Global downscaled projections for climate impacts research (GDPCIR): preserving ~~extremes~~ quantile trends for modeling future climate impacts

Diana R. Gergel[1,2], Steven B. Malevich[2], Kelly E. McCusker[2], Emile Tenezakis[2], Michael T. Delgado[2], Meredith Fish[3], and Robert E. Kopp[3]

[1]BlackRock, 601 Union Street, Seattle, WA 98101 USA
[2]Rhodium Group, 5 Columbus Circle, New York, NY 10019 USA
[3]Department of Earth and Planetary Sciences and Rutgers Institute of Earth, Ocean and Atmospheric Sciences, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854 USA

Correspondence: Diana R. Gergel (dgergel@gmail.com)

Abstract.

Global climate models (GCMs) are important tools for understanding the climate system and how it is projected to evolve under scenario-driven emissions pathways. Their output is widely used in climate impacts research for modeling the current and future effects of climate change. However, climate model output remains coarse in relation to the high-resolution climate data needed for climate impacts studies, and it also exhibits biases relative to observational data. Treatment of the distribution tails is a key challenge in existing bias-adjusted and downscaled climate datasets available at a global scale; many of these datasets used quantile mapping techniques that were known to dampen or amplify trends in the tails. In this study, we apply the ~~trend-preserving~~ Quantile Delta Mapping (QDM) ~~bias-adjustment method (Cannon et al., 2015) and develop a new~~ method (Cannon et al., 2015) for bias adjustment. After bias adjustment, we apply a new spatial downscaling method called ~~the~~ Quantile-Preserving Localized-Analog Downscaling (QPLAD)~~method that also preserves~~ , designed to preserve trends in the distribution tails. Both methods are integrated into a transparent and reproducible software pipeline, which we apply to global, daily ~~model output for surface variables~~ GCM surface variable outputs (maximum and minimum temperature and total precipitation) from the Coupled Model Intercomparison Project Phase 6 (CMIP6) experiments (O'Neill et al., 2016) for the historical experiment and four future emissions scenarios ranging from aggressive mitigation to no mitigation: SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 (Riahi et al., 2017). We use European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 ~~(Hersbach et al., 2018)~~ (Hersbach et al., 2020) temperature and precipitation reanalysis ~~data~~ as the reference dataset over the Sixth Intergovernmental Panel on Climate Change (IPCC) Assessment Report (AR6) reference period, 1995–2014. We produce bias-adjusted and downscaled data over the historical period (1950–2014) and ~~for four~~ the future emissions pathways (2015–2100) for 25 ~~models~~ GCMs in total. The output dataset ~~of this study~~ is the Global Downscaled Projections for Climate Impacts Research (GDPCIR), a global, daily, 0.25° horizontal-resolution product which is publicly available and hosted on Microsoft AI for Earth's Planetary Computer (https://planetarycomputer.microsoft.com/dataset/group/cil-gdpcir/).

1 Introduction

Global climate models (GCMs) are essential for studying the climate system and how it will evolve in the future. Simulations from the Coupled Model Intercomparison Project (CMIP) ~~experiments~~ are widely used in climate impact studies, exploring human health (e.g., Carleton et al., 2022), energy (e.g., Rode et al., 2021), labor productivity (e.g., Parsons et al., 2022), agriculture crop yields (e.g., Müller et al., 2021), and the impacts of climate change on GDP losses globally (e.g., Warren et al., 2021). However, despite progress in climate modeling, GCM simulations often exhibit systematic error (bias) relative to observations (François et al., 2020a) due to coarse spatiotemporal resolution, simplified physics, thermodynamic schemes, and incomplete and/or poorly understood representation of climate system processes (Sillmann et al., 2013). GCM simulations, relative to historical observations, can have large errors in their means and variance, and even larger biases in extreme values (Cannon et al., 2015). ~~All of these~~ These biases are challenging to impacts studies examining the future evolution of local climate impacts. This challenge is magnified when trying to understand how a particular climate ~~extreme~~ signal will affect a given outcome, for example, how changes in extreme temperatures will affect mortality rates in a location~~expected to experience large temperature increases throughout the twenty-first century~~. To explore these questions, it is necessary to have high-resolution climate projections for multiple emissions pathways with a statistical distribution consistent with historical observations.

To fill this need for climate impact assessments, statistical bias adjustment (BA) and downscaling methods ~~are used to adjust~~ have been applied to reduce biases and add high-resolution spatial information to ~~the coarse resolution of GCM simulations~~ GCM simulations (Pierce et al., 2015). BA methods adjust the difference in statistical properties between model simulations and observations ~~. In this context, downscaling is the process of moving from the coarse resolution of the GCM to the high-resolution local information needed to use as inputs for impacts models. The majority of statistical BA methods adjust the GCM simulation distribution by operating~~ or reanalysis data. Methods vary widely in complexity, from simpler parametric methods that operate only on the mean ~~, variance , higher moments, or quantiles (François et al., 2020a). These methods , particularly traditional quantile mapping ones, are known to affect trends in extreme quantiles differently than trends in the mean, thus degrading results at the distribution tails (Maurer and Pierce, 2014; Lehner et al., 2021; Holthuijzen et al., 2022).~~ ~~Standard~~ or the mean and variance to trend-preserving methods (Casanueva et al., 2020; Iturbide et al., 2022; François et al., 2020b) or methods that use deep learning neural networks (Baño-Medina et al., 2021). Other BA methods have been developed and applied extensively as well, such as the cumulative distribution function transform (CDF-t) (e.g., Michelangeli et al., 2009) and equidistant quantile mapping (e.g., Li et al., 2010), and compared with other methods over Europe in the VALUE study (Gutiérrez et al., 2019). A key result from the VALUE study was that the time window used in calibration was one of the most influential factors. Generally, quantile mapping (QM) methods ~~were used to create the~~ have been widely used in climate impacts studies, and particularly at the global scale due to their lower computational expense relative to other methods (Pierce et al., 2015). A quantile mapping approach that only corrects for the mean and variance, the BCSD method, was used for example in the popular NASA Earth Exchange (NEX) Global Daily Downscaled Projections (GDDP) global daily CMIP5 dataset (Thrasher et al., 2012)~~, which uses the Bias Correction and Spatial Disaggregation (BCSD) QM approach~~. However, QM methods that operate only on the mean, such as BCSD, may affect trends in high (and low) quantiles differently than trends

in the mean, often degrading results at the distribution tails (Maurer and Pierce, 2014; Lehner et al., 2021; Holthuijzen et al., 2022; Sanabria .

To mitigate this, QM approaches that are trend-preserving in the quantiles have been developed (Casanueva et al., 2020). A key example of these methods - and the bias adjustment method we apply in this study - is the Quantile Delta Mapping (QDM) method (Cannon et al., 2015). Although generally trend-preserving methods have been found to better preserve the climate change signal for climate change impacts indices, they also rely heavily on the observations or reanalysis dataset used for reference (Casanueva et al., 2020), and there is not a consensus in the literature that trend-preserving methods necessarily perform better for climate extremes(Iturbide et al., 2022). However, the majority of studies have supported the need for trend-preserving methods (e.g., Casanueva et al., 2019; Qian and Chang, 2021) to better represent the temperature extremes that have the most severe impacts. Moreover, Lehner et al. (2023) found that QDM was one of the ~~method differentially affects trends in extreme quantiles , degrading~~ best-performing BA methods for representing changes in threshold metrics. An additional question also worthy of mention and subject to extensive debate is whether or not the climate signal from the GCMs should even be preserved, as the ~~tails of the distribution. BCSD for example adjusts GCM simulations to have the same cumulative distribution function (CDF) as the reference dataset (for each day of the year) and then imposes a 9-year running monthly mean trend from the GCM on the adjusted-day value (Thrasher et al., 2012). This does not preserve trends in the tails of the distribution because the GCM trend imposed is the mean-monthly trend. Maurer and Pierce (2014) found that QM modifications of projected trends in seasonal mean model precipitation could be as large as the actual GCM-projected changes.~~ future signal is of course not known (Pierce et al., 2015). However, Casanueva et al. (2020) found that QDM in particular performed better in preserving trends for moderate to extreme climate indices, which was one of our key goals in designing this study.

~~Downscaling~~ Statistical downscaling faces similar challenges to ~~bias-adjustment~~ BA methods (Cannon et al., 2020). ~~Downscaling in BCSD, for example, dampens trends in the tails of the higher-resolution gridcells because the method involves bilinearly interpolating scaling factors computed as the difference (or ratio) of GCM to reanalysis climatologies on a per-pixel multi-decade basis.~~ Because of these challenges, many studies in the impacts literature stop short of downscaling (Maraun, 2016). ~~Others, such as Lange (2019), combine trend-preserving bias adjustment with statistical downscaling, but the final resolution of the downscaled data remains relatively coarse (0.5°). This effect is undesirable for climate impacts modeling because it dampens or amplifies trends in the tails, which are crucial to understanding how climate extremes and their associated impacts will evolve for various emissions pathways (Sanabria et al., 2022; Lanzante et al., 2020). Returning to the mortality impacts example, it is not sufficient to project the future mean rise in temperature at a given location. Trends for the hottest days must be preserved to understand and project mortality impacts.~~

~~Several~~ Notwithstanding, several CMIP6 ~~downscaling~~ bias-adjusted and downscaled datasets produced in the past ~~several~~ few years have attempted to address these issues, but they have either been limited in geographic scope (e.g., Supharatid et al., 2022), global but at a coarse spatial resolution (e.g., Xu et al., 2021), or global but preserving only mean trends (e.g., Thrasher et al., 2021). ~~Moreover,~~ Jupiter Intelligence (https://jupiterintel.com/), a climate risk-focused company in the private sector, has made a bias-adjusted CMIP6 dataset available for commercial applications~~. Unfortunately, its methods are neither published nor transparent~~ , however, its methods have not been published and the dataset is not publicly available

(Hacker, 2021). The ~~Intersectoral Impact Model Intercomparison Project (ISIMIP)~~ ISIMIP downscaled dataset (Lange, 2019) uses quantile trend-preserving bias adjustment and downscaling approaches on daily data at a global scale ~~, but it is only available for a limited number of GCMs,~~ for a larger set of variables than GDPCIR but a smaller set of GCMs at a 0.5° spatial resolution (Lange, 2021). In the past year, ~~several downscaled datasets for CMIP6 at a higher spatial resolution have been released. NASA updated~~ NASA released an updated version of the NASA-NEX dataset using CMIP6 projections ~~and released~~ (Thrasher et al., 2022). However, the new dataset ~~in early 2022 (Thrasher et al., 2022) but~~ still relies on the BCSD method ~~. Additionally, the updated NASA-NEX GDDP dataset still relies on the~~ and uses the Global Meteorological Forcing Dataset (GMFD) (Sheffield et al., 2006) as a reference dataset, a reanalysis dataset that ~~now dated and~~ is no longer maintained and is no longer widely used in bias adjustment and downscaling (Hassler and Lauer, 2021). CarbonPlan, a not-for-profit organization focused on climate and carbon capture research, has also released a global downscaled CMIP6 dataset using four distinct statistical downscaling methods with publicly available code (https://docs.carbonplan.org/cmip6-downscaling) ~~. While this is an important contribution to method transparency and comparison, the monthly resolution of the dataset is prohibitively coarse for many impacts modeling applications and the dataset is only available for a subset of six GCMs .~~ at a monthly resolution and for a subset of six GCMs. These datasets are key contributions to impacts research but a gap remains for a global product that preserves GCM quantile changes and is available at a high temporal and spatial resolution for a broad set of CMIP6 GCMs and emissions scenarios. This study aims to fill that gap.

~~To ameliorate these challenges for impacts modelers, this study uses statistical bias adjustment and downscaling methods that explicitly preserve relative changes in GCM simulation quantiles (Cannon et al., 2015). We use the quantile delta mapping (QDM)~~ Consequently, in this study we used the QDM method (Cannon et al., 2015) for bias adjustment ~~. For downscaling , we introduce~~ and for downscaling we designed the Quantile-Preserving Localized-Analog Downscaling (QPLAD) method, a ~~novel~~ statistical downscaling algorithm that applies a local analog-mapping approach to preserve quantile trends at the fine resolution. We explain the method and implementation further below. We have made the QDM and QPLAD methods and code transparent and reproducible via tagged code releases for the full pipeline, available in Github (https://github.com/ClimateImpactLab/downscaleCMIP6) and archived via Zenodo (https://doi.org/10.5281/zenodo.6403794). The dataset described herein, titled Global Downscaled Projections for Climate Impacts Research (GDPCIR), is, to our knowledge, the most comprehensive and high-resolution dataset that exists for CMIP6 that preserves quantile trends. ~~The preservation of quantile trends makes the dataset better suited for impacts modeling than other downscaled CMIP6 datasets since high and low tail trends are not dampened or amplified by only accounting for mean projected changes~~We hope that the publicly available and transparent code and pipeline infrastructure will be helpful for researchers who wish to bias-adjust and downscale additional variables, GCMs, or experiments. Alternatively, if additional meteorological variables, such as longwave and shortwave radiation, surface pressure and relative and specific humidity are needed for a given impacts modeling application, or subdaily temperature and precipitation projections, a meteorological disaggregation method can be used (Bennett et al., 2020).

The remainder of the paper is structured as follows. In Section 2, we describe the climate simulations and reference dataset. In Section 3, we describe the QDM-QPLAD bias adjustment and downscaling methods. Section 4 describes our downscaling pipeline and efforts to make ~~our pipeline~~ its implementation on commercial cloud computing platforms transparent and repro-

ducible. In Section 5, we explore trends and ~~extremes~~ quantile changes in the dataset at the global, city, and "admin1" (country) levels. ~~In the final section, we detail applications for econometric research for climate risk and other impacts modeling areas.~~

2 Climate data

2.1 Simulation data

We used the CMIP6 ~~GCM~~ historical and ScenarioMIP experiments (Eyring et al., 2016; O'Neill et al., 2016) as simulation data~~. We obtained the data~~ , obtained from the Google Cloud CMIP6 collection (https://pangeo-data.github.io/pangeo-cmip6-cloud/). This contains a subset of CMIP6 output migrated from the Earth System Grid Federation (ESGF) as part of a collaboration between the Pangeo Consortium (https://pangeo.io/), Lamont-Doherty Earth Observatory (LDEO) and Google Cloud. The migration to Google Cloud included converting data from NetCDF format (https://www.unidata.ucar.edu/software/netcdf/) to the cloud-optimized Zarr store format (https://zarr.readthedocs.io/en/stable/api/storage.html), and standardizing across dimensions, coordinates, and grids to ensure that ~~model~~ GCM output would be analysis-ready and cloud-hosted for streamlined use in scientific analysis ~~Abernathey et al. (2021)~~(Abernathey et al., 2021). CMIP6 ~~output~~ GCMs available through the ESGF but not in the CMIP6 Google Cloud collection ~~was excluded because it was~~ were excluded because they were not analysis-ready and cloud-optimized, and as such, could not run through our cloud-based downscaling pipeline. We also excluded GCMs included in the CMIP6 Google Cloud collection for which daily output was not available or other issues were found. Similarly, if an SSP is missing for a given GCM, that indicates that it was either not available in the CMIP6 Google Cloud collection or issues with the available data were found. Table B1 lists all GCMs with ScenarioMIP and CMIP experiment output participating in CMIP6 and details why certain GCMs were excluded. The GCMs included in the GDPCIR dataset provide broad coverage across the spread of CMIP6 models, including GCMs with high equilibrium climate sensitivity (ECS) such as CanESM5, HadGEM3-GC31-LL, and UKESM1-0-LL, and those with low ECS such as INM-CM4-8 and INM-CM5-0 (Meehl et al., 2020).

In addition to the last 65 years of the historical CMIP experiment, we included four 21st century ScenarioMIP experiments so as to span a range of possible future climate trajectories. These trajectories are defined by a combination of Shared Socioeconomic Pathways (SSPs) and Representative Concentration Pathways (RCPs): SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 (Riahi et al., 2017) ~~.~~and make up the "Tier 1", or top priority, experiments in CMIP6. For each GCM, we select a single ensemble member. When it was available in the Google Cloud CMIP6 collection, we used the *r1i1p1f1* ensemble member (see Table A1 for ensemble members for each GCM and Table B1 for further information). We did not include simulations that ~~have~~ had output populated with NaNs for some years or did not have complete spatiotemporal coverage. For example, the Hammoz-Consortium ~~model~~ GCM is not included because its temperature output available through the Google Cloud CMIP6 collection did not extend past 2055. We also ~~do~~ did not include the Community Earth System Model from the National Center for Atmospheric Research (NCAR) because there was no historical daily surface variable output available through NCAR for the historical experiment. A full list of reasons why some GCMs were excluded for quality control can be found in Table B1. We perform bias adjustment and downscaling on a subset of the historical CMIP experiment (1950–2014) and ScenarioMIP

scenarios (2015–2100) with a historical training period from 1995 to 2014, consistent with the IPCC AR6 reference period. The full dataset includes 25 GCMs (Table 1), with downscaled output for all four SSPs available for the majority of those GCMs. ~~If an SSP is missing for a given GCM, that indicates that it was either not available in the CMIP6 Google Cloud collection or we found issues with the data available. If a GCM from a modeling center that participated in the CMIP6 experiments is missing, that indicates that the GCM did not have daily surface variable output available for maximum and minimum temperature and surface precipitation in the CMIP6 Google Cloud collection as of 15 November 2021, or the output that was available contained data issues as discussed above.~~

We standardize calendars across all GCMs included in the dataset by converting them to a 365-day (e.g., "no-leap") calendar. Leap days are removed for GCMs with 366-day calendars. For the two GCMs on 360-day calendars (the Hadley Centre models), we follow the method in ~~Pierce et al. (2014)~~the downscaled CMIP5 LOCA dataset Pierce et al. (2014) described on the LOCA website (Pierce, 2021). Five days per year are chosen randomly to add to the calendar, each in a given fifth of the year. Feb. 29th is always missing. For each of the days that are added, a day value is produced by averaging the adjacent days. For example, if Feb. 16th is the day added in the first fifth of the year for a given year, it will be the average of Feb. 15th and Feb. 17th. Choosing a random day in a fifth of the year versus the same five days every year mitigates overall undesired effects on the statistics of particular days of the year or annual cycle statistics when converting from a 360-day to 365-day calendar.

GCM	**Institution**	**Ensemble member**	**SSPs**			
			SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5
ACCESS-ESM1-5	Commonwealth Scientific and Industrial Research Organisation, Aspendale, Victoria, Australia	r1i1p1f1	✓	✓	✓	X
ACCESS-CM2	Commonwealth Scientific and Industrial Research Organisation, Aspendale, Victoria, Australia	r1i1p1f1	X	✓	✓	X
BCC-CSM2-MR	Beijing Climate Center, Beijing, China	r1i1p1f1	✓	✓	✓	✓
CanESM5	Canadian Centre for Climate Modelling and Analysis, Victoria, BC	r1i1p1f1	✓	✓	✓	✓
CMCC-CM2-SR5	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Lecce, Italy	r1i1p1f1	✓	✓	✓	✓

Table 1 continued from previous page

GCM	Institution	Ensemble member	SSPs			
			SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5
CMCC-ESM2	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Lecce, Italy	r1i1p1f1	✓	✓	✓	✓
EC-Earth3	EC-Earth-Consortium	r1i1p1f1	✓	✓	✓	✓
EC-Earth3-AerChem	EC-Earth-Consortium	r1i1p1f1	X	X	✓	X
EC-Earth3-CC	EC-Earth-Consortium	r1i1p1f1	X	✓	X	✓
EC-Earth3-Veg	EC-Earth-Consortium	r1i1p1f1	✓	✓	✓	✓
EC-Earth3-Veg-LR	EC-Earth-Consortium	r1i1p1f1	✓	✓	✓	✓
FGOALS-g3	Chinese Academy of Sciences, Beijing, China	r1i1p1f1	✓	✓	✓	✓
GFDL-CM4	NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA	r1i1p1f1	X	✓	X	✓
GFDL-ESM4	NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA	r1i1p1f1	✓	✓	✓	✓
HadGEM3-GC31-LL	Met Office Hadley Centre, Exeter, Devon, United Kingdom	r1i1p1f3	✓	✓	X	✓
INM-CM4-8	Russian Academy of Science, Moscow, Russia	r1i1p1f1	✓	✓	✓	✓
INM-CM5-0	Russian Academy of Science, Moscow, Russia	r1i1p1f1	✓	✓	✓	✓
MPI-ESM1-2-HR	Deutscher Wetterdienst, Offenbach am Main, Germany	r1i1p1f1	✓	X	X	✓
MPI-ESM1-2-LR	Max Planck Institute for Meteorology, Hamburg, Germany	r1i1p1f1	✓	✓	✓	✓

Table 1 continued from previous page

GCM	Institution	Ensemble member	SSPs			
			SSP1-2.6	SSP2-4.5	SSP3-7.0	SSP5-8.5
MIROC-ES2L	Japan Agency for Marine-Earth Science and Technology, Kanagawa, Japan	r1i1p1f1	✓	✓	✓	✓
MIROC6	Japan Agency for Marine-Earth Science and Technology, Kanagawa, Japan	r1i1p1f1	✓	✓	✓	✓
NESM3	Nanjing University of Information Science and Technology, Nanjing, China	r1i1p1f1	✓	✓	X	✓
NorESM2-LM	NorESM Climate Modeling Consortium, Oslo, Norway	r1i1p1f1	✓	✓	✓	✓
NorESM2-MM	NorESM Climate Modeling Consortium, Oslo, Norway	r1i1p1f1	✓	✓	✓	✓
UKESM1-0-LL	Met Office Hadley Centre, Exeter, Devon, United Kingdom	r1i1p1f2	✓	✓	✓	✓

Table 1: Full list of Coupled Model Intercomparison Project (CMIP6) GCMs included in the GDPCIR dataset along with their corresponding institutions and the available SSPs for each GCM.

2.2 Reference data

We use the ~~ECMWF~~ European Center for Medium-Range Weather Forecasting (ECMWF) Reanalysis v5 (ERA5) ~~(Hersbach et al., 2018)~~ as the historical reference dataset for bias adjustment and downscaling ~~.~~ (Hersbach et al., 2018, 2020). While there are shortcomings for any reanalysis dataset, our goal was to select a reference dataset that performed well in comparison to observations and other reanalysis datasets particularly for extreme temperatures and precipitation in highly populated areas. Sheridan et al. (2020) compared observed extreme temperature days in the United States and Canada to three reanalysis products and found that ERA5 matched station data most closely, even in comparison to its higher-resolution counterpart, ERA5-Land. Other studies (e.g., Mistry et al., 2022; McNicholl et al., 2022) compared ERA5 temperatures globally to station observations and found that it performed well, with some reduced performance in tropical areas. Similar biases for precipitation in the tropics have also been noted; Hassler and Lauer (2021) and Tarek et al. (2020) found that ERA5 overestimated precipitation rates over the Atlantic Ocean and Indian Ocean. Nevertheless, the bias in ERA5 was lower than in other reanalyses products.

In addition to the performance of ERA5 in relation to other reanalysis datasets, it is also operationally maintained in near-real-time by ECMWF and cloud-optimized, available as a zarr store from Google Cloud and AWS. ERA5 reanalysis data is produced and archived on a reduced Gaussian grid with a resolution of N320, meaning that there are 320 quasi-regularly spaced latitude points from pole to equator, at a 31 km (0.28°) resolution. We obtained global, hourly temperature and precipitation estimates from 1979 through 2018 on a regular (latitude-longitude) Gaussian grid at the same resolution to minimize the impact of interpolation from the Copernicus Data Service regridder, particularly on precipitation. We ~~derive~~ derived daily maximum and minimum temperatures by taking the daily maximum and minimum of the hourly values and total daily precipitation by taking the sum of hourly values. ERA5 hourly precipitation values represent cumulative precipitation during the preceding hour, thus cumulative daily precipitation for a given day is the sum of hourly values minus the first hour and including the first hour of the following day. We then ~~subset~~ subsetted the ERA5 daily surface variables to 1995–2014 to be consistent with the historical reference period used in Masson-Delmotte et al. (2021), finally, we ~~remove~~ removed leap days. We ~~use~~ used the resulting 20-year ERA5 dataset as the historical reference data for bias adjustment and downscaling.

