporting Information S1 compares test R^2 scores from this leave-one-year-out approach with training, cross-validation and test R^2 scores from separate models which use distinct time periods for training (1979–2005), cross-validation (2006–2015), and testing (2016–2022). These results indicate that overfitting or high-bias issues are confined primarily to ocean regions (which are not the focus of our analysis) and performance scores are broadly comparable between the test data (2016–2022) and test scores from the leave-one-year-out approach over land. In addition, we note that there are at least 9 months gap between the last month of 1 year to JJA in the next year. Ultimately, the strongest out-of-sample test is the evaluation under the final decades of SSP5-8.5, which is not only distant in time but also poses an extrapolation challenge with respect to climate change for the CMIP model-consistent functions, providing a difficult test case for our methodology.

10.1029/2025JH000698

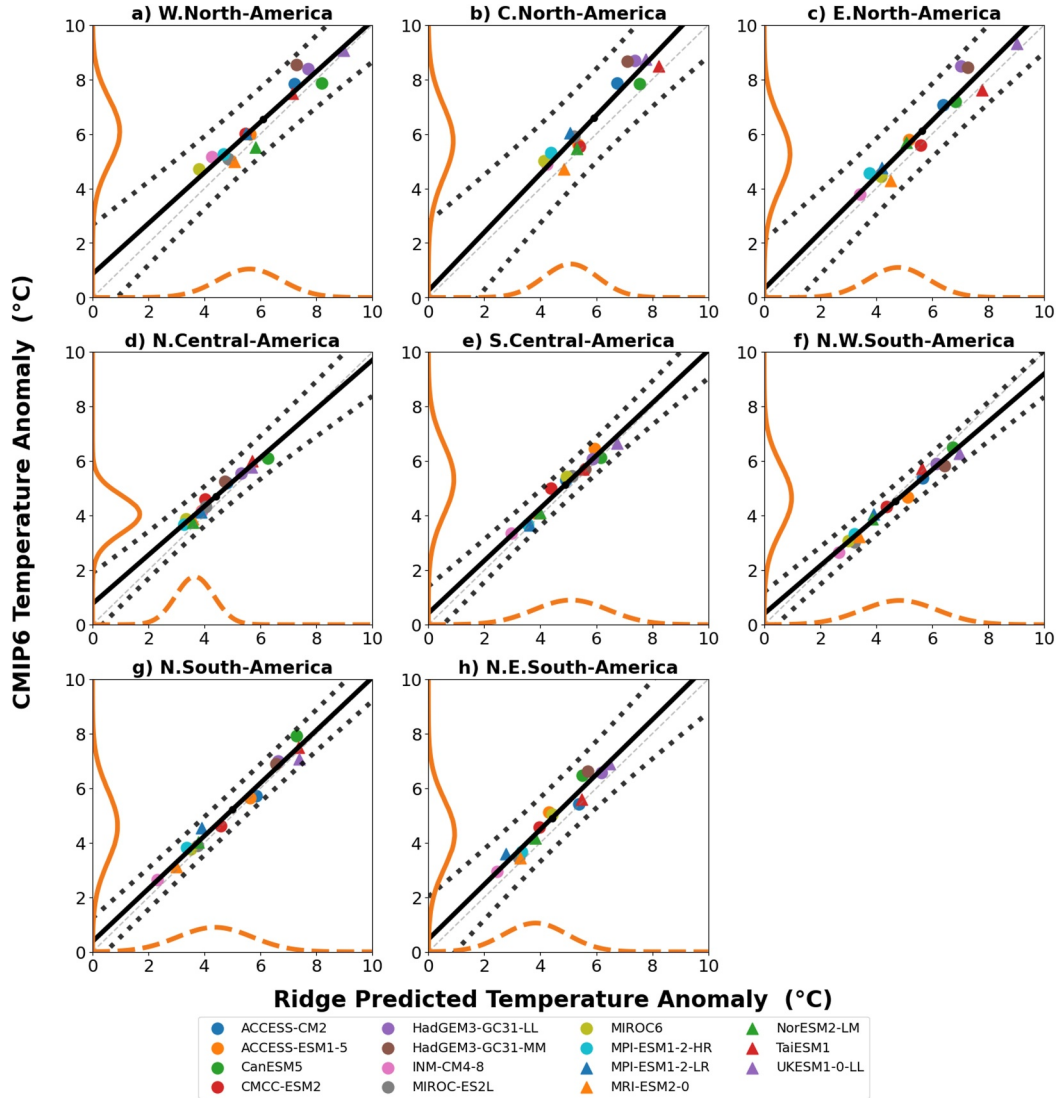

Figure 3. Averaged on a regional basis over land grid-cells in Northern Hemisphere AR6 regions in North, Central, and South America: Projected mean daily temperature change (by 2070–2100) under *SSP5-8.5* during JJA from CMIP6 (*y*-axis) plotted against predicted change by Ridge-CMIP (*x*-axis) with least squares fit to points (black line) & prediction interval calculated according to (Wilks, 2006), refer to Supporting Information S1 for further details (black dashed line) for different AR6 regions. Probability distributions (orange lines) for Ridge-ERA5 future predictions given *SSP5-8.5* inputs (*x*-axis) convolved with Ridge prediction error for final constraint (*y*-axis).

