Title

StableClim: Continuous estimates of climate stability for temperature and rainfall from the Last Glacial Maximum to the end of the 21st Century

Authors

Stuart C. Brown[1*], Tom M. L. Wigley[1,2], Bette L. Otto-Bliesner[2], Damien A. Fordham[1]

Affiliations

1. The Environment Institute and School of Biological Sciences, University of Adelaide, South Australia 5005, Australia

2. Climate and Global Dynamics Laboratory, National Center for Atmospheric Research, Boulder, CO 80307-3000 USA

*corresponding author: Stuart C. Brown (s.brown@adelaide.edu.au)

Abstract

Climate-model-based paleoclimatic time-series data are being used increasingly in ecological and evolutionary models to improve our understanding of the biogeographical processes that drive patterns of biodiversity across time. When models calibrated in past are driven by simulations of future climate change, the subsequent results can strengthen decisions regarding the future state of biodiversity. This, however, requires harmonised simulations of past and future temperature and precipitation changes. Here we provide harmonised, continuous global coverages of trend, variability, and signal-to-noise ratio for air temperature and precipitation from the Last Glacial Maximum to the end of the 21st Century. Thresholds of natural variability in trends in annual, area-weighted regional- and global-mean temperature enable users to identify and subset the dataset to periods in Earth's history when climatic conditions were extremely unstable. By providing access to continuous simulated estimates of climate stability, validated against paleoclimate reconstructions and observational data, *StableClim*, will improve our understanding of the roles of climate in shaping past, present-day and future patterns of biodiversity.

Background & Summary

A strengthened understanding of the relationships between climate change and species distributions can improve capacity to anticipate, and potentially manage responses of biodiversity and ecosystems to global change[1]. Approaches that combine macroecological models with inferences from paleoclimate simulations, paleoecology, and paleogenomics are being used with increasing frequency to establish the patterns and biogeographic processes that govern biodiversity in shifting climates[2]. This research has shown that a primary factor constraining the distributions and diversity of species at macro-scales is climate stability[3-5], with hotspots of biodiversity often occurring in regions that have experienced stable temperatures and variable rates of precipitation during the late Pleistocene and Holocene[6-9].

Unravelling the mechanisms that have shaped ancient and current-day patterns of biodiversity requires spatially and temporally consistent datasets of paleo climatic change[10]. While there is a growing library of high spatial and temporal resolution paleo climate datasets available to researchers[11-15], issues relating to spatiotemporal coverage and continuity persist. Furthermore, a lack of paleo climate simulations that have been harmonised (spatially and temporally) with independently derived future projections is preventing a wider use of paleo-archives in model forecasts of biodiversity change, including guiding conservation decisions regarding desired states of ecological systems under global warming[16,17]. Although there have been attempts to overcome this problem[18,19], a lack of temporal continuity in simulations that extend from the past into the future remains[15].

Data for centennial trends and variability are needed to calculate spatiotemporal change in climate stability from the last glacial maximum to the end of the 21st century, providing important opportunities for eco-evolutionary impacts of climate change to be quantified[6]. Here we provide continuous gridded global-scale simulations of centennial trend, variability and signal-to-noise ratio (SNR) in temperature and precipitation between 21,000 BP and 2100. We do this by harmonising, at 2.5° spatial resolution (~250km at the equator), three distinct data sets: paleoclimate simulations from the TraCE-21ka coupled atmosphere-ocean-general-circulation-model (AOGCM)[20], historical runs from 18 CMIP5 AOGCMs, and future projections from the same CMIP5 18 AOGCMs under 4 Representative Concentration Pathways (RCPs)[21,22].

We use pre-industrial control runs from CMIP5 AOGCMs to identify centuries of past and future extreme temperature change (given expected natural variability) at a global scale and regional scales: separate for terrestrial and ocean realms, distinct IPCC AR5 climatic regions[23], and for terrestrial zoogeographic realms[30] . Therefore users can subset *StableClim* to periods of rapid warming (and cooling) at global and/or regional scales[24], enabling rapid climate events that occurred in the past to be identified in space and time and compared directly with those projected for the future. Regions that experienced past climate shifts that are of similar magnitude to future forecasts provide important opportunities for establishing the species, communities and ecosystems most vulnerable to pronounced and rapid climate change[2], and for connecting theory to the on-ground design and implementation of effective measures to protect biodiversity[25].

Furthermore, we provide continuous coverage of gridded monthly-mean temperature and total monthly precipitation between 1850 and 2100 at monthly time-step with 2.5° x 2.5° spatial resolution allowing end-users to create their own metrics of continuous climate stability at alternative temporal scales. When combined with PaleoView[26] this provides users with more than 21,100 years of harmonised monthly temperature and precipitation climate data.

Methods

Overview

An overview of the design of *StableClim* is provided in Figure 1. Broadly, 18 atmosphere-ocean general circulation models (AOGCMs), from the Coupled Model Inter-comparison Project phase 5 (CMIP5)[27] were used to calculate continuous estimates of trend, variability, and signal-to-noise ratios (SNR) in pre-industrial control, historical, and future climates under four different emissions scenarios. Simulated climate data from the TraCE-21ka[20] experiment was used to calculate the same metrics for paleo climates since the last glacial maximum. Global and regional estimates of trend for pre-industrial control temperatures can be used to identify past and future extreme centennial conditions.

Pre-industrial, paleo, historical, and future climate data

Data access

Pre-industrial, historical, and future climate datasets with global coverages of modelled monthly-mean surface temperature, and monthly precipitation were extracted from the CMIP5 Earth System Grid Federation data portal (https://esgf-node.llnl.gov/projects/esgf-llnl/) using customised bash scripts (available from https://github.com/GlobalEcologyLab/ESGF_ClimateDownloads). Paleoclimate data from the TraCE-21ka experiment was extracted from PaleoView[26] at a monthly time-step for the period 21,000 BP to 0 BP (1950 C.E.).

We used four different modelled climate datasets to generate *StableClim* (Figure 1):

1) Pre-industrial control runs for 18 AOGCMs from CMIP5 used to model natural climate variability,
2) Paleoclimate simulations from the TraCE-21ka experiment, which used the Community Climate System Model ver. 3 (CCSM3)[28,29]
3) Historical runs for the same CMIP5 18 AOGCM's which were used to generate pre-industrial control climates[27],
4) Representative Concentration Pathway (RCP)[21,22] 2.6, 4.5, 6.0, and 8.5 runs for the same CMIP5 18 AOGCM's used to generate pre-industrial control climates.

Pre-industrial climate

Pre-industrial control runs are multi-century unforced climate simulations, where the initial model conditions are set based on atmospheric gas concentrations prior to large-scale industrialisation[27]. They have non-evolving boundary conditions (e.g. non-evolving land use and greenhouse gas concentrations) relevant to the chosen start year[27]. We elected to use only the first realisation (r1i1p1) from each model for the pre-industrial control runs as all models, with the exception of the Community Climate System Model ver. 4 (CCSM4)[30], only had a single realisation (i.e. initial conditions) for pre-industrial control conditions. Furthermore, the additional pre-industrial realisations (r2i1p1 and r3i1p1) for the CCSM4 model only ran for 156 and 120 years, respectively. The shortest modelled duration of pre-industrial control runs used in this analysis was 240 years (i.e., HadGEM2-CC; see Online-only Table 1).

Paleoclimate

The TraCE-21ka experiment was chosen to represent paleo-climate conditions because (i) the data are available at a high temporal (monthly) and moderate spatial

(2.5° x 2.5°) resolution[26]; and (ii) the model has been independently validated at multiple temporal and spatial scales[26,31-33]. These independent validations have shown that the TraCE-21ka model effectively reconstructs important regional-to-global paleoclimatic fluctuations during the last deglaciation event[26,31-33] and accurately simulates present-day climate patterns[26].

Historical climate

The historical simulations cover the period 1850-2005 (in some extended cases they continue to 2012), with the beginning of the modelling period occurring before significant anthropogenic forcing and climate change. The historical climate simulations allow simulated climatic conditions to be validated against observed datasets[27]. The historical simulations differ from the pre-industrial control conditions as they are forced by observed atmospheric composition changes and aerosol emissions (for both anthropogenic and natural sources) and include time-evolving land cover. All available model realisations were used for the historical period as there can be significant differences in trends due to internal climate variability in the models[34]. We chose to include all model realisations, as there is no way to determine which of the realisations should be preferred over others, and each realisation will lead to a slightly different climate state[35]. For example, all members within an ensemble of historical runs (e.g. CCSM4 r1i1p1, r2i1p1, r3i1p1) are forced in the same way, but each is initiated at a different point in the pre-industrial control run[27]. The differences in initial conditions result in different trajectories, and multi-realisation averaging reduces this "noise"[36].

