STAR-ESDM: A Generalizable Approach to Generating High-Resolution Climate Projections through Signal Decomposition

Katharine Hayhoe[1,2], Ian Scott-Fleming[1], Anne Stoner[3] and Donald J. Wuebbles[3,4]

[1] Climate Center, Texas Tech University, Lubbock TX 79409 USA.

[2] The Nature Conservancy, 4245 Fairfax Dr., Arlington VA 22203 USA.

[3] Earth Knowledge, Inc., Tucson AZ 85751 USA.

[4] Dept. of Atmospheric Sciences, University of Illinois at Urbana-Champaign, Urbana IL 61801 USA.

Corresponding author: Katharine Hayhoe (katharine.hayhoe@ttu.edu)

Key Points:

- We've developed a rapid, flexible and generalizable approach to bias-correct & downscale climate model output to any observational dataset
- Model stationarity ensures projected changes in temperature and precipitation are similar to those generated by a high-resolution global model
- STAR-ESDM can be applied using predictors/predictands from global/regional climate models, satellites, reanalysis, and weather stations.

Abstract

High-resolution climate projections provide crucial insights into assessing climate risk and developing climate resilience strategies. The Seasonal Trends and Analysis of Residuals empirical statistical downscaling model (STAR-ESDM) is a computationally-efficient and flexible approach to generating high-resolution climate projections that can be applied globally using a broad range of predictands and predictors that can be sourced from weather stations, gridded datasets, satellites, reanalysis, and global or regional climate models. It uses signal processing combined with Fourier filtering and kernal density estimation techniques to decompose and smooth any quasi-Gaussian time series, gridded or point-based, into multi-decadal long-term means and/or trends; static and dynamic annual cycles; and probability distributions of high-frequency variability. Long-term predictor trends are bias-corrected and predictor components are used to map remaining predictand components to future conditions. Components are then recombined for each station or grid cell to produce a continuous, high-resolution bias-corrected and downscaled time series at the spatial and temporal scale of the original time series. Comparing STAR-ESDM output with high-resolution daily temperature and precipitation projections generated by a fully dynamical global model demonstrates that the method is extremely robust, capable of accurately reproducing projected changes for all but the most extreme temperature and precipitation values. For most continental areas, biases in 1-in-1000 hottest and coldest temperatures are less than 0.5°C and biases in the 1-in-1000 wet day precipitation amounts are less than 5 mm/day. As climate impacts intensify, STAR-ESDM represents a significant advance in generating consistent high-resolution projections to comprehensively assess risk and optimize resilience.

Plain Language Summary

The STAR-ESDM tool is able to quickly and accurately generate climate projections for individual weather stations and high-resolution grids anywhere in the world. It does this by breaking down global or regional climate model output into different components, from the long-term trend to the day-to-day variability, then merging modelled projected changes with observations. When tested against projections generated by a much more computationally expensive and complex fully dynamical global model, STAR-ESDM produced almost the same output even for very extreme temperature and precipitation values, at a fraction of the computational cost. Moreover, unlike most statistical downscaling models, this method isn't tied to any specific geographic area or predictand and/or predictor dataset. It can be applied to any regional or global dataset, whether generated by a climate or reanalysis model, derived from satellite observations, recorded at weather stations, and more. As climate impacts escalate, STAR-ESDM offers a flexible and effective way to generate the high-resolution climate projections needed to better gauge climate risk and enhance resilience anywhere in the world where reliable observational or quasi-observational data, including from reanalysis or satellites, are available.

1 Introduction

Climate is now changing faster than any time in human history due to human activities, primarily emissions of greenhouse gases (IPCC, 2021a). These changes are already impacting food production, water quality, and infrastructure (IPCC, 2022) as well as increasing the frequency and/or intensity of many types of extreme events including heatwaves and heavy downpours (IPCC, 2021a). Quantifying the risks these rapid changes pose to human society and the natural environment can provide critical input to resilience and adaptation planning while simultaneously highlighting the need for mitigation. As the impacts of climate change become increasingly evident around the world, the urgency of such assessments – and the need for approaches to generating projections that can be applied globally, especially in the most vulnerable regions which often lack abundant observational data or modeling capacity – is increasing rapidly.

Future projections typically begin with a range of plausible scenarios that describe the emissions resulting from a consistent set of human choices regarding climate policy, energy, land use, population and more (Hayhoe et al. 2017; Chen et al. 2021). These projections are then used as input to global climate models (GCMs) that divide the atmosphere, ocean, and land surface up into millions of discrete cells to solve numerical equations representing the physical, biological, and chemical phenomena in each, as well as inter-cell transfer of water, gases, energy, and more. As output, GCMs generate gridded projections of key variables including temperature, precipitation, wind direction and speed, humidity, and other variables that characterize the evolution of long-term climatic conditions as well as shorter-term variability.

The horizontal spatial resolution of GCMs has increased significantly over the past few decades, with grid cells for CMIP6 models typically ranging from approximately 50 to 260 km per side (IPCC, 2021b). However, most climate resilience and preparedness efforts require climate inputs at spatial and sometimes temporal scales much finer than the resolution of even the latest GCMs (Kotamarthi et al. 2021). In addition, both regional and global model output is often biased relative to observations due to both structural and parametric uncertainty in the models as well as the unavoidable fact that the average of a grid cell, particularly for temperature or precipitation extremes over land, is rarely representative of the value at an individual location or smaller area within the grid cell. Use of regionally homogenous and/or biased output to calculate the impacts of, for example, extreme heat on the human body, crop yields, or flood risk due to heavy precipitation, will yield errors that could in turn result in adaptation measures that are insufficient (if the projections under-estimate future change, creating a Type 2 or false negative error) or overly expensive and stringent (if projections over-estimate future change, a Type 1 or false positive error).

These challenges are not unique to climate modeling; they are also relevant to the field of numerical weather modeling, where they were initially addressed by combining model output with observations to create a series of multivariate regression corrections such as Model Output Statistics (Glahn & Lowry, 1972). In a similar way, biases can be removed from climate model output by combining statistical methods with observed data. Specifically, empirical statistical bias correction and downscaling of climate model projections introduces new information from observations and combines this information with model output to generate higher-resolution projections based on coarser-resolution fields consisting of local weather and climate characteristics like temperature, humidity, and precipitation.