Overall, the Ridge-ERA5 framework achieves very good predictive performance for near-surface temperature anomalies across the Northern Hemisphere mid-latitudes. A selection of performance measures for Ridge-ERA5 predictions on held-out ERA5 test data (data not seen during training) for JJA temperature anomalies are shown in Figure 2. R^2 scores (a measure of explained variance with optimum value 1 and no limit to worst performance) and mean absolute error are mapped in Figures 2a and 2b, respectively. Similar patterns of performance are visible across both metrics with the best performance scores measured over land and at mid-latitudes. R^2 scores for the seasonal anomaly predictions indicate high performance with values >0.7 across most land regions. Performance is unsurprisingly poorest over Northern Siberia, where temperature variance is high and the controlling factor setup does not include factors such as snow cover and related feedbacks which play a role in heat extremes in these high-latitude regions (Marquardt Collow et al., 2022; Sato & Nakamura, 2019; R. Zhang et al., 2020).

Figures 2c–2d show time series of predictions from Ridge-ERA5 for 1 year of the historical time period at three test locations. Here, the Ridge-ERA5 prediction time series shows excellent day-to-day agreement with the true

ERA5 time series (Figures 2c and 2d). The magnitude of most anomalies is accurately captured as well as the patterns of daily variation; even during exceptionally warm years such as the 2003 European heatwave. In areas where Ridge-ERA5 struggles, such as in Northern Siberia, there are instances both of correctly modeled patterns but also incorrect magnitudes and false patterns of variation (Figure 2e). Such high-latitude regions (Greenland/Iceland, Russian Arctic, Western Siberia, Eastern Siberia, Russian Far East, North-Western North America, and North-Eastern North America) are excluded from the observational constraint in the present study and we focus on Northern Hemisphere mid-latitude land regions given our choice of controlling factors.

4. The Observational Constraint

Having evaluated that the controlling factor functions can skillfully predict historical temperature variations, we now evaluate their climate-invariance in a perfect model framework by projecting model-consistent temperature responses. We focus on the mean temperature change by 2070–2100 compared with a historical period of 1979–2022 averaged over land grid cells in Northern Hemisphere AR6 regions. For a map and more detailed description of these regions, we refer to the Supporting Information S1. Ridge-CMIP predictions of this quantity (produced by feeding each Ridge-CMIP emulator with projections of the controlling factors from the corresponding CMIP model) for the 8 land AR6 regions in North, Central, and South America are plotted against the actual future projections from *SSP5-8.5* simulations for each model in Figure 3 (results for the remaining AR6 regions in the Northern Hemisphere mid-latitudes are provided in Figures S3 and S4 in Supporting Information S1). A closer match to the one-to-one line indicates greater climate-invariance in the relationships learned from historical data only.

In general, goodness of fit for these future projections corresponds to performance on historical test data. Where historical performance was worse, so is future projection performance, likely due to the missing processes relevant to temperature anomalies at higher altitudes and latitudes discussed in Section 3. Generally, Ridge-CMIP predictions show excellent agreement with corresponding CMIP6 temperature anomalies even under the most extreme future warming scenario where temperatures far exceed the historically observed range. Western North-America and Eastern North-America have the highest maximum projected future warming across North, Central, and South America with some CMIP6 models projecting an increase of more than $9°C$. Ridge-CMIP emulator models trained on historical climate model data are still able to make accurate predictions under these extrapolation conditions which can be seen by comparing the least squares fit to points (black line) with the one-to-one line (gray dashed line). In other regions where the temperature change is less extreme for example, Northern Central-America, Southern Central-America, and Northern South-America, performance is better, as expected when the model is not required to extrapolate as far beyond the range of temperatures seen during training.

The Ridge-ERA5 coefficients are then combined with model projections of the controlling factors from the future projection to produce observationally moderated future temperature projections (dashed orange curve in Figure 3). Finally, the above distributions are convolved with the prediction error between CMIP6 and Ridge-CMIP future temperatures to give the final constrained distribution accounting for potential systematic errors (solid orange curve in Figure 3) in the Ridge future predictions. We summarize the observational constraint for the various AR6 regions in Figure 4.

Whilst the CMIP6 ensemble, particularly the multi-model mean, in many cases agrees with our observational constraint, in several regions there is a substantial downward correction and/or narrowing of the distribution. Additionally, in those regions where the CMIP6 ensemble appears to be in alignment with our observational constraint, this is sometimes the result of compensating errors in the controlling factors which will be explored in more detail in the following section. Here, we highlight three tendencies of the observational constraint which arise in different regions: (a) a significant downshift in the range of projected temperatures; (b) a moderate downshift in the range of projected temperatures which excludes models which project the greatest degree of warming; (c) little change to the range of temperatures projected by CMIP6 models. For a version of Figure 4 which allows individual CMIP6 models to be identified and compared with the observational constraint, refer to Figures S5 and S6 in Supporting Information S1.

The regions where the greatest downshift occurs, that is, Mediterranean, West Central Asia, the Arabian Peninsula, the Sahara, East Central Asia, and North Central America (Figures 4a–4e), all tend to be hotter and drier climatic regions. In these regions, the CMIP6 multi-model means fall outside or at the upper boundary of the central 66% uncertainty range of the observational constraint indicating that models in these regions tend to