3 Methods

3.1 Statistical bias adjustment with the QDM method

In this study, our goal was to emphasize downscaling and ~~bias-adjustment~~ bias adjustment methods that better preserve the ~~extreme~~ high tails of distributions, but within the constraints of the level of method complexity that could be undertaken given the scale of this project. Though some multivariate statistical methods might have better preserved joint correlations between variables, such as Multivariate Bias adjustment (Cannon, 2018), the computational intensity of even running a univariate method at this scale precluded the choice of a multivariate method. Some studies have also found that multivariate methods may lead to degraded results for one or more variables (e.g., temperature) that are being jointly bias-adjusted and/or downscaled, and also may perform poorly under projected climate change due to bias nonstationarity (Van de Velde et al., 2020; François et al., 2020a). Choosing a method that would not degrade temperature projections was necessary given the role of temperature as a key driver of future climate impacts.

With these constraints in mind, and after evaluating a number of statistical methods and their effects on the distribution tails, we chose the QDM method. The QDM ~~statistical bias-adjustment method preserves changes~~ method preserves model-projected trends in quantiles by applying simulated changes in the quantiles on top of the historical reference distribution (Cannon et al., 2015). Absolute changes or relative changes are preserved for additive or multiplicative variables, respectively. As a result, treatment of the tails is ~~better than in standard quantile mapping as well as in~~ improved over other forms of quantile mapping such as empirical quantile mapping (EQM), detrended quantile mapping (DQM), and various parametric and non-parametric variants of each (Qian and Chang, 2021). ~~This rationale, combined with the QDM methodbeing relatively computationally inexpensive compared~~ A limitation of the method, however, is that it is highly sensitive to the choice of reference dataset, especially for precipitation, and extreme temperature and precipitation indices (Casanueva et al., 2020). As a result, the biases in the reference data presented in Section 2.2 are transferred to the bias-adjusted and downscaled dataset, which is a limitation

of the final dataset. Results presented here should be taken in that context. Nonetheless, its performance at the tails and relatively inexpensive compute footprint in comparison to multivariate quantile mapping or machine learning-based methods ~~,~~ makes it a favorable method choice for a project of this scope and aim.

The QDM method adjusts the bias in projected values for a historical or future time period by first shifting the distribution to be consistent with the reference dataset and then imposing the relative ~~GCM-projected~~ model-projected trend, resulting in a bias-adjusted projection that has a distribution consistent with that of the reference dataset and also has a relative trend consistent with the source ~~GCM~~model, for a given quantile. In detail, following the notation in Cannon et al. (2015), let $F_{m,p}[\cdot]$, $F_{m,h}[\cdot]$ and $F_{o,h}[\cdot]$ denote, respectively, the CDF from model m in future period p, the CDF from model m in the historical period h and the CDF from the reference data o in the historical period h. Let $x_{m,p}$ be a modeled future value at time t (for example, maximum temperature on 13 March 2025), and let $x^*_{m,p}$ be the associated adjusted value for the same future date. In addition, let $\tau_{m,p}$ denote the non-exceedance probability associated with $x_{m,p}$, i.e $\tau_{m,p} = F_{m,p}[x_{m,p}]$. $F^{-1}[\cdot]$ represents the inverse CDF. The adjusted value is defined as follows for an additive variable:

$$x^*_{m,p}(t) = x_{m,p}(t) + (F^{-1}_{o,h}[\tau_{m,p}(t)] - F^{-1}_{m,h}[\tau_{m,p}(t)]) \tag{1}$$

Rearranging the right-hand side shows that Equation 1 is equivalent to introducing the ~~GCM-projected~~ model-projected change at a given quantile ($\tau_{m,p}$) on top of the reference data value at that quantile:

$$x^*_{m,p}(t) = \underbrace{F^{-1}_{o,h}[\tau_{m,p}(t)]}_{\text{reference value at model quantile}} + \underbrace{(x_{m,p}(t) - F^{-1}_{m,h}[\tau_{m,p}(t)])}_{\text{model trend in quantile}} \tag{2}$$

For a multiplicative variable such as precipitation, the right-hand side in equations (1) and (2) becomes multiplicative rather than additive, i.e., Equation 1 becomes $x^*_{m,p} = x_{m,p} * F^{-1}_{o,h}[\tau_{m,p}]/F^{-1}_{m,h}[\tau_{m,p}]$. This results in ~~GCM~~ model projections that preserve each ~~GCM~~model's change in distribution shape (including ~~extremes~~high and low quantiles) while simultaneously making the training-period distribution consistent with the reference dataset.

3.2 Statistical trend-preserving downscaling with the QPLAD method

A key goal of downscaling for climate impacts is increasing spatial resolution in a way that both preserves climate trends and introduces realistic local climatology and variability. In observations, the climate signal at a coarser scale will always – by definition – represent a smoothed version of local climate trends. Similarly, high-resolution climate projections need to have a distribution that is consistent with locally observed climate. Downscaling may break consistency with the original GCM dynamics, but this is necessary to produce the spatial heterogeneity required for modeling climate impacts (Maraun and Widmann, 2018). ~~Traditional downscaling~~ Downscaling methods typically work by introducing the climatological fine reference spatial pattern to the coarse resolution simulated data, as a difference or ratio between fine and coarse. This can have the effect of modifying trends and spatial patterns in the tails of the simulated distribution. To address this, we developed

the QPLAD method. The QPLAD method uses the difference in empirical quantiles of the reference data ~~–~~ each quantile is a given day, or "analog" of the reference training period – at coarse and fine resolution to downscale the coarse resolution GCM simulations. The outcome is a downscaled dataset that preserves the changes in coarse GCM quantiles in time while also reflecting the within-coarse-grid cell spatial heterogeneity from the fine reference data. As a result, localized, extreme changes in the downscaled data are consistent with the GCM projections.

Formally, QPLAD involves computing and applying "adjustment factors" for each quantile in the reference data over the training period. First, an empirical CDF, $F_{o,h,c}[\cdot]$, of the reference data o is calculated, over the training period h at the relatively "coarse" resolution c at which bias adjustment was applied to GCMs (1° in this study). The method described here in the GDPCIR pipeline assumes that QDM bias adjustment was performed at a coarser resolution than the target resolution for downscaling, but theoretically, one could apply QPLAD to unadjusted GCM simulations as well. Further detail on our implementation can be found in Section 4.3. The number of empirical quantiles q is equal to the number of ~~time steps~~ timesteps in the training period n (e.g., a training period of 20 years with a 31-day rolling window has $n = q = 20 * 31 = 651$, since each empirical quantile corresponds to a day in the training period). Next, the reference data at "fine" resolution is sorted into the same order as the coarse resolution empirical CDF, $B_{o,h,f}[\cdot]$, where the set B represents the fine reference ~~time steps~~ timesteps (days) sorted the same as the coarse CDF $F_{o,h,c}[\cdot]$ and f refers to the fine resolution. Adjustment factors are then calculated as the difference or ratio (for an additive or multiplicative variable, respectively) between the fine and coarse resolution values for each historical analog day in the sorted data (i.e., for each empirical quantile). For an additive variable, adjustment factors af are as follows:

$$af(q_c) = B^{-1}_{o,h,f}(q_c) - F^{-1}_{o,h,c}(q_c) \tag{3}$$

for all coarse empirical quantiles q_c, where $B^{-1}[\cdot]$ represents the fine reference values (rather than quantiles) in sorted order. Similar to QDM detailed above, the adjustment factors are applied to coarse resolution simulations by first determining the quantile of a given time step's value, $F_{m,p,c}(x_{m,p,c}) = \tau_{m,p}$ where $\tau_{m,p}$ is the non-exceedance probability associated with the value $x_{m,p,c}$. For an additive variable, the downscaled value for a given time step t in the projection simulation is defined as:

$$\tilde{x}_{m,p,f}(t) = x_{m,p,c}(t) + af_{q_c} \tag{4}$$

This results in high-resolution, downscaled projections where the subgrid cell heterogeneity from the original coarse resolution contains the more extreme days from the higher-resolution reference data. By definition, all of the target fine-resolution grid cells encompassed by the coarse-resolution grid cell will have downscaled values that average to the value for the coarse grid cell. No spatial smoothing is applied in order to maintain the original GCM quantile changes. In this way, "quantile-preserving" refers to maintaining the quantile information from the coarse-resolution day, and "localized" refers to the fine-resolution historical analogs located within a coarse-resolution grid cell. The method produces downscaled projections that add high-resolution information from the reference data training period and ensure that the fine-resolution spatial make-up of

more extreme days from the coarse simulations are coherent and analogous to those found in the reference data. Thus, extreme days are also preserved in the downscaled projections in a relative sense (in a similar manner to QDM). Note that the QDM and QPLAD methods, which explicitly preserve changes in the quantiles, do not necessarily preserve model-projected changes in the mean due to using empirical CDFs, which is a non-parametric approach. Taking a parametric approach and using an analytical CDF would preserve changes in the mean, but would also impose a distribution to the CDFs. As Lehner et al. (2021) discuss, the question of whether to take a parametric or non-parametric approach in bias adjustment is an active area of research, but the non-parametric approach in the QDM and QPLAD methods is more common and generally preferred.

3.3 Wet day frequency adjustment

~~The discrete and continuous nature of the~~ In bias-adjusting and downscaling daily precipitation data~~needs to be addressed when applying bias adjustment and downscaling. Moreover,~~ , the skewness of precipitation distributions must be accounted for (Maraun, 2013). GCMs are known to have a "drizzle day" problem where the frequency of wet days with low precipitation in GCMs has a high positive bias relative to observations (Dai, 2006). To ~~account for these issues~~address this issue, we apply a "pre" wet day frequency (WDF) adjustment to both ~~reference and daily~~ daily reference and GCM data after regridding both datasets to the 1° ~~bias adjustment~~ grid and before bias adjusting. We apply a second "post"" WDF adjustment after QPLAD downscaling where all downscaled daily precipitation values below 1.0 mm day^{-1} are replaced by 0 mm day^{-1}.

~~Our approach~~ The approach here is modified from Cannon et al. (2015). For daily reanalysis and GCM precipitation before bias adjustment, all values at the 1° grid that are less than a specified threshold are replaced by nonzero uniform random values less than the threshold. Initially, we used the same threshold and nonzero uniform random values as Cannon et al. (2015). However, we found that in grid cells where the seasonality and magnitude of daily precipitation values differed by a large amount between model and reanalysis, using the Cannon et al. (2015) threshold (0.05 mm day^{-1}) and adjustment could result in those grid cells having bias-adjusted precipitation values that were not physically realistic for the season and geographic location. Thus we raised the threshold to 1.0 mm day^{-1} (similar to Hempel et al., 2013) and the lower bound of the uniform random distribution from 0 to 0.5 mm day^{-1}. ~~After applying QPLAD downscaling, we then apply a "post" WDF adjustment where all downscaled daily precipitation values below 1.0 mm day^{-1} are replaced by 0 mm day^{-1}.~~

4 ~~Implementing the~~ Bias adjustment and downscaling pipeline implementation

In this section, we describe the pipeline ~~we created~~ for ingesting CMIP6 global, daily surface variable output from the CMIP6 Google Cloud collection, and applying statistical ~~bias-adjustment~~ bias adjustment and downscaling methods to produce a global, daily gridded dataset at a 0.25° horizontal resolution for four emissions pathways, 25 GCMs and three surface variables. The steps to produce the dataset are as follows~~and diagrammed in Figure 1~~: We first standardize the reference dataset and ~~climate model~~ GCM output. We then apply a modified version of the QDM ~~bias-adjustment~~ bias adjustment method at the 1° grid resolution. Next, we apply the QPLAD ~~downscaling~~ method to the bias-adjusted output ~~in order~~ to downscale the data to a 0.25° grid resolution. For precipitation, we apply a wet day frequency adjustment ~~both~~ before bias adjusting and after

downscaling. We apply additional post-processing for all surface variables after downscaling. These steps are diagrammed in Figure 1 and detailed in the remainder of Section 4.

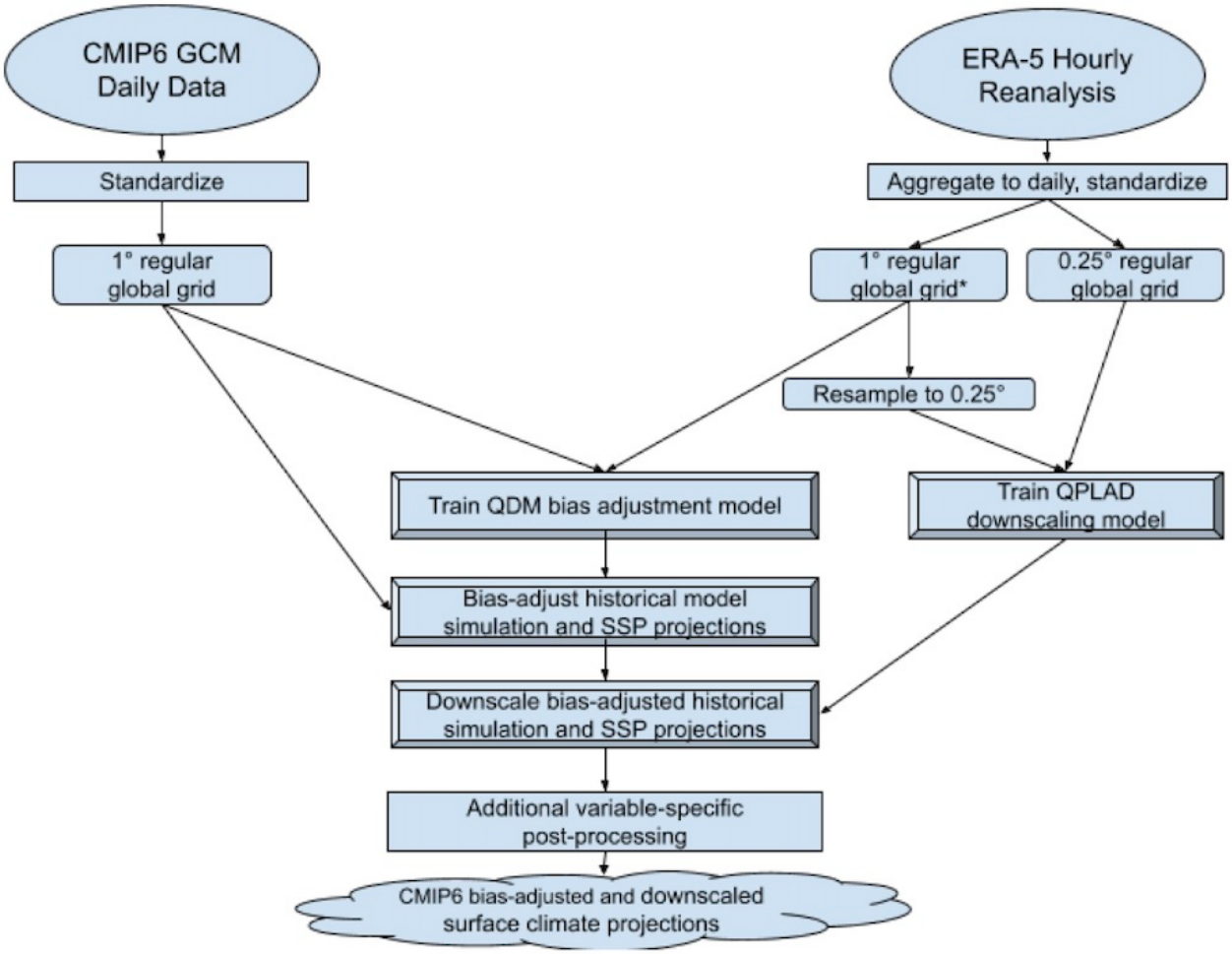

Figure 1. Diagram of CMIP6 bias adjustment and downscaling pipeline.

4.1 Standardizing simulation and reference data

Although the modeling centers participating in the CMIP6 experiments follow Climate and Forecast (CF) conventions (https://cfconventions.org/), significant differences remain in how GCM output is archived. The native resolution of GCMs also varies considerably. For example, four EC-Earth Consortium models have a relatively high resolution (spectral grids approximately 0.7° x 0.7°) and the CCCma CanESM5 GCM has a relatively low resolution (2.5° x 2.5°). Consequently, we begin by standardizing naming, dimensions, and coordinates for all ~~models~~ GCMs and removing leap days. Daily GCM outputs are regridded from the models' native resolution to a regular 1° x 1° global lat-lon grid using the xESMF Python regridding pack-

age (https://xesmf.readthedocs.io/). We use the bilinear regridding method for maximum and minimum surface temperature and first-order conservative area remapping for precipitation to conserve total precipitation between the native GCM grid and the 1° x 1° regular lat-lon grid. Bilinear regridding was chosen for temperature variables since they are continuous quantities, whereas first-order conservative-area regridding was chosen for precipitation for its ability to conserve quantities, thereby not introducing or destroying water. However, it should be noted that generally, any regridding method applied to precipitation alters its statistical properties and can have some undesirable impact on high quantiles (Rajulapati et al., 2021), a caveat that is unavoidable when standardization across GCMs is required.

The same standardization is applied to daily ERA5 reanalysis at the regular Gaussian, F320 grid. We prepare three versions of ERA5 that are used in the QDM-QPLAD method. For QDM bias adjustment, ERA5 is regridded from the F320 grid to the 1° x 1° regular lat-lon grid using the regridding ~~method described above~~ methods described above (bilinear for temperature variables and conservative-area remapping for precipitation). For downscaling ~~,~~ with QPLAD, the same methods are applied to regrid ERA5 ~~is regridded~~ from the F320 grid to the 0.25° x 0.25° regular lat-lon grid ~~using the same regridding methods as in the GCM output~~ ($ERA5_{fine}$), which is the final grid of the GDPCIR dataset. Then, ~~the~~ for use in computing the QPLAD adjustment factors, the 1° x 1° version of ERA5 used in bias adjustment is resampled (e.g., nearest-neighbor regridded) to the 0.25° x 0.25° regular lat-lon grid ($ERA5_{coarse}$).

4.2 Implementation of QDM bias adjustment

~~We bias adjust~~ GCM projections for each variable, GCM, experiment, pixel, year, and day at a 1° x 1° resolution are bias adjusted using the xclim Python package QDM implementation (Logan et al., 2021). To do this, ~~we first train~~ QDM models for each pixel and day of the year ~~using~~ are trained on a rolling 31-day centered window (± 15 days) ~~on~~ of daily ERA5 and GCM historical data ~~.~~ from 1995 to 2014. For ERA5 reference data, we include the last 15 days from 1994 and the first 15 days from 2015 such that each day group contains 620 values (20 years x 31 days)~~for ERA5 reference data~~. For CMIP6 historical data, since the simulation ends in 2014, we do not include the additional 15 days from 2015, ~~nor from~~ or 1994 for consistency. Each trained QDM model (per pixel ~~/~~ and day of year) has 100 equally spaced quantiles in our implementation. We used an additive adjustment for maximum and minimum temperature and a multiplicative adjustment for precipitation. Each variable was bias-adjusted separately.

~~One pitfall with this approach is that minimum temperatures may be larger than maximum temperatures on the same day in some parts of the world with very low diurnal temperature ranges, such as at high latitudes (Thrasher et al., 2012). As a post-processing step, we swapped minimum and maximum temperatures for the small number of pixels and days when the minimum temperature exceeded the maximum temperature after downscaling. This post-processing is described further in Section 4.3.1. We initially tried to avoid this issue by adjusting the maximum temperature using an additive adjustment , separately adjusting the diurnal temperature range (DTR) using a multiplicative adjustment and then deriving the minimum temperature by subtracting DTR from the maximum temperature (following Agbazo and Grenier (2020)). However, we found that this led to unrealistically large DTR values in some parts of the globe, particularly at higher latitudes. Additionally, some raw GCM data had a small number of minimum temperatures greater than the corresponding maximum temperatures, more~~

~~often in polar regions. Bias adjustment then inflated this undesirable behavior. Therefore, we bias-adjusted and downscaled maximum and minimum temperatures separately rather than bias-adjusting DTR.~~

~~We apply the~~ We apply the adjustment factors from the trained QDM models to historical ~~CMIP~~ GCM simulations and future GCM projections for each SSP on a per variable/GCM/pixel/year/day basis. For each year in the GCM data, ~~we group daily data~~ daily data are grouped using a 21-year rolling window and a rolling 31-day window (as in the training step, with ± 15 days). ~~For~~ When adjusting the historical CMIP experiments, ~~we concatenate~~ the first eleven years (2015–2025) of the SSP3-7.0 ~~projection period simulation~~ simulation are concatenated so that the full historical period input dataset encompasses the years 1950–2025 ~~.~~ to accommodate the rolling window in the year 2014. We use SSP3-7.0 to best simulate the current trajectory of emissions since 2015. If SSP3-7.0 output is ~~not available~~ unavailable for a given GCM, we ~~then~~ use SSP2-4.5. For the few ~~models~~ GCMs in which neither SSP3-7.0 nor SP2-4.5 output is available, we use SSP1-2.6. ~~For~~ When adjusting each SSP, ~~we concatenate the~~ the historical simulation's last eleven years (2004–2014) ~~of the CMIP model simulation~~ is concatenated so that the full projection period input dataset encompasses the years 2004–2100 ~~for the rolling 21-year window. Historical GCM~~ to accommodate the rolling window. At the beginning and end of the historical + projection time periods, fewer days can be included in the adjustment step resulting in historical years 1950–1960 ~~have~~ having fewer days in their rolling ~~window, as do projection period~~ windows and projection years 2090–2100, with the exception of GCMs for which ~~model~~ output was available past 2100 in the CMIP6 Google Cloud collection at run-time. For the beginning (ends) of ~~the time period~~each year's 21-year adjustment window, an additional 15 days from the previous (following) year is included such that each day group contains 651 values (21 years x 31 days). We use 100 equally-spaced quantiles as in the training step; adjustment factors for quantiles within the range [0.005, 0.995] are linearly interpolated from the ~~neighboring quantiles and linear~~ nearest computed adjustment factor and constant extrapolation is used to extend the range to 0 and 1 for accommodating the extreme tails. This method is based on the "QMv1" method evaluated by Themeßl et al. (2012) and means that new extreme values can occur in the future period or in the historical period outside of the calibration period. Because this method can rarely result in physically unrealistic extremes, we apply an additional post-processing step described in Section 4.3.1.

4.3 ~~Implementation of QPLAD downscaling~~

One pitfall of applying QDM separately to maximum and minimum temperatures is that minimum temperatures may be larger than maximum temperatures on the same day in some parts of the world with very low diurnal temperature ranges, such as at high latitudes (Thrasher et al., 2012). As a post-processing step, we swapped minimum and maximum temperatures for the small number of pixels and days when the minimum temperature exceeded the maximum temperature after bias adjustment and downscaling. This post-processing is described further in Section 4.3.1. We initially tried to avoid this issue by adjusting the maximum temperature using an additive adjustment, separately adjusting the diurnal temperature range (DTR) using a multiplicative adjustment and then deriving the minimum temperature by subtracting DTR from the maximum temperature, following Agbazo and Grenier (2020). However, we found that this led to unrealistically large DTR values in some parts of the globe, particularly at higher latitudes. Additionally, some raw GCM data had a small number of minimum temperatures greater than the corresponding maximum temperatures, most often in polar regions. Bias adjustment of DTR then further inflated

this undesirable behavior. Therefore, we bias-adjusted and downscaled maximum and minimum temperatures separately rather than bias-adjusting DTR.

