Earth Systems and Environment
https://doi.org/10.1007/s41748-024-00410-3

ORIGINAL ARTICLE

The Selection of Global Climate Models for Regional Impact Studies Should Consider Information from Historical Simulations and Future Projections

A. N. Rohith[1] · **Alfonso Mejia**[2] · **Raj Cibin**[1,2]

Received: 8 November 2023 / Revised: 8 April 2024 / Accepted: 23 May 2024

Abstract

The selection of Global Climate Models (GCMs) based on their ability to represent precipitation patterns of a region is required for hydrological climate change impact studies to address time and computational constraints. Generally, the selection of GCMs is determined based on their ability to reproduce observed climate statistics in historical simulations, assuming they will continue to perform well in the future. However, the performance of GCMs varies over time in ways that are not sensitive to their historical performance, indicating that GCMs' selection needs to consider historical simulation and future projection information. We propose a framework to account for future GCM projection convergence to and divergence from the ensemble mean, along with historical performance, to select the GCMs that are applicable to a particular regional climate impact study. The framework uses Reliability Ensemble Averaging (REA) with 30 Coupled Model Intercomparison Project-6 (CMIP6) GCMs to select GCMs based on the ensemble mean and variability of projections. We demonstrate the framework using three climate indices (annual maximum precipitation, annual total precipitation, and wet day precipitation intensity) in the Chesapeake Bay watershed of the United States. Our analysis shows that using only the GCM performance during the historical period could result in the selection of GCMs that are extreme outliers due to an inherent underprediction of precipitation extremes by all GCMs and requires an efficient bias correction before selection. There was also no significant correlation between the historical period performance of GCMs and future GCM convergence for more than 95% of the cases in the study region. This highlights the need to consider convergence and divergence information from climate projections when selecting GCMs for practical and computationally intensive applications. The proposed framework can be adapted to any study region and can help identify GCMs for computationally intensive climate change impact studies.

Highlights

- Both historical performance and future projections need accounting in selecting GCMs.
- The future convergence of GCMs is not sensitive to their historical performance.
- Efficient bias correction can benefit the identification of GCMs.
- A subset of GCMs were identified for impact studies in the Chesapeake Bay watershed.

Keywords GCM selection · Climate change · Bias correction · Extreme precipitation · Reliability ensemble averaging

1 Introduction

The impacts of global climate change on the Earth's environment have been among the most researched areas in the last few decades, owing to its threats to sustainable development. The attribution of several recent environmental disasters and extreme hydroclimatic occurrences to climate change (Smiley et al., 2022; Vargo et al., 2020) has drawn

✉ Raj Cibin
craj@psu.edu

A. N. Rohith
rfa5378@psu.edu

[1] Department of Agricultural and Biological Engineering, The Pennsylvania State University, University Park, PA, USA

[2] Department of Civil Engineering, The Pennsylvania State University, University Park, PA, USA

Published online: 10 June 2024

the attention of policymakers to investigate the impact of climate change on natural resources and human populations. It is estimated that 80% of the land and 85% of the total population may have been affected by climate change (Callaghan et al. 2021). Consequently, it is essential to scientifically project and study the future consequences of climate change for better preparedness.

The Intergovernmental Panel on Climate Change (IPCC) has been working on understanding the global-to-regional consequences of climate change since its establishment in the late 1980s. The latest reports (IPCC, 2022a, b, 2021) released by the IPCC as part of the sixth Assessment Report (AR6) are the most comprehensive and up-to-date knowledge of the Earth's climate change and serve as a basis for deriving adaptation strategies. Numerical climate simulations from global climate models (GCMs), released under the Coupled Model Intercomparison Projects (CMIP), are the primary data source for climate impact studies. The performance of these simulations varies with GCM and region (Baghel et al. 2022; Chhin and Yoden, 2018) due to various uncertainties, such as uncertainty of the model structure, uncertainty of unknown future scenarios, and random internal climate variability (Hawkins and Sutton, 2011; Rupp et al., 2013). Impact studies must be aware of the uncertainties or variability present in climate model simulations, and performance evaluation is required to gain confidence in climate-related decision-making (Rupp et al., 2013). Crucially, the high computational demands of hydrological/hydraulic models used in impact studies make it challenging to use all the available GCMs in decision-making and require the identification of a subset of GCMs. Performance evaluations of GCMs to simulate historically observed climate are generally used to rank the GCMs and find the most suitable ones for impact studies (Raju et al. 2017; Raju and Kumar 2020).

Recent studies have simplified the GCM identification and ranking problem by selecting models based on a single criterion: the models' ability to reproduce historically observed statistics. For example, Baghel et al. (2022) derived sector-specific ranks of 16 GCMs from CMIP6 using several climate indices, Perkins et al. (2007) evaluated the performance of 14 GCMs from CMIP3 using probability density functions, Khadka et al. (2022) considered 28 GCMs from CMIP5 and 32 from CMIP6 along with 25 metrics, and Anil et al. (2021) ranked 24 GCMs of CMIP6 using multicriteria decision making. A fundamental assumption in these studies is that GCMs that perform well in the historical simulations will perform similarly in future projections.

There have been ongoing debates on whether projections from every model are an equally valid and likely depiction of the future (Knutti 2010). Giorgi and Mearns (2002) argued that model performance could vary over time, and an outlier GCM in the future projections may not be the one with higher bias in reproducing current-day climate. They also propose a Reliability Ensemble Averaging (REA) technique with two criteria to estimate the average, uncertainty, and range of GCM simulations: model performance (historical) and model convergence (future). Their method assigns reliability weights to each GCM based on the GCMs' bias in historical simulations and the distance to the ensemble average in future projections. The weighting technique used in these uncertainty assessment studies (Giorgi and Mearns 2002; Xu et al. 2010) shows that the GCMs that perform well in historical periods may not converge well with ensemble projections in the future. Yet, the selection and ranking of GCMs in the literature are often based only on the historical performance of GCMs, ignoring information available in future projections.

Further, it has also been highlighted in several studies that the subselection of GCMs should consider model independence and climate change signal diversity to account for uncertainty in the projections (Di Virgilio et al. 2022; Merrifield et al. 2023; Vautard et al. 2021). However, the selection of all the GCMs that project diverse climate change signals or do not resemble each other may not solve the challenges of computational demand if most GCMs are distinct. One approach in such cases could be to select one GCM to represent the ensemble mean and two to represent the upper and lower bounds of uncertainty from an ensemble of projections. This could provide a way for selecting the subset of GCMs that represent the mean and variability of the ensemble that is useful for practical applications. Accordingly, the overall objectives of this study are (1) to evaluate possibilities of changes in GCMs reliability to replicate historically observed climate and their ability to converge towards ensembles mean in the future projections and (2) to develop a framework to identify suitable GCMs for hydrological impact studies considering GCM performance in the historical period and their convergence and divergence in future periods.