Future climate

The RCP scenarios describe a set of possible climate outcomes as a result of changes in emissions and land use, developed specifically to allow assessment of future climates over a wide range of warming scenarios[22]. The RCP numerical designation indicates the radiative forcing level reached at the end of the century (e.g. RCP 8.5 is a high emissions warming scenario with radiative forcing level reaching approximately 8.5 W/m^2 by 2100)[22]. The two intermediate scenarios (RCP 4.5 and RCP 6.0), feature a peak-and-stabilise scenario, whereby the radiative warming peaks at the given level before stabilising by 2100. The low emission RCP 2.6 scenario has radiative forcing peaking in the middle of the 21st century before decreasing to an eventual nominal

level of 2.6 W/m^2 [22]. As with the historical climate simulations, assessment of the RCP scenarios utilised all available model realisations to reduce inter-model noise in the ensemble average.

Pre-processing of climate data

To address the different temporal extents and spatial resolutions of the AOGCMs used to generate *StableClim***,** a number of pre-processing steps were required to ensure that each dataset had an adjoining timeframe for the period of interest (i.e., paleo, historical, and future), and that the data were on a spatially consistent grid.

Modelled years for TraCE-21ka simulations were constrained to the period 21,000 BP to 100 BP (1850 C.E.) to limit the influence of wide scale industrialisation on the paleoclimate simulations[37]. The start-date for the CMIP5 historical simulations is 1850. An end-date of 2005 was chosen because of the low number of models (n = 3; BCC-CSM1.1, CNRM-CM5, and MIROC5) with simulations extending beyond this time period. The RCP scenarios simulate possible future climates between 2005 and 2100 and are initialised using the climate conditions at the end of the historical period (1850 – 2005). As the RCP simulations are essentially continuations of the historical simulations[22] and we needed to have continuous centennial trends between the paleo, historical, and future periods, we temporally harmonised the historical and RCP simulations. Following Santer, et al. [38] we spliced the historical simulations to the beginning of the RCP simulations, with the exception of CNRM-CM5 (Online-only Table 1) which was spliced using a combination of the historical (1850-2005) and the historical extension (2006-2012) simulation to minimise the effect of differences in volcanic aerosol forcings on our modelled temperatures. More specifically, for the CNRM-CM5 RCP simulations, the data was subset to 2013 – 2100, before the spliced historical (1850-2005) / historicalExt (2005-2012) simulation was spliced to the start. For all other models the historical simulation was simply spliced to the beginning of the RCP simulation. This pre-processing resulted in a continuous data set from 21k BP – 2100 C.E (paleo = 21k BP – 1850 C.E., spliced historical/RCP = 1850 C.E – 2100 C.E.),

To enable spatially consistent comparisons with the TraCE-21ka simulation, the CMIP5 data were re-gridded to a 2.5° x 2.5° (latitude/longitude) global grid using bilinear interpolation. Re-gridding of the CMIP5 datasets to match the resolution of

the TraCE-21 data using bilinear interpolation was chosen because (i) the source and destination grids were rectilinear, (ii) precipitation and temperature varies smoothly spatially, and (iii) bilinear interpolation (more or less) retains the integrity and limitations of the original model output data, where orography is highly smoothed relative to the real-world[26]. Furthermore, the 2.5° x 2.5° grid cell resolution corresponds to the resolution of the TraCE-21ka data as documented in PaleoView[26], (downscaled to 2.5° x 2.5° from its nominal original resolution of ~3.75°[26]), and the resolution of projections from MAGICC/SCENGEN[39]. Surface temperatures and precipitation were then converted to °C (from Kelvin) and mm/year (from kg m^2 s^1) respectively.

Calculating trends in global mean temperature

Continuous estimates of trends in global-mean temperature through time allows comparisons between rates of change between key periods in Earth's history and the future (Figure 2). Pre-industrial control-runs can be used as a baseline for identifying high magnitude and rapid changes in global mean temperature ("extreme" events) that occurred in the past and that are likely to occur in the future[6,40].

We determined linear trends in area-weighted global-mean surface temperature associated with natural variability[41] for maximally overlapping century long windows for each of the CMIP5 pre-industrial control runs. Maximally overlapping means that, for a time series 1,2,3 … N, the 100-year windows would be years 1-100, 2-101, 3-103, etc. We calculated the latitudinally weighted global mean temperature, for each year using the formula:

$$global\ mean\ temperature\ = \frac{\sum_{i\ldots j}^{z}\left(t_i \times cos\left(x_i\ \times \left(\frac{\pi}{180}\right)\right)\right)}{\sum_{i\ldots j}\left(cos\left(x_i\ \times \left(\frac{\pi}{180}\right)\right)\right)}$$

where t_i is the temperature (°C) at cells i to j, for a given year z, and x_i is the latitude (°) of the grid cell i centroid.

Trends for annual area-weighted global-mean temperatures were then calculated using Generalised Least Squares (GLS) regression with AR(1) errors. The GLS models were calculated using the '*nlme*' package[42] for R (version 3.5.1)[43]. GLS regression with an AR(1) error structure was chosen to minimise any effect of temporal auto-correlation in the model residuals[42]. The resulting global natural trends

(i.e., the slope of the regression) for surface temperature were used to generate a multi-model, pre-industrial cumulative distribution function (CDF) using signed slopes.

Because the number of years varied between pre-industrial control runs from different AOGCMs (Online-only Table 1) we used a bootstrap procedure to ensure that all models had equal weights in the CDF (i.e. we did not want to bias the CDF towards models that had longer simulations, or higher/lower modelled global-mean temperatures). The bootstrap procedure involved first selecting the slopes for all overlapping windows for each model, and then randomly selecting slopes from each model (with replacement) equal to the difference between the number of overlapping windows for a given model, and the maximum number of overlapping windows across all models (n = 952). The bootstrap procedure was repeated 1000 times before building the CDF, ensuring that all intra-model variability was accounted for, while the effect of longer simulation runs was negated. A multi-model average global mean trend was then calculated by averaging across all bootstrap samples within each model (e.g. averaging ACCESS 1.3 bootstraps 1-1000, averaging GFDL-CM3 bootstraps 1-1000), and then averaging across all models.

For the paleo (21k BP-1850 C.E.) and spliced historical/future climate (1850 – 2100 C.E.) we calculated trends in global-mean temperature using the methods described above (Figure 2). However, we did not use a bootstrap approach here, as in the case of the paleo period we only had a single simulation (TraCE-21ka), and the historical and future simulations were a multi-model ensemble subset to a consistent temporal window negating the need for a bootstrap. As such, multi-model averages for the spliced historical/future climate was calculated by averaging across all realisations (e.g. r1i1p1, r2i1p1) within each model, before averaging across all models (n = 18). This process allowed us to effectively calculate robust measures of global mean temperature that accounted for intra- and inter-model variability[34,38,44].

Calculating trends in regional mean temperature

We also determined linear trends in area-weighted regional-mean surface temperature associated with natural variability for maximally overlapping century long windows for each of the CMIP5 pre-industrial control runs. The regions were defined by 18 distinct IPCC AR5 climatic regions[23], 19 Wallace Zoogeographic zones[24], and 11

zoogeographic realms[24] . Temperatures were extracted for any grid-cell that intersected the boundary of the region regardless of the percentage of overlap (i.e. cells were not just selected if their centroid was in the region). Weights for the regions were calculated as above. The 18 IPCC AR5 climatic regions are an amalgamation of the terrestrial regions defined by Working Group 1 for the IPCC Fifth Assessment Report[45]. The Wallace Zoogeographic zones and realms follow Holt, et al. [24] however the Polynesian zone was removed due to its small size (average island size in the Polynesian zone is ~118 km^2, or approximately 0.002 % of the area of our grid-cells). The IPCC regions, the Wallace zones, and the zoogeographic realms are available in the code repository for the analysis (see *Code Availability*).

Identifying thresholds of extreme climate change

To identify periods of rapid climate change at global and regional scales we used the ensemble averaged pre-industrial CDF at 1, 2.5, and 5% increments (e.g. 1%, 2.5%, 5%, 10%, 15%...90, 95%, 97.5%, 99%) to identify rates of change that could correspond to different "thresholds" of stable climate (at lower thresholds) or rapid climate change[46] (at higher thresholds). The 25/75% quartile of the pre-industrial CDF has previously been used to identify periods of change in global mean temperature that had either low rate and magnitude (25%) or high rate and magnitude (75%) since 21,000 BP[6].

Thresholds were calculated at a range of scales and for different regions and realms. Globally, thresholds are provided for combined terrestrial and oceanic realms, and also as separate terrestrial and oceanic realms for users interested in either domain. For climate focused studies, thresholds are provided at regional scales using the IPCC AR5 climatic regions described above. For biogeographical or ecological focussed work, thresholds are provided at two scales: (i) 19 smaller scale Wallace Zoogeographic "zones" and (ii) 11 broader scale terrestrial zoogeographic realms, both described above.