4.3 Implementation of QPLAD

After applying QDM bias adjustment, we downscale ~~GCM~~ projections for each variable, ~~model~~GCM, experiment, pixel, year, and day ~~at~~ to a 0.25° x 0.25° resolution, ~~similar to our handling of~~ using a similar approach to the QDM bias adjustment. To facilitate this, we implemented the QPLAD method in a forked version of the ~~xclim~~ `xclim` Python package (Logan et al., 2021) ~~in order~~ to leverage the existing parallelization that we used for QDM~~, and we are in the process of adding the method to the package~~. Before downscaling, ~~we resample~~ the bias-adjusted projections are resampled from the 1° x 1° ~~bias adjustment grid resolution~~ grid to the 0.25° x 0.25° target resolution. ~~For all variables, the~~ The method is consistent ~~to ensure that~~ across variables as each of the 16 0.25° ~~gridcells~~ grid cells contained within each 1° ~~gridcell has~~ grid cell must have the same value. Reanalysis data preparation for QPLAD is described in Section 3.1. ~~Although the QPLAD implementation assumes that bias adjustment was performed at a coarser resolution than the target resolution for downscaling, one could apply QPLAD to unadjusted GCM simulations as well.~~

As in bias adjustment, we use a rolling 31-day window (± 15 days) for each day of the year over the training period for each pixel. ~~We include the~~ The last 15 days from 1994 and the first 15 days from 2015 are included such that each day group contains 620 values (20 years x 31 days). We then downscale historical and future ~~model~~ GCM simulation data using the QPLAD adjustment factors (described in Section 3.2) for each variable, ~~model~~GCM, and experiment on a per pixel ~~/~~ per day basis. Since ~~we use~~ 100 empirical quantiles are used in QDM bias adjustment and 620 ~~quantiles~~ in QPLAD (each corresponding to an analog day), there is ~~not a~~ no 1:1 match between the QDM and QPLAD quantiles. Consequently, for a given day, the ~~closest quantile in QPLAD to the quantile~~ quantile assigned during bias adjustment is ~~selected~~ used to select the nearest QPLAD quantile from the 620 possible adjustment factors for that day of year and pixel. Figure 2 demonstrates the temporal and spatial dimensions of the QPLAD method for maximum temperatures around Miami, Florida. Due to its coastal location, the QPLAD adjustment factors for these pixels will show strong land-sea spatial variation, making it an ideal location to demonstrate the method. Panel 2a shows the sixteen spatial analogs (e.g., adjustment factors) for 15 August from the fine reference data (within one 1° ~~gridcell~~grid cell) corresponding to $\tau_m = 0.33$ and the location of Miami, Florida. It is important to note that the "spatial analogs" are only spatial within a single 1° grid cell. By design, the downscaled values for these sixteen gridcells will average to the bias-adjusted value at the 1° resolution x_m with that quantile for that day of year. Panel 2b zooms in on the 0.25° grid cell containing Miami, Florida, and shows all possible ~~spatial analogs for the same quantile but for all days of the year~~adjustment factors for all quantiles and all days. For most days of the year, the adjustment factor ~~for that day of year~~ is moderating the bias-adjusted value, which is expected given the coastal location of Miami~~and the relatively low quantile~~. Panel 2c ~~shows~~ is a slice of Panel 2b showing all possible analog days for 15 August, e.g.~~all possible~~, all 620 analogs. Finally, Panel 2d shows the bias-adjusted and downscaled time series of maximum temperatures for 2080 with the 15 August values highlighted. The analog day for that quantile ($\tau_m = 0.33$) is -1.5° and was applied additively to the bias-adjusted maximum

temperature value for that day, thus that value is the difference between the bias-adjusted and downscaled temperatures for 15 August 2080 shown in ~~Fig.~~ panel 2d.

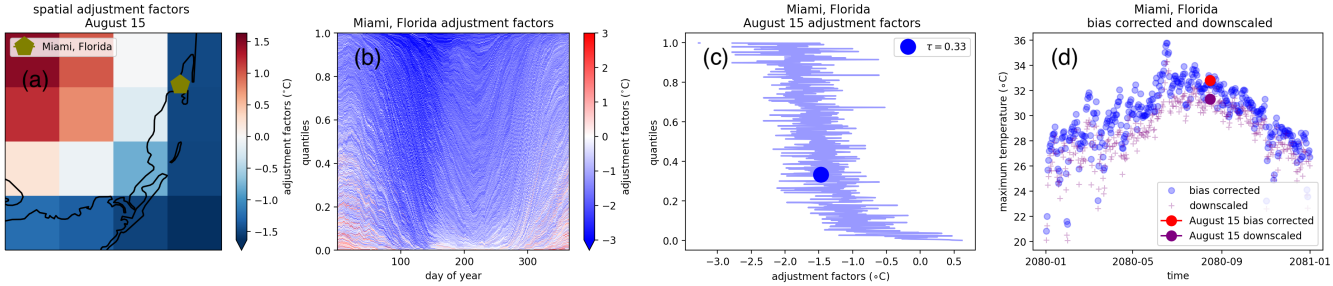

Figure 2. Diagram of QPLAD downscaling method applied to maximum temperature. 15 August is used as an example day grouping with $\tau = 0.33$ corresponding to the actual quantile for 15 August 2080 in the bias-adjusted output for SSP2-4.5. (a) shows spatial analogs for $\tau = 0.33$ for 15 August, (b) shows analogs (adjustment factors) for each day of the year for Miami, Florida, (c) shows all possible adjustment factors for 15 August, and (d) shows the bias-adjusted and downscaled maximum temperature data for 2080 and the difference between the bias-adjusted and downscaled values for 15 August before and after the analog-based adjustment factor for $\tau = 0.33$ has been applied. The example bias-adjusted and downscaled ~~model~~ data comes from the HadGEM3-GC31-LL GCM, produced by the United Kingdom Meteorological Office Hadley Centre.

4.3.1 Additional post-processing

After QPLAD downscaling, we apply an additional post-processing step that is variable-dependent. When DTR is very low in the source GCM, we found that minimum temperature ~~may be greater than~~ could exceed maximum temperature after bias adjustment and downscaling. For the small number of ~~time steps~~ timesteps and gridcells that have this behavior, we swap maximum and minimum temperatures. We found that these conditions ~~occurred infrequently~~ infrequently occurred in high-population areas, being concentrated in the polar oceans, and that this swap did not have a significant effect on seasonal or annual cycle statistics. Figure A1 shows the number of daily timesteps with maximum and minimum temperatures swapped over a 21-year period outside of the calibration period (1960–1980) for all GCMs. The concentration of this in the Arctic and Antarctic and the heterogeneity of spatial patterns across GCMs is apparent. Figure A2 shows the same metric except for SSP3-7.0, 2080–2100.

Precipitation requires a more complex additional bias adjustment for a limited number of grid cells and ~~time steps~~ timesteps globally. Adjustment factors from QDM bias adjustment at higher quantiles (e.g., above the 95th quantile) could become physically unrealistic when seasonal cycle behavior and precipitation magnitudes differed significantly between reanalysis reference data and the GCMs. If the GCM was biased low relative to reanalysis, this bias increased the adjustment factors further. ~~Moreover,~~ Figures A3 and A4 illustrate this behavior for two cities, Delhi, India and Cairo, Egypt, for a single GCM, MIROC6, and for a single scenario, SSP2-4.5. Both figures show full precipitation time series for the reference, raw GCM, bias-adjusted and downscaled GCM, and bias-adjusted, downscaled, and post-processed GCM. The magnitude, as well as the

infrequent occurrence, is particularly apparent in Figure A3. We found that adjustment factors would dramatically increase if the GCM had a strong increase in precipitation signal or if ~~values were very~~ total daily precipitation values were close to zero. However, an increasing signal did not need to be present to incur such a dramatic increase; we also found ~~that this~~ this behavior in the historical period outside of the training period if a given historical period either a) had a trend that was different from the training period trend or b) contained out-of-sample values that were not present in the training period. The confluence of these biases was insidious for GCMs that were ~~downward biased~~ downward-biased relative to reference data and had seasonal precipitation cycles different than those in reference data in the same areas. This was noticeable in the intertropical convergence zone (ITCZ). To correct for these issues in a robust way, we applied a per-pixel post-downscaling adjustment ~~factor~~ at the target resolution that was based on the maximum values of precipitation in the reference data and the fractional (SSP-dependent) increase in maximum precipitation between the historical and projected GCM simulations. Specifically, the maximum precipitation constraint for each pixel is defined as:

$$P_{max}(model, SSP, t) = max(P_{reference,t_1}) \times max\left(1, \frac{max(P_{model,SSP,t_2})}{max(P_{model,historical,t_1})}\right) \quad (5)$$

where t refers to a given day, t_1 is defined as the training period (1995–2014), $model$ refers to a given GCM, SSP represents one of the SSP trajectories, t_2 corresponds to the maximum precipitation in a 21-year rolling window centered on the year that t is in, and $P_{max}(model, SSP, t)$ refers to the maximum allowed precipitation at time t for a given ~~model~~ GCM and SSP. Scaling by the ratio of maximum precipitation in a future 21-year rolling window to historical precipitation allows for the scaling factor to increase during the projection period if the ~~model~~ GCM has an increase in the rolling 21-year maximum daily precipitation for that pixel. However, if the corresponding maximum daily precipitation decreases in the future (e.g., a scaling factor less than 1), the maximum precipitation value in the reference period for that pixel forms the constraint. After this daily constraint term is estimated for each pixel, year, experiment, and ~~model~~GCM, the final result is set equal to the minimum of the original bias-adjusted and downscaled value and this constraint. Figure A5 shows the number of daily timesteps that were clipped in a 21-year historical period (1960–1980) for precipitation. The number ranges from approximately 10-20 timesteps across GCMs. Figure A6 shows the same metric for SSP3-7.0 end-of-century, 2080–2100. The clipping pattern in and near the ITCZ is much more pronounced in this figure, with significant variation across GCMs in the number of clipped timesteps.

4.4 GDPCIR dataset standardization and technical guidelines for users

We save bias-adjusted and downscaled output for each GCM and scenario as a separate zarr store, chunked in time and space to facilitate analysis-ready use. In preparing the final output, we followed Climate and Forecast (CF) convention standards (Hassell et al., 2017) where possible but did not explicitly enforce them in our variable attributes. However, the metadata for each zarr store and variable contains extensive information on source GCM, source URL, and other attributes that may be of interest to the user. Metadata for each zarr store inherits all metadata from its source GCM, such as experiment id, native grid information, ensemble member id, source id, institution id, etc, and then we add additional metadata pertaining to the pipeline,

denoted by the prefix "dc6". Additional metadata fields specific to the pipeline include method information, creation date, licensing information, downscaling pipeline grid details, and pipeline versioning for reproducibility.

In total, the GDPCIR dataset is 23 TB. It is publicly available via Microsoft's Planetary Computer, and notebooks for example usage are provided as well that utilize the Planetary Computer's API (https://planetarycomputer.microsoft.com/dataset/group/cil-gdpcir/). Hosting the GDPCIR dataset via the Planetary Computer allows it to be used in conjunction with a number of other publicly available geospatial datasets.

4.5 Transparency and reproducibility with commercial cloud computing

Our bias-adjusting and downscaling pipeline is novel because it was developed and run entirely with commercial cloud computing infrastructure. Prototypes of the pipeline were built and run on Microsoft Azure, while later production runs used Google Cloud Platform. As such, we wanted the pipeline to be reasonably replicable, open, and not bound to the proprietary hardware or software of a single cloud-computing vendor.

We ran steps of the pipeline in containerized software applications. These containers are a common way to hold software applications with their dependencies so that the application can run reliably on different machines. We orchestrate the containers with Argo Workflows (https://argoproj.github.io/argo-workflows/) on Kubernetes (https://kubernetes.io/), an open-source platform for managing containerized applications on a robust computer cluster that can quickly scale up or down depending on the computing resources needed. Kubernetes is ubiquitous across cloud vendors, helping us to avoid vendor lock-in. The source code for the containers and manifests orchestrating the workflow steps are both available online under an open-source license in public GitHub repositories.

Infrastructure is an additional challenge as it can be practically impossible to make cloud infrastructure truly replicable because commercial cloud vendors iterate their products and platforms very quickly. Despite this, we wanted to be transparent about the cloud infrastructure used for the most intense stages of this pipeline. We provisioned and configured the cloud infrastructure and the Kubernetes clusters from the project's public GitHub repository. This means that pipeline infrastructure and configuration were stored as code and automatically provisioned directly from the repository. We provisioned Google Cloud and Azure resources, including storage and a Kubernetes cluster, using Terraform (https://www.terraform.io/). Terraform is a common open-source tool for provisioning computer infrastructure. Once provisioned, the software on the Kubernetes clusters was managed with ArgoCD (https://argo-cd.readthedocs.io), another open-source tool to deploy Kubernetes resources from the repository in near real-time. Additional information on computing resources is described in Appendix C.

5 Results

In this section, we evaluate the GDPCIR dataset and assess the robustness and performance of the QDM and QPLAD methods. ~~First, we examine behavior in high seasonal quantiles outside of the time windows at which, by design, absolute or relative (for an additive or multiplicative variable, respectively) changes are preserved. Next, we look at the method's performance over highly populated cities and regions that are particularly important for impacts research. We explore how bias-adjustment and~~

~~downscaling change~~ The QDM and QPLAD methods, as applied, preserve changes in GCM quantiles on any given future day, where that day's quantile is determined by the ±15 day and ± 10 year time window from the raw GCM. However, because the bias adjustment and downscaling are applied on a rolling, daily basis, it means that the adjustment factors are varying every day and year. Thus, when evaluating the final resulting bias-adjusted and downscaled GCM time series, there will likely be some aggregate modification to the quantile changes. In this section, we evaluate the extent to which quantile changes are preserved and how well the historical distribution~~relative to the reference distribution, and how trends are preserved for select cities across the globe. We then turn to understanding the performance of the method, specifically around trend modification, for moderate and more extreme climate indices as well as seasonal and annual aggregated metrics commonly used in impacts analysis, including seasonal mean maximum and minimum temperature and seasonal and annual total precipitation. We examine the performance of the QDM-QPLAD methods for these metrics first for a coastal city (Miami, Florida) and then at the state and country level , which we then weight by population.~~ 's biases are corrected by examining city-level, state-level, and country-level metrics.

5.1 Preserving quantile trends globally

~~One of the key advantages of the QDM method is its ability to preserve changes in daily model-projected extremes due to how the method imposes the model-projected change for each quantile, rather than purely the mean~~ Here we examine the preservation of changes in higher quantiles at a seasonal frequency. For each GCM, season, and pixel, we compute the change in the 95th percentile of daily maximum temperature in the raw GCM, the bias-adjusted GCM, and the bias-adjusted and downscaled GCM over the period 2080–2100 relative to 1995–2014 for SSP3-7.0. Figure 3 shows the comparison of quantile change across these stages of processing and averaged over all GCMs in the GDPCIR dataset and indicates the level at which the source GCM quantile changes are maintained or modified. Note that in addition to the rolling adjustment factors mentioned above, here the window over which the 95th percentile is computed (e.g. ~~, standard quantile mapping) . The QPLAD method provides an additional fine-scale analog-based adjustment factor approach to layering on analog day extremes present in the fine-scale reference data. Notwithstanding, it is expected that the methods will not perfectly preserve GCM-projected changes at a temporal aggregation different from that which the methods were applied. In other words, since our empirical CDFs are computed for each day of year with a 31-day window, the GCM data distributions are adjusted with the same grouping. Similarly, model-projected changes, computed as the difference or ratio between a rolling 21-year average and the historical period, will be preserved as applied. Consequently, examining temporal aggregations outside of these windows will show behavior that may dampen or inflate model-projected changes in the low and high quantiles to a minor degree. However, with a 21-year grouping of any 31-day window, corresponding to the QDM and QPLAD method-specified temporal grouping, model-projected trends would be preserved exactly. In this section, we examine trend preservation in higher quantiles at a seasonal frequency. Figure 3 shows for a selected model (NorESM2-LM) the change in the~~ for each season) is also different than the QDM-QPLAD method application, implying some further, albeit minor, differences. We opted to look at this metric because it better demonstrates how the methods have modified the original data in more aggregate terms which are commonly used in impacts modeling and therefore may be more useful to potential users.

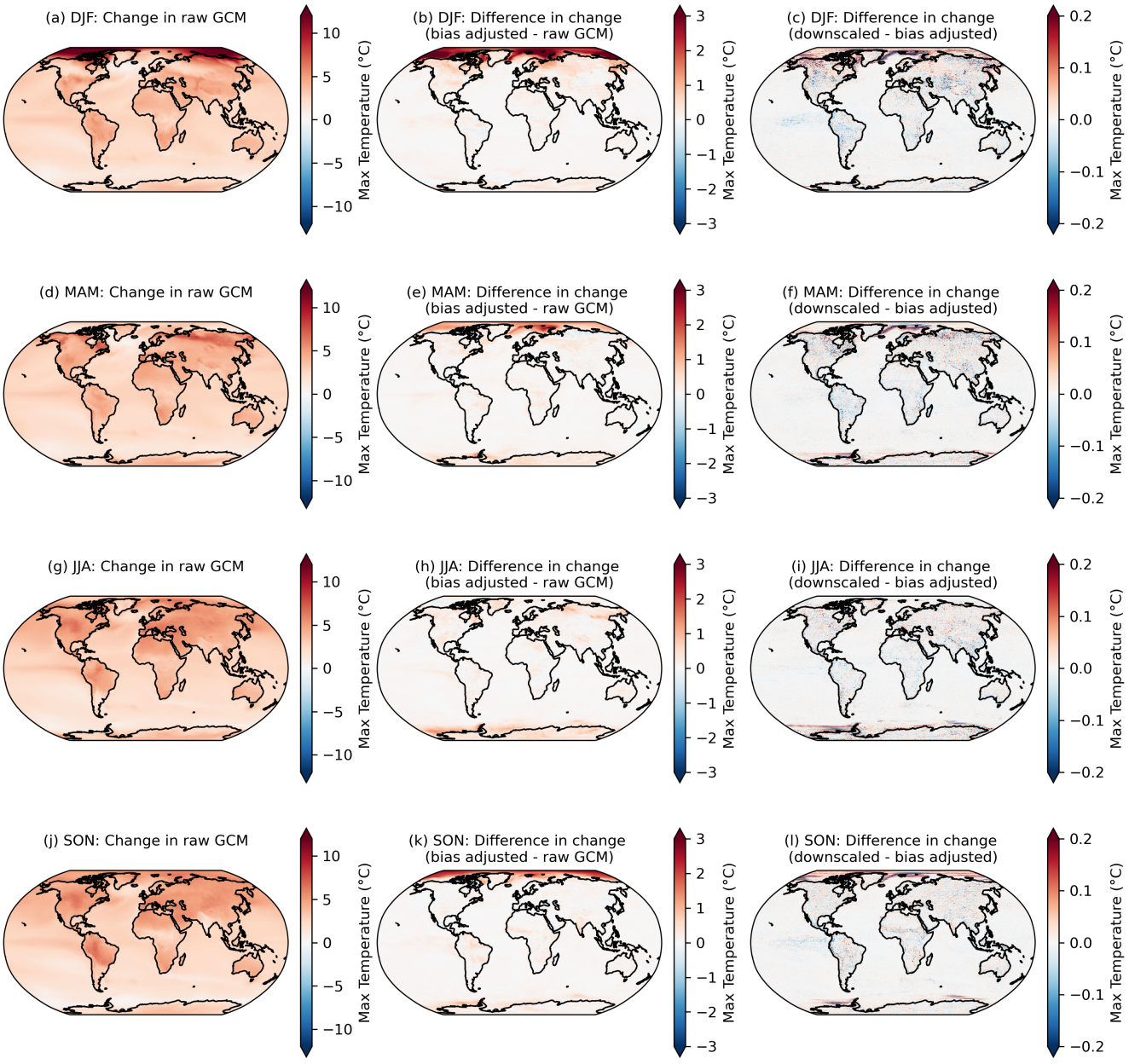

Figure 3. Changes in the 95th percentile ~~Northern Hemisphere summer (JJA) maximum temperature days between~~ of seasonal daily maximum temperature in 2080–2100 ~~and the historical period (~~relative to 1995–2014 ~~) in the source model data (a) and in panel (b)~~ in the raw GCMs (panels a, d, g, j), the difference in ~~this~~ the 95th percentile change between the bias-adjusted ~~model data and the source model data for the same time period and percentile.~~

~~**Figure 3.** Change in~~ and the raw, GCMs (panels b, e, h, k), and the difference in the 95th percentile ~~JJA maximum temperature trends globally between 2080–2100 and the training period (1995–2014) in the model (a)and the difference in~~

this trend between the change between the downscaled and the bias-adjusted data and the model (bGCMs (panels c, f, i, l) for seasons DJF (panels a–c), MAM (panels d–f), JJA (panels g–i), and SON (panels j–l). Results are shown for the model NorESM2-LM and the shown are the mean across the GCM ensemble for the scenario SSP3-7.0.

The Although the post-processing described in Section 4.3.1 is only applied to downscaled output within the downscaling pipeline, it is applied separately to bias-adjusteddata shown in Fig. 2b was post-processed according to the approach described in the methods section so that it is consistent with the bias-adjusted and downscaled data, pre-downscaled results shown here (e.g. Figure 3, second column) such that bias-adjusted and downscaled results are handled consistently for the purposes of this comparison. As noted above, it is expected that there will be slight modifications in the model-projected changes, raw GCM-projected changes. Moreover, here we show the analytical 95th percentile of Northern Hemisphere (NH) summer days days within each season and averaged over GCMs, rather than using an empirical CDF that corresponds corresponding to the actual bias-adjustment applied. Additionally, the bias-adjusted day closest to the 95th percentile. The raw, cleaned GCM data is at the original resolution of the GCM output and bias-adjusted GCM data is at a 1° resolution, whereas the downscaled data is at a 0.25° resolution, so the bias-adjusted data is coarser and, by construction, less extreme than the downscaled data. Indeed, parts of eastern Canada and Siberia exhibit amplifications in maximum temperature trends at the Some broad features emerge in Figure 3: the first column shows that generally in the raw, cleaned GCMs, the 95th percentile of every season is increasing everywhere, and more so on land and over the Arctic (except for in MAM when sea ice extent is at a maximum and surface temperatures remain near the freezing point over ice). The bias adjustment tends to increase the 95th percentile , as well as Antarctica. Much of these high-latitude areas that show amplification are also areas where the GCM-projected change in temperature is already high relative to other parts of the globe, consistent with the Arctic amplification that is already underway due to climate change (Previdi et al., 2021). We can infer that adjusting the GCM distribution to be consistent with the reference dataset is also contributing to this amplification of NH summer maximum temperatures. Supplemental figure 2 shows the difference in change between changes by a modest amount on average. Although the magnitude and extent vary by season, the vast majority of bias-adjusted percentile changes are within approximately 1°C of the raw, cleaned GCM changes (Figure 3, second column). The downscaling step adds fine resolution information that slightly modifies the change in 95th percentile in the bias-adjusted dataand , however in general changes between the bias-adjusted and downscaled data, and in comparing that to Fig. 3b we can infer that the amplification in model-projected changes is happening at the bias-adjustment step rather than the downscaling step, which figure A2 further confirms. However, amplification is generally very small in comparison to the actual magnitude of change projected by the model (3a)in those areasdata before and after downscaling is applied are on the order of a tenth of a degree Celsius. The largest differences appear over regions with large and variable (over the GCM ensemble) temperature gradients, such as near the edges of sea ice coverage. A comparable figure for the 99th percentile is included in Appendix A (Figure A7) and shows a similar story with slight increased magnitudes (e.g. bias adjustment increases the 99th percentile by a bit more than the 95th percentile).