2 Materials and Methods

2.1 Study Area: Chesapeake Bay Watershed

The Chesapeake Bay, located in the Mid-Atlantic region of the eastern United States, is the largest and the first estuary in the nation targeted for restoration and protection. The Chesapeake Bay watershed (Fig. 1) covers a 166,000 km^2 drainage area within six states and the District of Colombia. The watershed is under significant pressure to meet the Total Daily Maximum Load mandates from the US Environmental Protection Agency (US EPA 2009). Climate change

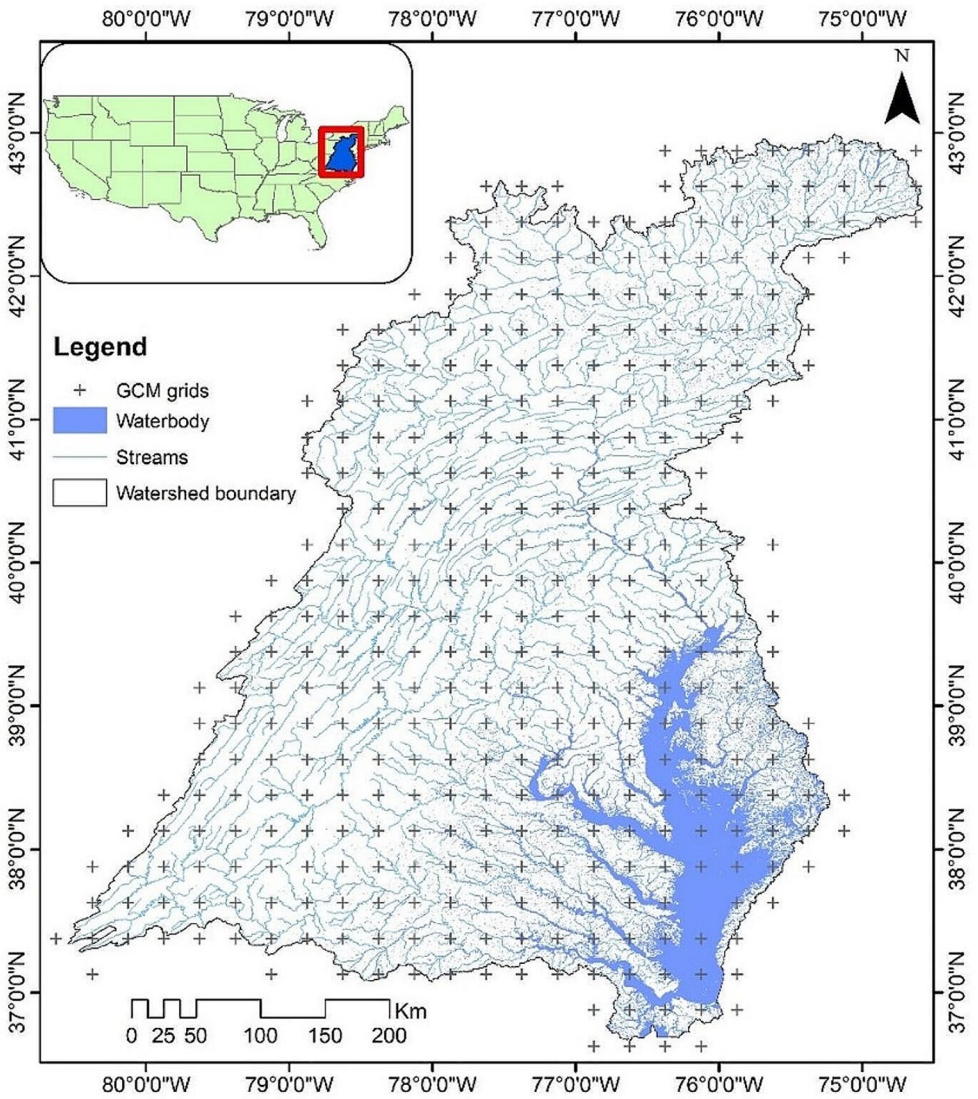

Fig. 1 Study area (Chesapeake Bay watershed, USA) along with all the GCM grids at which downscaled data is available and used to demonstrate selection of GCMs for climate impact studies

could greatly affect the Bay's recovery and economy. Heavy storms and higher temperatures could increase soil erosion, nutrient loss, flooding, and increased water temperature, stressing fish in the Bay watersheds. There are many efforts to evaluate potential climate change impacts and explore adaptation strategies in the Bay watershed (Hanson et al. 2022; Ansari et al. 2024; Maloney et al. 2020; Najjar et al. 2010; Saha et al. 2023).

2.2 Climate Data

National Aeronautics Space Administration (NASA) has recently downscalled and bias-corrected GCM data (Thrasher et al. 2022) from CMIP6 experiments through the project NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP-CMIP6). Its predecessor, NEX-GDDP-CMIP5 (https://www.nccs.nasa.gov/services/data-collections/land-based-products/nex-gddp), has been widely used throughout the world in climate impact studies (Jain et al. 2019; Li et al. 2020; Liao et al. 2019; Singh et al. 2019; Vigliano et al. 2018; Wang et al. 2020; Wu et al. 2020; Zhao et al. 2021). In this study, downscaled precipitation from 30 CMIP6 GCMs for the historical and future scenario SSP245 (Supplementary Table 1) (https://www.nccs.nasa.gov/services/data-collections/land-based-products/nex-gddp-cmip6, accessed in October 2022) was used. The NEX-GDDP-CMIP6 provides the dataset for only one variant in each GCM, which had r1i1p1f1 for 23 GCMs, r1i1p1f2 for 5 GCMs, and r3i1p1f1and r1i1p1f3 for one GCM each. More details about the GCMs used and their variants are provided in the Supplementary File). The Global Meteorological Forcing Datasets (GMFD; https://rda.ucar.edu/datasets/ds314.0/, accessed in October 2022) served as the reference observed data for the bias correction and evaluations. Both datasets are available at 0.25° resolution.

2.3 Bias Correction of Climate Projections

The climate projections from NEX-GDDP-CMIP6 are already downscaled and bias corrected to 0.25° resolution to match GMFD. However, we noticed that the extreme precipitation was still substantially underestimated by the downscaled GCMs, and the magnitude of the bias varied substantially (Fig. 2). Hence, an Extremes Weighted Empirical Quantile Mapping (EW-EQM; Rohith and Cibin 2024) bias correction was applied to overcome this underestimation. The EW-EQM is similar to the widely used conventional EQM (Boé et al. 2007) as given in Eq. (1), with the difference that separate bias correction was applied to extremes and non-extremes.

$$P_{bc} = F_{obs}^{-1}\left(F_{GCM}\left(P_{GCM}\right)\right) \tag{1}$$

where P_{GCM} and F_{GCM} are probabilities and the empirical distribution of GCM simulated precipitation, respectively, and F_{obs}^{-1} is the inverse of the empirical distribution of observed precipitation. In this study, a high precipitation threshold was set to differentiate extremes such that the number of extremes is equal to the number of years of data used in each analysis period (30 years). The extreme precipitation days were removed (or converted to zero) before bias correction was applied to non-extreme precipitation occurrences. The extracted extremes were bias-corrected separately and inserted back into the non-extremes time series.