Calculating local trends, variability, and SNR

For both temperature and precipitation, we calculated 'local' measures of trend at centennial and near centennial time scale trends for each grid-cell (n = 10,368 cells) for the paleo, and the spliced historical/RCP simulations. Trend was again defined as

the slope of a GLS(AR1) regression. We also calculated grid-cell estimates of variability, where variability was defined as the standard deviation of the residuals about the local trend[47]. Trend and the variability around the trend are the primary components of climate stability[40], and they provide a distinction between low frequency (trend) and high frequency (standard deviation) climate stability. To further account for the large inter-model differences in spatial resolution, forcings, physics, and sensitivities within each of the AOGCMS[39], we used a simple pattern scaling approach[48] where we standardised the local (cell-based) "raw" trends in temperature and precipitation by the trend in global-mean temperature for the matching window.

We also calculated a signal-to-noise ratio (SNR = abs(standardised trend)/variability)[49], for both temperature and precipitation. We opted to consider SNR in addition to trend and variability, because the SNR is a composite measure of the trend given background variability. Furthermore, the SNR can be utilised to estimate continuous estimates of climatic stability which is useful in comparing climate stability at different locations and times[6].

For the spliced historical/future climates, estimates of trend, standardised trend, variability, and SNR were determined by averaging across all realisations within each model and then across all models.

Ensemble estimates of monthly temperature and precipitation

While our estimates of trend, variability, and SNR provide continuous global coverages for air temperature and precipitation from the Last Glacial Maximum to the end of the 21st Century at centennial and near-centennial time scales, we recognise that some researchers may want to work with datasets that cover different temporal periods, e.g. seasonal or decadal trends. Therefore, we also provide ensemble mean estimates of monthly temperature and precipitation for the historical and future climates at the same spatial resolution of our continuous trend, variability, and SNR estimates. These ensembles allow end users to create their own estimates of trend, variability, and SNR at time scales suitable for their purposes (e.g. seasonal or decadal). These ensemble means are provided only for the spliced historical/future climate period (1850 – 2100 C.E). We opted not to provide ensemble means for the pre-industrial control runs as these simulations are not reconstructions of temporally explicit pre-industrial climate (unlike, e.g., TraCE-21ka), but are used to simulate

internal model variability, which can be used as a proxy for natural (unforced) climate variability. When combined with the data in PaleoView, the historical/future ensemble means provide a spatiotemporally harmonised monthly temperature and precipitation climate dataset from 21,000 BP to 2100 C.E.

To produce the monthly ensemble estimates, each individual model realisation (e.g. CCSM4 r1i1p1, CCSM4 r2i1p1) was re-gridded using bilinear interpolation to a common 2.5° x 2.5° grid. Model averages were then calculated across all realisations, before a multi-model average was calculated across all models.

Data Records

Gridded results and outputs are stored in netcdf format, with thresholds provided in .RDS format data frames. All outputs are available on request from the corresponding author. Paleo-climate data is available in PaleoView (https://github.com/GlobalEcologyLab/PaleoView).

Technical Validation

The TraCE-21ka simulation has previously been well validated across multiple spatial and temporal scales with regards to its ability to simulate known rapid climate change events[26,31-33], and to accurately model contemporary climates[26]. As such, we have done no additional technical validation on the raw temperature or precipitation data extracted from the TraCE-21ka simulation. Validations have however, been done on our estimates of SNR (see below for details).

The CMIP5 pre-industrial and RCP simulations are built using the same model structure as for the historical simulations but with altered forcing and boundary conditions[27]. An assessment of agreement between historical multi-model ensemble-averaged projections of temperature and precipitation, and observed temperature and precipitation provides confidence that our trends, variability, and SNR measures are an accurate representation of recent and future climates[26]. The tacit assumption made here is that changes in grid-cell temperatures (and variability) scale approximately linearly with changes in global-mean temperature.

The thresholds of extreme change we provide to subset continuous estimates of global-mean temperature trend, variability, and SNR to periods of rapid climate change have been validated recently. Brown, et al. [6] identified past centuries of rapid

change in global-mean temperature, over the period 21,000 BP to 100 BP as those having absolute global-mean temperature trends greater than the 90th percentile of the pre-industrial control CDF. To check that their definition of rapid climate change was appropriate, they ran two tests: 1) Brown, et al. [6] calculated the CDF for trends from the TraCE-21ka model and compared these to the CDF based on periods of rapid climate change from the pre-industrial control simulations; and 2) they determined the amount of time a calendar millennium was considered to be experiencing rapid rates of climate change by calculating the % of time that a millennium was characterised by trends ≥ 90th percentile of the pre-industrial control run trends. This confirmed that known large-scale climatic events during the last deglaciation (e.g. Bølling–Allerød) were being correctly identified as periods of rapid climate change in their analysis (see Supplementary Fig. 6 in Brown, et al. [6]).

Signal to Noise Ratio

To validate our method of calculating signal-to-noise ratio (SNR), we calculated estimates of SNR for Antarctica and Greenland using latitudinally weighted temperatures and compared these to estimates based on the Vostok[50] and NGRIP[51,52] ice-cores. The temporal resolution and timing of temperature estimates was matched between the TraCE-21 simulation and the ice-core data by sub-setting the (annual) TraCE-21 data to the same time steps as the Vostok (~150 years) and NGRIP data (~20 years). This allowed us to calculate estimates of SNR respectively at millennial and centennial timescales between observed (ice-core) and simulated (TraCE-21) datasets that were directly comparable. Boxplots of SNR values for four different windows during which known rapid climate change events occurred at the poles (21-15k BP; 15-11k BP; 11-3k BP; >3k BP) were constructed for visual interpretation, before the SNR values were statistically compared using PERMDISP[53] and PERMANOVA[54] on a Euclidean distance matrix. The four different windows were chosen as there are known major rapid climate change events that occur within each window: the oldest Dryas and the H1 Heinrich events occur in the period 21-15k[55-60], the Bolling-Allerod, Antarctic Cold Reversal, Younger Dryas, and the 11.7 event occur in the period 15-11k[55,56,59,61], with the 8.2k event occurring within the 11-3k window[62]. Both procedures had data source (TraCE-21 or ice-core) nested within window and used 999 permutations to generate P-values.

The PERMDISP results suggest there were significant differences in the dispersion of SNR values between sources (i.e. between the ice-core and simulated data) within windows for the Vostok core. However, after accounting for multiple comparisons[63], only one of the results was considered significant (15-11k comparison, adj. P = 0.01). Likewise, unadjusted P-values were significant for comparisons between the NGRIP core and our simulated estimate of SNR, but after adjusting for multiple comparisons none of the results were considered significant (all $P \geq 0.45$). These results suggest there were only significant differences in the dispersion of observed (Vostok) and simulated (TraCE-21) SNR during the period 15-11k BP. The PERMANOVA results suggested significant differences between sources within window for the Vostok core (pseudo-$F_{4,670} = 11.82$, P = 0.001; Figure 4), with pairwise comparisons confirming differences in the 15-11k (t = 3.16, adj. P = 0.008) and the 11-3k window (t = 2.81, adj. P = 0.013). There were no differences in the NGRIP ice core comparison (pseudo-$F_{3,4} = 1.64$, P = 0.188; Figure 4). These results suggest significant differences in mean SNR values between the observed and simulated datasets only for the Antarctic region in the 15-11k and 11-3k windows. In other words, for the NGRIP-TraCE21 comparisons there were no statistically significant SNR differences, while for the comparisons with Vostok data, and in particular the 15-11k window, the results were more equivocal.

Multi-model temperature and precipitation ensembles

To validate our finer scale ensemble mean historical temperature and precipitation datasets, we extracted gridded high resolution (0.5° x 0.5°, monthly time step) data between 1901 and 2005 from the Climatic Research Unit (CRU) time-series database[64]. The data were re-gridded to the same 2.5 x 2.5° grid of our ensemble monthly estimates and converted to annual average temperature and average total monthly precipitation. Annual average climatologies for temperature and precipitation were then calculated on both the CMIP5 ensemble mean historical dataset and the re-gridded CRU dataset, for a 50-yr period centered on 1980. To quantify the skill of our ensemble model to recreate observed temperature and precipitation conditions we used a combination of visual and statistical approaches. Figure 5 shows the relatively high correlations and low standard deviations between our ensemble estimates and the re-gridded CRU data at a global scale. The spread in inter-model correlations and standard deviations was, as expected, much higher for simulated precipitation than for

temperature[65]. We also calculated a range of statistical metrics to quantify the relationship between our ensembled data and the CRU data, namely: Percentage bend correlation[66], M-statistic[67], latitudinally weighted Root-Mean-Square-Error, ratio-of-standard-deviations, modified index of agreement[68], and percentage bias. These metrics were calculated globally and for four latitudinal bands: High-North (50°N – 90°N), Mid-North (20°N – 50°N), Mid-South (50°S – 20°S) and the High-Tropics (20°S – 20°N). Five IPCC AR5 regions[23] and 4 biogeographic realms[24] were also included in the validation (Table 1). All correlations were significant at $P < 0.001$ with correlation coefficient ranging between 0.67 (Neotropical realm) and 0.99 (Table 1, Figure 6). The M-statistic ranged between 40.5 and 91.3, with no clear relationship between scale and the resultant score indicating the ensemble estimate of climate has varying capacities to simulate observed conditions independent of scale. Percentage bias in precipitation varied between -3.4 and 31.5 % with the lowest values occurring in the tropics and the Mediterranean (Table 1). The ensemble mean precipitation was shown to over-estimate precipitation across all latitudinal bands. However, with the exception of the simulated precipitation in the Southern Africa and West Indian Ocean IPCC AR5 region, the %-bias values are still satisfactory[69]. On average, over a range of scales, precipitation was overestimated by ~11%.

