Bias correction of climate model outputs influences watershed model nutrient load predictions

Lorrayne Miralha a, Rebecca L. Muenich a,⁎, Donald Scavia b, Karlie Wells c, Allison L. Steiner c, Margaret Kalcic d, Anna Apostel d, Samantha Basile c, Christine J. Kirchhoff g

a School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ 85281, USA
b School for Environment and Sustainability, University of Michigan, Ann Arbor, MI 48104, USA
c Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109, USA
d Department of Food, Agricultural, and Biological Engineering, The Ohio State University, Columbus, OH 43210, USA
e Translational Data Analytics Institute, The Ohio State University, Columbus, OH 43210, USA
f National Climate Assessment, ICF, 1725 I St NW, Washington, DC 20006, USA
g Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT 06269, USA

HIGHLIGHTS

• The impact of bias correcting climate model output on watershed model nutrient load prediction was the focus of this study.
• Nitrogen estimations yielded a better fit compared to phosphorus when using climate outputs that were not bias-corrected.
• Bias correction was found to influence hydrological processes and historical and future nutrient load predictions.

ABSTRACT

Waterbodies around the world experience problems associated with elevated phosphorus (P) and nitrogen (N) loads. While vital for ecosystem functioning, when present in excess amounts these nutrients can impair water quality and create symptoms of eutrophication, including harmful algal blooms. Under a changing climate, nutrient loads are likely to change. While climate models can serve as inputs to watershed models, the climate models often do not adequately represent the distribution of observed data, generating uncertainties that can be addressed to some degree with bias correction. However, the impacts of bias correction on nutrient models are not well understood. This study compares 4 univariate and 3 multivariate bias correction methods, which correct precipitation and temperature variables from 4 climate models in the historical (1980–1999) and mid-century future (2046–2065) time periods. These variables served as inputs to a calibrated Soil and Water Assessment Tool (SWAT) model of Lake Erie’s Maumee River watershed. We compared the performance of SWAT outputs driven with climate model outputs that were bias-corrected (BC) and not bias-corrected (no-BC) for dissolved reactive P, total P, and total N. Results based on graphical comparisons and goodness of fit metrics showed that the choice of BC method impacts both the direction of change and magnitude of nutrient loads and hydrological processes. While the Delta method performed best, it should be used with caution since it considers historical variable relationships as the basis for predictions, which may not hold true under future climate. Quantile Delta Mapping (QDM) and Multivariate Bias Correction N-dimensional probability density function transform (MBCn) BC methods also performed well and work well for non-stationary climate scenarios.
1. Introduction

Water bodies in the US and around the globe have increasingly exhibited symptoms of eutrophication, including hypoxia and harmful algal blooms (HABs) (Taranu et al., 2015; Wurtsbaugh et al., 2019). The increase in frequency and extent of HABs is a threat to the health of aquatic ecosystems and to society because it disrupts food webs and can affect human and ecosystem health (Glibert and Burford, 2019; Menzi et al., 2010; Sönmez et al., 2016). Moreover, habitat alteration, such as wetland removal and deforestation, has contributed to increased N and P loads in Klamath Lake, Oregon USA (Paerl et al., 2018).

Climate change can exacerbate these impacts. For example, warming temperatures have been linked to the earlier emergence of HABs in shallow lakes located in distinct regions of South America and Europe (Kosten et al., 2012), and precipitation changes are largely controlling the Gulf of Mexico load variability (Donner and Scavia, 2007). Changes in rainfall variability may also increase bloom emergence in Australia (O’Neil et al., 2012). However, there remains a gap in our understanding of how a changing climate may impact future nutrient loads.

Watershed models are often used to evaluate the effectiveness of different agricultural best management practices (BMPs) as well as to explain factors (i.e., land use practice, climate change) that reduce (or drive) excess nutrients (Daloglu et al., 2012; Kalic et al., 2015a; Muenich et al., 2016). The Soil and Watershed Assessment Tool (SWAT) is an example of a watershed-scale model that has been widely applied to investigate hydrological responses and nutrient load changes based on land management, soil, climate, and topographic characteristics (Dagniew et al., 2019; Douglas-Mankin et al., 2010; Gassman et al., 2007). A recent study used multiple SWAT models to understand the impact of different management strategies that would decrease future P loads in Lake Erie, providing insights into how to achieve nutrient load reduction goals (Scavia et al., 2017). While management practices have been identified, the authors did not consider their performance under a changing climate. Little is known about how future climate may influence changes to lake nutrient loads. On one hand, studies have suggested that future nutrient loads may decrease in western Lake Erie, associated with increased temperatures and evapotranspiration (Kalic et al., 2019), as well as greater plant uptake under higher CO2 concentrations (Culbertson et al., 2016). On the other hand, studies indicate that nutrient loads might increase under future climate regimes as changing annual temperature and precipitation increase flow rates and runoff in the Maumee watershed (Verma et al., 2015). Therefore, uncertainties remain surrounding the impacts of future land-atmosphere interactions on regional nutrient loads.