10.1029/2025JH000698

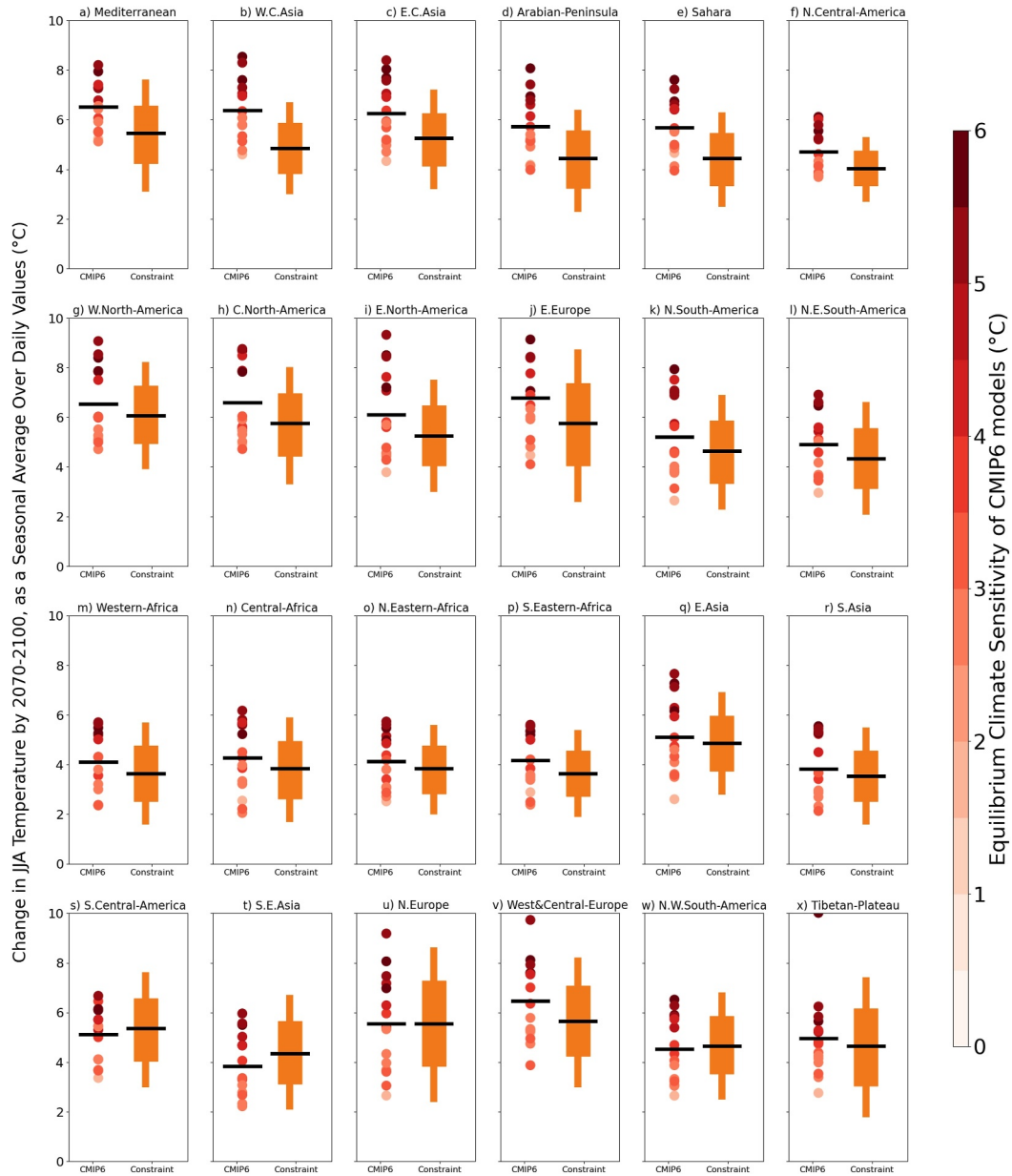

Figure 4. Mean temperature change during JJA projected by each CMIP6 model (red circles shaded according to model ECS) and the CMIP6 ensemble mean (black line) alongside observationally constrained mean (black line), 17th–83rd percentile (wide orange bar) and 5th–95th percentile (thin orange bar) intervals for each Northern Hemisphere AR6 region. In some regions, the observational constraint produces a significant downward correction (a–f), in other regions the downward correction is more moderate and excludes only the warmest CMIP6 models (g–r), in other regions the constraint broadly aligns with the CMIP6 ensemble (s–x).

project a greater degree of warming than is compatible with our observational constraint. The fact that these regions all share a similar climate indicates that particular processes associated with warm anomalies in these regions might be poorly simulated in GCMs, which we discuss in the next section.

In most other regions (Figures 4g–4r), the CMIP6 ensemble mean falls within the central 66% uncertainty range of the observational constraint although there is always a small downward correction. Additionally, the 2–3 models which project the greatest degree of warming in each region are almost always excluded by the observational constraint. The models excluded frequently include CanESM5, HadGEM3-GC31-LL/MM, and UKESM1-0-LL —some of the highest sensitivity models in the CMIP6 archive. Our observational constraint suggests that the

level of warming projected by these models is not compatible with observational evidence across much of the Northern Hemisphere mid-latitudes.

Regions where the constraint is widest relative to the CMIP6 range and results in little to no shift in the multi-model mean tend to correspond to mountainous regions, for example,: the Tibetan Plateau or North Western South America (Figures 4s–4x). Poor performance in these regions may be linked to the representation of topography with mountains in each region being represented differently in each CMIP model and in ERA5, effectively resulting in more noise when applying the Ridge-ERA5 functions to CMIP6 inputs. Additionally, some of these regions are also associated with poorer historical performance on test data (e.g., Northern Europe, West & Central Europe, Tibetan Plateau). The fact that this approach results in a wider uncertainty range in the regions of poorer performance is itself a positive feature of the method rather than obtaining an over-confident but narrower constraint.

5. Physical Drivers of Responses

We now attempt to identify the potential physical drivers behind our results by conducting a SHAP value analysis of the contributions from individual controlling factors. First, we consider the regions where we find the most significant constraints on the CMIP temperature range, that is, those in Figures 4a–4e. In these regions, we find that the controlling factor responses for near-surface relative humidity and the monthly means of near-surface relative humidity and 850 hPa specific humidity are weaker in the Ridge-ERA5 coefficients than the CMIP6 coefficients leading to the downward correction of the temperature range. This is demonstrated for the Mediterranean region in Figure 5a.