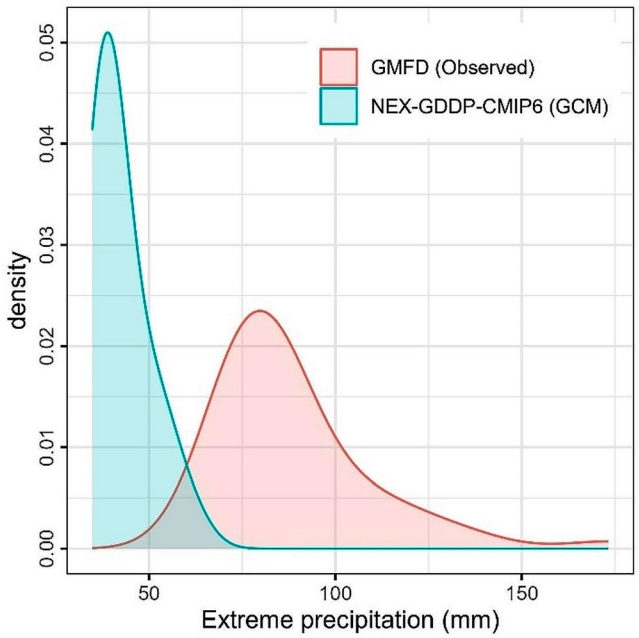

Fig. 2 Density plot of GMFD and NEX-GDDP-CMIP6 simulated extreme precipitation (50 extremes) at a location in the Mid-Atlantic region of the United States (US)

This method allows a better representation of extremes in the bias-corrected data and a more logical selection of GCMs for impact studies, as demonstrated in the Results and Discussion section. Further, the bias correction methodology employed assumes stationarity in the bias in both historical and future periods.

2.4 Reliability Ensemble Averaging

The framework developed in this study to identify suitable GCMs for impact study is based on the REA method extensively used to calculate the uncertainty, range, and ensemble mean of GCM projections (Giorgi and Mearns 2002, 2003). In the context of REA, ensemble refers to simulations/projections from different GCMs, not different realizations of the same GCM. REA facilitates comprehensive and quantitative assessment of the impact of climate change based on collective information from an ensemble of future projections. One of the primary goals of REA is to derive an ensemble average for future projections to act as a plausible representation of the ensemble in the absence of observations in the future period. The framework, as proposed by Giorgi and Mearns (2002), involves the evaluation of two criteria: (i) model performance criteria and (ii) model convergence criteria. The model performance criteria evaluate the ability of each GCM to replicate the historically observed climate patterns and assign reliability scores to them. For example, assuming P represents the total annual precipitation, a reliability score for P simulated by GCMs for the historical period is estimated as,

$$R_{B,i} = \frac{1}{RMSE\left(CDF\left(P_{h,m}\right), CDF\left(P_o\right)\right)} \tag{2}$$

where $P_{h,m}$ is the precipitation simulations by the GCM for the historical period, P_o is the observed precipitation, CDF is the cumulative distribution function, RMSE is the root mean square error, and $R_{B,i}$ is the reliability score. The $R_{B,i}$ is a function of the bias between the GCM simulated precipitation CDF and actual observed CDF in the historical period. The lower the bias, the higher the performance and reliability of the GCM to represent the current-day climate. A weight $w_{h,i}$ can be assigned to each GCM based on the reliability score to facilitate the calculation of the ensemble weighted average by giving a higher weight to GCMs witih higher reliability score than those with higher bias relative to the observations. The weights for N GCMs can be estimated as,

$$w_{h,i} = \frac{R_{B,i}}{\sum_{i=1}^{N} R_{B,i}} \tag{3}$$

The weights $w_{h,i}$ estimated for each GCM in the historical period are termed model performance weights, representing historical model performance. These weights can be used to estimate the ensemble average for the future period, assuming that the GCMs with higher reliability scores in the historical period are the ones with future projections that agree closely with the ensemble projections. However, the historical performance of GCM does not necessarily mean a higher degree of convergence in the projections (Giorgi and Mearns 2002).

The model convergence criterion evaluates how well projections from different GCMs converge. Similar to bias in the historical performance criteria, a distance was calculated between ensemble weighted average CDF and CDF from individual model projections to evaluate the model convergence criteria. The lower the distance of individual GCMs CDF to the weighted ensemble mean, the higher the GCM convergence. An initial ensemble weighted average was calculated as,

$$CDF\left(P_{f,m,avg,j}\right) = \sum_{i=1}^{N} w_{h,i} \times CDF\left(P_{f,m,i,j}\right), j = 1, \ldots . k \tag{4}$$

where $P_{f,m,i,j}$ is the projected precipitation for i^{th} GCM at j^{th} quantile and $P_{f,m,avg}$ is the initial ensemble weighted average for the future. A set of new GCM reliability scores and weights were estimated based on distances between ensemble weighted CDF and individual GCM CDF as follows:

$$R_{D,i} = \frac{1}{RMSE\left(CDF\left(P_{f,m}\right), CDF\left(P_{f,m,avg}\right)\right)} \tag{5}$$

$$w_{f,i} = \frac{R_{D,i}}{\sum_{i=1}^{N} R_{D,i}} \tag{6}$$

The procedure to estimate the ensemble weighted average and the weights was repeated until the weights converged. The final converged weights are termed convergence weights. If most of the GCMs were assigned similar weights, it indicates that the climate change projections are slightly sensitive to model differences. The GCMs with lower convergence weights are the outliers or the ones that do not agree with projections from most GCMs.

Further, we extend the methodology to select GCMs that represent the variability of projections, which is termed model divergence criteria. Similar to the identification of GCMs that are close to the ensemble mean, the GCMs closer to the 5 and 95 percentiles of the ensemble GCMs were identified to represent the variability or uncertainty in projections. This leads to the selection of at least 4 GCMs, one accounting for historical performance, one for future ensemble mean, and two for projected variability. The performance and convergence weights of the GCMs were compared across the watershed to evaluate if a GCM convergence in projections implies a better performance in the historical period (Giorgi and Mearns 2002).