Code Availability

Code used to generate the data contained in *StableClim* is available on request from the corresponding author, with bash scripts to download the CMIP5 data from ESGF available at https://github.com/GlobalEcologyLab/ESGF_ClimateDownloads

Figures

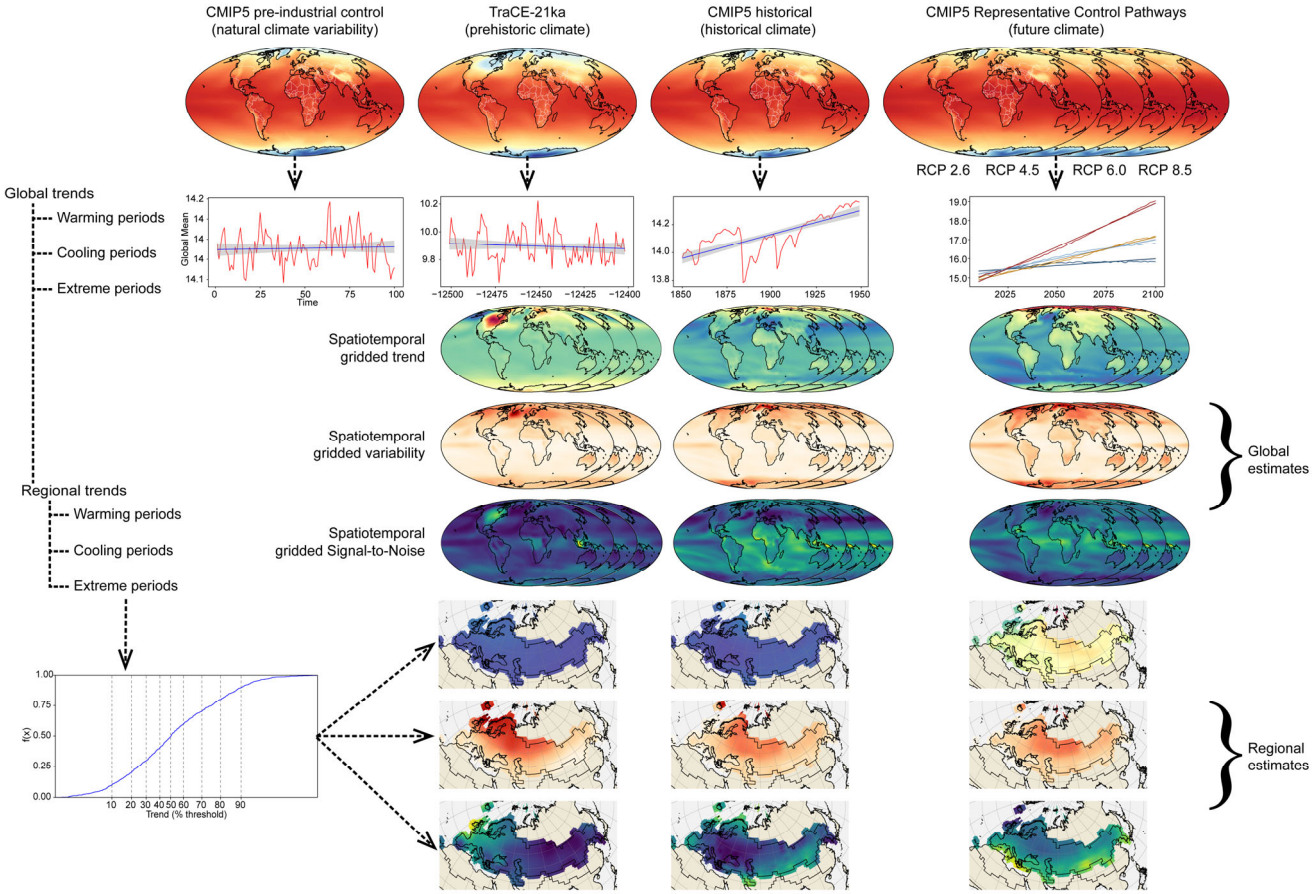

Figure 1: Overview of the *StableClim* database. Simulated climate data for temperature and precipitation for pre-industrial, historical, and future climates come from 18 CMIP5 climate models. Paleo climatic conditions come from the TRaCE-21ka simulation. Trends in global means of both variables were then determined for 100-year windows for each climate scenario. Temporally continuous grid (n = 10,368 cells) based trends, variability, and signal-to-noise ratio were calculated globally for the same size windows. Trends were also calculated for regional mean temperatures using the pre-industrial climate simulations – these trends can be subset based on a series of thresholds allowing extraction of gridded estimates of trend, variability, and signal-to-noise at regional scales during periods of regionally rapid climate change.

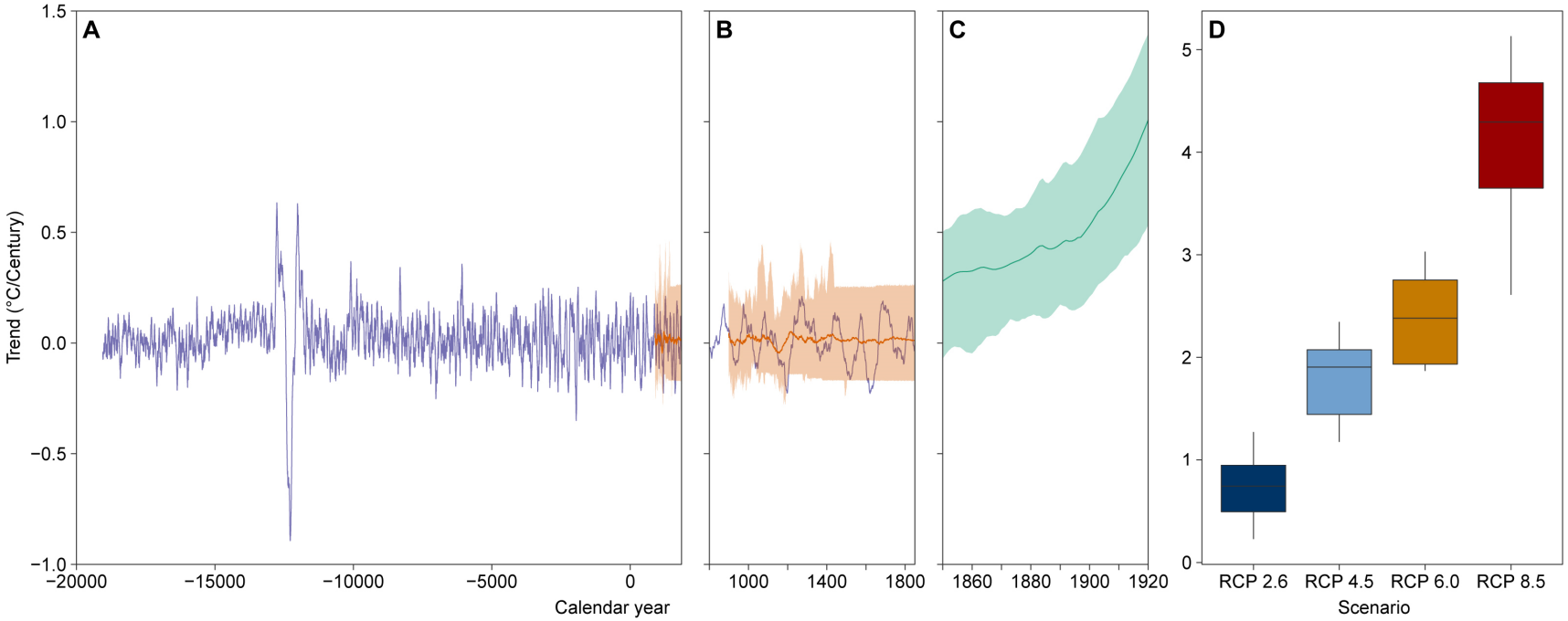

Figure 2: Trends in global mean temperatures calculated at centennial time scales for each of the four periods. **A** paleoclimate as simulated by TraCE-21ka, **B** pre-industrial control simulations, **C** spliced historical/future climate simulations to 1920, **D** future climate (2001-2100) under four different RCP scenarios. The shaded area in **A**, shows the overlap between the pre-industrial control simulations and the end of the paleo period. The shaded areas in **B** and **C** show the intra-model variability in trend estimates (95% CI). Boxplots in **D** show the variability in global mean trends under the four different RCP scenarios (*n.b.* the difference in scale between **A-C** and **D**). The plots show the continuous nature of the *StableClim* database from the last-glacial maximum to the end of the 21st century. Years in **A** have been converted from Years BP to calendar years to aid interpretation.