The application of statistical methods to bias correcting GCM output was first proposed by Karl et al. (1990) as "a method, called climatological projection by model statistics, to relate GCM grid-point free-atmosphere statistics, the predictors, to these important local surface observations." Since then, hundreds of ESDMs have been developed and published, each describing a separate method or additional development of statistical bias correction and downscaling method. Previously independent methods have been combined and additional advanced statistical and computational techniques applied, including neural networks, clustering methods such as expectation-maximization algorithms, combined statistical-dynamical approaches, machine learning techniques, and deep learning models (e.g., Coulibaly et al. 2005; Vrac et al. 2007a; Walton et al. 2015; Sachindra et al. 2018; Hernanz et al. 2022; Wang & Tian, 2022). This abundance of methods, a number of which we were involved in developing or evaluating (e.g. Vrac et al., 2007b; Stoner et al., 2012; Barsugli et al. 2013; Dixon et al. 2016), begs the obvious question: Why is another needed? Why STAR-ESDM?

Despite the plethora of statistical methods and the more than 30-year period over which they have been developed, stakeholders still do not have a tool that fully encompasses all that the scientific modeling community can provide. With the notable exception of the Statistical DownScaling Model (SDSM, Wilby et al. 2002), most ESDMs are developed for a specific geographic region. This has led to a plethora of different ESDMs being applied to generate projections for watersheds, regions, and countries around the world; differences between those projections, especially at the tails of the distribution, are difficult to resolve without digging into the nuances of each model. In addition, the predictors and predictands for most ESDMs are hard-wired into the model. It can be difficult and time-consuming to update them with new CMIP simulations or different predictands, and typically that new data must be in the same format as the original predictor and predictand for which they were designed: gridded or station-based. Other ESDMs may demonstrate high reliability in simulating important features such as wet and dry spells, but were designed more as a proof of concept than a tool that can be used by stakeholders on a regular basis (e.g. Vrac et al. 2007a).

As discussed in Kotamarthi et al. (2021), there is a significant and growing demand for projections that are: (1) robust, with clearly quantifiable accuracy; (2) generalizable and flexible, applicable to any region of the world and any observational or modelled dataset; (3) capable of rapidly bias-correcting and downscaling large suites of global or regional climate model output for multiple scenarios; and (4) can be used with confidence to calculate increasing risks for applications where absolute values are required, such as extreme heat, changing energy demand, shifting crop yields, changes in water demand and supply, and more. Many existing methods meet one or more of these criteria, but few if any meet them all.

Our objective in developing STAR-ESDM is to create a highly functional and generalizable ESDM that addresses these four stakeholder concerns, and that can be applied broadly around the world using any predictand dataset in which the user has confidence. Building on previous research (e.g. Hayhoe et al., 2004, 2008; Stoner et al., 2012) and evaluation techniques (Dixon et al. 2016) we have developed a demonstrably flexible, computationally efficient, and robust statistical model that is capable of downscaling any atmospheric variable that is measured on a daily basis as long as it has, or can be transformed into, an approximately Gaussian distribution. This approach can be applied globally and to a broad range of climate and weather data sources, from global and regional model output to satellite data, gridded observations, and weather station records. Section 2, Model Development, describes the statistical basis of the model and refinements that improve its ability to downscale global model outputs. Section 3, Model Performance and Evaluation, describes how the model's ability to simulate temperature and precipitation extremes across the globe was tested using the perfect model framework. Finally, Section 4 summarizes the findings of this analysis, its application to climate impact assessments, and future model development objectives.

2 STAR-ESDM Design and Development

STAR-ESDM is a MATLAB-based code that combines signal decomposition with Fourier filtering and kernel density estimation to create an effective and computationally-efficient bias-removal and spatial disaggregation technique that can be used to analyze and translate a coarser-resolution time series of any quasi-Gaussian variable into a finer-resolution time series of the same variable. It is designed for application to climatological data: specifically, (1) to analyze biases in GCM and RCM simulations compared to observational or quasi-observational datasets over timescales relevant to the statistics of climate and weather, and (2) to use this information to bias-correct and spatially disaggregate predictor projections to the scale of the input predictands. Predictands can consist of gridded or station-based observations, quasi-observations such as satellite datasets or reanalysis, or higher resolution model output at any spatial scale finer than that of the predictor.

The first part of the acronym, STAR (seasonal trends and analysis of residuals) refers to the decomposition of a time series into components based on temporal variability: (1) long-term decadal average and/or trend, (2) climatological annual cycle (average over the historical period), (3) dynamical annual cycle (changing over time), and (4) high frequency daily anomalies. The second part of the acronym, ESDM (empirical-statistical downscaling model) refers to the steps taken once a modeled predictor and an observed or quasi-observed predictand has been decomposed into these components. The predictor signal is bias-corrected relative to the observed or predictand signal for a historical period and the projected changes then applied to observed values, resulting in bias-corrected and downscaled high-resolution projections at the spatial scale of the observations.

In empirical statistical bias-correction and downscaling, stationarity is a primary concern (e.g. Dixon et al. 2016). Will statistical relationships developed based on historical data hold true under potentially very different climatic conditions in the future? In cases where human intervention directly or indirectly, through climate change, significantly alters the characteristics of the land surface (e.g. by expanding an urban area, through large-scale deforestation, or when the timing of snowpack melt shifts), the answer is clearly no. The only way to account for these in climate modeling, whether statistical or dynamical, is to either explicitly include the change or build in the capacity to predict the change. In other cases, however, it may be possible to increase the stationarity of the model by increasing its ability to resolve different components of a signal that may be changing differently over time: and this is the hypothesis on which STAR-ESDM is based.

Why might signal decomposition improve the stationarity of a statistical model? Biases and errors in GCM simulations, which arise due to both structural and parametric uncertainties in the model, are typically physical in origin and relate to a process or a component not accurately represented in the model for that region or aspect of the climate system. When the GCM signal is considered in sum, as an 'analogue' signal, it increases the likelihood that some of the biases interact and even cancel each other out. A statistical method that cannot distinguish between the various sources of bias must assume that not only the biases, but how they interact with each other, remain stationary over time. Decomposing the signal by the timescale over which various types of biases may occur allows them to be better quantified and a statistical correction developed that is appropriate to the timescale at which they are relevant: effectively 'digitizing' the signal, to a certain extent. Through decomposing the signal, each component of which may be biased due to different sources of structural or parametric uncertainty within the GCM, these can be corrected independently before the signal is recombined. Previous evaluations of a beta version of STAR-ESDM v1 using the perfect model framework (Dixon et al. 2016) shows that the manner in which it allows distributions of daily anomalies and annual climatology to change over time relaxes the stationarity assumption that underlies most ESDMs, significantly reducing the bias, particularly at the tails of the distribution where extreme events may be relatively rare but have a proportionally greater impact. Based on that preliminary analysis, we have now developed a fully operational STAR-ESDM v2 can be used to bias-correct and downscale a broad range of coarser-resolution predictor datasets to finer-resolution predictands, and evaluate its stationarity here.