3 Results and Discussions

Three climate change indices were evaluated: annual maximum 1-day precipitation, average annual precipitation, and average annual wet-day precipitation intensity. RA1max (annual maximum 1-day precipitation) is critical for flood risk assessments and stormwater management designs. RAsum (average annual precipitation) and SDDI (Simple Precipitation Intensity Index) are critical for water resources management, agricultural water management, and drought risk assessments. An empirical CDF of all three climate indices from observed data, all the 30 downscaled simulations and projections, along with the selected GCM from both historical performance criteria and future convergence criteria, are plotted in Fig. 3 for a randomly selected location (longitude = -75.375 W and latitude = 41.375 N). The GCM simulations in the historical period were underestimated entirely (more than 50%) for RA1max and SDII (Fig. 3a and c). They had 0% overlap with the observed CDFs, while the agreement is better for RAsum because the GCM simulations were bias-corrected to match only the RAsum. Since there is a consistent underestimation of RA1max and SDII by GCMs, the selected GCM for the region from the conventional performance criteria is simply the one with higher magnitudes (red lines in Fig. 3) for these indices. Figure 3 (red lines) shows that the GCM, which had the highest magnitude in the historical period, may continue to simulate higher magnitudes for SDII in the future (Fig. 3f). Hence, the higher magnitude GCM may not be the most probable case but the most extreme GCM.

Further, it should also be noted that RA1max and SDII of the historical downscaled GCM data did not preserve the observed interannual variability (Fig. 3a and c) and could have also led to lower interannual variability in the future (Fig. 3d and f). Here, the interannual variability was inferred based on the spread of CDF plots, as the variables plotted are at an annual scale. In contrast, the interannual variability of RAsum, which influences water availability-based impact decisions, was well reproduced by the downscaled historical GCM simulations, and the GCM selection worked reasonably well as the selected GCMs were close to the observed values in the historical period and ensemble averages in the future period. The underestimation of both mean and interannual variability of RA1max and SDII could substantially influence decisions related to extreme events, such as floods and droughts. This highlights the need for advanced bias

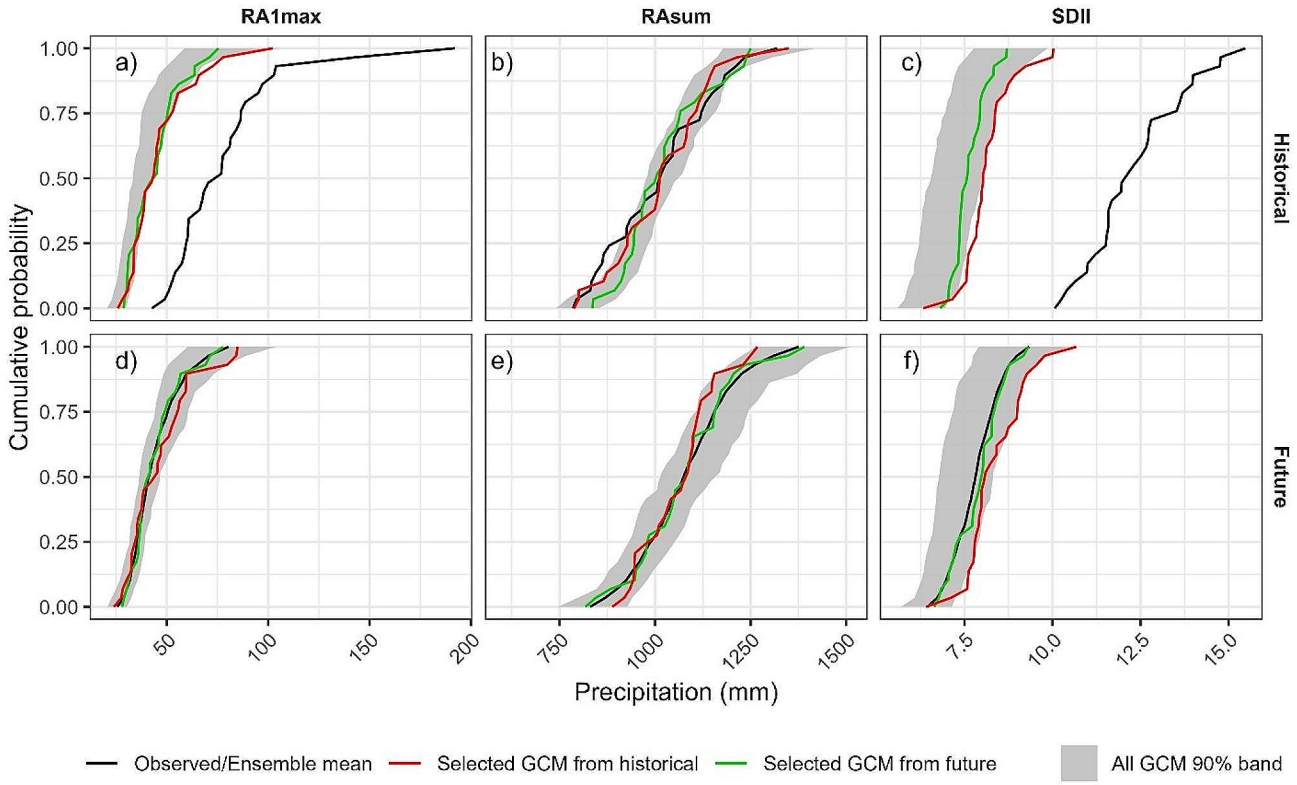

Fig. 3 Cumulative distribution function (CDF) plot of the daily annual maximum (RA1max), total annual precipitation (RAsum), and average wet-day intensity (SDII) climate indices at a location (longitude = -75.375 W and latitude = 41.375 N) in the Chesapeake Bay watershed in the historical (1976–2005) and future (2021–2050) periods. The climate simulations and projections are the NEX-GDDP data for CMIP6. The black line indicates the observed precipitation in the historical period and the reliability ensemble average in the future period; the grey band indicates the 90% band of simulated data from 30 GCMs; the red line indicates the CDF of the selected GCM based on historical performance criteria for each climate index; and the green line indicates the CDF of the selected GCM based on the future convergence criteria for each climate index

correction methods that can account for multiple indices in bias correction.

The GCM selected based on future convergence criteria (green lines in Fig. 3) falls very close to the ensemble mean and is 100% within the 90% band of all GCM in both the historical and the future periods. However, the initial weights that were used in the future convergence criteria to obtain the ensemble mean were based on historical performance weights. Performance weights were higher for the GCMs with higher magnitudes for RA1max and SDII, which may not be the most reliable GCMs. Hence, EW-EQM was applied to NEX-GDDP data to correct the bias in RA1max and SDDI. From Fig. 4, all three climate indices are well represented by GCMs after EW-EQM bias correction, and the observed CDF is more than 96% within the 90% band of all GCMs. The selected GCM based on the performance criteria was close to the observed and within the 90% band of GCMs in both historical and future periods. However, historical performance-based selected GCMs were not the closest to the ensemble average of the future projection (Fig. 4d and e, & 4f), and future convergence-based selected GCMs were not the closest to the observed in the historical period (Fig. 4a and b, & 4c). This indicates that a GCM that performed well in the historical period may not be the one that converges well in the future period.