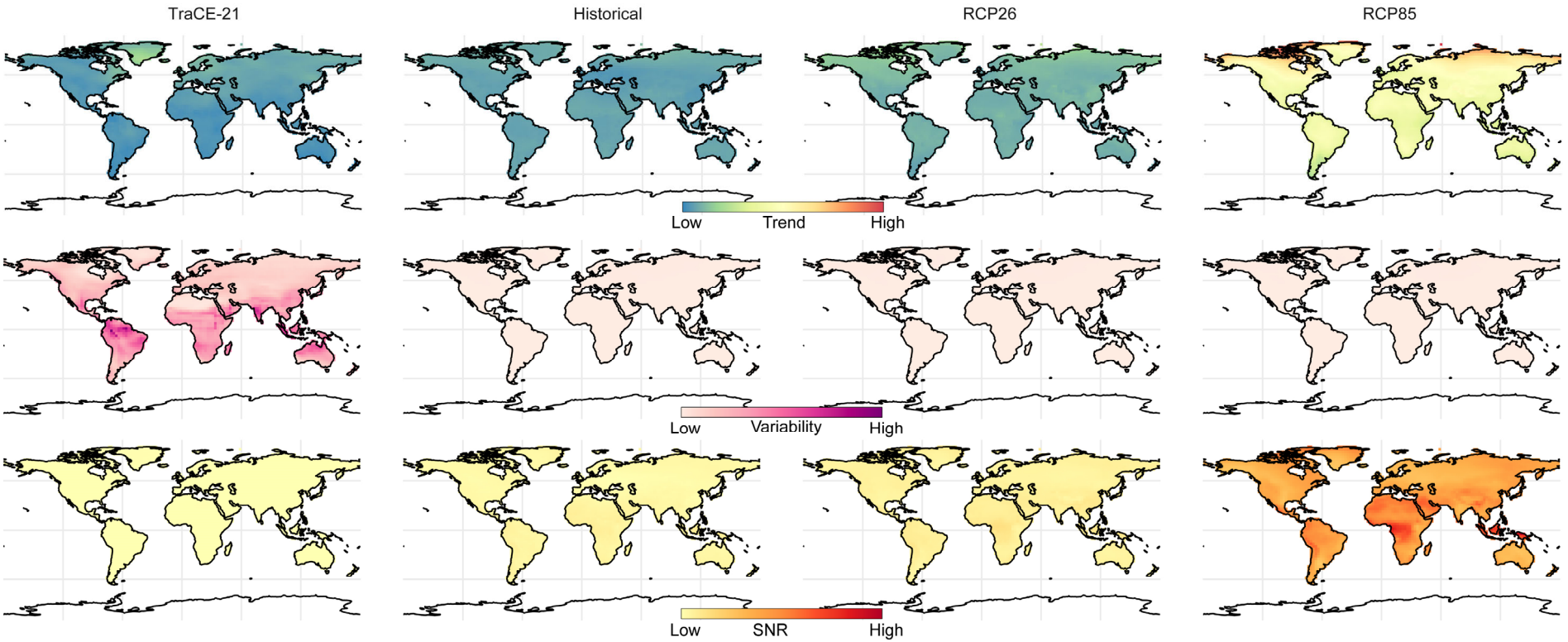

Figure 3: Maps showing trend, variability, and signal-to-noise ratio for air temperature across time. TraCE-21 estimates are the median estimates for overlapping century windows during the Bolling-Allerod (14.7-14.2k BP[61]); Historical estimates are the median estimates for century windows between 1850 and 2005. RCP 2.6 and 8.5 are single-window snapshots between 2001 and 2100.

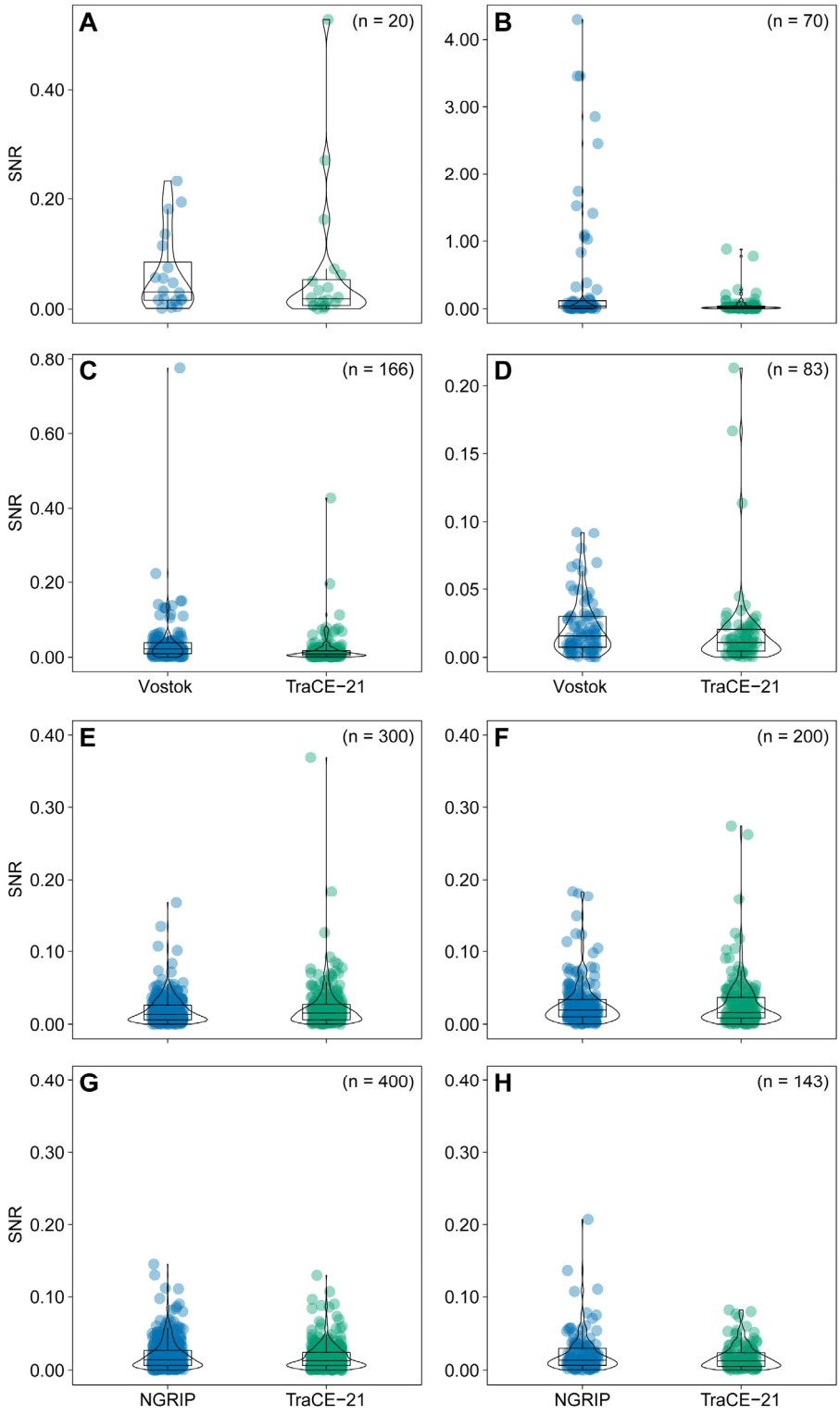

Figure 4: Validation of our modelled Signal-to-Noise ratio (SNR) against SNR calculated for the Vostok (**A-D**) and NGRIP (**E-H**) ice-cores[50-52]. Differences between the shape of the distributions and the SNR values were significant in **B**, with significant differences in mean SNR for **B** and **C**, but non-significant in all other windows based on PERMDISP and PERMANOVA results.