Applying the STAR-ESDM framework to bias-correct and spatially disaggregate coarser predictor simulations (which could be derived from any dataset a user wishes to bias-correct and downscale, but are most likely to consist of GCM or RCM output) relative to a higher-resolution predictand dataset consists of three main steps (Figure 1). First, each time series is disaggregated into individual components. Second, predictand components are used to bias-correct predictor components, and predictor components are used to map predictand statistics to future conditions. Lastly, bias-corrected and downscaled components are recombined to create a single continuous time series at the spatial scale of the predictand but covering the temporal range of the predictor.

The first component is the long-term trend, determined by fitting an optimized linear or third-order trend to the entire time series. This represents the climatological change resulting from human choices. Based on analyses comparing model-simulated with observed multi-decadal trends at the regional to global scale (IPCC, 2021a), the framework currently assumes that the absolute value of the predictand and the long-term normalized trend of the predictor are accurate.

The second component is the mean or climatological annual cycle over the historical period. It is extracted by averaging each day of the year over the historical period, then smoothing the resulting curve using a low-pass Fourier filter to remove noise. This was used in place of a conventional smoothing filter such as a rolling mean to prevent dampening the extremes. The historical period is flexible, automatically determined based on the beginning and end of the predictand data being used. Thus, if a longer dataset or a more recently updated one is being used, predictand values beginning as far back as 1900 or ending as recently as the latest full year of predictand data will be considered part of the historical "training" period; whereas if a relatively short predictand dataset such as 1971-2000 is being used, then only this period will be used to determine the static annual cycle as well as the components that follow. The difference between the predictand and predictor annual cycle is used to bias-correct the predictor's dynamic annual cycle (how it varies from year to year).

The third component is the annually-varying or dynamical annual cycle. This component is smoothed both along the year axis and the day-of-year axis to create a climatological surface that changes over time, but with the variability and the long-term trend over both day to day and year to year timescales removed. This is then used to adjust the predictand annual cycle to account for future changes. It captures the change in the shape of the climatology over time, accounting for how summer may be broadening, for example, and shoulder seasons shrinking, or how the timing of the monsoon season may be shifting, independent of long-term trends in average annual values or short-term daily variability. Changes in the predictor-simulated annual cycle over time are then used to extend the predictand climatology into the future.

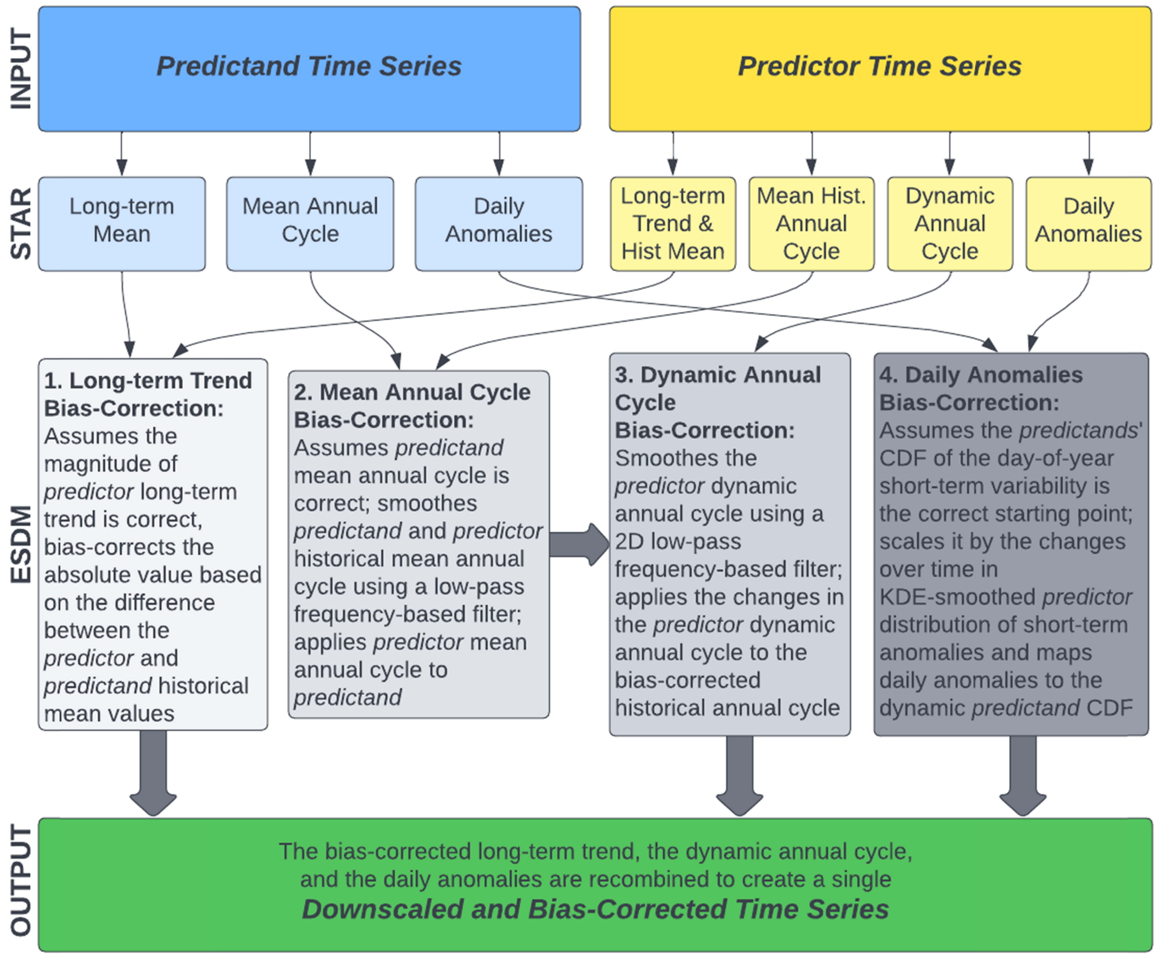

Figure 1. The STAR-ESDM framework decomposes the predictor (model) and predictand (observed or quasi-observed) time series into components which are then used to individually bias-correct predictor and/or adjust predictand components as shown here. Components are then recombined to generate a continuous high resolution time series of climate projections at the spatial scale of, and matching the statistical properties of, the predictand dataset but covering the time period of the predictor dataset.

To correct biases in and downscale projected changes in the annual cycle, we make the following assumptions which allow for relaxation of certain stationarity assumptions. First, we assume that while the shape of the predictor's annual cycle most likely contains biases, predicted changes in the shape of the annual cycle are a reasonable approximation to how the predictand's climatology would change over time, given the assumptions underlying the predictor and its inputs. Similarly, we assume that the changes over time to the shape of the predictor's daily probability distributions, described as cumulative distribution functions (CDFs), are a reasonable estimate of the likely changes in the predictand's CDFs over time as well. The former allows the bias-corrected annual signal to vary over time, and the latter allows mapping of the predictor's daily anomalies from the predictor's time-varying CDFs to an estimate of the predictand's time-varying CDFs, rather than to a static historical CDF.