To better understand whether there is a relationship between the historical performance of GCMs and their future convergence, a correlation coefficient was calculated between REA performance and convergence weights of all the GCMs for each grid in the watershed and all three variables. A scatter plot of REA performance and convergence weights is presented in Fig. 5, along with a histogram of the correlation coefficient across all the grids for each index. The correlation coefficient is mostly less than 0.5 for all the indices and was statistically significant at only less than 3.8%, 5.2%, and 4.1% of grids at 95% confidence for RA1max, RAsum, and SDII, respectively. This clearly indicates that the GCMs that performed well in the historical period may not always converge well with ensembles in the future. This warrants consideration of historical performance and future

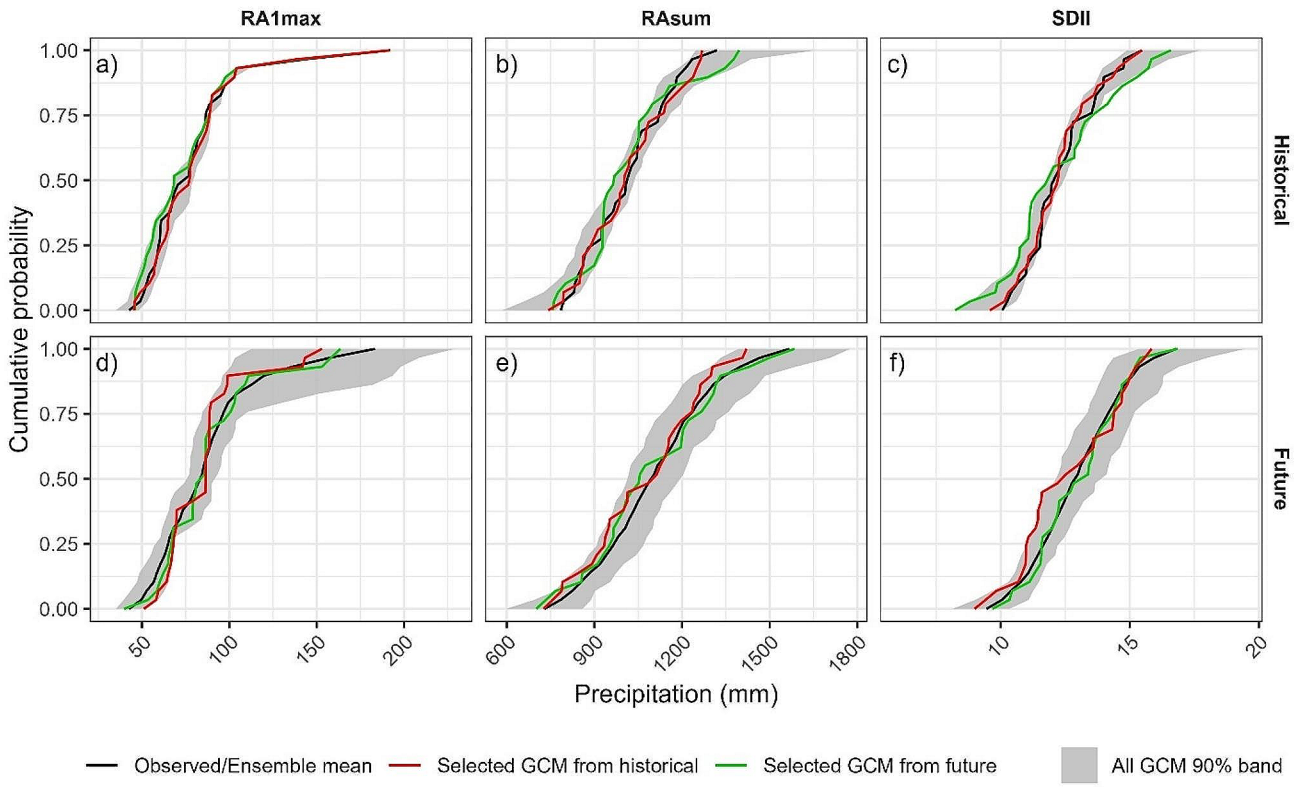

Fig. 4 Cumulative distribution function (CDF) plot of the daily annual maximum (RA1max), total annual precipitation (RAsum), and average wet-day intensity (SDII) climate indices at a location (longitude = -75.375 W and latitude = 41.375 N) in the Chesapeake Bay watershed in the historical (1976–2005) and future (2021–2050) periods. The climate simulations and projections are the Extremes-Weighted Empirical Quantile Mapping bias-corrected NEX-GDDP data for CMIP6. The black line indicates the observed precipitation in the historical period and the reliability ensemble average in the future period; the grey band indicates the 90% coverage of simulated data from 30 GCMs; the red line indicates the CDF of the selected GCM based on the historical performance criteria for each climate index; and the green line indicate the CDF of the selected GCM based on the future convergence criteria for each climate index

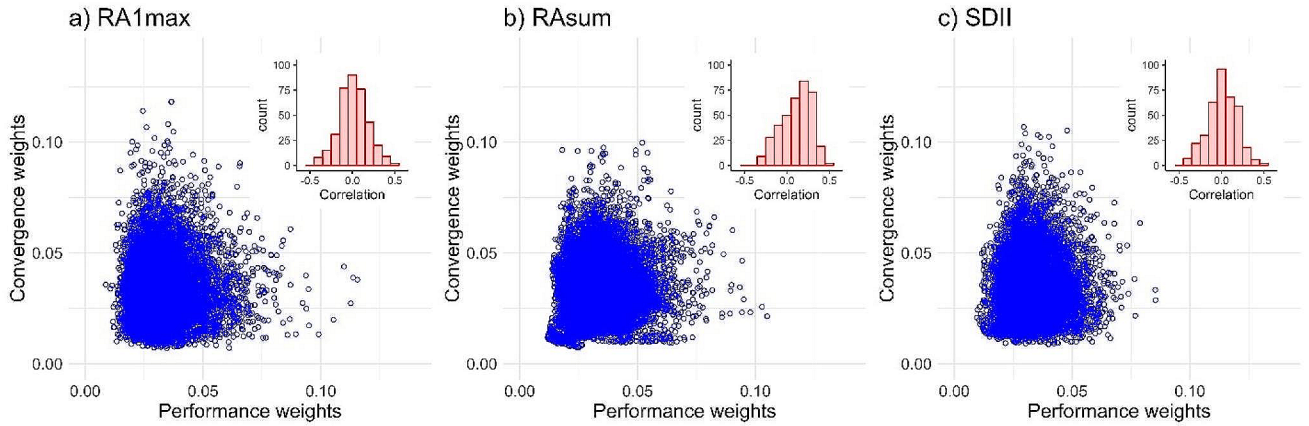

Fig. 5 Scatter plot between REA weights of historical performance and future convergence for each of 30 GCMs and 372 grids and three indices (a) RA1max, (b) RAsum, and (c) SDII in the Chesapeake Bay watershed. The annotated histogram inside each scatter plot represents the correlation coefficient of REA performance and convergence weights across GCMs at each point of the watershed grid