Precipitation has a similar but more nuanced and complex story. A longstanding challenge with bias adjustment of precipitation at a global scale is dealing with the disagreement in the seasonal migration and magnitude of precipitation in the

~~intertropical convergence zone (ITCZ)~~ ITCZ between reanalysis and GCMs. The ITCZ is a tropical "belt" where deep convection and heavy precipitation occur due to convergence of the trade winds, and ~~the belt of heavy precipitation~~ it migrates between 9°N and 2°N ~~as a result of~~ due to annual warming of sea surface temperatures (van Hengstum et al., 2016). GCMs ~~still~~ exhibit bias in simulating tropical precipitation and this bias differs widely between CMIP6 models (Hagos et al., 2021; Tian and Dong, 2020). ~~Figure~~ Similar to Figure 3, Figure 4 shows the 95th percentile ~~NH summer~~ of daily precipitation for ~~the same model and scenario (~~each season averaged across the GDPCIR ensemble for SSP3-7.0. Days with total precipitation less than 1 mm day^{-1} are not included so as to only include wet days in the analysis. We also include the same figure showing the 99th percentile of daily precipitation for each season in Appendix A (Figure A8). ~~Panels A and B show 95th percentile seasonal Northern Hemisphere summer precipitation for the reference and the source model data respectively, over the historical period. Panel C shows the model-projected change in 95th percentile NH summer precipitation between 2080–2100 and the historical period in the source model. Finally, panels D and E show the ratio of the bias-adjusted data to the source model data trends the ratio of the bias-adjusted and downscaled data to the bias-adjusted data, respectively. The NH summer ITCZ is pronounced and markedly different in panels A and B; for this particular GCM and percentile, differences~~ In comparing seasonal precipitation in reference data versus the ensemble mean before bias adjustment and downscaling (panels b, g, l, q), there is broad disagreement on the ITCZ present year-round but particularly strong in Northern hemisphere summer and fall. Differences are notable in both the shape ~~(e.g., the ITCZ is shifted southwards in the GCM off the Brazilian coast, relative to reanalysis)~~ and the strength of the ITCZ~~(e. g., the GCM overestimates the intensity of the summer monsoon over southeastern Asia and in particular over India). These~~ . The climate change signal (panels c, h, m, r) show 95th percentile changes generally increasing over most land areas in the raw GCM ensemble mean and over the ITCZ, with broad decreases in precipitation over subtropical oceans that vary by season. These ITCZ biases result in slight modifications in preserving GCM-projected relative changes in the quantiles. ~~For example, panel 4D shows that inland and just off the coastlines of India,~~ Although the biggest modifications of the change in the 95th quantile primarily occur over the oceans (due primarily to the ITCZ bias), there are also some modifications in drier areas, such as Sub-Saharan Africa and parts of the Middle East. In these areas bias adjustment results in a mild amplification of the already-increasing signal from the GCMs, again driven by differences in seasonality and magnitude between reanalysis and the GCMs. For example, the ~~bias adjustment amplifies the projected change from the source GCM. Mechanically, because we choose to consider relative changes, the areas exhibiting the largest relative trends (panel C) or trend alterations (panel D) are also the driest areas (~~gridcells in Sub-Saharan Africa in reference data shown in white (e.g. ~~Brazil, Namibia, the Arabic peninsula in panel D). Finally, it appears, as expected, that statistical downscaling is not significantly altering the model-projected relative change signals (panel E) in comparison to the bias adjustment step, as we noted for temperature as well~~zero precipitation) have low but non-zero precipitation in the GCMs, an illustration of the "drizzle day" GCM problem (Dai, 2006). We apply the WDF correction discussed in Section 3.3 to mitigate the effects of this disagreement but it does not completely solve the issue in the results. In comparing changes in the bias-adjusted data to changes in the bias-adjusted and downscaled data (panels e, j, o, t), changes are most noticeable in Sub-Saharan Africa as well, where the "post" WDF is applied to bias-adjusted and downscaled data in our pipeline but not to the bias-adjusted, pre-downscaled data. Thus the right column, in essence, illustrates the effects of the WDF. To further understand the effects of the WDF as well

as modification of seasonal changes in more arid regions, we show the same analysis as in Figure 4 and Figure A8 for daily precipitation < 10 mm day^{-1}, shown for the 95th percentile (Figure A9) and the 99th percentile (Figure A10).

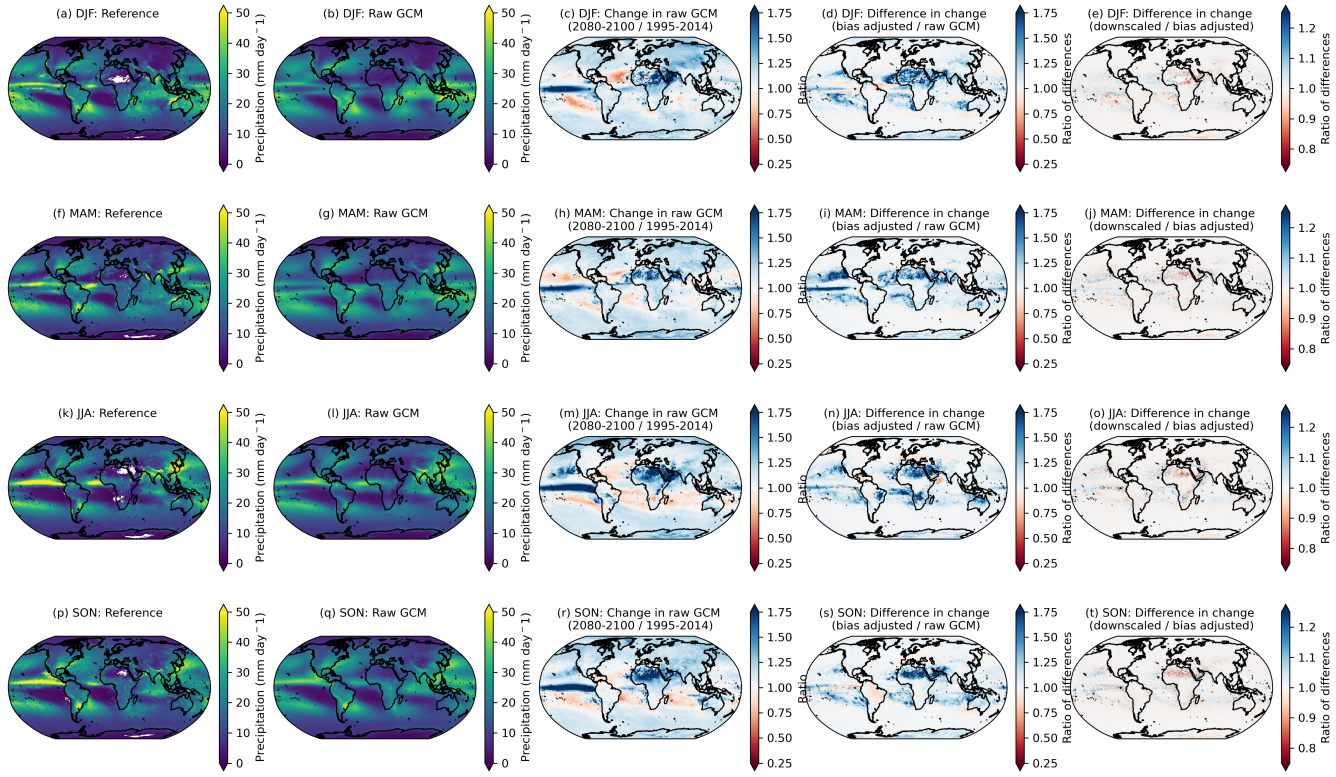

Figure 4. The 95th percentile ~~JJA precipitation over the historical (1995–2015) period in the reference data (A) and in the model data (B), along with the relative change, of the percentile in the~~ of seasonal daily total precipitation for the reference (panels a, f, k, p) and raw, cleaned GCM (b, g, l, q) over the training period, 1995–2014. The change in the 95th percentile of seasonal daily total precipitation in 2080–2100 ~~period to the same percentile in the historical period in the model data (C),~~ relative to 1995–2014, as a ratio, in the raw, cleaned GCMs (panels c, h, m, r), the ratio of the ~~relative change in~~ 95th percentile change between the bias-adjusted ~~data to the relative change in the model data (D),~~ and the raw, cleaned GCMs (panels d, i, n, s), and the ratio of the ~~relative change in the~~ 95th percentile change between the downscaled and the bias-adjusted ~~and downscaled data to the relative change in the bias-adjusted data (E) . White areas in panel (E)are due to values equal to zero in the bias-adjusted and downscaled data in the historical period after the post-processing described in the methods section is applied. Results are shown for~~ GCMs (panels e, j, o, t) for seasons DJF (panels a–e), MAM (panels f–j), JJA (panels k–o), and SON (panels p–t). Results shown are the average for wet days across the GCM ensemble for the ~~model NorESM2-LM and the~~ scenario SSP3-7.0.

5.2 Historical and future method performance for selected cities and regions

We further quantify the bias adjustment and trend preservation ~~, here for selected aggregations and areas~~error for highly populated cities and selected aggregated regions containing the cities. Following the analysis in Bürger et al. (2012) and Cannon et al. (2015), we assess the performance of the QDM ~~method~~ and QPLAD methods by comparing the distributions of various CCI/CLIVAR/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI) metrics (Karl et al., 1999) ~~metrics~~ as well as other aggregated ~~metrics~~ values widely used in impacts research~~(Table 2)~~ , listed in Table 2. We compute these over the historical period in the bias-adjusted and downscaled ~~model data against their distribution~~ data and compare against their distributions in the reanalysis ~~dataset, for a single city. The selected indices encompass~~ reference dataset. For the initial city analysis, we use a set of 17 highly populated cities: Paris, France; Shanghai, China; Lagos, Nigeria; Delhi, India; Dhaka, Bangladesh; Mexico City, Mexico; Cairo, Egypt; Moscow, Russia; São Paulo, Brazil; Miami, Florida; New York City, New York; Manila, Philippines; Istanbul, Turkiye; Mumbai, India; Buenos Aires, Argentina; Tokyo, Japan; and London, United Kingdom. The first eight cities are inland cities and the latter nine coastal cities. Later in this section, we examine the same indices for aggregated regions in which each of the cities is located. The selected ETCCDI indices and the additional metrics include maximum and minimum ~~temperatures and~~ temperature-based values as well as values that are derived from total precipitation, ensuring that all variables included in the GDPCIR dataset are tested. We examine the performance of these metrics across all ~~models~~GCMs included in the GDPCIR dataset, given the heterogeneity of temperature and precipitation signals~~among models. Then, we focus on a single model, a selection of 17 metropolises globally along with their corresponding lower resolution regions and assess both bias adjustment and trend preservation.~~ .

Name	Description
summer days	Annual count of days when daily maximum temperature >25°C
tropical nights	Annual count of days when daily minimum temperature >20°C
frost days	Annual number of days under 0°C
days over 90	Annual number of days over 90°F
days over 95	Annual number of days over 95°F
seasonal maximum temperature	Mean seasonal maximum temperature for each year
seasonal minimum temperature	Mean seasonal minimum temperature for each year
wet days	Annual count of wet days (daily total precipitation >1mm)
wet days with a specified threshold	Annual count of moderate precipitation days (daily total precipitation >10mm)

Table 2 continued from previous page

Name	Description
consecutive dry days	Annual maximum number of consecutive dry days (daily total precipitation <1mm)
annual precip	Annual precipitation
seasonal precip	Total precipitation summed over seasons each year

Table 2: Selected moderate and extreme metrics for analyzing bias adjustment and downscaling algorithm performance over cities and admin1 (state/province) regions.

5.2.1 Historical extremes indices

To check the historical distributions of the ~~downscaled models, we examine "moderate" and "extreme" extremes detailed in Table 2 by computing~~ bias-adjusted and downscaled GCMs, we compute the selected indices listed in Table 2 on an annual basis over the ~~training period and over a separate validation period~~ historical period for the raw GCM, bias-adjusted and downscaled GCM and reanalysis for ~~a single city, Miami, Florida. The chosen metricsare checks on the distributions of all variables included in the dataset. Some of the metrics~~the 17 selected metropolises. The ETCCDI metrics, such as summer days, tropical nights, ~~and~~ annual wet days~~represent more moderate extremes less~~ , and consecutive dry days represent extremes affected by threshold behavior. ~~Others~~Other more extreme temperature metrics not classified as ETCCDI indices, such as ~~consecutive dry days , days~~ days over 35°C ~~,~~ and days over 32.2°C, are even more affected by threshold behavior. ~~Others, such as seasonal temperature means and total precipitation , while~~ While those more extreme temperature metrics and the seasonal and annual temperature and precipitation metrics are not classified as ETCCDI indices, they are widely used as input data to sector-specific impacts modeling and thus are included here to guide users of the dataset.

~~We calculated distributions~~ Distributions of the indices ~~on~~ are computed using the raw GCM output and on the bias-adjusted and downscaled GCM and each are compared against the reanalysis distribution of the same index using a two-sample ~~Kolmogorov-Smirnov~~ Kolmogorov–Smirnov (K-S) test at a 0.05 significance level. The null hypothesis is that the two samples (e.g., raw GCM and reanalysis or bias-adjusted and downscaled GCM and reanalysis) are drawn from the same distribution. A ~~model~~ GCM is considered to pass the K-S test, either for the raw GCM or the bias-adjusted and downscaled GCM, if the null hypothesis is not rejected, in other words, if the p-value ~~<~~ > 0.05. This is a slight modification of the usage of K-S tests in Cannon et al. (2015) and Bürger et al. (2012), where the authors use the D statistic rather than the p-value as a diagnostic. The p-value is used here for significance due to the effects of disagreement in seasonality between reanalysis and the GCM on the D statistic versus the p-value. We compute the K-S tests over ~~two time periods: a calibration period (1995–2014) and a validation period (1979-1994). The validation period used is shorter because aggregated reanalysis data was only available~~ a climatological historical period from 1979 ~~to near-real-time and had not yet been extended back to 1950 by ECMWF and made available when hourly data was first downloaded and aggregated to daily for use in the bias adjustment and downscaling pipeline. For metrics that use precipitation data, the validation periodis~~ – 2014 for temperature variables and for precipitation

we use a slightly shorter historical period, 1984–1994, because quality control showed that precipitation data for 1983 contained errors.

In Figure 5, the results of the K-S tests for the twelve selected indices for a ~~single city, Miami, Florida,~~ subset of inland cities around the globe are shown for the ~~downscaled~~ bias-adjusted and downscaled GCMs and raw GCMs~~for the calibration and validation periods. The selected index that never passes K-S tests is frost days (the number of annual days below the freezing point) because there are so few years in the historical period in both the raw GCMs and the reanalysis that contain days below freezing. Prior to~~ . The same analysis for coastal cities around the globe can be found in Figure A11). For nearly all of the inland cities, bias adjustment and downscaling ~~, around half of the GCMs pass K-S tests in the calibration period, and afterward, all GCMs pass the K-S tests except for frost days. During the validation period, a considerably smaller~~ shows a significant improvement in the number of K-S tests ~~pass in the raw GCMs, while all pass after bias adjustment and downscaling, showing that the distributions of the selected moderate and extreme metrics have been effectively adjusted by the bias adjustment and downscaling algorithms. A few models are notable in the lack of passing K-S tests before~~ passing over the source GCM distributions. The notable two exceptions to this are Mexico City and Moscow. For Mexico City, this can be explained by its high elevation relative to the other cities; it is at an elevation of 2240 m above sea level and located in a valley. Moscow's relative lack of improvement from bias adjustment and downscaling ~~, including the BCC-CSM2-MR, GFDL-ESM4, and CCMC-ESM2 models~~can be explained by its colder climate relative to other inland cities and, therefore, lack of occurrences for the maximum temperature metrics, as well as a strong urban heat island effect (Lokoshchenko, 2014). By contrast, ~~INM-CM4-8 is notably high before downscaling in the calibration period, indicating that its distributions of moderate and more extreme extrema were already closer to reanalysis for this location prior to bias adjustment. It is worth noting that because the location shown here is a coastal city,~~ the ~~additional benefit of downscaling versus solely bias adjustment is more profound, and thus the adjusted distributions shown here benefit both from QDM and QPLAD adjustments. Had an inland city been shown, the additional effect of downscaling in adjusting the distributions to the higher-resolution reanalysis distribution would have likely been less significant (unless it were in an area with complex topography)~~coastal cities (Figure A11) show a markedly different side of the narrative, illustrating the limitations of bias adjustment and downscaling for coastal areas in some parts of the world. Miami, Manila, and Mumbai, in particular, show little improvement between the raw GCM and bias-adjusted and downscaled GCM, which points to the inherent challenges of GCM representations of coastlines as well as limitations with coastal areas in reanalysis data.

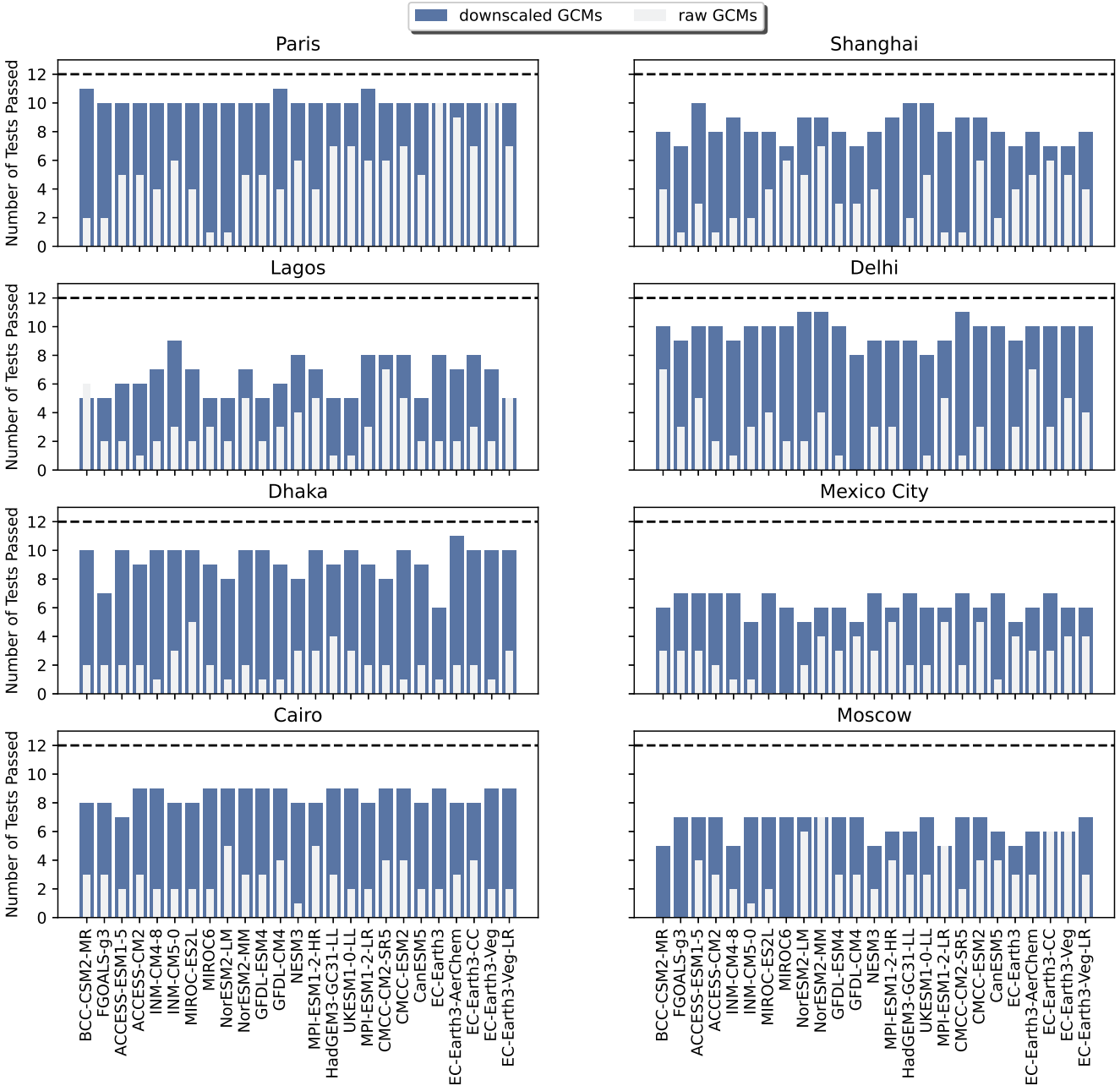

Figure 5. Bar ~~plot~~ plots showing the number of ~~K-S~~ Kolmogorov-Smirnov tests passed for the twelve selected indices for the ~~downscaled model and raw model~~ bias-adjusted and downscaled GCM and raw GCM (overlain) for each of the GCMs included in the GDPCIR dataset for ~~a single coastal city, Miami, Florida (USA). The calibration period is shown in (a) and the validation period in (b)~~eight inland cities around the globe. The dashed line shows the maximum possible number of K-S tests.

~~Index Description Surface Variable 0.5extx_days_above Annual count of days when daily maximum temperature > 25°C maximum temperature 0.5extn_days_above Annual count of days when daily minimum temperature > 20°C minimum temperature 0.5exwet_days Annual count of wet days (daily total precipitation > 1.0mm) total precipitation 0.5exwet_days_prop Annual count of moderate precipitation days (daily total precipitation > 10.0mm) total precipitation 0.5exseasonal minimum temperature Mean seasonal minimum temperature for each year minimum temperature 0.5exseasonal maximum temperature Mean seasonal maximum temperature for each year maximum temperature 0.5exseasonal precipitation Total precipitation summed over seasons each year total precipitation 0.5exannual precipitation Annual precipitation total precipitation 0.5exdays over 32.2°C Annual number of days over 90°F maximum temperature 0.5exdays over 35°C Annual number of days over 95°F maximum temperature 0.5exfrost days Annual number of days under 0°C minimum temperature 0.5exconsecutive dry days Annual maximum number of consecutive dry days (daily total precipitation < 1.0mm) total precipitation 0.5ex0.5exSelected moderate and extreme metrics for analyzing bias adjustment and downscaling algorithm performance over cities and admin1 (state/province) regions.~~

5.2.2 Bias adjustment and relative trend preservation

To further examine the performance of the bias adjustment and downscaling algorithms, we ~~examine seasonal aggregated metrics (mean maximum and minimum temperature and total precipitation) across selected~~ compute median absolute errors in bias adjustment and downscaling and trend preservation across the same highly populated cities ~~globally for a single model, BCC-CSM2-MR, by computing the median absolute error in bias adjustment compared to the mean seasonal error in~~ for all GCMs included in the GDPCIR dataset for the selected projection period, SSP3-7.0, and for all variables. We compute the error over daily 21-year climatologies after smoothing the daily data with a 31-day rolling window mean. Median absolute error in bias-adjusted and downscaled data is computed over the historical period (1995–2014) and compared to trend preservation between the raw ~~and downscaled GCMs. Following~~ GCMs and bias-adjusted and downscaled GCMs for 2080–2100. Based on the method used by Lange (2019), we define absolute error in bias adjustment as:

$$e = |y_{hist}^{sim} - x_{hist}^{obs}| \tag{6}$$

where y_{hist}^{sim} represents ~~bias-adjusted historical simulations~~ bias-adjusted and downscaled historical daily climatological GCM data from 1995–2014 and x_{hist}^{obs} represents historical ~~reference data.~~ daily climatological reference data over the same time period. Median bias adjustment and downscaling errors are computed as the median of the error for all days of the year. We then define median absolute error in trend preservation as:

$$e = |(y_{fut}^{sim} - y_{hist}^{sim}) - (x_{fut}^{sim} - x_{hist}^{sim})| \tag{7}$$

where y_{fut}^{sim} represents ~~bias-adjusted projections~~bias-adjusted and downscaled daily climatological projection data from 2080–2100 for SSP3-7.0, y_{hist}^{sim} represents ~~bias-adjusted~~ daily climatological bias-adjusted and downscaled historical simulations, x_{fut}^{sim} represents ~~future projections (unadjusted)~~ daily climatological future projection data from the raw GCM over the

same future period and x_{hist}^{sim} represents ~~historical simulations (unadjusted).~~ daily climatological historical data from the raw GCM for the same historical period. As with bias adjustment and downscaling error, trend preservation error is also computed as the median of the error for all days of the year.