Another important feature of STAR-ESDM is that creates probability surfaces of the annual cycles rather than discrete monthly CDFs. These surfaces are generated for both the historical period and future conditions using a rolling window that steps forward through time. Two-dimensional frequency-based (Fourier domain) filtering of the probability surfaces ensures they vary smoothly, with unique values for each day of the year. Similarly, one-dimensional Fourier filtering is used to smooth the static or climatological annual cycles for the predictand's and the predictor's historical period data, and two-dimensional filtering is used to calculate dynamic annual cycle surfaces which evolve over time by combining the observations' historical and model future data. Thus, the dynamic annual cycle surface represents the expected value of the data variable over time, while the probability surfaces represents the variability of the data variable, both in the historical period and over time. The predictor's annual cycle is adjusted to match the mean value and shape of the predictand's in the historical period, and allowed to vary over time as estimated by the predictor through the full time period from past into the future.

The frequency range of the filters is selected to allow separating the signal into the seasonally varying component (climatology) and the short-term variability (weather). Seasonally-varying filters capture 90 to 95 percent of typical seasonal-shape changes, including events that repeat annually but excluding short-term random events. Once these three components have been removed from the signal, what remains primarily reflects high-frequency daily variability in the time series which may not repeat from year to year. This last component characterizes the magnitude and frequency of extremes, which are the most challenging aspect of future projections for ESDMs to accurately resolve, and simultaneously the most relevant to many stakeholder applications in building climate resilience to threats including heatwaves, droughts, floods and more.

Daily anomaly values are bias-corrected by calculating a quantile value for each data point and mapping between the predictor's set of rolling CDFs and the predictand's set of rolling estimated CDFs. Filtering of daily variability is also done in the Fourier domain as it is much faster computationally than using time-domain convolution, and computational efficiency is another key consideration in development of this model. Finally, one last round of Fourier filtering combined with Kernal Density Estimation (KDE) smoothing is used to transform the two-dimensional histograms of the anomalies into PDF surfaces, which are then integrated along the probability axis to create CDF surfaces; KDE being an approach determined by McGinnis et al. (2015) to perform well in bias correcting climate model output.

These steps are generalizable to any variable with a quasi-Gaussian distribution or that can be transformed into a quasi-Guassian form. As discussed in the next section, however, through iterative application of the Perfect Model framework, the method has been optimized to daily maximum and minimum temperature and precipitation (and in the future, will be similarly optimized to apply the model to humidity, solar radiation, and more).

3 Model Performance and Evaluation

We evaluate the stationarity of the STAR-ESDM framework using a "Perfect Model" framework. As described in Dixon et al. (2016), the name of this approach is not intended to contradict the aphorism often attributed to statistician George Box, "All models are wrong but some are useful." Rather, "Perfect Model" describes a methodology that uses a high-resolution GCM simulation (in this case, a 25km resolution GFDL simulation for a historical period and for an end-of-century 2086-2095 RCP8.5 scenario) and a coarsened version of the same simulation as the predictand and predictor variables to train an ESDM in a "pseudo-reality context," as Erlandsen et al. (2020) describe it. The resulting ESDM is then applied to the coarsened GCM simulations in the future and the output compared to the fully dynamical simulations for the same time period. Differences between the statistically bias-corrected and downscaled ESDM and dynamical GCM output for the same future time period reveal structural uncertainties in the ESDM that prevent it from generating the information that a much more complex (but less flexible and more computationally demanding) dynamical high-resolution global climate model would. Locations where biases are minimal indicate that the ESDM can generate virtually identical values to those of a fully dynamical model (with the benefits of greater flexibility and significantly reduced computational cost, as well as bias-correction).

To illustrate the improvement offered by the signal processing approach, we compare the biases in STAR-ESDM output with the biases in projections generated using the Asynchronous Regional Regression Model, a parametric quantile mapping approach to bias correction and downscaling that has been used in a number of regional climate assessments across the United States (Stoner et al. 2012). As shown in Figure 2, while ARRM performed well over the contiguous United States at simulating daily maximum temperature for quantiles ranging between 0.1 and 99.9, it displayed biases in maximum temperature values ranging from 3 to 5°C for both extreme hot and cold temperatures along most major coastlines and biases in mean values averaging around 3°C across much of Central and South America and central Africa that would preclude its use in those regions. In contrast, STAR-ESDM displays almost no biases in mean values across any continental area other than a bias of between 1-2°C at the highest elevations of the Rocky and Himalayan mountain ranges. Even out to the tails of the distribution, at the 0.1 and 99.9th quantiles, biases are significantly reduced in both extent and magnitude relative to ARRM. For the 1-in-1000 coldest temperatures, STAR-ESDM warm biases occur primarily along Arctic coastlines, while for the 1-in-1000 highest temperatures, significant warm and cold biases are present only in the northern Andes and the highest elevation region of the Himalayas. Table S1 summarizes the locations where STAR-ESDM use is not recommended for temperature projections, and Figure S2 shows the same results as in Figure 1, except for daily minimum temperature.