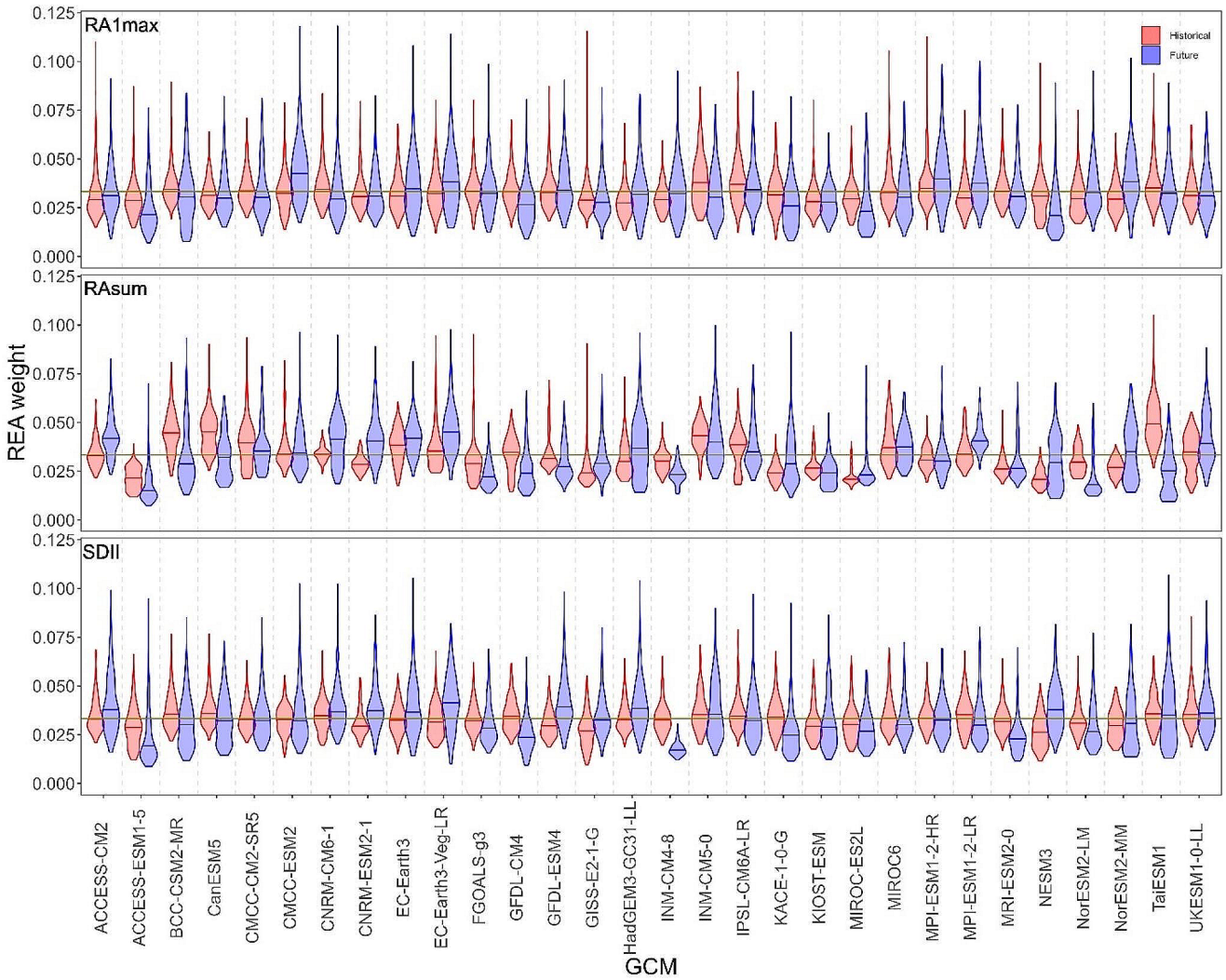

Fig. 6 Violin plot representing the density of REA weight for each GCM and in each criterion (performance and convergence) across 372 grids of the Chesapeake Bay watershed. The yellow horizontal line indicates a value of REA weight equal to 0.0333, which is 1/number of GCMs (30 in this study), indicating a value of GCM weight if all the GCMs are equally reliable. The number of data points used in the violin plot for each index, GCM, and time period was 372 (one for each grid in the watershed)

convergence in identifying suitable GCMs for computationally intensive hydrological impact studies.

The selection of GCMs demonstrated in Figs. 3 and 4 was for a single location (grid point) in the watershed. However, actual hydrological applications require the identification of GCMs for larger regions, such as watersheds. One way to identify that could be by taking the median of REA weights for each GCM across the watershed, and the GCM with a higher median would be a more suitable GCM across the watershed. In the violin plot (Fig. 6), it was observed that there are a few GCMs that have REA weights consistently higher than others across the watershed. For example, from the performance criteria, TaiESM1, BCC-CSM2-MR, and CanESM5 had their REA performance weight consistently higher than others for RAsum. Similar observations can be found with EC-Earth3-Veg-LR and EC-Earth3 for future convergence criteria. For RA1max, INM-CM5-0 and IPSL-CM6A-LR were more suitable than others for the performance criteria, and CMCC-ESM2 and MPI-ESM1-2-HR were suitable in the convergence criteria. Similarly, for SDII, CanESM5 and TaiESM1 were more suitable in the performance criteria and EC-Earth3-Veg-LR, GFDL-ESM4, and HadGEM3-GC31-LL in the convergence criteria. A violin plot of REA weights of downscaled GCM divergence (closest to 5 and 95 percentiles) is presented in Figure S1 (Supplementary).

The final GCMs identified (in the order of performance, convergence, 5 percentile, and 95 percentile) were INM-CM5-0, CMCC-ESM2, CanESM5, and GFDL-CM4 for RA1max, TaiESM1, EC-Earth3-Veg-LR, NorESM2-LM, and TaiESM1 for RAsum, and CanESM5, EC-Earth3-Veg-LR, MIROC-ES2L and EC-Earth3 for SDII. It can be noted

that TaiESM1, which was the selected GCM based on the historical performance criteria for RAsum, was also the 95 percentile variability GCM in the future projections. A ranking of GCM based on historical performance is also presented in Figure S2.