Second, for other regions where there is a moderate downward correction of the ensemble mean and the constraint excludes the CMIP6 models which project the most warming (Figures 4g–4r) the constraint also appears to be driven primarily by differences in the monthly humidity responses, see for example, the SHAP value decomposition for Western Africa in Figure 5b. We found that the inclusion of these monthly mean humidity variables was essential to enable Ridge-ERA5 and the Ridge-CMIP emulators to simulate the degree of the future warming. As monthly mean values, these variables capture longer term information about the state of the local climate, providing information about the background warming state on top of which other effects such as soil moisture feedbacks or dynamical variables influence temperature anomalies.

In some regions where the constraint results in the small downward correction described above, we also see an opposite effect in the soil moisture response. However, this effect is smaller in magnitude compared with the humidity responses, hence the overall downshifting of the uncertainty range. In Europe and North America, the soil moisture response in the climate models appears to be too weak while in Ridge-ERA5 the same soil moisture anomaly results in a greater degree of warming. This is demonstrated in Figure 5c where, for Eastern North America, the SHAP value difference is positive for soil moisture contributions.

In these regions, the shift in the median implies that model uncertainty in the current controlling factor set-up might be encoded in the controlling factor responses (which are not constrained), rather than in the relationship between the controlling factor and target variable response (which we do constrain). Consequently, future work could focus on defining new controlling factor set-ups that better redistribute uncertainties away from the raw controlling factor responses, in particular by finding alternative proxies for processes contained in the specific humidity predictors. Nevertheless, even in cases where model uncertainty is not constrained by our approach, our decomposition of the constraint into contributions from individual predictor variables indicates that the soil moisture feedback in some CMIP6 models may be underestimated and that compensating biases in underlying processes contribute to temperature projections in some regions.

In the regions where the observational constraint appears to be in alignment with the CMIP range of responses (Figures 4g–4r), this is sometimes the result of bias in one variable being offset by an opposite bias in another variable. For example, for the Northern Europe region, the constraint (range and median value) is broadly in line with the CMIP range (Figure 4u) but an analysis of the controlling factors in Figure 5d indicates this may be down to a compensation of monthly mean specific humidity by other factors such as the role of soil moisture feedbacks. While the soil moisture feedback response appears to be too weak across the CMIP6 models in Northern Europe, this effect is compensated for by a monthly mean specific humidity response which is too strong. For SHAP value decompositions in the remaining regions, refer to Figures S7–S10 in Supporting Information S1.

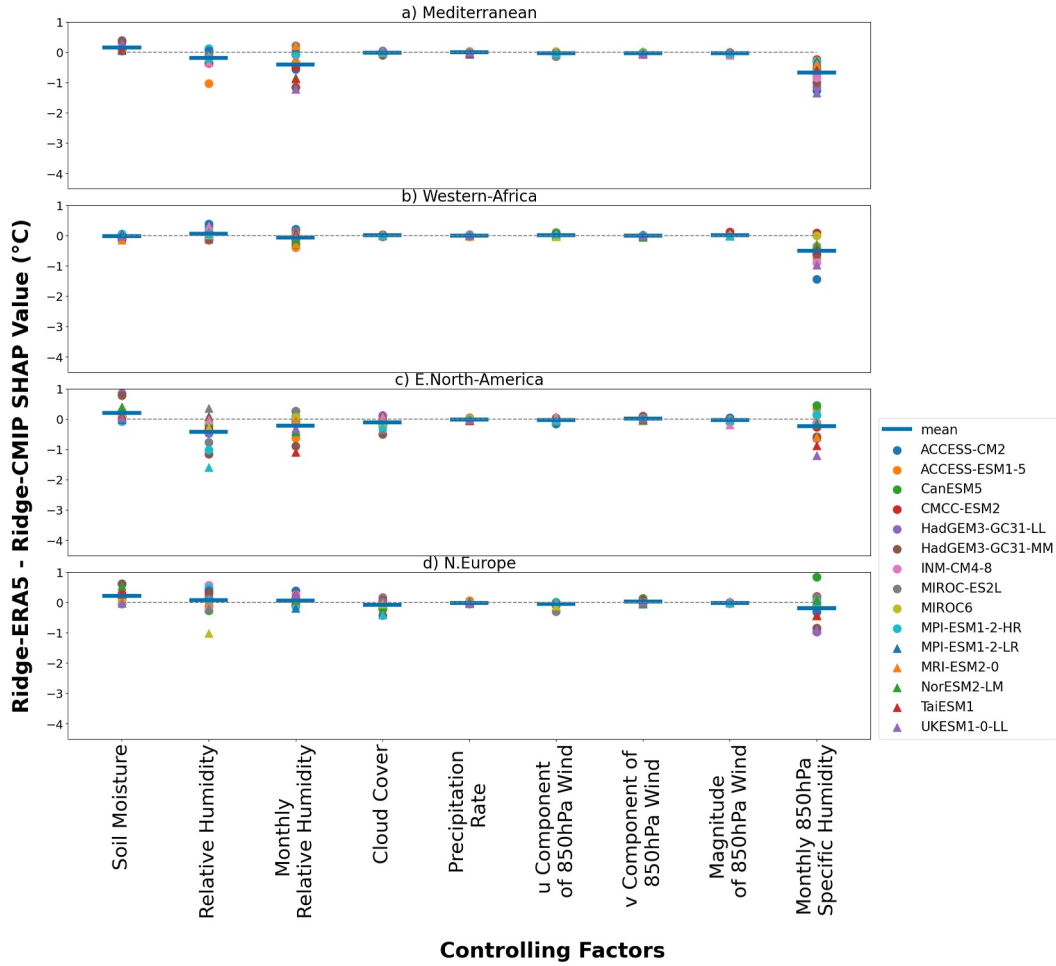

Figure 5. Average difference in SHAP values (Ridge-ERA5 minus Ridge-CMIP) for end-of-century temperature predictions for each Ridge predictor variable over the Mediterranean (a), Western Africa (b), Eastern North America (c), and Northern Europe (d). A positive (negative) value of $N°C$ for a particular predictor variable indicates that the variable contributes an average $N°C$ of warming (cooling) more in Ridge-ERA5 than the same predictor variable in Ridge-CMIP.