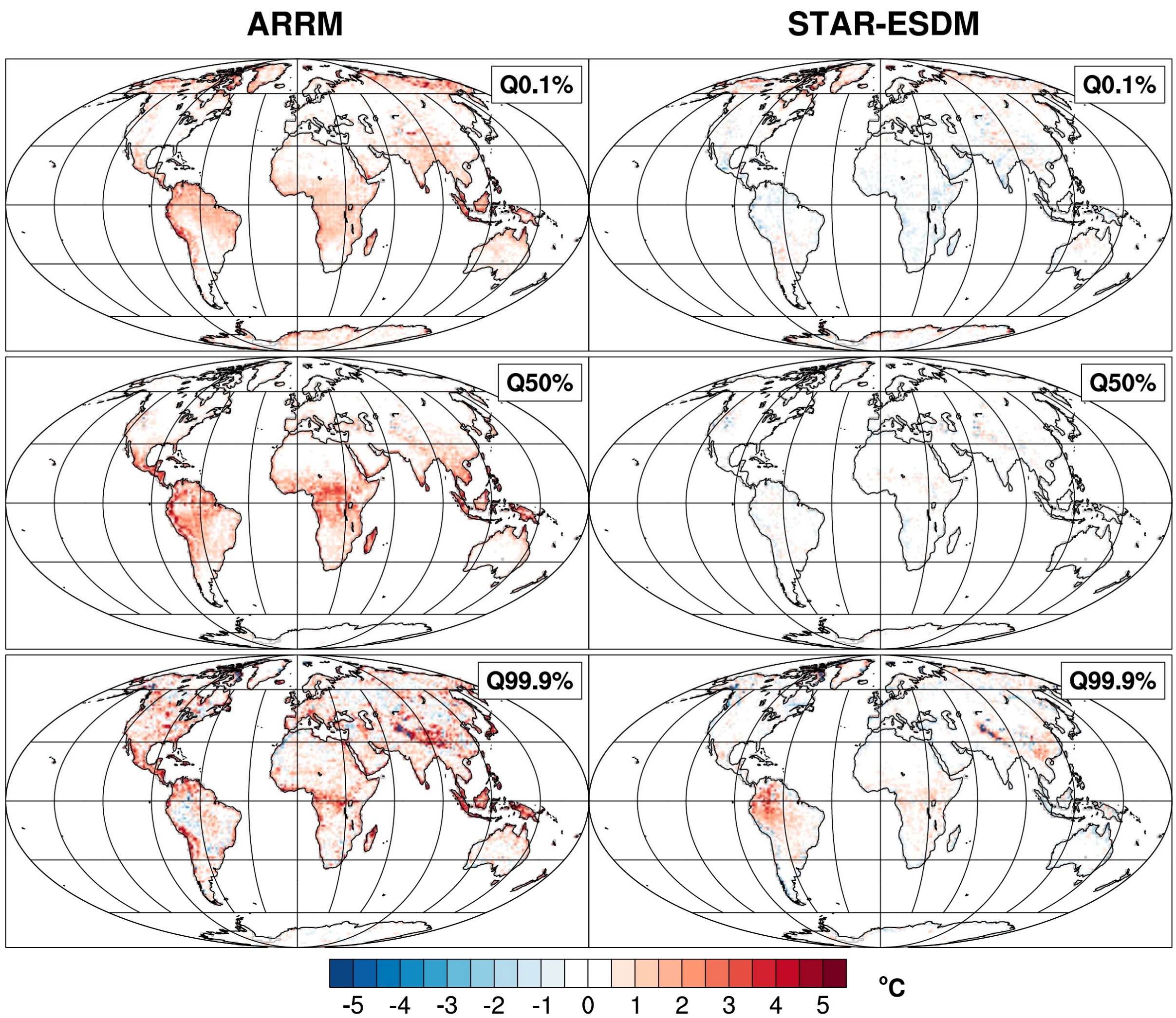

Figure 2. Biases over land areas in ARRM (left) and STAR-ESDM (right) daily maximum temperature (degrees C) compared to high resolution global climate model output for 2086-2095 under the higher RCP8.5 scenario for the 1-in-1000 coldest, mean, and 1-in-1000 hottest days illustrate the significant improvements the signal processing approach uses over the parametric quantile mapping approach.

Improvements in precipitation are even more pronounced. (Here, precipitation quantiles are calculated based on wet days only. As such, they represent a more extreme assessment than temperature quantiles, which are calculated based on all days.) As shown in Figure 3, ARRM is capable of simulating projected changes in precipitation over the contiguous United States, northern Africa, the Middle East, and Europe up to the 90th quantile of wet day precipitation with minimal biases. Biases of 20-30 mm/day occurred over South America, central Africa, and southeast Asia. At the 99th quantile, however, positive biases ranging from 20 to 40 mm per day across the contiguous United States and up to 100mm per day across other continental areas occurred. By the 99.9th quantile, positive biases ranging from 50 up to greater than 100mm per day occurred across most continental areas. STAR-ESDM biases are less than 10 mm per day at the 90th quantile across all land areas. At the 99th quantile, small negative biases on the order of

10 to 20 mm per day occur across equatorial regions (within about 10 degrees of the equator) while biases across the rest of the world remain below 10 mm per day. At the 99.9th quantile, while negative biases approaching 50 mm emerge in equatorial regions, biases across most of North America, Europe, and central Asia remain below 10 mm. This indicates that even for the wettest few days in a decade, the statistical model is able to reproduce the values that a high-resolution dynamical model would provide at that quantile. Once again, Table S1 summarizes the locations, quantiles and seasons where STAR-ESDM use is not recommended for precipitation.