The inferences of this study establish that information from both historical simulations and future projections of GCMs needs to be accounted for when identifying the most suitable GCMs for regional impact studies. The framework developed facilitates the selection of at least 4 GCMs among the pool, that represent historical performance, future convergence, and future variability in the projections. This selection allows to quantify uncertainty in future projections without the use of ensemble GCMs in computationally intensive hydrological impact studies. It should also be noted that the selection of GCMs presented here is not directly based on raw GCM simulations or NEX-GDDP outputs but bias-corrected data of NEX-GDDP outputs, as explained earlier. Using different bias correction techniques, a different source of data, climate, or performance indices may result in different selected GCMs. This work aims to demonstrate the need to consider future convergence and divergence information along with the historical performance in the selection of GCMs. Hence, we recommend that specific applications use the methodology to select GCMs specific to their case. The methodology can also be seamlessly extended to a sector-specific ranking that accounts for multiple indices simultaneously (Baghel et al. 2022), multiple approaches of weighting and constraining projections (Brunner et al. 2020), and multiple regions.

4 Conclusions

This study develops an approach to account for the GCMs' ability to converge to the ensemble mean, their divergence information in the future, and their historical performance in order to identify suitable GCMs for impact studies. The historical performance of the GCM was quantified based on its ability to simulate observed climate indices. Future convergence ability was assessed based on the multi-model ensemble mean estimated from Reliability Ensemble Averaging. The GCM divergence information was quantified based on their closeness to the 5 and 95 percentiles of the ensemble in the projections. Bias-corrected simulations/projections for 30 GCMs of CMIP6 provided by NEX-GDDP were used to demonstrate the methodology in the Chesapeake Bay watershed. The results show that if only conventional performance criteria are used, the consistent underestimation of extremes by the GCM could lead to a selection of GCMs that suggest the most extreme cases in the ensemble. The accuracy of the bias correction method can greatly benefit the historical performance-based ranking and selection of GCMs. The proposed model convergence and divergence criteria overcome the assumption that a more reliable GCM in the historical period will continue to converge well with the ensemble in the future. Many GCMs that performed better in the historical period did not converge well on the future projections. Finally, the selection of at least four GCMs was recommended: one for historical performance, one for future convergence, and two for future variability. This generic approach can be extended to any region and sector and help identify suitable GCMs for computationally intensive climate change-related decision-making applications.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41748-024-00410-3.

Acknowledgements R. Cibin is supported, in part, USDA National Institute of Food and Agriculture and Hatch Appropriations under Project #PEN04967 and Accession #7006666. A. Mejia acknowledges support from the NSF CMMI program, grant number 2332169.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Anil S, Manikanta V, Pallakury AR (2021) Unravelling the influence of subjectivity on ranking of CMIP6 based climate models: a case study. Int J Climatol 41:5998–6016. https://doi.org/10.1002/JOC.7164

Ansari AH, Mejia A, Cibin R (2024) Flood teleconnections from levees undermine disaster resilience. npj Nat Hazards 1:2. https://doi.org/10.1038/s44304-024-00002-1

Baghel T, Babel MS, Shrestha S, Salin KR, Virdis SGP, Shinde VR (2022) A generalized methodology for ranking climate models based on climate indices for sector-specific studies: an application to the Mekong sub-basin. Sci Total Environ 829:154551. https://doi.org/10.1016/J.SCITOTENV.2022.154551

Boé J, Terray L, Habets F, Martin E (2007) Statistical and dynamical downscaling of the Seine basin climate for hydro-meteorological studies. Int J Climatol 27:1643–1655. https://doi.org/10.1002/JOC.1602

Brunner L, McSweeney C, Ballinger AP, Befort DJ, Benassi M, Booth B, Coppola E, Vries H, De, Harris G, Hegerl GC, Knutti R, Lenderink G, Lowe J, Nogherotto R, O'Reilly C, Qasmi S, Ribes A, Stocchi P, Undorf S (2020) Comparing methods to Constrain Future European Climate projections using a consistent Framework. J Clim 33:8671–8692. https://doi.org/10.1175/JCLI-D-19-0953.1

Callaghan M, Schleussner CF, Nath S et al (2021) Machine-learning-based evidence and attribution mapping of 100,000 climate impact studies. Nat Clim Chang 11:966–972. https://doi.org/10.1038/s41558-021-01168-6

Chhin R, Yoden S (2018) Ranking CMIP5 GCMs for Model Ensemble Selection on Regional Scale: Case Study of the Indochina

Region. J Geophys Res Atmosph 123:8949–8974. https://doi.org/10.1029/2017JD028026

Contribution of Working Group I to the Sixth Assessment (2021) In: Zhai VP, Pirani A, Connors SL, Péan C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis MI, Huang M, Leitzell K, Lonnoy E, Matthews JBR, Maycock TK, Waterfield T, Yelekçi O, Yu R, Zhou B (eds) Climate Change 2021: the physical science basis. Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, p 2391. https://doi.org/10.1017/9781009157896.

Di Virgilio G, Ji F, Tam E, Nishant N, Evans JP, Thomas C, Riley ML, Beyer K, Grose MR, Narsey S, Delage F (2022) Selecting CMIP6 GCMs for CORDEX Dynamical Downscaling: Model Performance, Independence, and Climate Change signals. Earths Future 10. https://doi.org/10.1029/2021EF002625. e2021EF002625

Giorgi F, Mearns LO (2002) Calculation of average, uncertainty range, and reliability of Regional Climate Changes from AOGCM simulations via the Reliability Ensemble Averaging (REA) Method. J Clim 15:1141–1158. https://doi.org/10.1175/1520-0442(2002)015>1141:COAURA<2.0.CO;2

Giorgi F, Mearns LO (2003) Probability of regional climate change based on the Reliability Ensemble Averaging (REA) method. Geophys Res Lett 30:2–5. https://doi.org/10.1029/2003GL017130

Hanson J, Bock E, Asfaw B, Easton ZM (2022) A systematic review of Chesapeake Bay climate change impacts and uncertainty: watershed processes, pollutant delivery and BMP performance. CBP/TRS-330-22. https://bit.ly/BMP-CC-synth

Hawkins E, Sutton R (2011). The potential to narrow uncertainty in projections of regional precipitation change. Clim Dyn 37:407–418. https://doi.org/10.1007/S00382-010-0810-6/FIGURES/8

IPCC, Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O (2022a) Climate Change 2022: Impacts, Adaptation and Vulnerability. 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 University Press, Cambridge, UK and New York, NY, USA, 3056 pp., https://doi.org/10.1017/9781009325844

IPCC (2022b) Summary for Policymakers. In: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Diemen, D. McCollum, M. Pathak, S. Some, P. Vyas, R. Fradera, M. Belkacemi, A. Hasija, G. Lisboa, S. Luz, J. Malley, (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA. https://doi.org/10.1017/9781009157926.001

Jain S, Salunke P, Mishra SK, Sahany S, Choudhary N (2019) Advantage of NEX-GDDP over CMIP5 and CORDEX Data: Indian summer Monsoon. Atmos Res 228:152–160. https://doi.org/10.1016/J.ATMOSRES.2019.05.026