While the SHAP value decomposition of the constraint can provide some insight into the key variables involved, care must still be taken in the interpretation of this analysis. The relationships, which Ridge-ERA5 represents between predictor and target variables, are not necessarily one-way or causal (Nowack et al., 2020). Whilst these relationships have been demonstrated to hold strong predictive power, in a complex, inter-connected climate system many of the predictor variables both influence and are influenced by the target variable—temperature. That being said, SHAP value analysis is still a powerful tool in this context for highlighting controlling factors and associated processes for further investigation.

6. Conclusions

We have introduced a first controlling factor analysis for observational constraints on regional near-surface temperature projections. Ridge-ERA5 is able to make accurate, out-of-sample test predictions for daily temperature anomalies to the mean seasonal cycle based on input information from a range of climatic predictor variables at most Northern Hemisphere land grid cell locations. Climate invariance of the Ridge coefficients is demonstrated by predictions from Ridge-CMIP emulators (which learn from historical climate model data) comparing very well with end of century temperature anomalies in corresponding CMIP6 climate projections.

The range of the constraint produced by the controlling factor analysis can exceed the range of CMIP6 temperature projections because the constraint is based upon relationships learned from observational data. Thus the

constraint represents a physical, process-based constraint, without many problems of other statistical approaches, such as emergent constraint frameworks, where indirect links between historical changes and future projections may sometimes result from coincidental correlations in the data rather than representing physically based out-of-sample predictive connections in the climate system.

A key finding of the Ridge-ERA5 observational constraint is the exclusion, across Northern Hemisphere land, of those models which project the greatest degree of warming on a regional basis by the end of the century. There has been much discussion surrounding the increased equilibrium climate sensitivity between the previous generation of climate models in the CMIP5 ensemble and those in CMIP6 (Meehl et al., 2020). Our results indicate that some models may have warmed too much by the 2020s, and are incompatible with observed relationships between temperature and other driving factors in our framework, and are consistent with a growing body of evidence suggesting that the sensitivity of some CMIP6 models may not be compatible with the observed historical climate (Brunner et al., 2020; Jiménez-de-la Cuesta & Mauritsen, 2019; Nijsse et al., 2020; O'Reilly et al., 2024; Tokarska et al., 2020; Zhu et al., 2020).

Results have been demonstrated here for Northern Hemisphere land areas and for summertime temperature anomalies. This methodology could be applied in other regions, seasons, and to other variables by tailoring the controlling factors selected to describe local physical drivers of the climate variable of interest. Performance of the methodology was demonstrated on the most extreme future warming scenario in CMIP6, SSP5-8.5, as this represented the most extreme out-of-sample extrapolation test for the Ridge-CMIP6 models. The same approach could be applied to other less extreme scenarios where we would expect models with an ECS that is incompatible with our observational constraint to continue to be excluded.

Data Availability Statement

All reanalysis and CMIP6 data used in this study are publicly available. A specific list of data sets and variables accessed, data pre-processing steps, and Python code for analysis are archived in Zenodo and available at: https://doi.org/10.5281/zenodo.15115895 (Wilkinson, 2025).

Acknowledgments

S.W. was supported by a University of East Anglia Faculty of Science PhD studentship. P.N. and M.J. were supported by the UK Natural Environment Research Council (Grant NE/V012045/1). We acknowledge the WCRP, which, through its Working Group on Coupled Modeling, coordinated and promoted CMIP6. We thank the climate-modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access and the multiple funding agencies that support CMIP6 and ESGF. We further used the JASMIN postprocessing system (Lawrence et al., 2013) operated by the Science and Technology Facilities Council on behalf of the UK Natural Environment Research Council. The analysis presented in this paper was mostly carried out on the High Performance Computing Cluster supported by the Research and Specialist Computing Support service at the University of East Anglia.

References

Allen, M. R., & Tett, S. F. (1999). Checking for model consistency in optimal fingerprinting. *Climate Dynamics*, *15*(6), 419–434. https://doi.org/10.1007/S003820050291

Andersen, H., Cermak, J., Douglas, A., Myers, T. A., Nowack, P., Stier, P., et al. (2023). Sensitivities of cloud radiative effects to large-scale meteorology and aerosols from global observations. *Atmospheric Chemistry and Physics*, *23*(18), 10775–10794. https://doi.org/10.5194/ACP-23-10775-2023

Brunner, L., Lorenz, R., Zumwald, M., & Knutti, R. (2019). Quantifying uncertainty in European climate projections using combined performance-independence weighting. *Environmental Research Letters*, *14*(12), 124010. https://doi.org/10.1088/1748-9326/AB492F

Brunner, L., Pendergrass, A. G., Lehner, F., Merrifield, A. L., Lorenz, R., & Knutti, R. (2020). Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics*, *11*(4), 995–1012. https://doi.org/10.5194/ESD-11-995-2020

Ceppi, P., & Nowack, P. (2021). Observational evidence that cloud feedback amplifies global warming. *PNAS*, *118*(30). https://doi.org/10.1073/pnas.2026290118

Cos, J., Doblas-Reyes, F., Jury, M., Marcos, R., Bretonnière, P. A., & Samsó, M. (2022). The Mediterranean climate change hotspot in the CMIP5 and CMIP6 projections. *Earth System Dynamics*, *13*(1), 321–340. https://doi.org/10.5194/ESD-13-321-2022

Diffenbaugh, N. S., & Barnes, E. A. (2023). Data-driven predictions of the time remaining until critical global warming thresholds are reached. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(6). https://doi.org/10.1073/PNAS.2207183120

Douville, H. (2023). Robust and perfectible constraints on human-induced Arctic amplification. *Communications Earth & Environment*, *4*(1), 1–9. https://doi.org/10.1038/s43247-023-00949-5

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/GMD-9-1937-2016

Ghafarianzadeh, M., & Monteleoni, C. (2013). Climate prediction via matrix completion. *AAAI Workshop - Technical Report, WS-13–17*, 35–37.