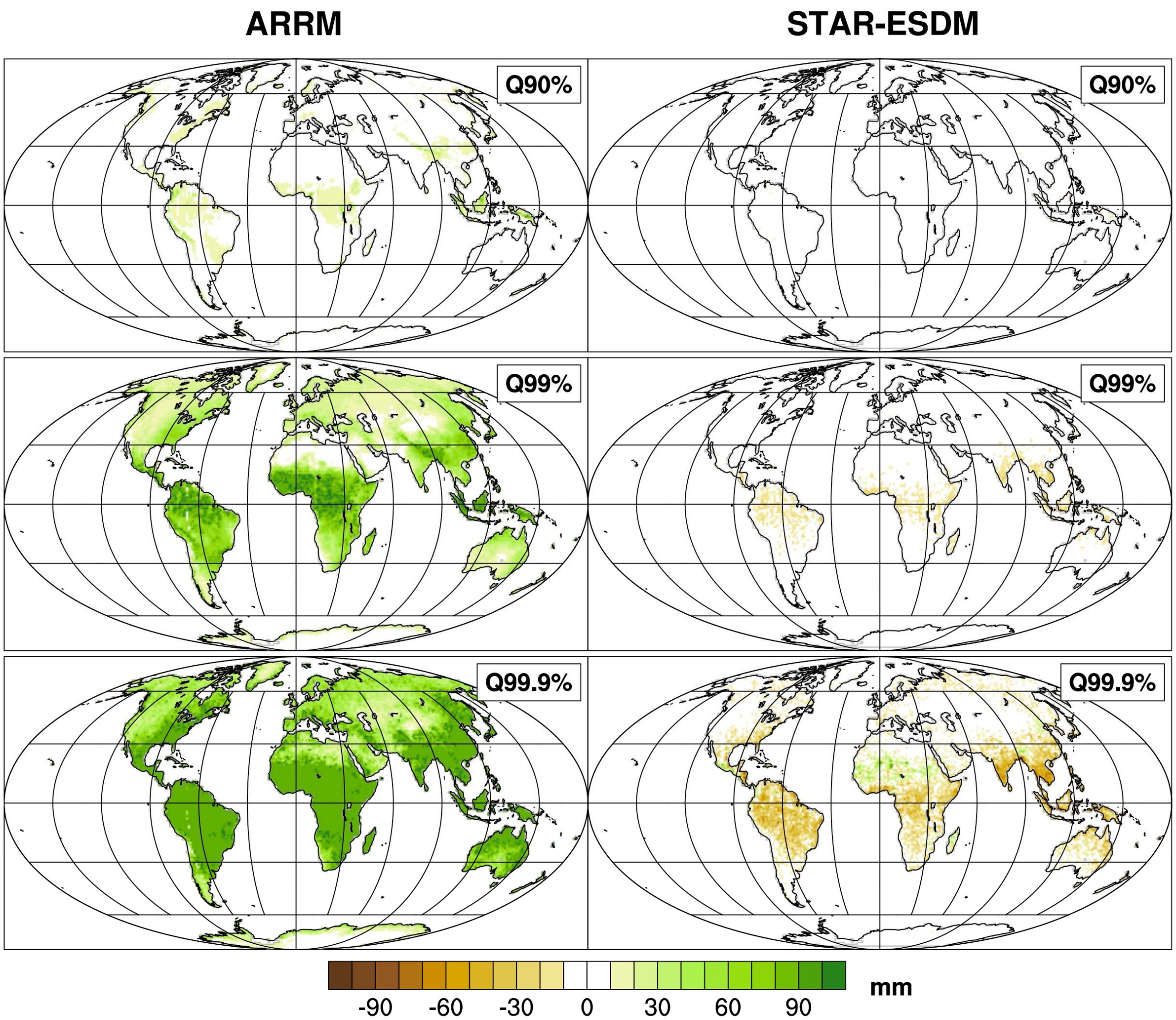

Figure 3. Biases over land areas in ARRM (left) and STAR-ESDM (right) daily cumulative precipitation values (mm per day) compared to 25km global climate model output for 2086-2095 under the higher RCP8.5 scenario for the 1-in-10, 1-in-100, and 1-in-1000 wettest days of all wet days at that location illustrate the stationarity of the STAR approach and the significant improvements the signal processing approach uses over the parametric quantile mapping approach.

As discussed in the section above, the signal processing approach used by STAR-ESDM already offers significant improvements relative to previous statistical methods. However, a second unique aspect of this work is that STAR-ESDM was not developed linearly. Rather, we used the Perfect Model framework to iteratively and interactively optimize the original code for daily maximum and minimum temperature and 24-hour cumulative precipitation. This enabled us to identify areas with high biases, and explore statistical approaches to reducing those biases that were: (a) consistent with the likely physical basis of the bias, and (b) globally generalizable, to avoid over-fitting.

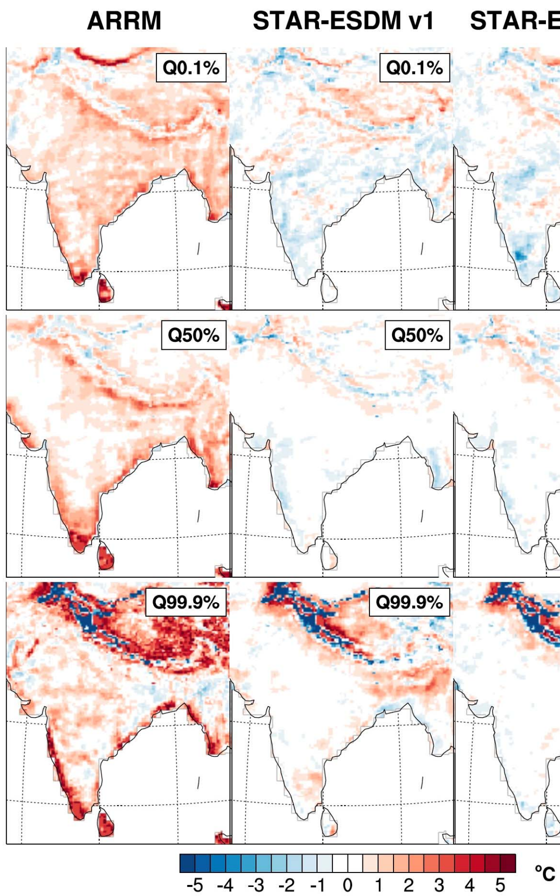

Figure 4. Biases in ARRM (left), STAR-ESDM v1 (center) and STAR-ESDM v2 (right) daily maximum temperature (degrees C) compared to high resolution global climate model output for 2086-2095 under the higher RCP8.5 scenario for the 0.1, 50th and 99.9th quantiles of the distribution over the Indian subcontinent show how both the statistical method and the further refinement using the Perfect Model approach reduce biases compared with previous methods.

When comparing STAR-ESDM v1 (the original generic code) with v2 (the code that has been iteratively optimized) at the global scale, there are few obvious improvements in maximum and minimum temperature biases, other than a reduction in the magnitude of high quantile biases in equatorial regions (Figures S1 and S2). At a finer scale, however, the added value of these refinements can be more evident. In high-elevation locations, for example, shifts in the timing of snow melt lead to large positive and negative temperature biases in spring and summer. Figure 4 compares biases in maximum temperature projections over the Indian subcontinent corresponding to ARRM, STAR-ESDM v1 and STAR-ESDM v2. While the greatest improvements are still obtained by changing the statistical approach, at the 99.9th quantile it can be seen that an outlier adjustment that was part of the original v1 design (scaling daily extremes beyond 2.5 sigma) produced slightly greater biases then simply smoothing the entire time series using the KDE. Removing this scaling further reduced the bias at mid elevations and in the foothills and expanded the geographic area over which this method can effectively be used in v2.

Precipitation is a more challenging variable to characterize than temperature and as a result, for this variable the Perfect Model optimization yielded greater benefits. As shown in Figure S4, STAR-ESDM v1 output for the 99th and 99.9th quantile was characterized by large bands of positive and negative precipitation biases across equatorial regions, while continental biases were less but still notable, ranging from -20 mm per day across the Gulf Coast up to 50 mm per day at high elevations in the Andes and Himalayas. We first explored the use of different methods to transform precipitation into an approximately Gaussian distribution, and identified power mapping as the most consistent. However, originally v1 used the same empirically-determined power mapping for all locations. Adding an interative search to identify the optimal equation for each individual location reduced precipitation biases in equatorial regions in terms of both geographic extent and magnitude, while biases across continental areas in the northern hemisphere were reduced to less than +/- 5 mm per day. While a dry bias persists across South America, southern Africa and southeast Asia in v2, the magnitude of the bias ranges from -5 to -15 mm per day, nearly an order of magnitude reduction compared to STAR-ESDM v1.