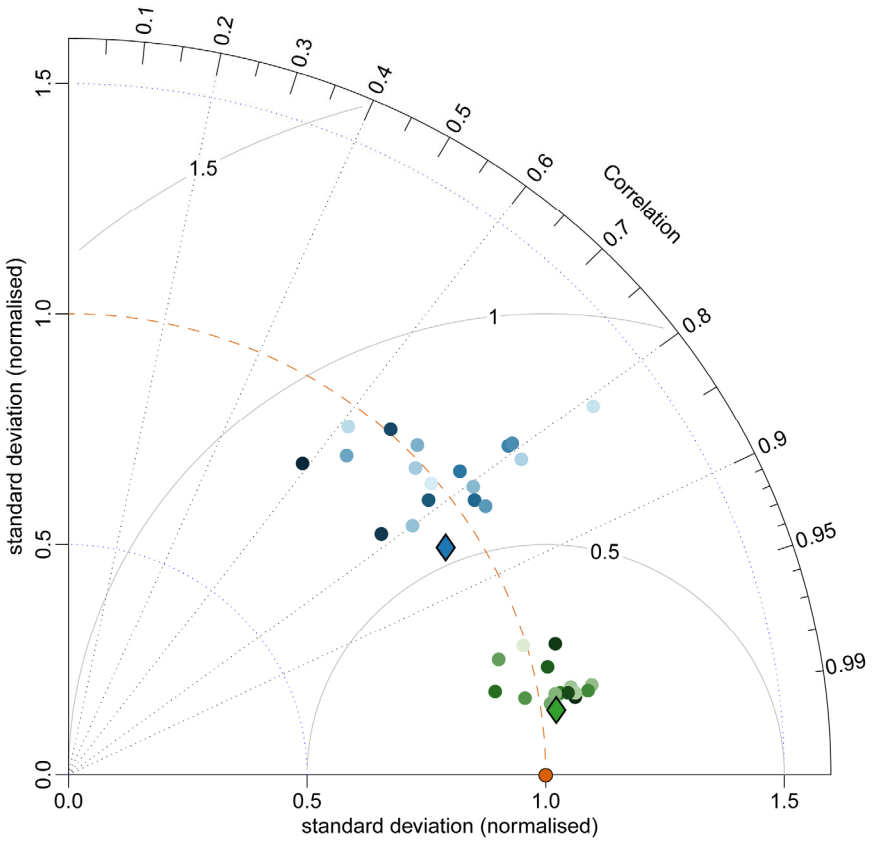

Figure 5: Taylor diagram[70] showing the relationship between our ensemble estimates of temperature (green points), precipitation (blue points), and the CRU dataset (orange point). Ensemble means are shown by the diamonds. The reference (CRU) climatology is shown by the orange circle, with SD values normalised to 1.

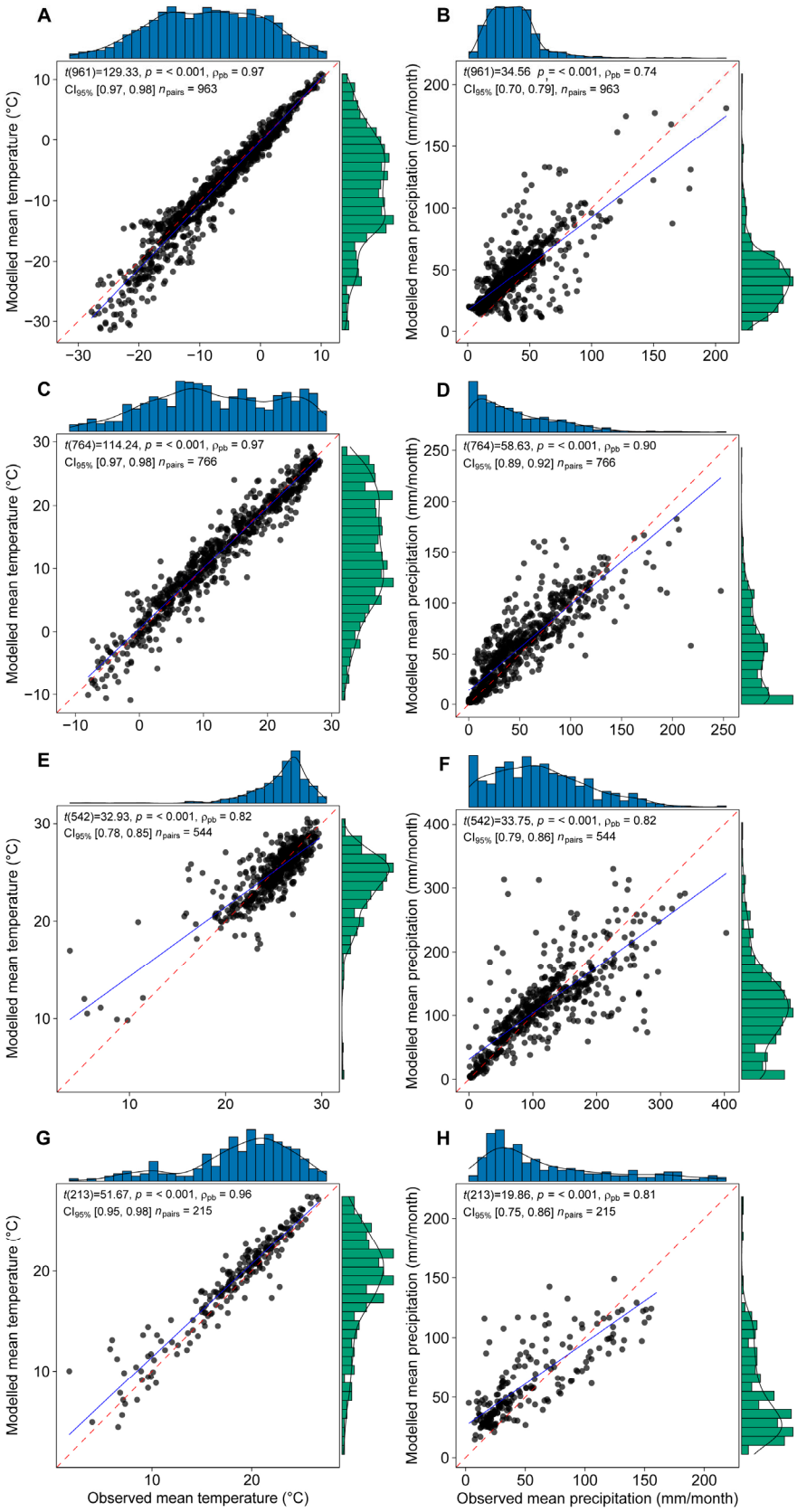

Figure 6: Independent validation of simulated historical global mean temperatures and precipitation from the ensemble mean estimates at different latitudes: High-north (50°:90°; **A, B**), Mid-north (20°:50°; **C, D**), High-tropics (-20°:20°; **E, F**), and Mid-south (-20°:-50°; **G, H**). All Percentage bend correlations are significant at P < 0.001.

Tables

Online-only Table 1: The 19 models used for the analysis of paleoclimate from the last glacial maximum through to pre-industrialisation (21,000 BP – 100 BP), 'natural' climate conditions simulated with the CMIP5 pre-industrial control runs, historical climate simulations (1850-2005), and future simulated climate (2010-2100) under four Representative Concentration Pathways (2.6, 4.5, 6.0, and 8.5).

Model	Ensemble*	Institution/s	Atmospheric Resolution (°)		Number of years†
			Lat	Long	
ACCESS 1.3	r1i1p1	Commonwealth Scientific and Industrial Research Organisation / Bureau of Meteorology	1.9	1.2	500
BCC-CSM1.1	r1i1p1	Beijing Climate Center, China Meteorological Administration	2.8	2.8	500
CanESM2	r1i1p1, r2i1p1, r3i1p1, r4i1p1, r5i1p1	Canadian Centre for Climate Modelling and Analysis	2.8	2.8	996
CCSM4	r1i1p1, r2i1p1, r3i1p1, r4i1p1, r5i1p1, r6i1p1	National Center for Atmospheric Research	0.9	1.3	1,051
CESM1(CAM5)	r1i1p1, r2i1p1, r3i1p1	National Center for Atmospheric Research	0.9	1.3	320
CNRM-CM5	r1i1p1, r2i1p1, r3i1p1, r4i1p1, r5i1p1, r6i1p1, r7i1p1, r8i1p1, r9i1p1, r10i1p1	Centre National de Recherches Meteorologiques / Centre Europeen de Recherche et Formation Avancees en Calcul Scientifique	1.4	1.4	850
CSIRO-Mk3.6.0	r1i1p1, r2i1p1, r3i1p1, r4i1p1, r5i1p1, r6i1p1, r7i1p1, r8i1p1, r9i1p1, r10i1p1	Commonwealth Scientific and Industrial Research Organisation / Queensland Climate Change Centre of Excellence	1.9	1.9	500
GFDL-CM3	r1i1p1	National Oceanic and Atmospheric Administration Office of Oceanic and Atmospheric Research - Geophysical Fluid Dynamics Laboratory	2.0	2.5	500
HadGEM2-CC	r1i1p1, r2i1p1, r3i1p1	Met Office Hadley Centre	1.3	1.9	240
HadGEM2-ES	r1i1p1, r2i1p1, r3i1p1, r4i1p1	Met Office Hadley Centre	1.3	1.9	575
INM-CM4	r1i1p1	Institute for Numerical Mathematics	1.5	2.0	500
IPSL-CM5A-LR	r1i1p1, r2i1p1, r3i1p1, r4i1p1	Institut Pierre-Simon Laplace	1.9	3.8	1,000
IPSL-CM5A-MR	r1i1p1	Institut Pierre-Simon Laplace	1.3	2.5	300

Model	Ensemble*	Institution/s	Atmospheric Resolution (°)		Number of years†
			Lat	Long	
MIROC5	r1i1p1	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies / Japan Agency for Marine-Earth Science and Technology	1.4	1.4	670
MIROC-ESM	r1i1p1, r2i1p1, r3i1p1, r4i1p1, r5i1p1	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo) / National Institute for Environmental Studies	2.8	2.8	630
MPI-ESM-LR	r1i1p1, r2i1p1, r3i1p1	Max Planck Institute for Meteorology	1.9	1.9	1,000
MRI-CGCM3	r1i1p1	Meteorological Research Institute	1.1	1.1	500
NorESM1-M	r1i1p1	Norwegian Climate Centre	1.9	2.5	501
TraCE-21ka#	N/A	National Centre for Atmospheric Research	2.5	2.5	21,000

*All pre-industrial control simulations only utilised ensemble r1i1p1. See methods for details. †Number of years of pre-industrial control simulation. All RCP model runs were extended back from 01/2006 to 01/2000 by splicing the historical simulation to the start of the RCP simulation[38]. #TraCE-21ka data was pre-processed for PaleoView[26].