Giorgi, F., & Mearns, L. O. (2002). Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "reliability ensemble averaging" (REA) method. *Journal of Climate*, *15*(10), 1141–1158. https://doi.org/10.1175/1520-0442(2002)015⟨1141:COAURA⟩2.0.CO

Hall, A., Cox, P., Huntingford, C., & Klein, S. (2019). Progressing emergent constraints on future climate change. *Nature Climate Change*, *9*(4), 269–278. https://doi.org/10.1038/s41558-019-0436-6

Hall, A., & Qu, X. (2006). Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophysical Research Letters*, *33*(3), 3502. https://doi.org/10.1029/2005GL025127

Hasselmann, K. (1997). Multi-pattern fingerprint method for detection and attribution of climate change. *Climate Dynamics*, *13*(9), 601–611. https://doi.org/10.1007/s003820050185

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1107. https://doi.org/10.1175/2009BAMS2607.1

Hegerl, G. C., Hasselmann, K., Cubasch, U., Mitchell, J. F., Roeckner, E., Voss, R., & Waszkewitz, J. (1997). Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Climate Dynamics*, *13*(9), 613–634. https://doi.org/10.1007/S003820050186/

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/QJ.3803

Jiménez-de-la Cuesta, D., & Mauritsen, T. (2019). Emergent constraints on Earth's transient and equilibrium response to doubled CO2 from post-1970s global warming. *Nature Geoscience*, *12*(11), 902–905. https://doi.org/10.1038/s41561-019-0463-y

Kettleborough, J. A., Booth, B. B., Stott, P. A., & Allen, M. R. (2007). Estimates of uncertainty in predictions of global mean surface temperature. *Journal of Climate*, *20*(5), 843–855. https://doi.org/10.1175/JCLI4012.1

Klein, S. A., Hall, A., Norris, J. R., & Pincus, R. (2017). Low-cloud feedbacks from cloud-controlling factors: A review. *Surveys in Geophysics*, *38*(6), 1307–1329. https://doi.org/10.1007/S10712-017-9433-3

Knutti, R., Rugenstein, M. A., & Hegerl, G. C. (2017). Beyond equilibrium climate sensitivity. *Nature Geoscience*, *10*(10), 727–736. https://doi.org/10.1038/ngeo3017

Lawrence, B. N., Bennett, V. L., Churchill, J., Juckes, M., Kershaw, P., Pascoe, S., et al. (2013). Storing and manipulating environmental big data with JASMIN. *Proceedings - 2013 IEEE International Conference on Big Data, Big Data 2013*, 68–75. https://doi.org/10.1109/BIGDATA.2013.6691556

Li, J., Huo, R., Chen, H., Zhao, Y., & Zhao, T. (2021). Comparative assessment and future prediction using CMIP6 and CMIP5 for annual precipitation and extreme precipitation simulation. *Frontiers in Earth Science*, *9*, 687976. https://doi.org/10.3389/FEART.2021.687976

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4768–4777.

Mamalakis, A., Barnes, E. A., & Hurrell, J. W. (2023). Using explainable artificial intelligence to quantify "climate distinguishability" after stratospheric aerosol injection. *Geophysical Research Letters*, *50*(20), e2023GL106137. https://doi.org/10.1029/2023GL106137

Marquardt Collow, A. B., Thomas, N. P., Bosilovich, M. G., Lim, Y. K., Schubert, S. D., & Koster, R. D. (2022). Seasonal variability in the mechanisms behind the 2020 siberian heatwaves. *Journal of Climate*, *35*(10), 3075–3090. https://doi.org/10.1175/JCLI-D-21-0432.1

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J. F., Stouffer, R. J., et al. (2020). Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models. *Science Advances*, *6*(26). https://doi.org/10.1126/SCIADV.ABA1981

Murphy, J., Harris, G., Sexton, D., Kendon, E., Bett, P., Clark, R., et al. (2018). UKCP18 land projections: Science report. *Met Office Report*.

Murphy, J., Sexton, D., Jenkins, G., & Booth, B. (2009). *UK climate projections science report: Climate change projections*. Met Office Hadley Centre.

Nijsse, F. J., Cox, P. M., & Williamson, M. S. (2020). Emergent constraints on transient climate response (TCR) and equilibrium climate sensitivity (ECS) from historical warming in CMIP5 and CMIP6 models. *Earth System Dynamics*, *11*(3), 737–750. https://doi.org/10.5194/ESD-11-737-2020

Nowack, P., Ceppi, P., Davis, S. M., Chiodo, G., Ball, W., Diallo, M. A., et al. (2023). Response of stratospheric water vapour to warming constrained by satellite observations. *Nature Geoscience*, *16*(7), 577–583. https://doi.org/10.1038/s41561-023-01183-6

Nowack, P., Runge, J., Eyring, V., & Haigh, J. D. (2020). Causal networks for climate model evaluation and constrained projections. *Nature Communications*, *11*(1), 1415. https://doi.org/10.1038/s41467-020-15195-y