However, we depart from the Lange (2019) method by computing the ~~absolute error over seasonal means or sums (for temperature and precipitation, respectively) and for cities globally~~ median absolute error for highly-populated cities around the globe (e.g., at the pixel level) rather than at multiple spatial resolutions. ~~Because we are computing the error on an annual basis rather than over a 21-year rolling window, the difference in trend between the raw GCM and downscaled GCM is non-zero (our QDM implementation perfectly preserves relative seasonal trends between a given future 21-year rolling window and the historical training period)~~ Some artifacts of regridding affect the analysis; bias-adjusted and downscaled data, raw GCM data, and reanalysis data are necessarily at different resolutions: 0.25°, native GCM grid, usually around 1° (with some exceptions), and the native N320 (regular Gaussian) ERA5 grid, respectively. Figure 6 shows ~~these results over selected cities globally. Overall, bias-adjustment and trend-preservation errors are lower for minimum temperature than for maximum temperature but are generally low for both surface variables~~boxplots for the median absolute error across all GDPCIR GCMs for maximum and minimum temperature and precipitation. Trend preservation error represents the error for 2080–2100 for a single scenario, SSP3-7.0. Bias adjustment error ~~for precipitation is low for the majority of cities with the notable exception of São Paulo, which has a median error of over 14. This comparatively larger error in bias adjustment can be explained by both a) a difference in seasonal precipitation magnitudes for that location in reanalysis vis-a-vis the GCM, and b) significantly larger interannual variability for São Paulo versus other cities shown. A shortcoming of only showing bias-adjusted and downscaled error in Figure 6 versus also showing the same analysis for bias-adjusted data for the same model and cities means that we cannot attribute the bias adjustment error in São Paolo to (a) or (b), but it is likely that both play a large role, given the issues with precipitation in the tropics discussed earlier~~represents the error after QDM bias adjustment and QPLAD downscaling have been performed. Overall, the range of error for both bias adjustment and downscaling and trend preservation is lower for precipitation than for minimum temperature variables. A small subset of coastal cities show a much higher range in trend preservation error across GCMs, particularly Miami and New York and São Paulo to a lesser extent. Mexico City also shows a higher range, similar to the previous section due to its high elevation and the complex topography surrounding the city. Larger trend preservation error for these cities is unsurprising; it is well-known that GCMs struggle with capturing the land-sea interface. However, the modification of the change signal represented by the trend preservation error should not be interpreted as undesirable behavior; Iturbide et al. (2022) found that bias adjustment amplified the climate change signal (up to a factor of two in some regions), which resulted in an improvement in modeling future heat-related threshold indices. In Lange (2019), the author conducted similar error analysis for surface variables over different CMIP5 GCMs (MIROC5, IPSL-CM5A-LR, and GFDL-ESM2M) at a coarser resolution (2°) and found similar magnitudes of error in trend preservation, with slightly smaller errors in bias adjustment. ~~The larger values in bias adjustment shown here are likely due to the fact that the errors are being calculated at a higher resolution for cities (with raw GCM, reanalysis, and downscaled data drawn from the nearest 0.25° gridcell) and thus there are larger errors between the coarse-resolution GCM and the higher-resolution reanalysis, and the preservation of climate change signal in quantiles is implemented on a 21-year rolling window in our study.~~

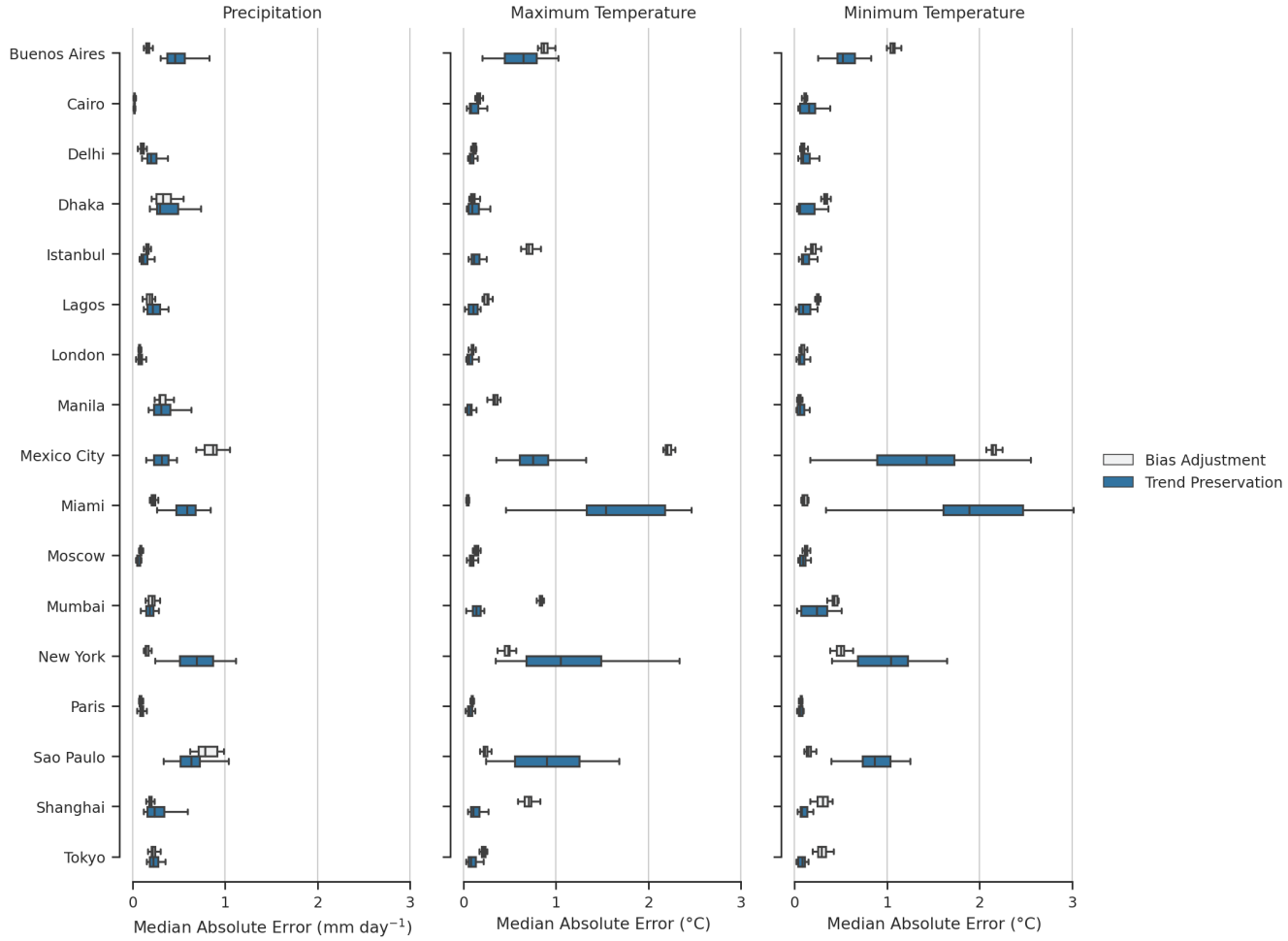

Figure 6. ~~Efficacy of bias adjustment and relative trend preservation for seasonal maximum and minimum temperature and seasonal total precipitation computed annually for a single GCM (BCC-CSM1-2) for selected highly populated cities globally. The historical period used is 1995-2014, while the projection period is~~ Range of median absolute error for bias-adjusted and downscaled historical data (1995–2014) and the range of median absolute error for trend preservation (SSP3-7.0~~2080-2100. More precisely, for temperature variables, the values on the x-axis are the median over the seasonal absolute (additive) bias adjustment errors , and the values on the y-axis are the median over the seasonal absolute (additive) error in (additive) trend preservation. For precipitation, errors are multiplicative and instead of taking the mean over seasons, we take the sum over seasons~~, 2080–2100) for the 17 metropolises globally. Values shown represent the range of median absolute errors for all GCMs included in the GDPCIR dataset. Absolute error is computed over daily 21-year climatologies (historical or future) with the median taken over all days of the year.

5.2.3 Relative trend preservation in selected regions

One of the key considerations in developing a method and dataset for use in the study of the human impacts of climate change is the performance of the given method when the data is reconfigured, transformed, or re-weighted by the users of the data. Impacts research frequently uses weighted, aggregated extreme value measures, such as crop-output-weighted frost-day counts for a given agricultural zone ~~,~~ or population-weighted counts of hot nights for a given census region. To understand the performance of our data under such circumstances, we use the same set of diagnostic cities examined above to understand the preservation of moderate and extreme trends for several of the moderate and extreme ETCCDI indicators at varying levels of aggregation. Following the regional aggregation method described in Rode et al. (2021), these comparisons use a 30-arcsecond population raster dataset (CIESIN, 2018) to determine the weight of each grid cell in the climate dataset within each region's total, based on whether the population grid cell is contained within each region~~'s shapefile~~. Data is aggregated to either admin0 or admin1 regions after computing the ETCCDI metrics on gridded data. An admin1 region is a generic term that refers to ~~the~~ a country's largest subnational administrative unit~~of a country~~; for example, a state in the US or a prefecture in Japan. An admin0 region refers to national boundaries, e.g. the US or Japan. ~~Shapefiles that define~~ Polygons defining these region boundaries are taken from the Natural Earth dataset (Natural Earth, 2022), and are further subset to include the admin0 or admin1 region, which includes each of the diagnostic cities listed above.

For the analysis in this section, we use the same temporal aggregation as in the method implementation such that any modification of trend is not due to the effects described earlier but instead due to aggregation or weighting effects. Because the method exactly preserves quantile trends within a 31-day window during bias adjustment ~~,~~ and preserves trends in minimum temperature, maximum temperature, and log(precipitation) for a given quantile on an average basis across 0.25° gridcells within each coarse 1° cell, discrepancies between trends in seasonal and annual mean minimum temperature and maximum temperature are due solely to differences between area and population weights, and due to the effects of ~~Gaussian~~ interpolation from the native GCM grid to the regular 1° grid used for bias adjustment. This behavior can be seen in the very high degree of agreement between source GCM and ~~bias adjusted~~ bias-adjusted and downscaled trends at both the admin0 and admin1 level for maximum temperature in Figure 7. Here, we calculate trend using the difference between the ~~1995-2014~~ 1995–2014 period average and the ~~2079-2099~~ 2079–2099 period average; the year 2100 is not included because it is ~~not available~~ unavailable in all GCMs. Panels ~~a-e~~ a–e in Figure 7 show the change in period average annual and seasonal maximum temperature for admin0 regions (e.g. countries) and for admin1 regions (e.g. states/provinces) in panels f-j. The admin0 and admin1 regions shown correspond to the regions where each ~~of the cities~~ city is located, and results are shown for all GCMs and all scenarios (SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5). Both admin0 and admin1 regions have an r^2 value of at least 0.9 for ~~both~~ annual temperature and all seasons, showing extremely minimal trend modification.

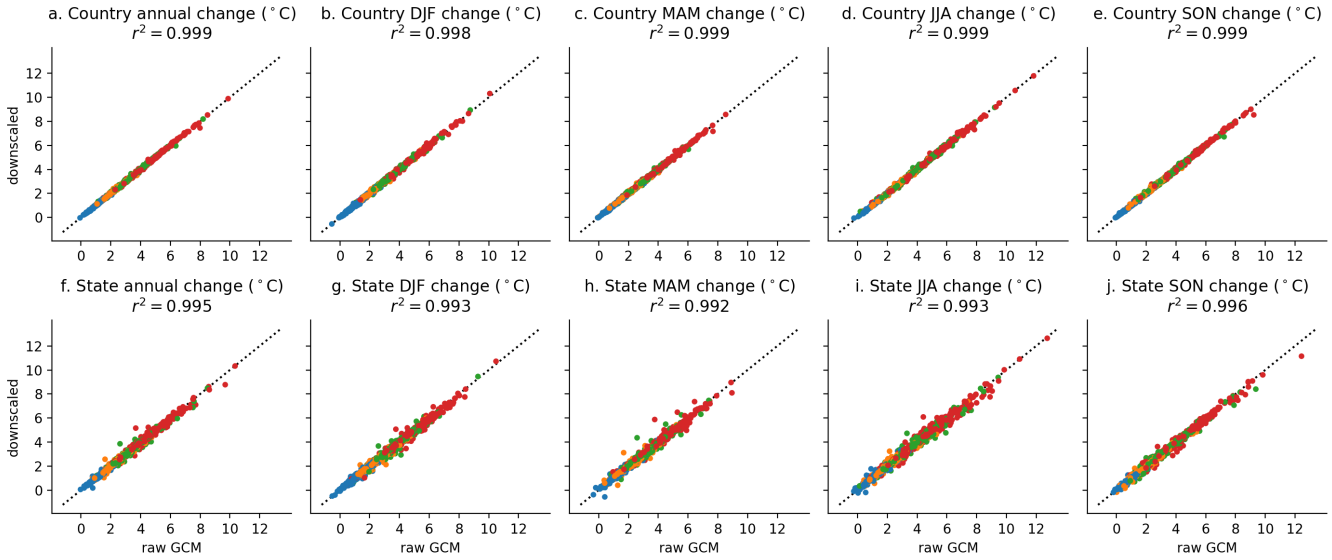

Figure 7. Change in period average annual and seasonal maximum daily T_{max} from ~~1995-2014 to 2079-2099~~1995–2014 to 2079–2099, for countries (top row; panels ~~a-e~~a–e) and states/provinces (bottom row; panels ~~f-j~~f–j) containing the 17 diagnostic cities. All GCMs and scenarios are shown; with SSP1-2.6 (blue), SSP2-4.5 (orange); SSP3-7.0 (green), and SSP5-8.5 (red).

However, because precipitation adjustments are multiplicative, 21-year seasonal and annual totals are not preserved exactly when aggregated. Fidelity to the source ~~model~~ GCM trend in the downscaled data is closer when comparing trends in log(21-year annual average precipitation) or log(21-year seasonal average precipitation), which can be seen in comparing the first and second rows in Figure 8. Figure 8 shows annual and seasonal precipitation for the countries containing the 17 selected global cities for all GCMs and scenarios, with the change in period average precipitation shown in panels ~~a-e~~ a–e and log(period average annual and seasonal precipitation) in panels ~~f-j~~f–j. As expected, the higher emissions scenarios SSP3-7.0 and SSP5-8.5 appear far more often as outliers, which is expected given their relatively larger change signals in precipitation.

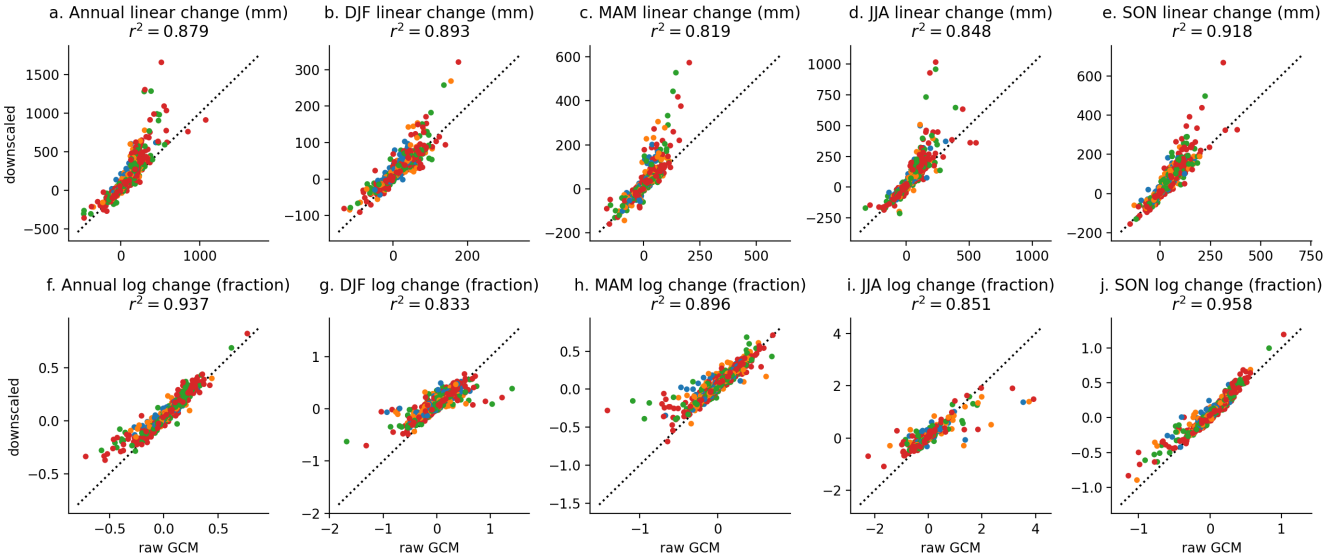

Figure 8. Change in period average annual and seasonal precipitation from ~~1994-2014 to 2079-2099~~1994–2014 to 2079–2099, (top row; panels ~~a-e~~a–e) and the change in log(period average annual and seasonal precipitation) (bottom row; panels ~~f-j~~f–j) for the countries containing the 17 diagnostic cities. All GCMs and scenarios are shown; with SSP1-2.6 (blue), SSP2-4.5 (orange); SSP3-7.0 (green), and SSP5-8.5 (red).

To understand trend preservation among extreme metrics, we computed the count of days above or below various thresholds, shown in Figure 9. The method does not explicitly preserve the GCM signal in such metrics, as anomalies in temperatures, even at extreme quantiles, will cross a threshold with different frequencies after a linear or multiplicative adjustment. This behavior is consistent with the fact that, while trends in extreme values measured as quantiles will be preserved within any 31-day window from the GCM to the final result, trends in any absolute measure, such as counts of days above or below a threshold, will be affected by the bias adjustment and may be significantly different in the result depending on the metric.

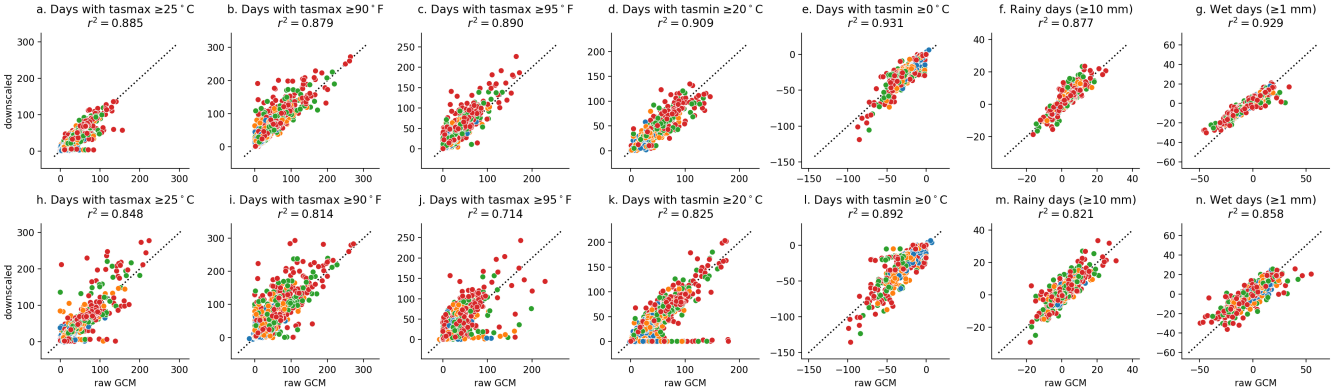

Figure 9. Change in period average threshold counts from 1994-2014 to 2079-2099, for countries (top row; panels ~~a–g~~a–g) and state/provinces (bottom row; panels ~~h–n~~h–n) containing the 17 diagnostic cities. All GCMs and scenarios are shown; with SSP1-2.6 (blue), SSP2-4.5 (orange); SSP3-7.0 (green), and SSP5-8.5 (red).

6 Conclusions

We hope that the GDPCIR dataset will be a useful contribution for climate impacts research in its scope, resolution and in the methods applied that were specifically tailored to understanding the tail risks associated with future emissions pathways. The QDM-QPLAD bias adjustment and downscaling algorithms preserve quantile trends~~and therefore allow users to better understand~~ , allowing users to understand better and model the effects of different emissions pathways on sector-specific and aggregate climate impacts. The 0.25° resolution of the GDPCIR dataset allows for its use in econometric models that require high-resolution surface climate data for estimating response functions. Errors in bias adjustment and trend preservation are low, with some exceptions for precipitation due to issues already discussed. ~~Figure A1~~ Appendix D goes into further detail on this, with Figure D1 showing land-weighted changes in temperature and precipitation signals in CMIP6 raw GCMs and the bias-adjusted and downscaled GDPCIR GCMs. We expect that the dataset will have broad use in ~~a variety of~~ climate impacts modeling, from estimating econometric dose-response functions to hydrology and ecology to modeling ecosystem services and natural capital.

Code availability. The R/CIL GDPCIR dataset codebase containing notebooks, pipeline architecture, and infrastructure is publicly available at https://github.com/ClimateImpactLab/downscaleCMIP6 and archived at https://doi.org/10.5281/zenodo.6403794. The software container and all code used for individual downscaling pipeline tasks is publicly available at https://github.com/ClimateImpactLab/dodola and archived at https://doi.org/10.5281/zenodo.6383442, and our production pipeline was run with release v0.19.0.

Data availability. The GDPCIR dataset is publicly available and hosted on the Microsoft Planetary Computer (https://planetarycomputer.microsoft.com/dataset/group/cil-gdpcir/).

Appendix A: Supplementary figures

A1 ~~Global temperature and precipitation changes~~

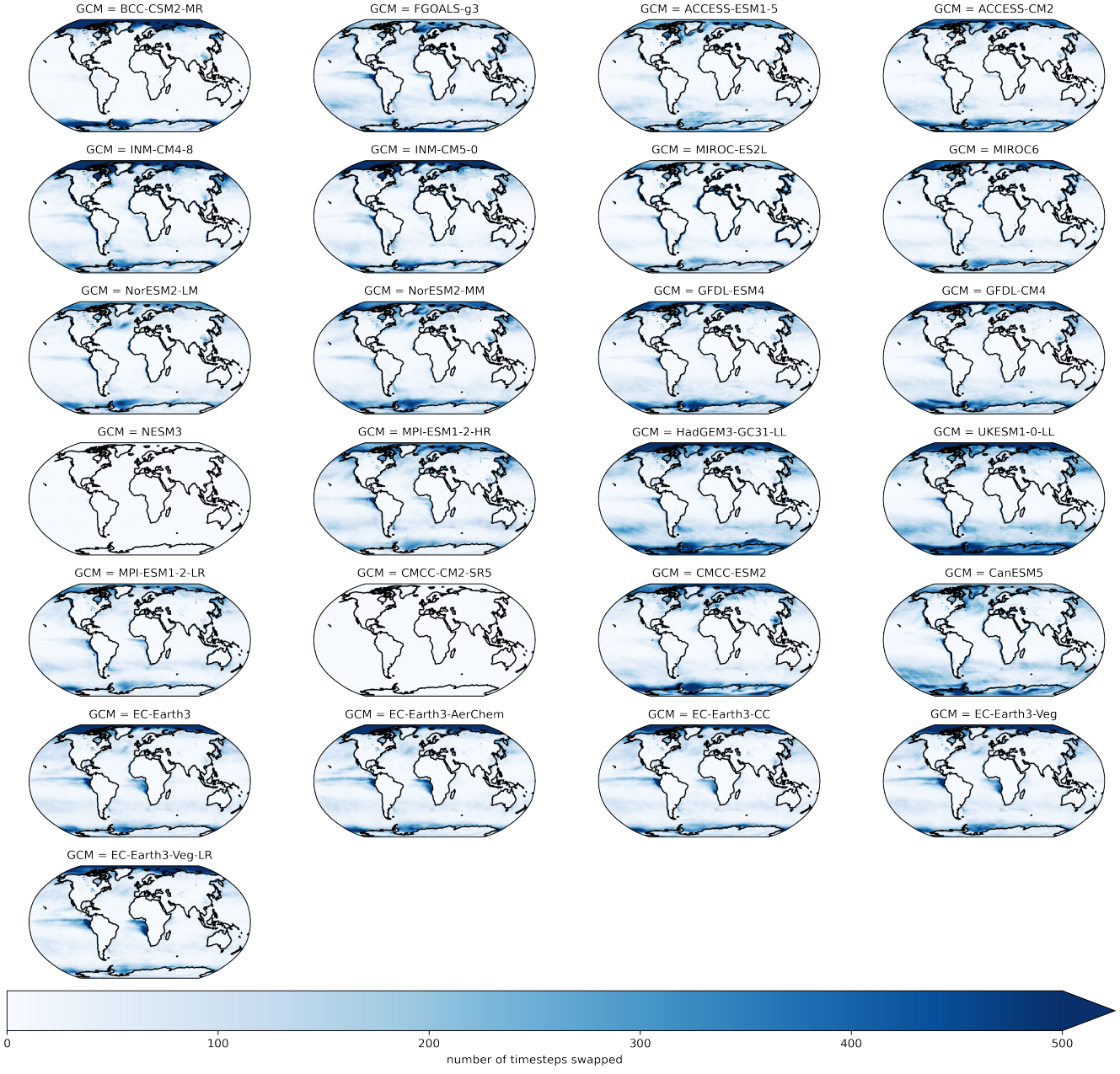

Figure A1. Number of daily timesteps where maximum and minimum temperature were swapped in the bias-adjusted and downscaled GCMs over a 21-year climatological historical period (1960–1980) for all GCMs included in the GDPCIR dataset. For these timesteps, minimum temperature exceeded maximum temperature.