Region	Temperature					Precipitation				
	ρ_{pb}	M	$RMSE_w$	rSD	md	ρ_{pb}	M	%-bias	rSD	md
Global	0.99	91.3	1.98	1.03	0.95	0.89	64.5	10.9	0.93	0.76
High-North	0.97	83.3	1.99	1.08	0.89	0.75	51.5	21.0	0.98	0.61
Mid-North	0.97	84.3	2.17	0.99	0.89	0.91	63.8	15.5	0.99	0.78
Mid-South	0.96	78.2	1.70	0.97	0.85	0.81	52.8	20.9	0.90	0.65
High-Tropics	0.81	61.1	1.83	0.86	0.72	0.83	57.4	0.10	0.92	0.73
High Latitudes*	0.96	78.9	2.28	1.12	0.87	0.72	66.1	13.6	0.94	0.61
Mediterranean and Sahara*	0.96	81.2	1.53	0.87	0.85	0.97	77.1	-2.1	0.97	0.86
North America (East)*	0.99	89.0	0.99	0.94	0.92	0.89	62.5	6.4	0.79	0.72
Southern Africa and West Indian Ocean*	0.79	60.4	1.72	0.89	0.75	0.90	48.1	31.5	0.76	0.55
Australia and New Zealand*	0.99	78.9	1.42	0.94	0.83	0.93	66.1	13.2	0.79	0.71
Neotropical#	0.95	79.1	2.10	0.91	0.88	0.67	40.5	-2.1	0.77	0.59
Oriental#	0.86	76.3	2.34	1.05	0.81	0.84	61.9	-3.4	0.95	0.75
Palearctic#	0.98	86.5	2.13	1.05	0.91	0.88	57.5	22.8	0.89	0.67

Table 1: Metrics used to assess the ability of our ensemble estimate of historical temperatures and precipitations to replicate observed conditions.
ρpb = percentage bend correlation[66], where higher values indicate more agreement between observed and simulated conditions; M = m statistic[67] (x100), where higher values indicate more agreement between observed and simulated conditions; $RMSE_w$ = Root-Mean-Square-Error weighted by latitude, lower values indicate better agreement between simulated and observed conditions; rSD = ratio of standard deviations, values closer to 1 indicate better agreement between simulated and observed conditions; md = modified index of agreement[68], values closer to 1 indicate better agreement between simulated and observed conditions; %-bias = percentage bias, the tendency of the simulated values to be larger or smaller than observed. * = IPCC AR5 regions from van Oldenborgh, et al. [23]. # = Biogeographic realms following Holt, et al. [24].

References

1 Fordham, D. A., Brook, B. W., Moritz, C. & Nogues-Bravo, D. Better forecasts of range dynamics using genetic data. *Trends Ecol Evol* **29**, 436-443, doi:10.1016/j.tree.2014.05.007 (2014).

2 Nogués-Bravo, D. *et al.* Cracking the Code of Biodiversity Responses to Past Climate Change. *Trends in Ecology & Evolution* **33**, 765-776, doi:10.1016/j.tree.2018.07.005 (2018).

3 Fine, P. V. A. Ecological and Evolutionary Drivers of Geographic Variation in Species Diversity. *Annu Rev Ecol Evol S* **46**, 369-392, doi:10.1146/annurev-ecolsys-112414-054102 (2015).

4 Rangel, T. F. *et al.* Modeling the ecology and evolution of biodiversity: Biogeographical cradles, museums, and graves. *Science* **361**, eaar5452, doi:10.1126/science.aar5452 (2018).

5 Lister, A. M. & Stuart, A. J. The impact of climate change on large mammal distribution and extinction: Evidence from the last glacial/interglacial transition. *Comptes Rendus Geoscience* **340**, 615-620, doi:10.1016/j.crte.2008.04.001 (2008).

6 Brown, S. C., Wigley, T. M. L., Otto-Bliesner, B. L., Rahbek, C. & Fordham, D. A. Persistent Quaternary climate refugia are hospices for biodiversity in the Anthropocene. *Nature Climate Change*, doi:10.1038/s41558-019-0682-7 (2020).

7 Fjeldså, J. & Lovett, J. C. Geographical patterns of old and young species in African forest biota: The significance of specific montane areas as evolutionary centres. *Biodivers Conserv* **6**, 325-346, doi:10.1023/A:1018356506390 (1997).

8 Haffer, J. Speciation in Amazonian Forest Birds. *Science* **165**, 131-137, doi:10.1126/science.165.3889.131 (1969).

9 Harrison, S. & Noss, R. Endemism hotspots are linked to stable climatic refugia. *Ann Bot* **119**, 207-214, doi:10.1093/aob/mcw248 (2017).

10 Fordham, D. A. & Nogues-Bravo, D. Open-access data is uncovering past responses of biodiversity to global environmental change. *Past Global Change Magazine* **26**, 77-77, doi:10.22498/pages.26.2.77 (2018).

11 Funk, C. *et al.* The Centennial Trends Greater Horn of Africa precipitation dataset. *Sci Data* **2**, 150050, doi:10.1038/sdata.2015.50 (2015).

12 Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas. *Sci Data* **4**, 170122, doi:10.1038/sdata.2017.122 (2017).

13 Werner, A. T. *et al.* A long-term, temporally consistent, gridded daily meteorological dataset for northwestern North America. *Sci Data* **6**, 180299, doi:10.1038/sdata.2018.299 (2019).

14 Brown, J. L., Hill, D. J., Dolan, A. M., Carnaval, A. C. & Haywood, A. M. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Sci Data* **5**, 180254, doi:10.1038/sdata.2018.254 (2018).

15 Lorenz, D. J., Nieto-Lugilde, D., Blois, J. L., Fitzpatrick, M. C. & Williams, J. W. Downscaled and debiased climate simulations for North America from 21,000 years ago to 2100AD. *Sci Data* **3**, 160048, doi:10.1038/sdata.2016.48 (2016).

16 Barnosky, A. D. *et al.* Merging paleobiology with conservation biology to guide the future of terrestrial ecosystems. *Science* **355**, doi:10.1126/science.aah4787 (2017).

17 Nogués-Bravo, D. *et al.* Amplified plant turnover in response to climate change forecast by Late Quaternary records. *Nature Climate Change* **6**, 1115-1119, doi:10.1038/nclimate3146 (2016).

18 Maiorano, L. *et al.* Building the niche through time: using 13,000 years of data to predict the effects of climate change on three tree species in Europe. *Global Ecology and Biogeography* **22**, 302-317, doi:10.1111/j.1466-8238.2012.00767.x (2013).

19 Blois, J. L., Zarnetske, P. L., Fitzpatrick, M. C. & Finnegan, S. Climate Change and the Past, Present, and Future of Biotic Interactions. *Science* **341**, 499, doi:10.1126/science.1237184 (2013).

20 Liu, Z. *et al.* Transient simulation of last deglaciation with a new mechanism for Bolling-Allerod warming. *Science* **325**, 310-314, doi:10.1126/science.1171041 (2009).

21 Meinshausen, M. *et al.* The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change* **109**, 213-241, doi:10.1007/s10584-011-0156-z (2011).

22 van Vuuren, D. P. *et al.* The representative concentration pathways: an overview. *Climatic Change* **109**, 5-31, doi:10.1007/s10584-011-0148-z (2011).

23 van Oldenborgh, G. J. *et al.* in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker TF *et al.*) 1311-1393 (Cambridge University Press, Cambridge, United Kingdom, 2013).

24 Holt, B. G. *et al.* An Update of Wallace's Zoogeographic Regions of the World. *Science* **339**, 74, doi:10.1126/science.1228282 (2013).