Nowack, P., & Watson-Parris, D. (2025). Opinion: Why all emergent constraints are wrong but some are useful – A machine learning perspective. *Atmospheric Chemistry and Physics*, *25*(4), 2365–2384. https://doi.org/10.5194/ACP-25-2365-2025

O'Reilly, C. H., Brunner, L., Qasmi, S., Nogherotto, R., Ballinger, A. P., Booth, B., et al. (2024). Assessing observational constraints on future European climate in an out-of-sample framework. *Climate and Atmospheric Science*, *7*(1), 1–14. https://doi.org/10.1038/s41612-024-00648-8

Qasmi, S., & Ribes, A. (2022). Reducing uncertainty in local temperature projections. *Science Advances*, *8*(41), 6872. https://doi.org/10.1126/SCIADV.ABO6872

Ribes, A., Boé, J., Qasmi, S., Dubuisson, B., Douville, H., & Terray, L. (2022). An updated assessment of past and future warming over France based on a regional observational constraint. *Earth System Dynamics*, *13*(4), 1397–1415. https://doi.org/10.5194/ESD-13-1397-2022

Ribes, A., Qasmi, S., & Gillett, N. P. (2021). Making climate projections conditional on historical observations. *Science Advances*, *7*(4), 671–693. https://doi.org/10.1126/SCIADV.ABC0671

Sanderson, B. M., Knutti, R., & Caldwell, P. (2015a). Addressing interdependency in a multimodel ensemble by interpolation of model properties. *Journal of Climate*, *28*(13), 5150–5170. https://doi.org/10.1175/JCLI-D-14-00361.1

Sanderson, B. M., Knutti, R., & Caldwell, P. (2015b). A representative democracy to reduce interdependency in a multimodel ensemble. *Journal of Climate*, *28*(13), 5171–5194. https://doi.org/10.1175/JCLI-D-14-00362.1

Sanderson, B. M., Pendergrass, A. G., Koven, C. D., Brient, F., Booth, B. B., Fisher, R. A., & Knutti, R. (2021). The potential for structural errors in emergent constraints. *Earth System Dynamics*, *12*(3), 899–918. https://doi.org/10.5194/ESD-12-899-2021

Sato, T., & Nakamura, T. (2019). Intensification of hot Eurasian summers by climate change and land–atmosphere interactions. *Scientific Reports*, *9*(1), 1–8. https://doi.org/10.1038/s41598-019-47291-5

Schlund, M., Eyring, V., Camps-Valls, G., Friedlingstein, P., Gentine, P., & Reichstein, M. (2020). Constraining uncertainty in projected gross primary production with machine learning. *Journal of Geophysical Research: Biogeosciences*, *125*(11), e2019JG005619. https://doi.org/10.1029/2019JG005619

Sexton, D. M., Murphy, J. M., Collins, M., & C, M. J. (2012). Multivariate probabilistic projections using imperfect climate models part I: Outline of methodology. *Climate Dynamics*, *38*(11–12), 2513–2542. https://doi.org/10.1007/S00382-011-1208-9

Shapley, L. S. (1953). A value for n-person games. In H. Kuhn & A. Tucker (Eds.), *Contributions to the theory of games (am-28)* (Vol. 2, pp. 307–318). Princeton University Press. https://doi.org/10.1515/9781400881970-018

Stott, P. A., & Kettleborough, J. A. (2002). Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature*, *416*(6882), 723–726. https://doi.org/10.1038/416723a

Stott, P. A., Tett, S. F., Jones, G. S., Allen, M. R., Mitchell, J. F., & Jenkins, G. J. (2000). External control of 20th century temperature by natural and anthropogenic forcings. *Science*, *290*(5499), 2133–2137. https://doi.org/10.1126/science.290.5499.2133

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in CMIP6 models. *Tech. Rep.*, *6*(12). https://doi.org/10.1126/sciadv.aaz9549

Weigel, A. P., Knutti, R., Liniger, M. A., & Appenzeller, C. (2010). Risks of model weighting in multimodel climate projections. *Journal of Climate*, *23*(15), 4175–4191. https://doi.org/10.1175/2010JCLI3594.1

Wilkinson, S. (2025). Code from "process-based machine learning observationally constrains future regional warming projections". https://doi.org/10.5281/zenodo.15115895

Wilks, D. S. (2006). *Statistical methods in the atmospheric sciences*. Elsevier Academic Press.
Wilson Kemsley, S., Ceppi, P., Andersen, H., Cermak, J., Stier, P., & Nowack, P. (2024). A systematic evaluation of high-cloud controlling factors. *Atmospheric Chemistry and Physics*, *24*(14), 8295–8316. https://doi.org/10.5194/ACP-24-8295-2024
Yu, Y., Mao, J., Wullschleger, S. D., Chen, A., Shi, X., Wang, Y., et al. (2022). Machine learning–based observation-constrained projections reveal elevated global socioeconomic risks from wildfire. *Nature Communications*, *13*(1), 1250. https://doi.org/10.1038/S41467-022-28853-0
Zhang, R., Sun, C., Zhu, J., Zhang, R., & Li, W. (2020). Increased European heat waves in recent decades in response to shrinking Arctic sea ice and Eurasian snow cover. *Climate and Atmospheric Science*, *3*(1), 1–9. https://doi.org/10.1038/s41612-020-0110-8
Zhang, S., & Chen, J. (2021). Uncertainty in projection of climate extremes: A comparison of CMIP5 and CMIP6. *Journal of Meteorological Research*, *35*(4), 646–662. https://doi.org/10.1007/S13351-021-1012-3
Zhu, J., Poulsen, C. J., & Otto-Bliesner, B. L. (2020). High climate sensitivity in CMIP6 model not supported by paleoclimate. *Nature Climate Change*, *10*(5), 378–379. https://doi.org/10.1038/s41558-020-0764-6