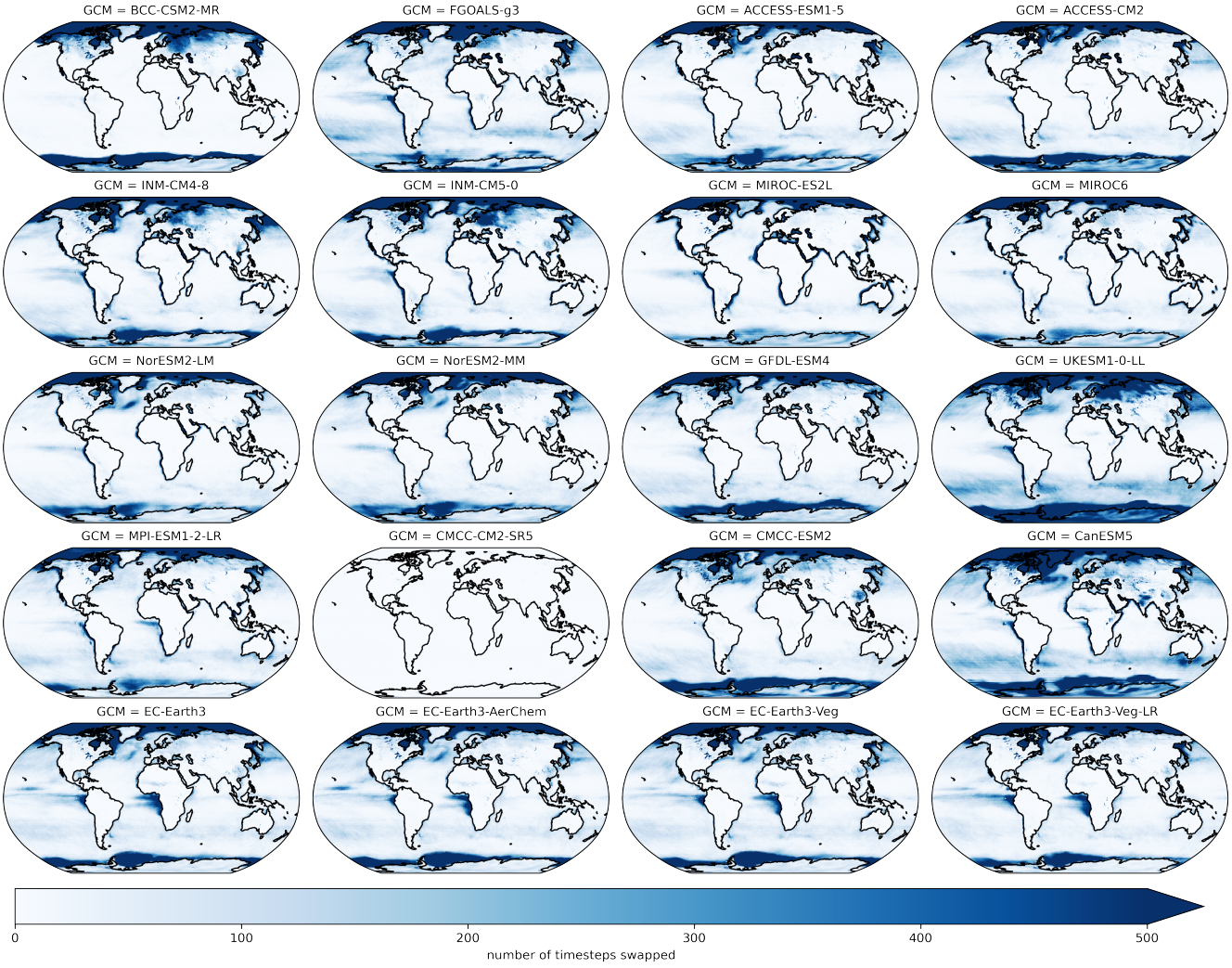

Figure A2. Number of daily timesteps where maximum and minimum temperature were swapped in the bias-adjusted and downscaled GCMs over a 21-year climatological future period (2080–2100) for all GCMs included in the GDPCIR dataset (for SSP3-7.0). For these timesteps, minimum temperature exceeded maximum temperature.

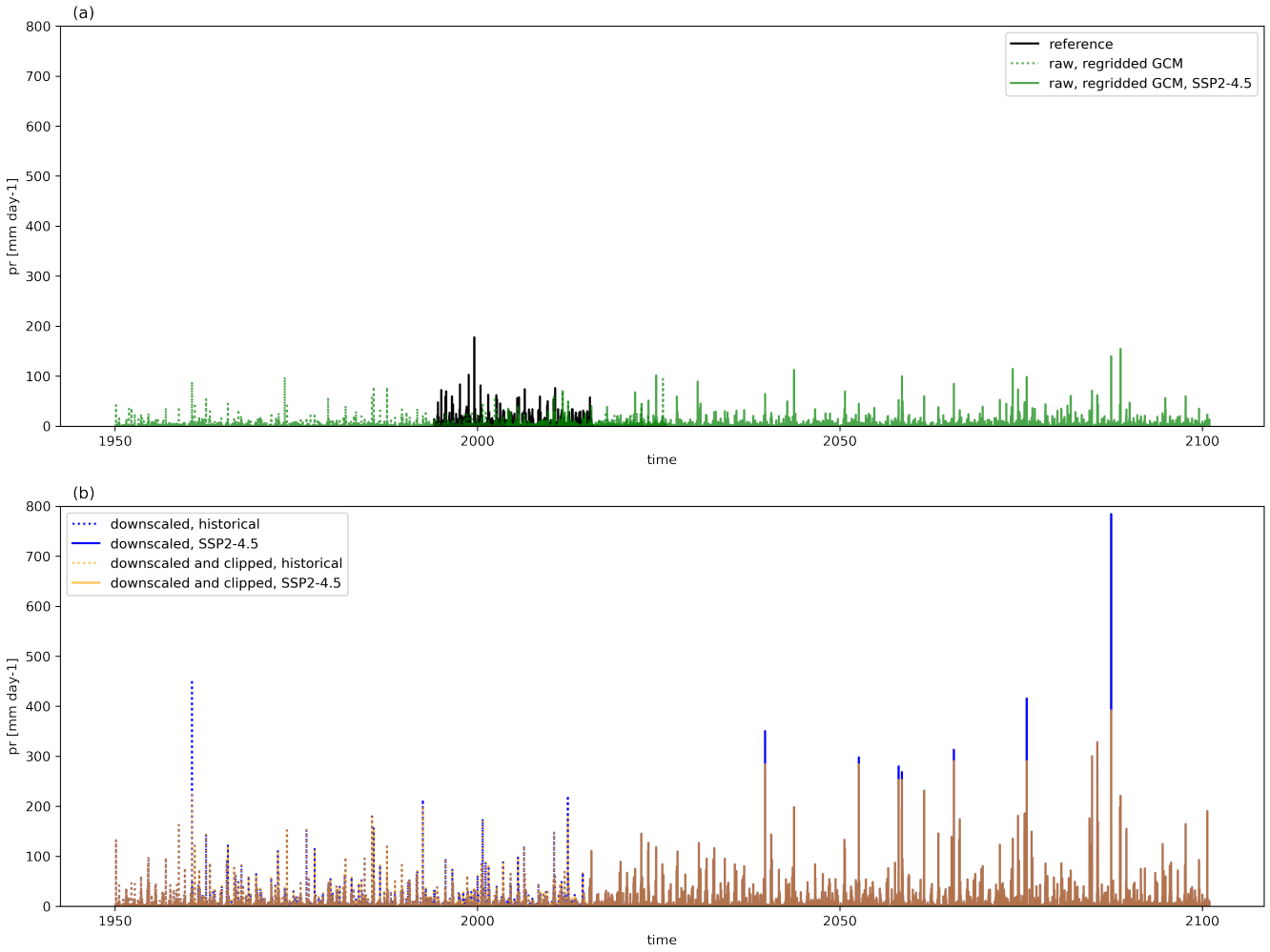

Figure A3. Time series of total daily precipitation for Delhi, India showing reference data and raw, regridded and cleaned GCM data for the historical period and SSP2-4.5 (panel a) and bias-adjusted and downscaled data for the historical period and SSP2-4.5 before after post-processing (panel b).

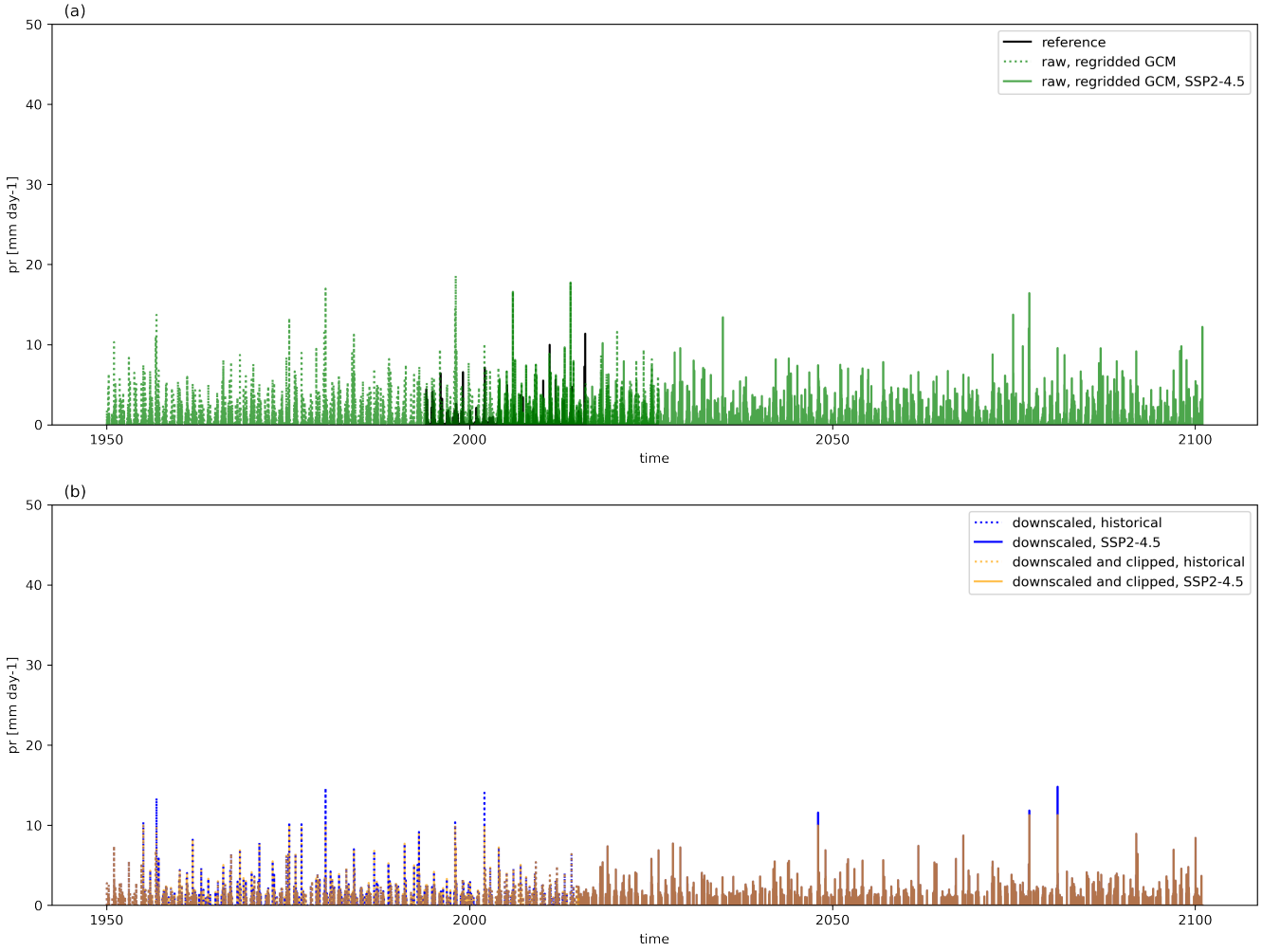

Figure A4. Time series of total daily precipitation for Cairo, Egypt showing reference data and raw, regridded and cleaned GCM data for the historical period and SSP2-4.5 (panel a) and bias-adjusted and downscaled data for the historical period and SSP2-4.5 before after post-processing (panel b).

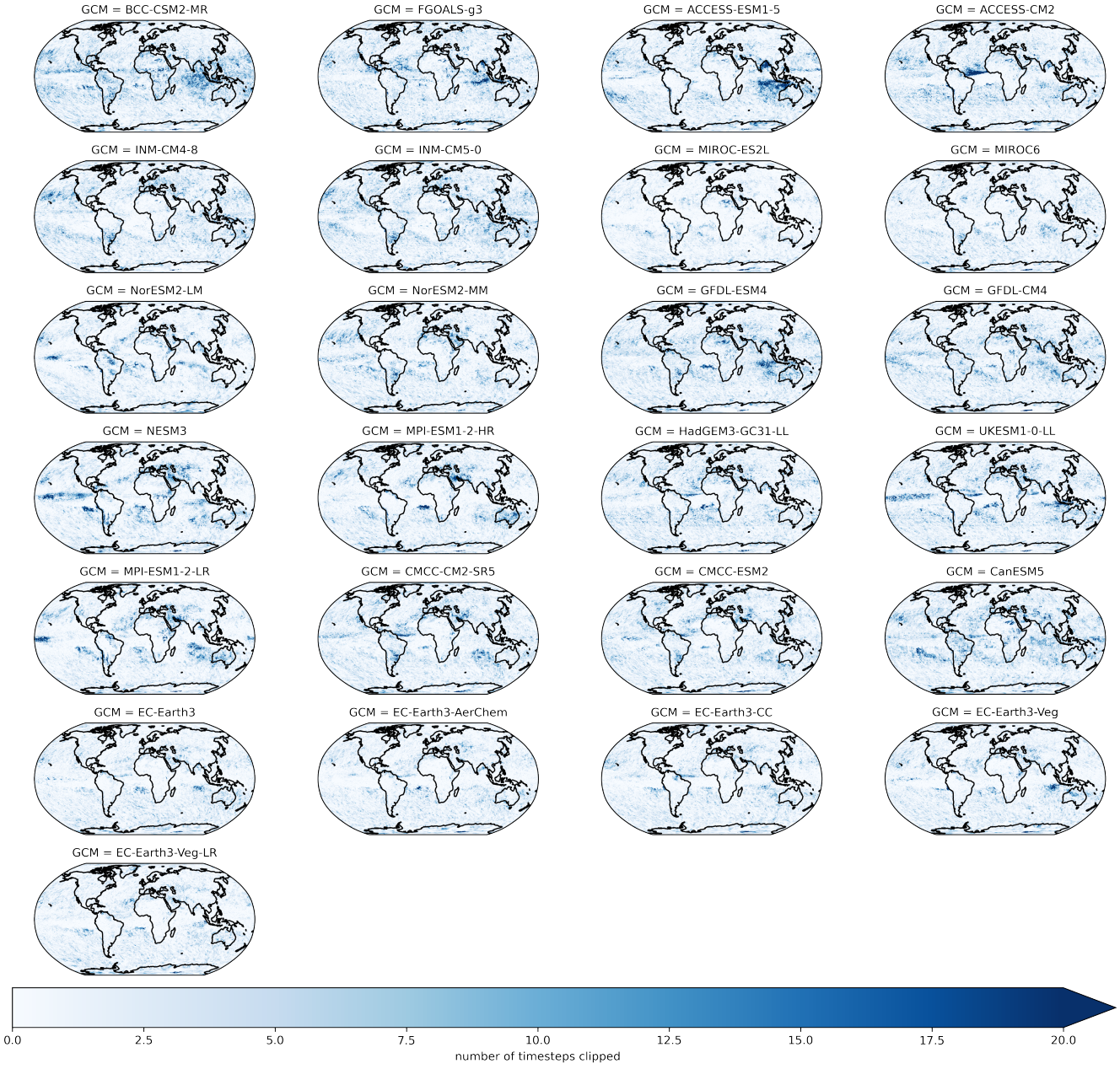

Figure A5. Number of daily timesteps where post-processing (e.g., clipping) was applied to precipitation values in the bias-adjusted and downscaled GCMs over a 21-year climatological historical period (1960–1980) for all GCMs included in the GDPCIR dataset.

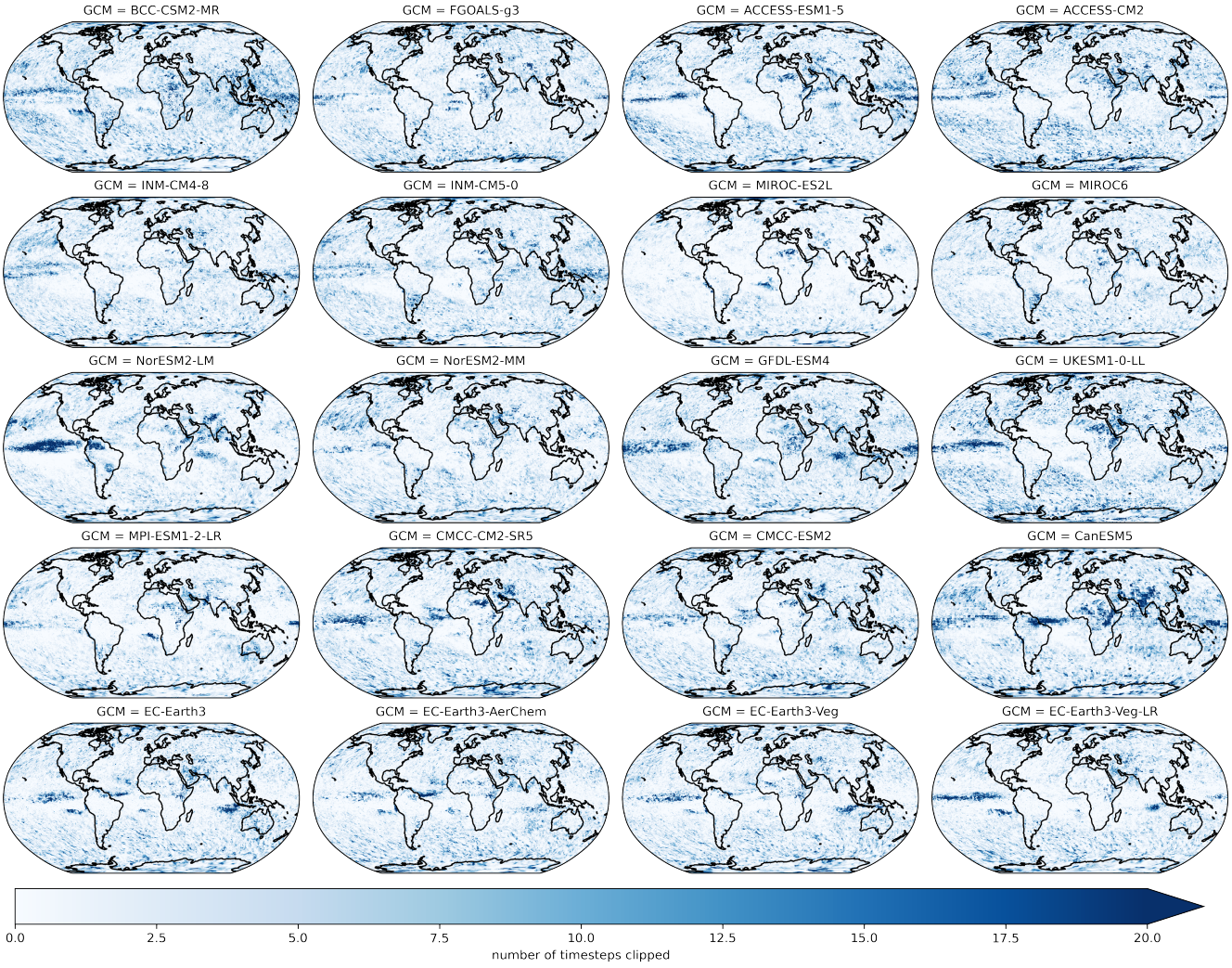

Figure A6. Number of daily timesteps where post-processing (e.g., clipping) was applied to precipitation values in the bias-adjusted and downscaled GCMs over a 21-year climatological future period (2080–2100) for all GCMs included in the GDPCIR dataset (for SSP3-7.0).

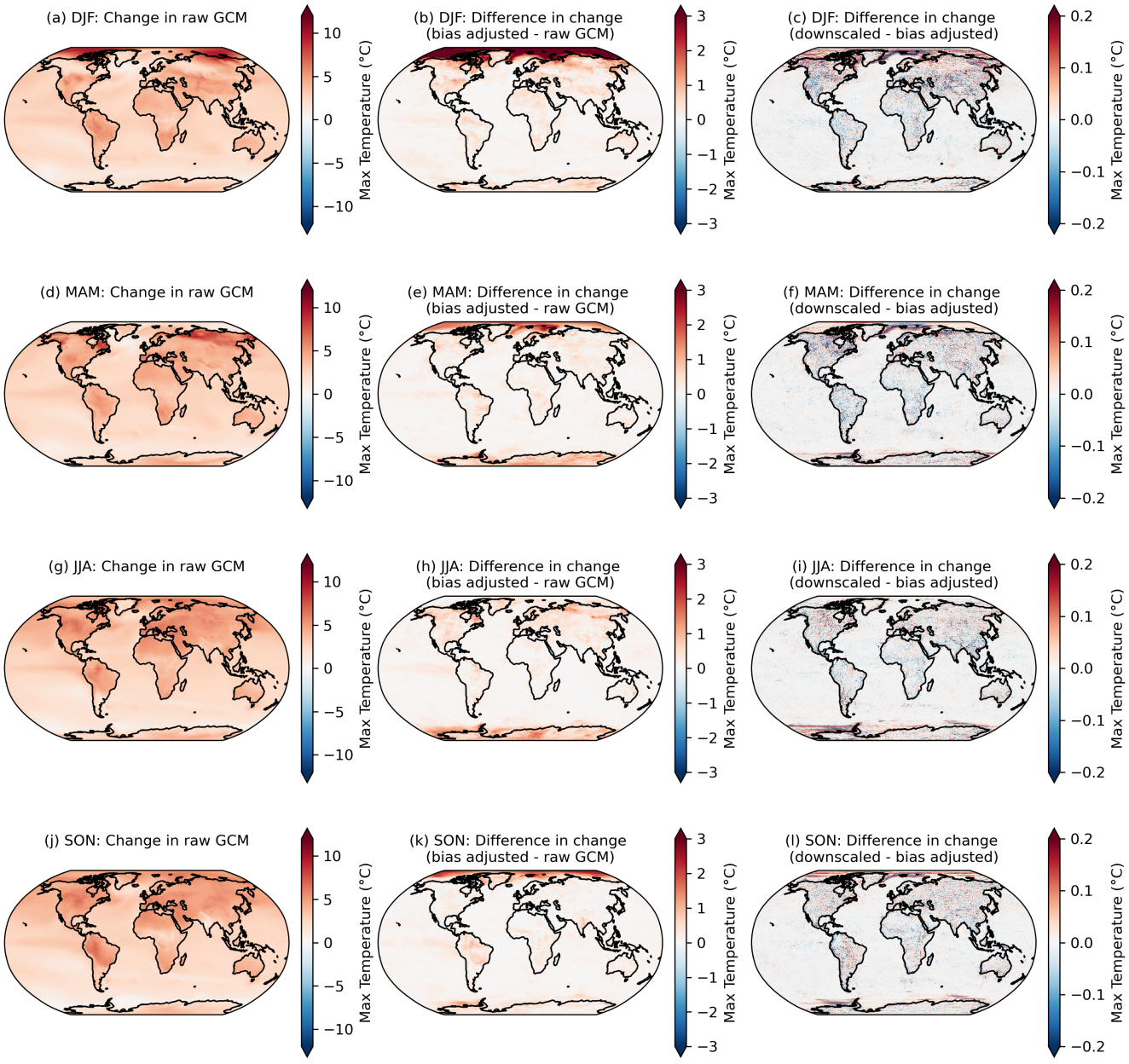

Figure A7. Changes in the 99th percentile of seasonal daily maximum temperature in 2080–2100 relative to 1995–2014 in the raw GCMs (panels a, d, g, j), the difference in the 99th percentile change between the bias-adjusted and the raw, GCMs (panels b, e, h, k), and the difference in the 99th percentile change between the downscaled and the bias-adjusted GCMs (panels c, f, i, l) for seasons DJF (panels a–c), MAM (panels d–f), JJA (panels g–i), and SON (panels j–l). Results shown are the mean across the GCM ensemble for the scenario SSP3-7.0.