25 Fordham, D. A. *et al.* Predicting and mitigating future biodiversity loss using long-term ecological proxies. *Nature Clim. Change* **6**, 909-916, doi:10.1038/nclimate3086 (2016).

26 Fordham, D. A. *et al.* PaleoView: a tool for generating continuous climate projections spanning the last 21 000 years at regional and global scales. *Ecography* **40**, 1348-1358, doi:10.1111/ecog.03031 (2017).

27 Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An Overview of CMIP5 and the Experiment Design. *B Am Meteorol Soc* **93**, 485-498, doi:10.1175/Bams-D-11-00094.1 (2012).

28 Otto-Bliesner, B. L. *et al.* Climate sensitivity of moderate- and low-resolution versions of CCSM3 to preindustrial forcings. *J Climate* **19**, 2567-2583, doi:10.1175/Jcli3754.1 (2006).

29 Collins, W. D. *et al.* The Community Climate System Model Version 3 (CCSM3). *J Climate* **19**, 2122-2143, doi:10.1175/jcli3761.1 (2006).

30 Gent, P. R. *et al.* The Community Climate System Model Version 4. *J Climate* **24**, 4973-4991, doi:10.1175/2011jcli4083.1 (2011).

31 Barker, S. *et al.* Interhemispheric Atlantic seesaw response during the last deglaciation. *Nature* **457**, 1097-1102, doi:10.1038/nature07770 (2009).

32 Carlson, A. E. in *The Encyclopedia of Quaternary Science* Vol. 3 (ed S.A. Elias) 126-134 (Elsevier, 2013).

33 Marsicek, J., Shuman, B. N., Bartlein, P. J., Shafer, S. L. & Brewer, S. Reconciling divergent trends and millennial variations in Holocene temperatures. *Nature* **554**, 92, doi:10.1038/nature25464 (2018).

34 Deser, C., Knutti, R., Solomon, S. & Phillips, A. S. Communication of the role of natural variability in future North American climate. *Nature Climate Change* **2**, 775, doi:10.1038/nclimate1562 (2012).

35 Lutz, A. F. *et al.* Selecting representative climate models for climate change impact studies: an advanced envelope-based selection approach. *International Journal of Climatology* **36**, 3988-4005, doi:10.1002/joc.4608 (2016).

36 Tebaldi, C. & Knutti, R. The use of the multi-model ensemble in probabilistic climate projections. *Philosophical transactions of the royal society A: mathematical, physical and engineering sciences* **365**, 2053-2075 (2007).

37 Kaplan, J. O. *et al.* Holocene carbon emissions as a result of anthropogenic land cover change. *Holocene* **21**, 775-791, doi:10.1177/0959683610386983 (2011).

38 Santer, B. D. *et al.* Identifying human influences on atmospheric temperature. *PNAS* **110**, 26-33, doi:10.1073/pnas.1210514109 (2013).

39 Fordham, D. A., Wigley, T. M. L., Watts, M. J. & Brook, B. W. Strengthening forecasts of climate change impacts with multi-model ensemble averaged projections using MAGICC/SCENGEN 5.3. *Ecography* **35**, 4-8, doi:10.1111/j.1600-0587.2011.07398.x (2012).

40 Fordham, D. A., Brown, S. C., Wigley, T. M. L. & Rahbek, C. Cradles of diversity are unlikely relics of regional climate stability. *Current Biology* **29**, R356-R357, doi:10.1016/j.cub.2019.04.001 (2019).

41 Sen Gupta, A., Jourdain, N. C., Brown, J. N. & Monselesan, D. Climate Drift in the CMIP5 Models. *J Climate* **26**, 8597-8615, doi:10.1175/Jcli-D-12-00521.1 (2013).

42 nlme: Linear and Nonliner Mixed Effects Models v. R package version 3.1-131 (2017).

43 R: A language and environment for statistical computing v. 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria, 2018).

44 Hawkins, E. & Sutton, R. Time of emergence of climate signals. *Geophys Res Lett* **39**, doi:10.1029/2011gl050087 (2012).

45 IPCC. in *Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* Vol. 1535 (eds Thomas F Stocker *et al.*) 1535 (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013).

46 Botta, F., Dahl-Jensen, D., Rahbek, C., Svensson, A. & Nogues-Bravo, D. Abrupt Change in Climate and Biotic Systems. *Curr Biol* **29**, R1045-R1054, doi:10.1016/j.cub.2019.08.066 (2019).

47 Nadeau, C. P., Urban, M. C. & Bridle, J. R. Coarse climate change projections for species living in a fine-scaled world. *Global Change Biology* **23**, 12-24, doi:10.1111/gcb.13475 (2017).

48 Santer, B. D., Wigley, T. M., M.E Schlesinge & J.F.B Mitchell. Developing climate scenarios from equilibrium GCM results. (Max Planck Institute for Meteorology, Hamburg, Germany, 1990).

49 Frame, D., Joshi, M., Hawkins, E., Harrington, L. J. & de Roiste, M. Population-based emergence of unfamiliar climates. *Nature Climate Change* **7**, 407, doi:10.1038/Nclimate3297 (2017).

50 Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429-436, doi:10.1038/20859 (1999).

51 Andersen, K. K. *et al.* The Greenland Ice Core Chronology 2005, 15–42ka. Part 1: constructing the time scale. *Quaternary Science Reviews* **25**, 3246-3257, doi:10.1016/j.quascirev.2006.08.002 (2006).

52 Rasmussen, S. O. *et al.* Synchronization of the NGRIP, GRIP, and GISP2 ice cores across MIS 2 and palaeoclimatic implications. *Quaternary Science Reviews* **27**, 18-28, doi:10.1016/j.quascirev.2007.01.016 (2008).

53 Anderson, M. J. Distance-Based Tests for Homogeneity of Multivariate Dispersions. *Biometrics* **62**, 245-253 (2006).

54 Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26**, 32-46, doi:10.1111/j.1442-9993.2001.01070.pp.x (2001).

55 Clark, P. U. *et al.* Global climate evolution during the last deglaciation. *PNAS* **109**, E1134, doi:10.1073/pnas.1116619109 (2012).

56 Liu, Z. *et al.* Younger Dryas cooling and the Greenland climate response to CO<sub>2</sub&gt. *PNAS*, doi:10.1073/pnas.1202183109 (2012).

57 McManus, J. F., Francois, R., Gherardi, J. M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428**, 834-837, doi:10.1038/nature02494 (2004).

58 Peck, V. L. *et al.* The relationship of Heinrich events and their European precursors over the past 60ka BP: a multi-proxy ice-rafted debris provenance study in the North East Atlantic. *Quaternary Science Reviews* **26**, 862-875, doi:10.1016/j.quascirev.2006.12.002 (2007).

59 Shakun, J. D. *et al.* Global warming preceded by increasing carbon dioxide concentrations during the last deglaciation. *Nature* **484**, 49, doi:10.1038/nature10915 (2012).

60 Carlson, A. E. & Winsor, K. Northern Hemisphere ice-sheet responses to past climate warming. *Nature Geoscience* **5**, 607-613, doi:10.1038/ngeo1528 (2012).

61 Renssen, H. & Isarin, R. F. B. The two major warming phases of the last deglaciation at ~14.7 and ~11.5 ka cal BP in Europe: climate reconstructions and AGCM experiments. *Global and Planetary Change* **30**, 117-153, doi:10.1016/S0921-8181(01)00082-0 (2001).

62 Alley, R. B. & Ágústsdóttir, A. M. The 8k event: cause and consequences of a major Holocene abrupt climate change. *Quaternary Science Reviews* **24**, 1123-1149, doi:10.1016/j.quascirev.2004.12.004 (2005).

63 Yekutieli, D. & Benjamini, Y. The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics* **29**, 1165-1188, doi:10.1214/aos/1013699998 (2001).

64 Harris, I., Jones, P. D., Osborn, T. J. & Lister, D. H. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *International Journal of Climatology* **34**, 623-642, doi:10.1002/joc.3711 (2014).

65 Wigley, T. M. *MAGICC/SCENGEN 5.3: User manual (version 2)*. (NCAR, 2008).

66 Wilcox, R. R. The percentage bend correlation coefficient. *Psychometrika* **59**, 601-616, doi:10.1007/BF02294395 (1994).

67 Watterson, I. G. Non-dimensional measures of climate model performance. *International Journal of Climatology* **16**, 379-391, doi:10.1002/(Sici)1097-0088(199604)16:4<379::Aid-Joc18>3.0.Co;2-U (1996).

68 Willmott, C. J. On the Validation of Models. *Physical Geography* **2**, 184-194, doi:10.1080/02723646.1981.10642213 (2013).

69 Moriasi, D. N. *et al.* Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE* **50**, 885-900, doi:10.13031/2013.23153 (2007).

70 Taylor, K. E. Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res-Atmos* **106**, 7183-7192, doi:10.1029/2000jd900719 (2001).