References From the Supporting Information

Bi, D., Dix, M., Marsland, S., O'Farrell, S., Rashid, H., Uotila, P., et al. (2013). The ACCESS coupled model: Description, control climate and evaluation. *Australian Meteorological and Oceanographic Journal*, *63*(1), 41–64. https://doi.org/10.1071/ES13004
Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography (Copenhagen)*, *36*(1), 27–46. https://doi.org/10.1111/J.1600-0587.2012.07348.X
Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., et al. (2020). Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*, *13*(5), 2197–2244. https://doi.org/10.5194/GMD-13-2197-2020
Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023a). ERA5 hourly data on pressure levels from 1940 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.bd0915c6
Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2023b). ERA5 hourly data on single levels from 1940 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*. https://doi.org/10.24381/cds.adbb2d47
Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634
Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cerezo-Mota, R., Cimadevilla, E., et al. (2020). An update of IPCC climate reference regions for subcontinental analysis of climate model data: Definition and aggregated datasets. *Earth System Science Data*, *12*(4), 2959–2970. https://doi.org/10.5194/ESSD-12-2959-2020
Jain, S., Scaife, A. A., Shepherd, T. G., Deser, C., Dunstone, N., Schmidt, G. A., et al. (2023). Importance of internal variability for climate model assessment. *Climate and Atmospheric Science*, *6*(1), 1–7. https://doi.org/10.1038/s41612-023-00389-0
Kuhlbrodt, T., Jones, C. G., Sellar, A., Storkey, D., Blockley, E., Stringer, M., et al. (2018). The low-resolution version of HadGEM3 GC3.1: Development and evaluation for global climate. *Journal of Advances in Modeling Earth Systems*, *10*(11), 2865–2888. https://doi.org/10.1029/2018MS001370
Law, R. M., Ziehn, T., Matear, R. J., Lenton, A., Chamberlain, M. A., Stevens, L. E., et al. (2017). The carbon cycle in the Australian Community Climate and Earth System Simulator (ACCESS-ESM1) - Part 1: Model description and pre-industrial simulation. *Geoscientific Model Development*, *10*(7), 2567–2590. https://doi.org/10.5194/GMD-10-2567-2017
Lee, W. L., Wang, Y. C., Shiu, C. J., Tsai, I. C., Tu, C. Y., Lan, Y. Y., et al. (2020). Taiwan Earth System Model Version 1: Description and evaluation of mean state. *Geoscientific Model Development*, *13*(9), 3887–3904. https://doi.org/10.5194/GMD-13-3887-2020
Lovato, T., Peano, D., Butenschön, M., Materia, S., Iovino, D., Scoccimarro, E., et al. (2022). CMIP6 simulations with the CMCC Earth System Model (CMCC-ESM2). *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002814. https://doi.org/10.1029/2021MS002814
Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038. https://doi.org/10.1029/2018MS001400
Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., et al. (2018). A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383–1413. https://doi.org/10.1029/2017MS001217
O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D., et al. (2015). The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, *42*. https://doi.org/10.1016/j.gloenvcha.2015.01.004
Po-Chedley, S., Fasullo, J. T., Siler, N., Labe, Z. M., Barnes, E. A., Bonfils, C. J., & Santer, B. D. (2022). Internal variability and forcing influence model–satellite differences in the rate of tropical tropospheric warming. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(47), e2209431119. https://doi.org/10.1073/pnas.2209431119
Schlund, M., Lauer, A., Gentine, P., Sherwood, S. C., & Eyring, V. (2020). Emergent constraints on equilibrium climate sensitivity in CMIP5: Do they hold for CMIP6? *Earth System Dynamics*, *11*(4), 1233–1258. https://doi.org/10.5194/ESD-11-1233-2020
Seland, Ø., Bentsen, M., Olivié, D., Toniazzo, T., Gjermundsen, A., Graff, L. S., et al. (2020). Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. *Geoscientific Model Development*, *13*(12), 6165–6200. https://doi.org/10.5194/GMD-13-6165-2020
Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., et al. (2019). UKESM1: Description and evaluation of the U.K. Earth System Model. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4513–4558. https://doi.org/10.1029/2019MS001739
Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. https://doi.org/10.5194/GMD-12-4823-2019
Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., et al. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. https://doi.org/10.5194/GMD-12-2727-2019
Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykosov, V. N., Gritsun, A. S., et al. (2017). Simulation of modern climate with the new version of the INM RAS climate model. *Izvestiya - Atmospheric and Ocean Physics*, *53*(2), 142–155. https://doi.org/10.1134/S0001433817020128
Volodin, E. M., Mortikov, E. V., Kostrykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., et al. (2017a). Simulation of the present-day climate with the climate model INMCM5. *Climate Dynamics*, *49*(11–12), 3715–3734. https://doi.org/10.1007/S00382-017-3539-7

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., et al. (2018). The met office global coupled model 3.0 and 3.1 (GC3.0 and GC3.1) configurations. *Journal of Advances in Modeling Earth Systems*, *10*(2), 357–380. https://doi.org/10.1002/2017MS001115

Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., et al. (2019). The Meteorological Research Institute Earth System Model Version 2.0, MRI-ESM2.0: Description and basic evaluation of the physical component. *Journal of the Meteorological Society of Japan. Ser. II*, *97*(5), 931–965. https://doi.org/10.2151/JMSJ.2019-051

Ziehn, T., Lenton, A., Law, R. M., Matear, R. J., & Chamberlain, M. A. (2017). The carbon cycle in the Australian Community Climate and Earth System Simulator (ACCESS-ESM1) - Part 2: Historical simulations. *Geoscientific Model Development*, *10*(7), 2591–2614. https://doi.org/10.5194/GMD-10-2591-2017