Khadka D, Babel MS, Abatan AA, Collins M (2022) An evaluation of CMIP5 and CMIP6 climate models in simulating summer rainfall in the southeast Asian monsoon domain. Int J Climatol 42:1181–1202. https://doi.org/10.1002/JOC.7296

Knutti R (2010) The end of model democracy? Clim Change. 102:395–404. https://doi.org/10.1007/S10584-010-9800-2/METRICS

Li C, Wang M, Liu K, Coulthard TJ (2020) Landscape evolution of the Wenchuan earthquake-stricken area in response to future climate change. J Hydrol (Amst) 590:125244. https://doi.org/10.1016/J.JHYDROL.2020.125244

Liao X, Xu W, Zhang J, Li Y, Tian Y (2019) Global exposure to rainstorms and the contribution rates of climate change and population change. Sci Total Environ 663:644–653. https://doi.org/10.1016/J.SCITOTENV.2019.01.290

Maloney KO, Krause KP, Buchanan C, Hay LE, McCabe GJ, Smith ZM, Sohl TL, Young JA (2020) Disentangling the potential effects of land-use and climate change on stream conditions. Glob Chang Biol 26:2251–2269. https://doi.org/10.1111/gcb.14961

Merrifield AL, Brunner L, Lorenz R, Humphrey V, Knutti R (2023) Climate model selection by Independence, Performance, and spread (ClimSIPS v1.0.1) for regional applications. Geosci Model Dev 16:4715–4747. https://doi.org/10.5194/GMD-16-4715-2023

Najjar RG, Pyke CR, Adams MB, Breitburg D, Hershner C, Kemp M, Howarth R, Mulholland MR, Paolisso M, Secor D, Sellner K, Wardrop D, Wood R (2010) Potential climate-change impacts on the Chesapeake Bay. Estuar Coast Shelf Sci 86:1–20. https://doi.org/10.1016/j.ecss.2009.09.026

Perkins SE, Pitman AJ, Holbrook NJ, McAneney J (2007) Evaluation of the AR4 climate models' simulated daily Maximum temperature, Minimum temperature, and precipitation over Australia using probability density functions. J Clim 20:4356–4376. https://doi.org/10.1175/JCLI4253.1

Raju KS, Kumar DN (2020) Review of approaches for selection and ensembling of GCMs. J Water Clim Change 11:577–599. https://doi.org/10.2166/WCC.2020.128

Raju KS, Sonali P, Nagesh Kumar D (2017) Ranking of CMIP5-based global climate models for India using compromise programming. Theor Appl Climatol 128:563–574. https://doi.org/10.1007/s00704-015-1721-6

Rohith AN, Cibin. R (2024) An extremes-weighted empirical quantile mapping for global climate model data bias correction for improved emphasis on extremes. Theor Appl Climatol. https://doi.org/10.1007/s00704-024-04965-z

Rupp DE, Abatzoglou JT, Hegewisch KC, Mote PW (2013) Evaluation of CMIP5 20th century climate simulations for the Pacific Northwest USA. J Geophys Res Atmosph 118:10884–10906. https://doi.org/10.1002/JGRD.50843

Saha A, Cibin R, Veith TL, White CM, Drohan PJ (2023) Water quality benefits of weather-based manure application timing and manure placement strategies. J Environ Manage 333:117386. https://doi.org/10.1016/J.JENVMAN.2023.117386

Singh V, Jain SK, Singh PK (2019) Inter-comparisons and applicability of CMIP5 GCMs, RCMs and statistically downscaled NEX-GDDP based precipitation in India. Sci Total Environ 697:134163. https://doi.org/10.1016/J.SCITOTENV.2019.134163

Smiley KT, Noy I, Wehner MF, Frame D, Sampson CC, Wing OEJ (2022) Social inequalities in climate change-attributed impacts of Hurricane Harvey. Nat Commun 13:3418. https://doi.org/10.1038/s41467-022-31056-2

Thrasher B, Wang W, Michaelis A, Melton F, Lee T, Nemani R (2022) NASA Global Daily Downscaled Projections, CMIP6. Scientific Data 2022 9:1 9, 1–6. https://doi.org/10.1038/s41597-022-01393-4

US EPA (2009) National Water Quality Inventory: Report to Congress, 2004 Reporting Cycle. Washington, DC 20460

Vargo LJ, Anderson BM, Dadić R, Horgan HJ, Mackintosh AN, King AD, Lorrey AM (2020) Anthropogenic warming forces extreme annual glacier mass loss. Nat Clim Chang 10:856–861. https://doi.org/10.1038/s41558-020-0849-2

Vautard R, Kadygrov N, Iles C, Boberg F, Buonomo E, Bülow K, Coppola E, Corre L, van Meijgaard E, Nogherotto R, Sandstad M, Schwingshackl C, Somot S, Aalbers E, Christensen OB, Ciarlo JM, Demory ME, Giorgi F, Jacob D, Jones RG, Keuler K, Kjellström E, Lenderink G, Levavasseur G, Nikulin G, Sillmann J, Solidoro C, Sørland SL, Steger C, Teichmann C, Warrach-Sagi K, Wulfmeyer V (2021) Evaluation of the Large EURO-CORDEX Regional Climate Model Ensemble. Journal of Geophysical Research: Atmospheres 126, e2019JD032344.https://doi.org/10.1029/2019JD032344

Vigliano PH, Rechencq MM, Fernández Mv, Lippolt GE, Macchi PJ (2018) Fish thermal habitat current use and simulation of thermal habitat availability in lakes of the Argentine Patagonian Andes under climate change scenarios RCP 4.5 and RCP 8.5. Sci Total Environ 636:688–698. https://doi.org/10.1016/J.SCITOTENV.2018.04.237

Wang G, Zhang Q, Yu H, Shen Z, Sun P (2020) Double increase in precipitation extremes across China in a 1.5°C/2.0°C warmer climate. Sci Total Environ 746:140807. https://doi.org/10.1016/J.SCITOTENV.2020.140807

Wu Y, Miao C, Duan Q, Shen C, Fan X (2020) Evaluation and projection of daily maximum and minimum temperatures over China using the high-resolution NEX-GDDP dataset. Clim Dyn 55:2615–2629. https://doi.org/10.1007/S00382-020-05404-1/FIGURES/11.

Xu Y, Gao X, Giorgi F (2010) Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections. Clim Res 41:61–81. https://doi.org/10.3354/cr00835

Zhao Y, Dong N, Li Z, Zhang W, Yang M, Wang H (2021) Future precipitation, hydrology and hydropower generation in the Yalong River Basin: projections and analysis. J Hydrol (Amst) 602:126738. https://doi.org/10.1016/J.JHYDROL.2021.126738