As alluded to earlier, quantifying biases using the Perfect Model approach also enables us to identify geographic locations where use of this framework is not recommended as the ESDM displays significant non-stationarity; these locations and quantiles are listed in Table S1. For example, shifts in the timing of monsoonal precipitation in a warmer world may be what leads to precipitation biases across Central America, Mexico and the southwestern United States that dominate during spring and summer seasons. Similarly, large biases in temperature values still occur under a set of specific conditions, including biases (a) in high temperatures for very high elevations in the Andes and Himalayas under a higher scenario by end of century, during the season that saw the highest level of snow melt during the historical period, and (b) in both high and low temperatures along Arctic coastlines where rapid melting of shoulder-season sea ice introduces significant non-stationarity relative to historical conditions. In these cases, the biases are likely due to the fact that the statistical relationships between predictor and predictand for future months are now representative of those of a different month in the historical time period. In future versions of STAR-ESDM, this can be addressed by introducing a phase shift into the bias corrections applied to the dynamical climatology and the high frequency variability. Effective reduction of other biases might require more granular information that is absent in the coarser-scale model, such as distinguishing between land and air temperatures for small islands below the spatial scale of the predictor grid cell, or for valleys and mountains in areas with rapidly-varying topography. In the future, we plan to explore whether these could be improved through incorporating high-resolution digital elevation maps, lapse rates, and more into the statistical model.

4 Conclusions and Next Steps

As anthropogenic climate change increasingly impacts food production, water quality, infrastructure integrity, and the frequency and intensity of extreme events, there is a critical need for accurate and high-resolution climate projections to inform sectoral and regional climate resilience planning. The STAR-ESDM model, developed within the Perfect Model framework, represents a significant advancement in the ability to generate robust high-resolution climate projections for regional to local-scale climate impact assessments. As described above, the STAR-ESDM model decomposes a predictand and predictor signal into multiple components which are then bias-corrected and adjusted individually before being recombined into a single coherent time series covering the time period of the predictor. The result is a stationary and computationally-efficient ESDM which was further iteratively developed within the Perfect Model framework to quantify model biases by variable, region and season. Additionally, it is extremely flexible, allowing for a range of predictor and predictand inputs, depending on what is available for that region.

Evaluating STAR-ESDM's ability to bias correct and downscale climate projections for end-of-century under a higher (RCP8.5) scenario by comparing its output to that of a high-resolution dynamical model demonstrates that, for quantiles ranging from 0.1 to 99.9 over most land areas, it is able to produce temperature and precipitation projections that are virtually identical to those that would be obtained from a high-resolution fully dynamical model. With a few exceptions, such as the Arctic coastline and areas with rapidly varying topography at high elevation such as the Himalayas, STAR-ESDM can be confidently applied to nearly any location in the world for which gridded or point-based predictand data is available. The granularity of this guidance, as summarized in Table S1, offers stakeholders and users a clear pathway to assessing the reliability of this information for informing their future assessments, depending on which quantiles of the distribution, geographic location, and season(s) are most relevant to the impacts with which they are concerned.

We have already applied the STAR-ESDM framework to generating high-resolution projections of daily maximum and minimum temperature and precipitation using SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5 simulations by 24 CMIP6 GCMs as predictors for the following geographic regions and predictand datasets: (1) the contiguous U.S. using the NClimGrid 5x5km observational dataset (Vose et al. 2014), (2) the contiguous U.S. using to the Livneh 1/16th degree observational dataset (Livneh et al. 2015), and (3) North and Central America and the Caribbean, using over 10,000 GHCNd long-term weather station records (Menne et al. 2012). We are currently extending this work to the global scale, using Sheffield et al. (2006) global 0.25 x 0.25 degree gridded dataset and the ERA5-Land 0.1 x 0.1 degree global dataset (Muñoz-Sabater et al., 2021) as predictands to generate global high-resolution bias-corrected projections. These datasets will be archived shortly; we anticipate we will be able to remove this sentence and include a DOI for each of these datasets by the time this manuscript is published.

In future model development, we propose to examine projected phase shifts in annual cycles to determine whether it is possible to reduce biases arising from to changes in the timing of monsoonal precipitation or snowpack melt at high elevations. We also plan to refine the weighting scheme used to select predictor grid cells, to better characterize conditions in areas of rapidly varying topography such as islands, coastlines, and mountainous areas, and explore the outcomes of including a high resolution digital elevation map. Finally, recognizing that a fundamental flaw in statistical methods is the separation and independent bias-correction of variables that are dynamically linked, long-term we propose to develop a version of STAR-

ESDM where the input is a hypersurface composed of average temperature, daily temperature range, humidity and precipitation, rather than a single variable.

As the need for accurate and detailed climate impact assessments increases, high-resolution climate projections will play a crucial role in helping societies worldwide adapt to and mitigate the impacts of climate change. STAR-ESDM provides a valuable tool for resilience and adaptation planning, especially in the most vulnerable regions which often lack abundant observational data or modeling capacity.

Acknowledgments

We thank Keith Dixon and John Lanzante for many fruitful discussions that greatly informed the direction of this work. In addition to their positions listed here, K. Hayhoe is the owner of ATMOS Research, a non-publicly traded entity and both K. Hayhoe and A. Stoner have served as consultants in developing and contributing to various regional climate assessments and climate action plans. K. Hayhoe, I. Scott-Fleming and A. Stoner were supported by Texas Tech University and A. Stoner and D. Wuebbles by Earth Knowledge in the development of this work.

Open Research

Version 2.0 of the STAR-ESDM MATLAB code, the development and evaluation of which is described in this manuscript, will be made freely available for download via a Github public repository prior to publication. A permanent link to the code repository and a DOI will be provided with article revisions.

References

Barsugli, J., G. Guentchev, R. M. Horton, A. Wood, L. O. Mearns, X.-L. Liang, J. Winkler, K. Dixon, K. Hayhoe, R. B. Rood, L. Goddard, A. Ray, L. Buja & C. Ammann (2013), The Practitioner's Dilemma: How to Assess the Credibility of Downscaled Climate Projections. *EOS, 94*(46), 424-425, doi:10.1002/2013EO460005

Chen, D., M. Rojas, B.H. Samset, K. Cobb, A. Diongue Niang, P. Edwards, S. Emori, S.H. Faria, E. Hawkins, P. Hope, P. Huybrechts, M. Meinshausen, S.K. Mustafa, G.-K. Plattner, & A.-M. Tréguier (2021) Framing, Context, and Methods. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 147–286, doi:10.1017/9781009157896.003.