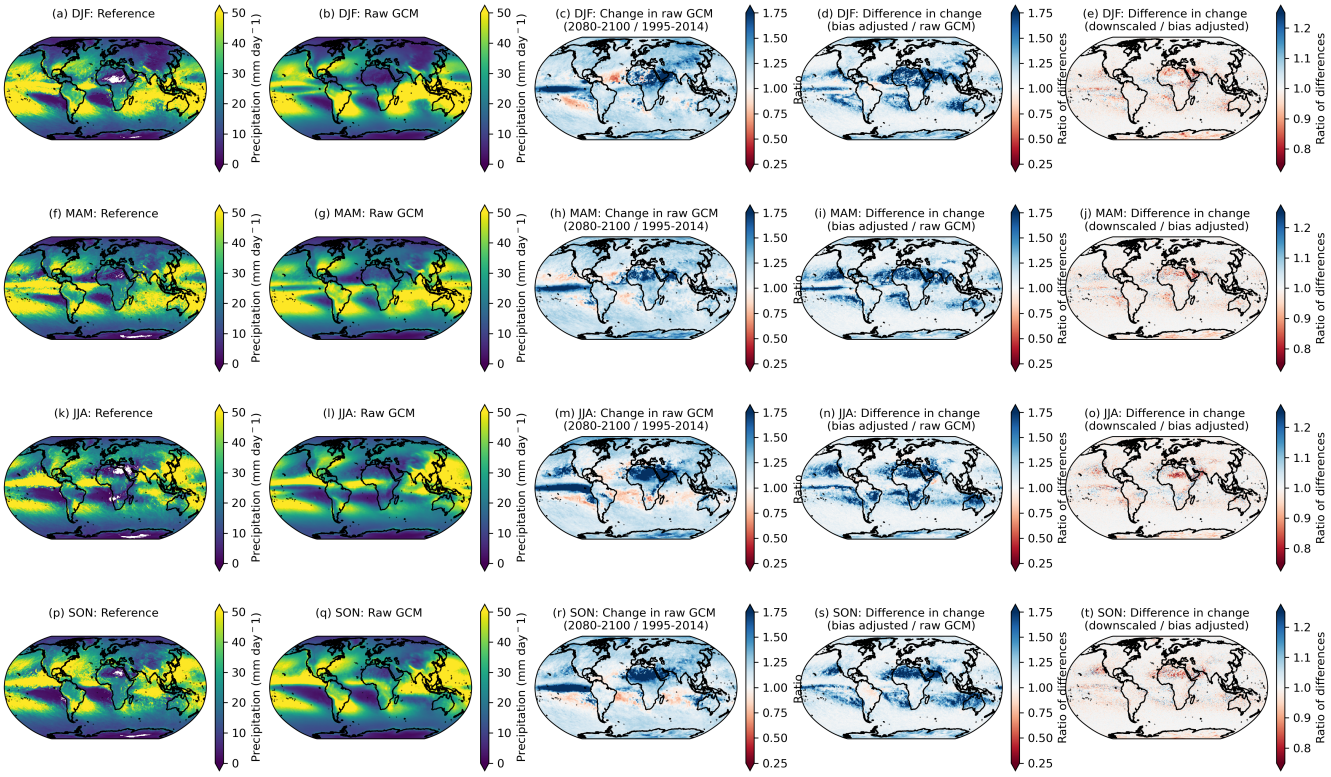

Figure A8. The 99th percentile of seasonal daily total precipitation for the reference (panels a, f, k, p) and raw, cleaned GCM (b, g, l, q) over the training period, 1995–2014. The change in the 99th percentile of seasonal daily total precipitation in 2080–2100 relative to 1995–2014, as a ratio, in the raw, cleaned GCMs (panels c, h, m, r), the ratio of the 99th percentile change between the bias-adjusted and the raw, cleaned GCMs (panels d, i, n, s), and the ratio of the 99th percentile change between the downscaled and the bias-adjusted GCMs (panels e, j, o, t) for seasons DJF (panels a–e), MAM (panels f–j), JJA (panels k–o), and SON (panels p–t). Results shown are the average for wet days across the GCM ensemble for the scenario SSP3-7.0.

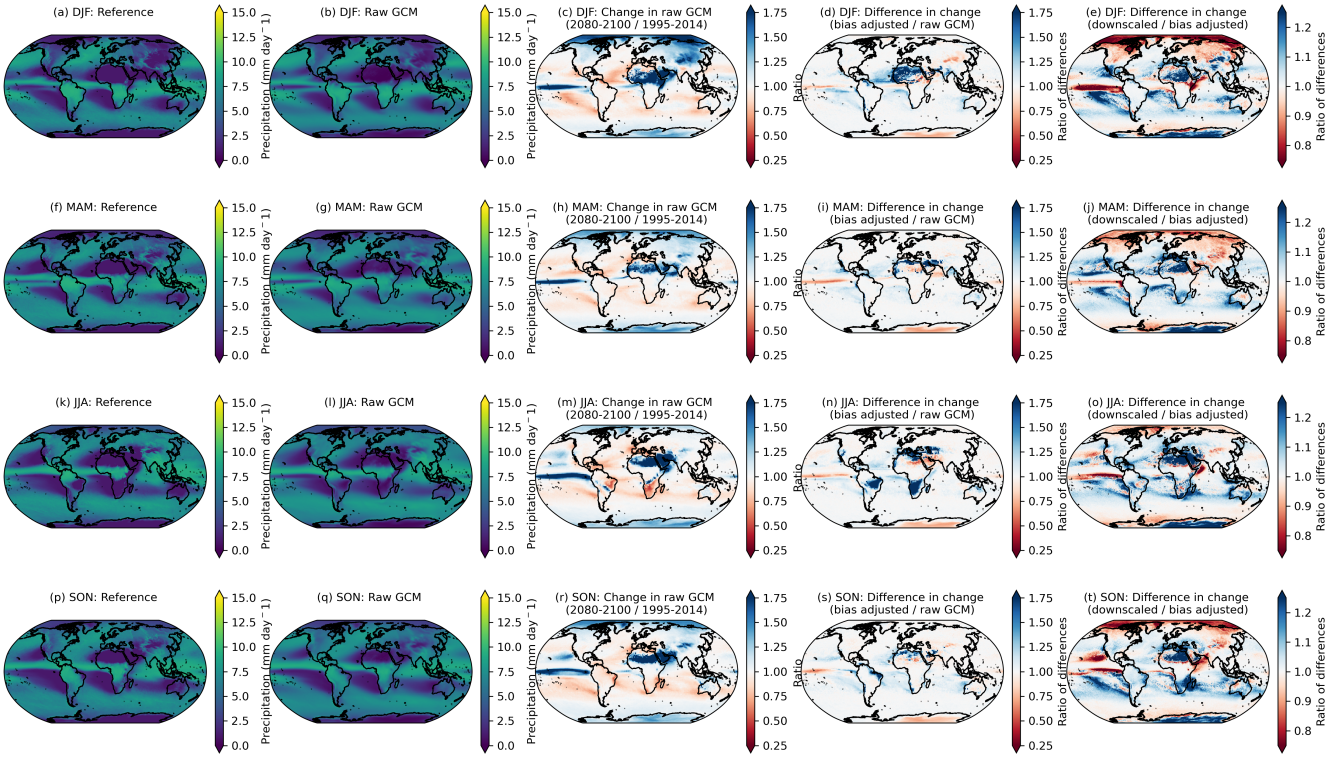

Figure A9. The 95th percentile of seasonal daily total precipitation for the reference (panels a, f, k, p) and raw, cleaned GCM (b, g, l, q) over the training period, 1995–2014. The change in the 95th percentile of seasonal daily total precipitation in 2080–2100 relative to 1995–2014, as a ratio, in the raw, cleaned GCMs (panels c, h, m, r), the ratio of the 95th percentile change between the bias-adjusted and the raw, cleaned GCMs (panels d, i, n, s), and the ratio of the 95th percentile change between the downscaled and the bias-adjusted GCMs (panels e, j, o, t) for seasons DJF (panels a–e), MAM (panels f–j), JJA (panels k–o), and SON (panels p–t). Results shown are the average for drier days (e.g., days with precipitation values < 10 mm day^{-1}) across the GCM ensemble for the scenario SSP3-7.0.

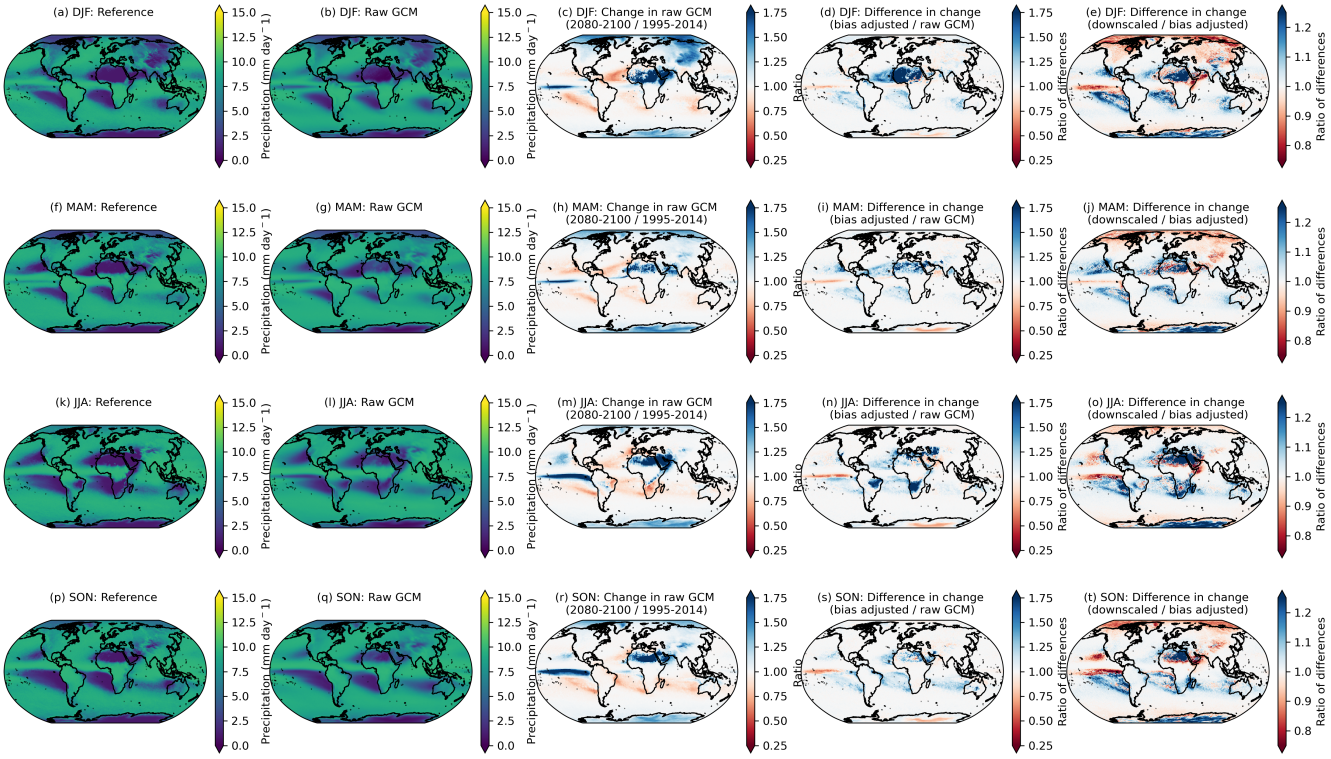

Figure A10. The 99th percentile of seasonal daily total precipitation for the reference (panels a, f, k, p) and raw, cleaned GCM (b, g, l, q) over the training period, 1995–2014. The change in the 99th percentile of seasonal daily total precipitation in 2080–2100 relative to 1995–2014, as a ratio, in the raw, cleaned GCMs (panels c, h, m, r), the ratio of the 99th percentile change between the bias-adjusted and the raw, cleaned GCMs (panels d, i, n, s), and the ratio of the 99th percentile change between the downscaled and the bias-adjusted GCMs (panels e, j, o, t) for seasons DJF (panels a–e), MAM (panels f–j), JJA (panels k–o), and SON (panels p–t). Results shown are the average for drier days (e.g., days with precipitation values < 10 mm day^{-1}) across the GCM ensemble for the scenario SSP3-7.0.

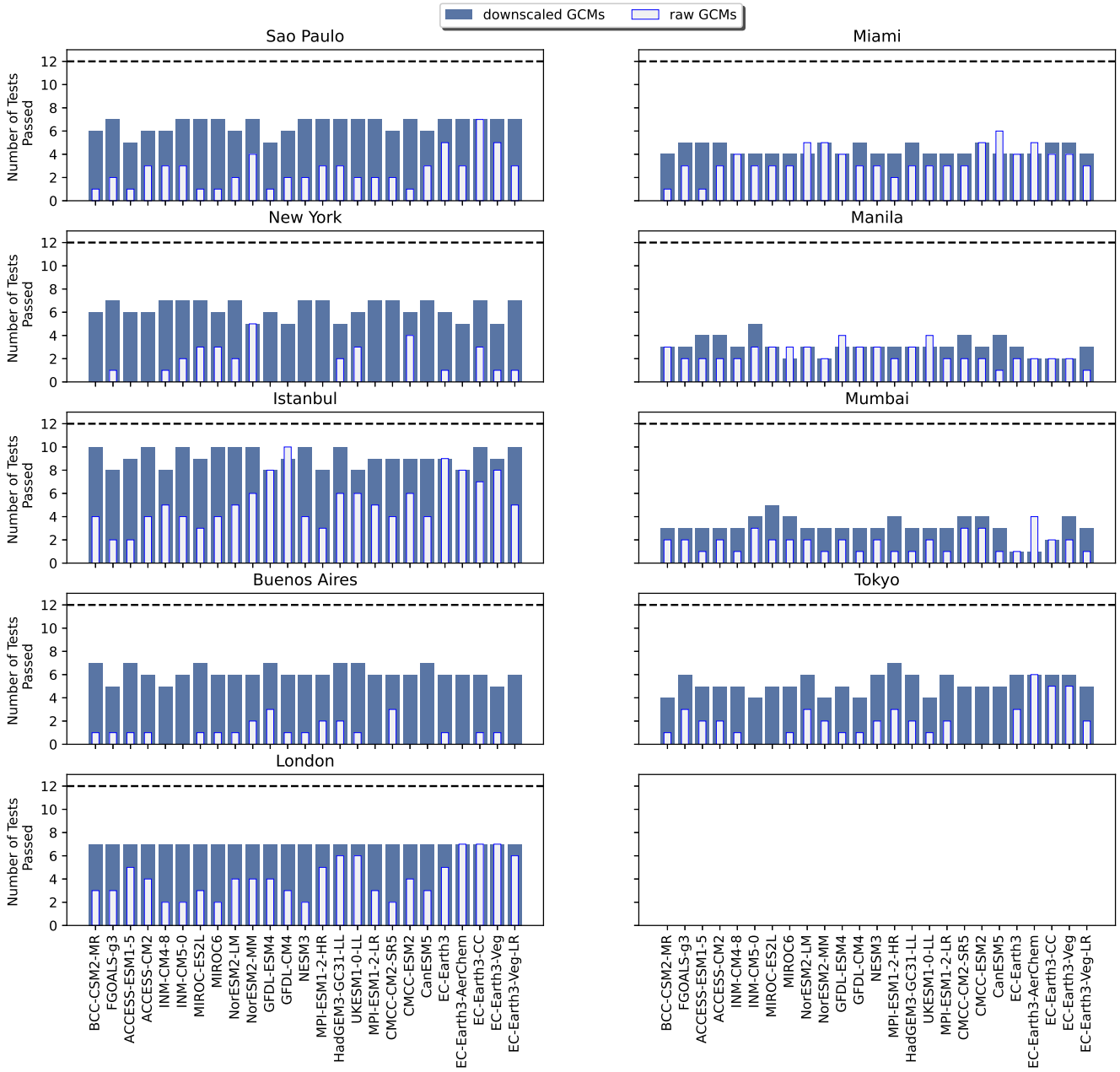

Figure A11. Bar plots showing the number of Kolmogorov-Smirnov tests passed for the twelve selected indices for the bias-adjusted and downscaled GCM and raw GCM (overlain) for each of the GCMs included in the GDPCIR dataset for nine coastal cities around the globe. The dashed line shows the maximum possible number of K-S tests.

Appendix B: Supplementary tables

B1 CMIP6 GCM Inventory

Table B1

Models	Institution	ensemble member	included in GDPCIR dataset	reason for exclusion from GDPCIR dataset
ACCESS–CM2	CSIRO-ARCCSS	r1i1p1f1	YES	
ACCESS–ESM1-5	CSIRO	r1i1p1f1	YES	
AWI-CM-1-1-MR	AWI	r1i1p1f1	NO	not available in GC CMIP6 collection
AWI-CM-1-1-LR	AWI	r1i1p1f1	NO	not available in GC CMIP6 collection
BCC-CSM2-MR	BCC	r1i1p1f1	YES	
BCC-ESM1	BCC	r1i1p1f1	NO	not available in GC CMIP6 collection
CAMS-CSM1-0	CAMS	r2i1p1f1	NO	not available in GC CMIP6 collection
CAS-ESM2-0	CAS	r1i1p1f1	NO	not available in ESGF
CESM2	NCAR	r4i1p1f1	NO	historical daily output not available in ESGF
CESM2-FV2	NCAR	r1i1p1f1	NO	not available in GC CMIP6 collection
CESM2-WACCM	NCAR	r1i1p1f1	NO	not available in GC CMIP6 collection
CESM2-WACCM-FV2	NCAR	r1i1p1f1	NO	not available in GC CMIP6 collection
CIESM	THU	r1i1p1f1	NO	not available in ESGF
CMCC-ESM2	CMCC	r1i1p1f1	YES	
CMCC-CM2-SR5	CMCC	r1i1p1f1	YES	
CNRM-CM6-1	CNRM-CERFACS	r1i1p1f2	NO	licensing issues for commercial use
CNRM-CM6-1-HR	CNRM-CERFACS	r1i1p1f2	NO	licensing issues for commercial use
CNRM-ESM2-1	CNRM-CERFACS	r1i1p1f2	NO	licensing issues for commercial use
CanESM5	CCCma	r1i1p1f1	YES	
EC-Earth3-Veg	EC-Earth-Consortium	r1i1p1f1	YES	
EC-Earth3	EC-Earth-Consortium	r1i1p1f1	YES	
EC-Earth3-AerChem	EC-Earth-Consortium	r1i1p1f1	YES	
EC-Earth3-Veg-LR	EC-Earth-Consortium	r1i1p1f1	YES	
FGOALS-f3-L	CAS	r1i1p1f1	NO	not available in GC CMIP6 collection
FGOALS-g3	CAS	r1i1p1f1	YES	

Table B1 continued from previous page

Models	Institution	ensemble member	included in GDPCIR dataset	reason for exclusion from GDPCIR dataset
FIO-ESM-2-0	FIO-QLNM	r1i1p1f1	NO	not available in ESGF
GFDL-CM4	NOAA-GFDL	r1i1p1f1	YES	
GFDL-ESM4	NOAA-GFDL	r1i1p1f1	YES	
GISS-E2-1-G	NASA-GISS	r1i1p1f1	NO	not available in GC CMIP6 collection
HadGEM3-GC31-LL	MOHC	r1i1p1f3	YES	
HadGEM3-GC31-MM	MOHC	r1i1p1f3	NO	Only SSP1-2.6/SSP5-8.5 available in GC CMIP6 collection
UKESM1-0-LL	MOHC	r1i1p1f2	YES	
IITM-ESM	CCCR-IITM	r1i1p1f1	NO	not available in GC CMIP6 collection
INM-CM4-8	INM	r1i1p1f1	YES	
INM-CM5-0	INM	r1i1p1f1	YES	
IPSL-CM6A-LR	IPSL	r1i1p1f1	NO	licensing issues for commercial use
KACE-1-0-G	NIMS-KMA	r1i1p1f1	NO	QA/QC pipeline found data issues
KIOST-ESM	KIOST	r1i1p1f1	NO	QA/QC pipeline found data issues
MCM-UA-1-0	UA	r1i1p1f1	NO	not available in ESGF
MIROC6	MIROC	r1i1p1f1	YES	
MIROC-ES2L	MIROC	r1i1p1f1	YES	
MPI-ESM1-2-HR	MPI-M	r1i1p1f1	YES	
MPI-ESM-1-2-HAM	HAMMOZ-Consortium	r1i1p1f1	NO	no data past 2055
MPI-ESM1-2-LR	MPI-M	r1i1p1f1	YES	
MRI-ESM2-0	MRI	r1i1p1f1	NO	QA/QC pipeline found data issues
NESM3	NUIST	r1i1p1f1	YES	
NorCPM1	NCC	r1i1p1f1	NO	not available in ESGF
NorESM2-LM	NCC	r1i1p1f1	YES	
NorESM2-MM	NCC	r1i1p1f1	YES	
SAM0-UNICON	SNU	r1i1p1f1	NO	not available in GC CMIP6 collection
TaiESM1	AS-RCEC	r1i1p1f1	NO	not available in GC CMIP6 collection

Appendix C: Pipeline computing resources

The downscaling pipeline was run on Kubernetes clusters with a flexible pool of preemptible ("spot") general-purpose machines. Each machine had between 8 and 32 CPUs using Intel Skylake, Broadwell, Haswell, Sandy Bridge, and Ivy Bridge CPU platforms. All machines were "high-memory" with 8 GB per CPU.

A downscaling run on a single GCM projection experiment for a single variable. For example, minimum daily air temperature in SSP2-4.5 from EC-Earth3-Veg-LR required approximately 500 CPU hours and 3,500 GiB hours. This completes with a wall time of 2–3 hours. This work can easily run in parallel to other downscaling jobs if preemptible machines are available to the cluster. The complete set of downscaling jobs could complete within 3 days.

Appendix D: Global temperature and precipitation changes

In this section we explore trends in global temperature and precipitation across ~~models~~GCMs. We report these trends for both the source data and the bias-adjusted and downscaled data in order to shed light on how these global trends are affected by QDM and QPLAD. To obtain global values, the data is averaged using land-weighting. Results are shown in Fig ~~A1~~D1. We find that when comparing the source data with the bias-adjusted and downscaled data, global trends in temperature are preserved: all the differences across models and scenarios are within $\pm$ 0.1°C. In contrast, changes in global precipitation have some amount of inflation across all models and scenarios. Going further, in the SSP2-4.5 and SSP3-7.0(respectively) source data, change in average annual mean maximum temperature across models ranges from 1.71°C (2.56°C) to 4.55°C (6.53°C) and in the bias-adjusted downscaled data this range is almost identical, from 1.71°C (2.84°C) to 4.55°C (6.54°C). In contrast, change in average annual total precipitation ranges from -0.11% (-2.47%) to 8.99% (9.61%) in the source data and is shifted upwards in the bias-adjusted and downscaled data, from 2.57% (-0.79%) to 12.6% (15.22%). For precipitation, the largest change is in the scenario SSP3-7.0, CanESM5 model, with a source trend of around 7.5% and a trend in our results of 15%. This model also has one of the highest precipitation trends in the source data, but there is no systematic relationship between the magnitude of the source trend and the magnitude of trend modification. For example, NorESM2-MM SSP2-4.5 has a trend close to zero in the source data and in the results the trend is around 4%, whereas BCC-CSM2-MR has a trend of around 2.5% in both scenarios and the alteration is very low at less than 0.2 percentage points in both scenarios.