Coulibaly, P., Y. Dibike & F. Anctil (2005), Downscaling Precipitation and Temperature with Temporal Neural Networks. *Journal of Hydrometeorology*, *6*(4), 483-496, doi:10.1175/JHM409.1

Dixon, K., J. Lanzante, M.J. Nath, K. Hayhoe, A. Stoner, A. Radhakrishnan, V. Balaji & C. Gaitán (2016), Evaluating the stationarity assumption in statistically downscaled climate projections: is past performance an indicator of future results? *Climatic Change, 135*(3-4), 395-408, doi:10.1007/s10584-016-1598-0

Erlandsen, H., K. Parding. R. Benestad, A. Mezghani & M. Pontoppidan (2020), A hybrid downscaling approach for future temperature and precipitation change. *Journal of Applied Meteorology and Climatology, 59*(11), 1793-1807, doi: 10.1175/JAMC-D-20-0013.1

Glahn, H. & D. Lowry (1972), The use of Model Output Statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology and Climatology*, *11*(8), 1203-1211, doi:10.1175/1520-0450(1972)011%3C1203:TUOMOS%3E2.0.CO;2

Hayhoe, K., D. Cayan, C. B. Field, P. C. Frumhoff, E. P. Maurer, N. L. Miller, S. C. Moser, S. H. Schneider, K. Nicholas, E. E. Cleland, L. Dale, R. Drapek, R. M. Hanemann, L. S. Kalkstein, J. Lenihan, C. K. Lunch, R. P. Neilson, S. C. Sheridan, & J. H. Verville (2004), Emissions pathways, climate change, and impacts on California. *Proceedings of the National Academy of Sciences, 101*(34), 12422-12427, doi:10.1073/pnas.0404500101

Hayhoe, K., C. Wake, B. Anderson, X-Z Liang, E. Maurer, J. Zhu, J. Bradbury, A. DeGaetano, A. Stoner & D. Wuebbles (2007), Regional climate change projections for the Northeast USA. *Mitigation and Adaptation Strategies for Global Change*, *13*, 425-436, doi:10.1007/s11027-007-9133-2

Hayhoe, K., J. Edmonds, R.E. Kopp, A.N. LeGrande, B.M. Sanderson, M.F. Wehner, & D.J. Wuebbles (2017), Climate models, scenarios, and projections. In: *Climate Science Special Report: Fourth National Climate Assessment, Volume I* [Wuebbles, D.J., D.W. Fahey, K.A. Hibbard, D.J. Dokken, B.C. Stewart, and T.K. Maycock (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, pp. 133-160, doi:10.7930/J0WH2N54.

Hayhoe, K., I. Scott-Fleming, A. Stoner and D. Wuebbles, (2023), STAR-ESDM Version 2.0 [Software], Github repository link and DOI TBD.

Hernanz, A. J. A. Garcia-Valero, M. Dominguez & E. Rodriguez-Camino (2022), A critical view on the suitability of machine learning techniques to downscale climate change projections: Illustration for temperature with a toy experiment. *Atmospheric Science Letters, 23*(6), e1087, doi:10.1002/asl.1087

IPCC (2021a), Summary for Policymakers. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

IPCC (2021b), Annex II: Models [Gutiérrez, J M., A.-M. Tréguier (eds.)]. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 2087–2138, doi:10.1017/9781009157896.016.

IPCC (2022), Summary for Policymakers [H.-O. Pörtner, D.C. Roberts, E.S. Poloczanska, K. Mintenbeck, M. Tignor, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem (eds.)]. In: *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3–33, doi:10.1017/9781009325844.001.

Karl, T., W-C Wang, M. Schlesinger, R. Knight & D. Portman (1990), A Method of Relating General Circulation Model Simulated Climate to the Observed Local Climate. Part I: Seasonal Statistics. *Journal of Climate, 3*(10), 1053-1079, doi:10.1175/1520-0442(1990)003%3C1053:AMORGC%3E2.0.CO;2

Kotamarthi, R., K. Hayhoe, L. Mearns, D. Wuebbles, J. Jacobs & J. Jurado (2021), *Downscaling Techniques for High-Resolution Climate Projections: From Global Change to Local Impacts*. Cambridge University Press, doi:10.1017/9781108601269

Livneh, B., Bohn, T., Pierce, D. et al. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Sci Data, 2*, 150042, doi:10.1038/sdata.2015.42

McGinnis, S., D. Nychka & L. O. Mearns (2015), A New Distribution Mapping Technique for Climate Model Bias Correction. In: Lakshmanan, V., Gilleland, E., McGovern, A., Tingley, M. (eds) *Machine Learning and Data Mining Approaches to Climate Science*. Springer, Cham. doi:10.1007/978-3-319-17220-0_9

Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, & T.G. Houston (2012), An overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Ocean Technology, 29,* 897-910, doi:10.1175/JTECH-D-11-00103.1.

Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., & Thépaut, J.-N. (2022), ERA5-Land: A state-of-the-art global reanalysis dataset for land applications, *Earth Syst. Sci. Data,13*, 4349–4383, 2021, doi:10.5194/essd-13-4349-2021.

Sachindra, D., K. Ahmed, M. M. Rashid, S. Shahid & B. Perera (2018), Statistical downscaling of precipitation using machine learning techniques. *Atmospheric Research, 212*, 240-258, doi:10.1016/j.atmosres.2018.05.022

Sheffield, J., G. Goteti & E. Wood (2006), Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling. *Journal of Climate, 19*(13), 3088-3111, doi:10.1175/JCLI3790.1

Vose, Russell S., Applequist, S., Squires, M., Durre, I., Menne, M. J., Williams, C. N. Jr., Fenimore, C., Gleason, K, and Arndt, D. (2014), *NOAA Monthly U.S. Climate Gridded Dataset (NClimGrid), Version 1.* NOAA National Centers for Environmental Information, doi:10.7289/V5SX6B56 [Jan 2023].

Vrac, M., M. Stein & K. Hayhoe (2007a), Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing. *Climate Research, 34*(3), 169-184, doi:10.3354/cr00696

Vrac, M., M. L. Stein, K. Hayhoe & X.-Z. Liang (2007b), A general method for validating statistical downscaling methods under future climate change. *Geophysical Research Letters, 34*(18), doi:10.1029/2007GL030295

Walton, D., F. Sun, A. Hall & S. Capps (2015), A Hybrid Dynamical–Statistical Downscaling Technique. Part I: Development and Validation of the Technique. *Journal of Climate, 28*(12), 4597-4617, doi:10.1175/JCLI-D-14-00196.1

Wang, F. & D. Tian (2022), On deep learning-based bias correction and downscaling of multiple climate models simulations. *Climate Dynamics, 59*, 3451-3468, doi:10.1007/s00382-022-06277-2

Wilby, R., C. Dawson & E. Barrow (2002). SDSM – a decision support tool for the assessment of regional climate change impacts. *Environmental Modelling & Software, 17*(2), 145-157, doi: 10.1016/S1364-8152(01)00060-3