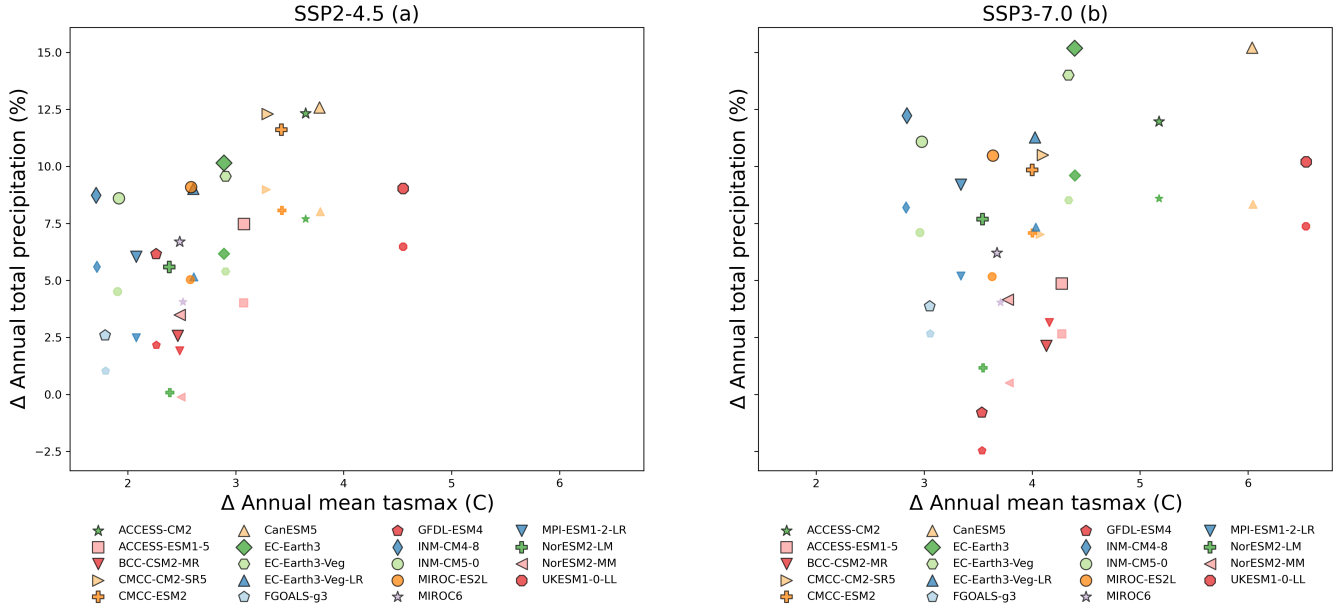

Figure A1D1. Changes in temperature and precipitation signals in CMIP6 source models and CMIP6 bias-adjusted and downscaled models. For each model, scenario and pixel, the annual average (x-axis) and the annual total (y-axis) is computed for each year of both the historical (1995-2015) and future (2080-2100) period. Then, the data is averaged over space with a land-weighting scheme (e.g. ocean pixels are assigned zero weights). Finally, the data is averaged over years for both the historical and future period separately and the difference between the future and historical global values (x-axis) or the percent change between the future and the historical global values (y-axis) is plotted. Data point symbols with transparent borders represent the source model data while those with black color borders represent the bias-adjusted and downscaled data. The list of models is restricted to those that have bias-adjusted and downscaled data for both SSP2-4.5 and SSP3-7.0.

~~**Figure A2.** Difference in the change in 95th percentile JJA maximum temperature trends globally between 2080-2100 and the training period (1995-2014) between the bias-adjusted and downscaled data and the bias-adjusted data. Results are shown for the model NorESM2-LM and the scenario SSP3-7.0.~~

Author contributions.

DRG~~led the design, methods development and collaborations for the CMIP6 downscaling project and also contributed extensively to code development and pipeline architecture. SBM was the lead data scientist and engineer, developed our downscaling pipeline and also contributed to methods development. KEMand MTD contributed to method and code developmentthroughout the project. ET worked on code development and helped with production runs and later worked on analysis and QC of results.~~

~~MAF led analysisand preparation of ERA5 reanalysis data for the production pipeline. REK provided technical advice and oversight on the project throughout.~~

~~DRGwrote the first complete draft of the article with contributions from SBM(reproducibility section) and MTD (results aggregated to state and country levels). DRG~~, KEM, SBM, MTD, and REK designed the methods and project architecture. DRG, SBM, KEM, MTD and ET contributed to code development. DRG, KEM, MTD, SBM, ET and MAF performed analysis. DRG, SBM, ET, MTD and KEM ~~made all figures. REK provided an extremely thorough review. DRG revised the article, iteratively with comments from KEM, ET and SBM~~drafted the manuscript, and DRG, SBM, KEM, MTD, ET and REK revised the final manuscript.

Competing interests.

The authors declare that no competing interests are present.

Acknowledgements. The authors acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and would like to thank the climate modeling groups for producing and making available their model output. They would further like to thank Lamont-Doherty Earth Observatory, the Pangeo Consortium, Google Cloud, and the Google Public Datasets program for making the CMIP6 Google Cloud collection possible. In particular, the authors would like to express gratitude to Ryan Abernathey, Naomi Henderson, and Charles Blackmon-Luca for ongoing collaboration on making the CMIP6 Zarr stores analysis-ready. The authors are grateful to the xclim developers, in particular, Pascal Bourgault and Travis Logan, for implementing the QDM ~~bias-adjustment~~ bias adjustment method in the xclim Python package, supporting the authors' QPLAD implementation into the package, and ongoing support integrating the dask parallel computing library into downscaling workflows. For support with securing funding, outreach and helpful insights throughout the project, the authors would like to express their gratitude to Hannah Hess. For method advice and useful conversations, the authors would like to thank Keith Dixon and Dennis Adams-Smith. Thanks to Dan Morris, Tom Augspurger, and the team at Microsoft AI for Earth and Azure for help coordinating computing infrastructure, public data release, and all the subtle steps in between.

Financial support.

This research has been supported by the Rockefeller Foundation and the Microsoft AI for Earth Initiative.

References

Abernathey, R. P., Augspurger, T., Banihirwe, A., Blackmon-Luca, C. C., Crone, T. J., Gentemann, C. L., Hamman, J. J., Henderson, N., Lepore, C., McCaie, T. A., Robinson, N. H., and Signell, R. P.: Cloud-Native Repositories for Big Scientific Data, Computing in Science & Engineering, 23, 26–35, https://doi.org/10.1109/MCSE.2021.3059437, 2021.

Agbazo, M. N. and Grenier, P.: Characterizing and avoiding physical inconsistency generated by the application of univariate quantile mapping on daily minimum and maximum temperatures over Hudson Bay, International Journal of Climatology, 40, 3868–3884, https://doi.org/10.1002/joc.6432, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.6432, 2020.

Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M.: On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections, Climate Dynamics, 57, 2941–2951, https://doi.org/10.1007/s00382-021-05847-0, 2021.

Bennett, A. R., Hamman, J. J., and Nijssen, B.: MetSim: A Python package for estimation and disaggregation of meteorological data, Journal of Open Source Software, 5, 2042, https://doi.org/10.21105/joss.02042, 2020.

Bürger, G., Murdock, T. Q., Werner, A. T., Sobie, S. R., and Cannon, A. J.: Downscaling Extremes—An Intercomparison of Multiple Statistical Methods for Present Climate, Journal of Climate, 25, 4366–4388, https://doi.org/10.1175/JCLI-D-11-00408.1, publisher: American Meteorological Society Section: Journal of Climate, 2012.

Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, Climate Dynamics, 50, 31–49, https://doi.org/10.1007/s00382-017-3580-6, 2018.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes?, Journal of Climate, 28, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1, publisher: American Meteorological Society Section: Journal of Climate, 2015.

Cannon, A. J., Piani, C., and Sippel, S.: Chapter 5 - Bias correction of climate model output for impact models, in: Climate Extremes and Their Implications for Impact and Risk Assessment, edited by Sillmann, J., Sippel, S., and Russo, S., pp. 77–104, Elsevier, https://doi.org/10.1016/B978-0-12-814895-2.00005-7, 2020.

Carleton, T., Jina, A., Delgado, M., Greenstone, M., Houser, T., Hsiang, S., Hultgren, A., Kopp, R. E., McCusker, K. E., Nath, I., Rising, J., Rode, A., Seo, H. K., Viaene, A., Yuan, J., and Zhang, A. T.: Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits, The Quarterly Journal of Economics, 137, 2037–2105, https://doi.org/10.1093/qje/qjac020, 2022.

Casanueva, A., Kotlarski, S., Herrera, S., Fischer, A. M., Kjellstrom, T., and Schwierz, C.: Climate projections of a multivariate heat stress index: the role of downscaling and bias correction, Geoscientific Model Development, 12, 3419–3438, https://doi.org/10.5194/gmd-12-3419-2019, 2019.

Casanueva, A., Herrera, S., Iturbide, M., Lange, S., Jury, M., Dosio, A., Maraun, D., and Gutiérrez, J. M.: Testing bias adjustment methods for regional climate change applications under observational uncertainty and resolution mismatch, Atmospheric Science Letters, 21, e978, https://doi.org/10.1002/asl.978, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asl.978, 2020.

CIESIN: Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11, https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-adjusted-to-2015-unwpp-country-totals-rev11, center for International Earth Science Information Network - CIESIN - Columbia University, Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), 2018.

Dai, A.: Precipitation Characteristics in Eighteen Coupled Climate Models, Journal of Climate, 19, 4605–4630, https://doi.org/10.1175/JCLI3884.1, publisher: American Meteorological Society Section: Journal of Climate, 2006.

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

François, B., Vrac, M., Cannon, A. J., Robin, Y., and Allard, D.: Multivariate bias corrections of climate simulations: which benefits for which losses?, Earth System Dynamics, 11, 537–562, https://doi.org/10.5194/esd-11-537-2020, publisher: Copernicus GmbH, 2020a.

François, B., Vrac, M., Cannon, A. J., Robin, Y., and Allard, D.: Multivariate bias corrections of climate simulations: which benefits for which losses?, Earth System Dynamics, 11, 537–562, https://doi.org/10.5194/esd-11-537-2020, publisher: Copernicus GmbH, 2020b.

Gutiérrez, J. M., Maraun, D., Widmann, M., Huth, R., Hertig, E., Benestad, R., Roessler, O., Wibig, J., Wilcke, R., Kotlarski, S., San Martín, D., Herrera, S., Bedia, J., Casanueva, A., Manzanas, R., Iturbide, M., Vrac, M., Dubrovsky, M., Ribalaygua, J., Pórtoles, J., Räty, O., Räisänen, J., Hingray, B., Raynaud, D., Casado, M. J., Ramos, P., Zerenner, T., Turco, M., Bosshard, T., Štěpánek, P., Bartholy, J., Pongracz, R., Keller, D. E., Fischer, A. M., Cardoso, R. M., Soares, P. M. M., Czernecki, B., and Pagé, C.: An intercomparison of a large ensemble of statistical downscaling methods over Europe: Results from the VALUE perfect predictor cross-validation experiment, International Journal of Climatology, 39, 3750–3785, https://doi.org/10.1002/joc.5462, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.5462, 2019.

Hacker, J.: The Essential Components of the Downscaling Toolbox, https://jupiterintel.com/wp-content/uploads/2021/04/Jupiter-Downscaling-Science-Insights.pdf, jupiter Intelligence Science Insights, 2021.

Hagos, S. M., Leung, L.-Y., Garuba, O. A., Demott, C., Harrop, B. E., Lu, J., and Ahn, M.-S.: The Relationship between Precipitation and Precipitable Water in CMIP6 Simulations and Implications for Tropical Climatology and Change, Journal of Climate, 34, https://doi.org/10.1175/jcli-d-20-0211.1, institution: Pacific Northwest National Lab. (PNNL), Richland, WA (United States) Number: PNNL-SA-152254 Publisher: American Meteorological Society, 2021.

Hassell, D., Gregory, J., Blower, J., Lawrence, B. N., and Taylor, K. E.: A data model of the Climate and Forecast metadata conventions (CF-1.6) with a software implementation (cf-python v2.1), Geoscientific Model Development, 10, 4619–4646, https://doi.org/10.5194/gmd-10-4619-2017, 2017.

Hassler, B. and Lauer, A.: Comparison of Reanalysis and Observational Precipitation Datasets Including ERA5 and WFDE5, Atmosphere, 12, 1462, https://doi.org/10.3390/atmos12111462, number: 11 Publisher: Multidisciplinary Digital Publishing Institute, 2021.

Hempel, S., Frieler, K., Warszawski, L., Schewe, J., and Piontek, F.: A trend-preserving bias correction – the ISI-MIP approach, Earth System Dynamics, 4, 219–236, https://doi.org/10.5194/esd-4-219-2013, publisher: Copernicus GmbH, 2013.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on single levels from 1979 to present, https://doi.org/10.24381/cds.adbb2d47, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (Accessed 14 April 2021), 2018.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Holthuijzen, M., Beckage, B., Clemins, P. J., Higdon, D., and Winter, J. M.: Robust bias-correction of precipitation extremes using a novel hybrid empirical quantile-mapping method, Theoretical and Applied Climatology, 149, 863–882, https://doi.org/10.1007/s00704-022-04035-2, 2022.

Iturbide, M., Casanueva, A., Bedia, J., Herrera, S., Milovac, J., and Gutiérrez, J. M.: On the need of bias adjustment for more plausible climate change projections of extreme heat, Atmospheric Science Letters, 23, e1072, https://doi.org/10.1002/asl.1072, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asl.1072, 2022.

Karl, T. R., Nicholls, N., and Ghazi, A.: Clivar/GCOS/WMO Workshop on Indices and Indicators for Climate Extremes Workshop Summary, Climatic Change, 42, 3–7, https://doi.org/10.1023/A:1005491526870, 1999.

Lange, S.: Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0), Geoscientific Model Development, 12, 3055–3070, https://doi.org/10.5194/gmd-12-3055-2019, publisher: Copernicus GmbH, 2019.

Lange, S.: ISIMIP3BASD, https://doi.org/10.5281/zenodo.4686991, 2021.

Lanzante, J. R., Adams-Smith, D., Dixon, K. W., Nath, M., and Whitlock, C. E.: Evaluation of some distributional downscaling methods as applied to daily maximum temperature with emphasis on extremes, International Journal of Climatology, 40, 1571–1585, https://doi.org/10.1002/joc.6288, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.6288, 2020.

Lehner, F., Nadeem, I., and Formayer, H.: Evaluating quantile-based bias adjustment methods for climate change scenarios, Hydrol. Earth Syst. Sci. Discuss., 2021, 1–26, https://doi.org/10.5194/hess-2021-498, publisher: Copernicus Publications, 2021.

Lehner, F., Nadeem, I., and Formayer, H.: Evaluating skills and issues of quantile-based bias adjustment for climate change scenarios, Advances in Statistical Climatology, Meteorology and Oceanography, 9, 29–44, https://doi.org/10.5194/ascmo-9-29-2023, publisher: Copernicus GmbH, 2023.

Li, H., Sheffield, J., and Wood, E. F.: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/10.1029/2009JD012882, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2009JD012882, 2010.

Logan, T., Bourgault, P., Smith, T. J., Huard, D., Biner, S., Labonté, M.-P., Rondeau-Genesse, G., Fyke, J., Aoun, A., Roy, P., Ehbrecht, C., Caron, D., Stephens, A., Whelan, C., Low, J.-F., and Lavoie, J.: Ouranosinc/xclim: v0.31.0, https://doi.org/10.5281/zenodo.5649661, 2021.

Lokoshchenko, M. A.: Urban 'heat island' in Moscow, Urban Climate, 10, 550–562, https://doi.org/10.1016/j.uclim.2014.01.008, 2014.

Maraun, D.: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue, Journal of Climate, 26, 2137–2143, https://doi.org/10.1175/JCLI-D-12-00821.1, publisher: American Meteorological Society Section: Journal of Climate, 2013.

Maraun, D.: Bias Correcting Climate Change Simulations - a Critical Review, Current Climate Change Reports, 2, 211–220, https://doi.org/10.1007/s40641-016-0050-x, 2016.

Maraun, D. and Widmann, M.: Statistical Downscaling and Bias Correction for Climate Research, Cambridge University Press, Cambridge, https://doi.org/10.1017/9781107588783, 2018.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, Ö., Yu, R., and Zhou, B., eds.: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, 2021.

Maurer, E. P. and Pierce, D. W.: Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean, Hydrology and Earth System Sciences, 18, 915–925, https://doi.org/10.5194/hess-18-915-2014, 2014.

McNicholl, B., Lee, Y. H., Campbell, A. G., and Dev, S.: Evaluating the Reliability of Air Temperature From ERA5 Reanalysis Data, IEEE Geoscience and Remote Sensing Letters, 19, 1–5, https://doi.org/10.1109/LGRS.2021.3137643, conference Name: IEEE Geoscience and Remote Sensing Letters, 2022.

Meehl, G. A., Senior, C. A., Eyring, V., Flato, G., Lamarque, J.-F., Stouffer, R. J., Taylor, K. E., and Schlund, M.: Context for interpreting equilibrium climate sensitivity and transient climate response from the CMIP6 Earth system models, Science Advances, 6, eaba1981, https://doi.org/10.1126/sciadv.aba1981, publisher: American Association for the Advancement of Science, 2020.

Michelangeli, P.-A., Vrac, M., and Loukos, H.: Probabilistic downscaling approaches: Application to wind cumulative distribution functions, Geophysical Research Letters, 36, https://doi.org/10.1029/2009GL038401, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2009GL038401, 2009.

Mistry, M. N., Schneider, R., Masselot, P., Royé, D., Armstrong, B., Kyselý, J., Orru, H., Sera, F., Tong, S., Lavigne, E., Urban, A., Madureira, J., García-León, D., Ibarreta, D., Ciscar, J.-C., Feyen, L., de Schrijver, E., de Sousa Zanotti Stagliorio Coelho, M., Pascal, M., Tobias, A., Guo, Y., Vicedo-Cabrera, A. M., and Gasparrini, A.: Comparison of weather station and climate reanalysis data for modelling temperature-related mortality, Scientific Reports, 12, 5178, https://doi.org/10.1038/s41598-022-09049-4, 2022.

Müller, C., Franke, J., Jägermeyr, J., Ruane, A. C., Elliott, J., Moyer, E., Heinke, J., Falloon, P. D., Folberth, C., Francois, L., Hank, T., Izaurralde, R. C., Jacquemin, I., Liu, W., Olin, S., Pugh, T. A. M., Williams, K., and Zabel, F.: Exploring uncertainties in global crop yield projections in a large ensemble of crop models and CMIP5 and CMIP6 climate scenarios, Environmental Research Letters, 16, 034 040, https://doi.org/10.1088/1748-9326/abd8fc, 2021.

Natural Earth: 1:10m Cultural Vectors, Version 5.0.1, https://naturalearthdata.com, 2022.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geoscientific Model Development, 9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, publisher: Copernicus GmbH, 2016.

Parsons, L. A., Masuda, Y. J., Kroeger, T., Shindell, D., Wolff, N. H., and Spector, J. T.: Global labor loss due to humid heat exposure underestimated for outdoor workers, Environmental Research Letters, 17, 014 050, https://doi.org/10.1088/1748-9326/ac3dae, publisher: IOP Publishing, 2022.

Pierce, D. W.: LOCA Statistical Downscaling (Localized Constructed Analogs), https://loca.ucsd.edu/loca-calendar/, 2021.

Pierce, D. W., Cayan, D. R., and Thrasher, B. L.: Statistical Downscaling Using Localized Constructed Analogs (LOCA), Journal of Hydrometeorology, 15, 2558–2585, https://doi.org/10.1175/JHM-D-14-0082.1, publisher: American Meteorological Society Section: Journal of Hydrometeorology, 2014.

Pierce, D. W., Cayan, D. R., Maurer, E. P., Abatzoglou, J. T., and Hegewisch, K. C.: Improved Bias Correction Techniques for Hydrological Simulations of Climate Change, Journal of Hydrometeorology, 16, 2421 – 2442, https://doi.org/https://doi.org/10.1175/JHM-D-14-0236.1, 2015.

Previdi, M., Smith, K. L., and Polvani, L. M.: Arctic amplification of climate change: a review of underlying mechanisms, Environmental Research Letters, 16, 093 003, https://doi.org/10.1088/1748-9326/ac1c29, publisher: IOP Publishing, 2021.

Qian, W. and Chang, H. H.: Projecting Health Impacts of Future Temperature: A Comparison of Quantile-Mapping Bias-Correction Methods, International Journal of Environmental Research and Public Health, 18, 1992, https://doi.org/10.3390/ijerph18041992, 2021.

Rajulapati, C. R., Papalexiou, S. M., Clark, M. P., and Pomeroy, J. W.: The Perils of Regridding: Examples Using a Global Precipitation Dataset, Journal of Applied Meteorology and Climatology, 60, 1561 – 1573, https://doi.org/https://doi.org/10.1175/JAMC-D-20-0259.1, place: Boston MA, USA Publisher: American Meteorological Society, 2021.

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., Kc, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Da Silva, L. A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J. C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., and Tavoni, M.: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview, Global Environmental Change, 42, 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009, 2017.

Rode, A., Carleton, T., Delgado, M., Greenstone, M., Houser, T., Hsiang, S., Hultgren, A., Jina, A., Kopp, R. E., McCusker, K. E., Nath, I., Rising, J., and Yuan, J.: Estimating a social cost of carbon for global energy consumption, Nature, 598, 308–314, https://doi.org/10.1038/s41586-021-03883-8, number: 7880 Publisher: Nature Publishing Group, 2021.

Sanabria, L. A., Qin, X., Li, J., and Cechet, R. P.: Bias correction of extreme values of high-resolution climate simulations for risk analysis, Theoretical and Applied Climatology, 150, 1015–1026, https://doi.org/10.1007/s00704-022-04210-5, 2022.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling, Journal of Climate, 19, 3088–3111, https://doi.org/10.1175/JCLI3790.1, publisher: American Meteorological Society Section: Journal of Climate, 2006.

Sheridan, S. C., Lee, C. C., and Smith, E. T.: A Comparison Between Station Observations and Reanalysis Data in the Identification of Extreme Temperature Events, Geophysical Research Letters, 47, e2020GL088 120, https://doi.org/10.1029/2020GL088120, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL088120, 2020.

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., and Bronaugh, D.: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate, Journal of Geophysical Research: Atmospheres, 118, 1716–1733, https://doi.org/10.1002/jgrd.50203, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jgrd.50203, 2013.

Supharatid, S., Aribarg, T., and Nafung, J.: Bias-corrected CMIP6 climate model projection over Southeast Asia, Theoretical and Applied Climatology, 147, 669–690, https://doi.org/10.1007/s00704-021-03844-1, 2022.

Tarek, M., Brissette, F. P., and Arsenault, R.: Large-Scale Analysis of Global Gridded Precipitation and Temperature Datasets for Climate Change Impact Studies, Journal of Hydrometeorology, 21, 2623 – 2640, https://doi.org/https://doi.org/10.1175/JHM-D-20-0100.1, 2020.

Themeßl, M. J., Gobiet, A., and Heinrich, G.: Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal, Climatic Change, 112, 449–468, https://doi.org/10.1007/s10584-011-0224-4, 2012.

Thrasher, B., Maurer, E. P., McKellar, C., and Duffy, P. B.: Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping, Hydrology and Earth System Sciences, 16, 3309–3314, https://doi.org/10.5194/hess-16-3309-2012, publisher: Copernicus GmbH, 2012.

Thrasher, B., Wang, W., Michaelis, A., and Nemani, R.: NEX-GDDP-CMIP6, https://doi.org/10.7917/OFSG3345, 2021.

Thrasher, B., Wang, W., Michaelis, A., Melton, F., Lee, T., and Nemani, R.: NASA Global Daily Downscaled Projections, CMIP6, Scientific Data, 9, 262, https://doi.org/10.1038/s41597-022-01393-4, number: 1 Publisher: Nature Publishing Group, 2022.

Tian, B. and Dong, X.: The Double-ITCZ Bias in CMIP3, CMIP5, and CMIP6 Models Based on Annual Mean Precipitation, Geophysical Research Letters, 47, e2020GL087 232, https://doi.org/10.1029/2020GL087232, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2020GL087232, 2020.

Van de Velde, J., Demuzere, M., De Baets, B., and Verhoest, N. E. C.: Impact of bias nonstationarity on the performance of uni- and multivariate bias-adjusting methods, Hydrology and Earth System Sciences Discussions, pp. 1–47, https://doi.org/10.5194/hess-2020-639, publisher: Copernicus GmbH, 2020.

van Hengstum, P. J., Donnelly, J. P., Fall, P. L., Toomey, M. R., Albury, N. A., and Kakuk, B.: The intertropical convergence zone modulates intense hurricane strikes on the western North Atlantic margin, Scientific Reports, 6, 21 728, https://doi.org/10.1038/srep21728, number: 1 Publisher: Nature Publishing Group, 2016.

Warren, R., Hope, C., Gernaat, D. E. H. J., Van Vuuren, D. P., and Jenkins, K.: Global and regional aggregate damages associated with global warming of 1.5 to 4 °C above pre-industrial levels, Climatic Change, 168, 24, https://doi.org/10.1007/s10584-021-03198-7, 2021.

Xu, Z., Han, Y., Tam, C.-Y., Yang, Z.-L., and Fu, C.: Bias-corrected CMIP6 global dataset for dynamical downscaling of the historical and future climate (1979–2100), Scientific Data, 8, 293, https://doi.org/10.1038/s41597-021-01079-3, number: 1 Publisher: Nature Publishing Group, 2021.