When considering the implementation of future climate model simulations in watershed modeling, a common approach includes bias correction (Cannon et al., 2020; Hakala et al., 2019) which is applied to correct systematic errors in climate model outputs, primarily due to the difference in scale (Hakala et al., 2019), that serve as inputs to watershed models. For example, most climate models are developed at either a global (GCMs; 100–500 km grid resolution) or regional (RCMs; 10–50 km grid resolution) scale (Leung et al., 2003), whereas watershed models operate with a full domain at typically an 8-digit hydrologic unit code scale (total model area ~ 100–1500 km2) or smaller (NRCS, 2020). Studies have investigated the impacts of bias correcting climate data for input to hydrological models in watershed processes, such as changes in flow, evapotranspiration, and rainfall (Cannon et al., 2015; Meyer et al., 2019; Teutschbein and Seibert, 2012). For instance, comparing two bias correction techniques, Quantile Mapping (QM) and Quantile Delta Mapping (QDM), Cannon et al. (2015) found that QM can inflate the magnitude of rainfall extremes in Canada. Another study compared 6 bias correction approaches (e.g. linear scaling (LS), distribution mapping (DM), QM) when correcting RCM precipitation and temperature outputs to simulate streamflow in an arid catchment in China. The authors found that while all methods improved raw RCM-simulated precipitation, there were variations in their corrected statistics, which resulted in differences when simulating streamflow via SWAT. Among all methods, LS overestimated flow by 100% in the simulation period and had the greatest bias (Fang et al., 2015).

However, there is a lack of understanding on how bias correction influences the prediction of nutrient loads. Kalic et al. (2019) explored the influence of climate change on nutrient load predictions in the Maumee River watershed using mid-century climate projections from one global and four regional climate models as input to SWAT. While their results suggested nutrient loads would decrease under a future climate, they did not bias correct the climate model outputs. Other studies have applied bias correction to predict future nutrient loads but did not investigate how different bias correction techniques would impact modeled nutrient outputs (Culbertson et al., 2016; Mehan et al., 2019). For example, Teutschbein et al. (2017) modeled changes in total inorganic N across 19 sites in Sweden. The authors combined GCM and RCM outputs and used only one bias correction method (i.e., distribution scaling) to correct for systematic bias. Results indicated significant increases in total inorganic N loads in the future (Teutschbein et al., 2017). A recent study in Germany on the effects of bias correcting climate model outputs to predict future changes in discharge suggested that predicted future flow differs by bias correction method (Wörner et al., 2019). This kind of investigation has yet to be done for nutrient load predictions. There remains a need to understand how bias correcting climate model outputs (i.e. precipitation and temperature) influences nutrient load model outputs to advance this field of study and guide decision-making processes.

This study evaluates the impacts of bias correcting precipitation and temperature from 4 GCMs from the Climate Model Intercomparison Project version 5 (CMIP5) model database (Taylor et al., 2012) on SWAT outputs of nutrient loads to the Western Lake Erie Basin (WLEB) from the Maumee River, Michigan USA. We modeled observed dissolved reactive P (DRP), total P (TP), and total N (TN) loads from 1985 to 1999, which were compared to SWAT loads driven by not bias-corrected and bias-corrected GCM outputs. The climate model scenario and bias correction method that most closely matched the observations were selected and used for the evaluation of hydrological processes and absolute nutrient load changes in the mid-century (2051–2065). Results from this study advance knowledge on the impacts of bias-corrected climate model outputs driving watershed modeled nutrient loads. It may also serve as a guide for watershed modelers worldwide on how bias correction techniques and climate model
choices may influence modeled nutrient outputs as well as serve as a cautionary tale for the decision-making process of developing future nutrient load management strategies.

2. Methodology

2.1. Study area

The study site for this work is the Maumee River basin located in parts of Michigan, Indiana, and Ohio, US (Fig. 1). This agriculturally dominated basin has flat topography characteristics as well as poorly drained soils, which requires installation of subsurface tile drains to ensure viability of agriculture. The basin is predominantly corn, soybean, and wheat crops, with a mix of dairy, swine, and poultry livestock operations. The watershed receives about 984 mm of rainfall on average, with an annual average temperature of 10 °C (Williams and King, 2020). Rainfall patterns have changed over the past 30 to 40 years influencing tributary discharge and P delivery in Lake Erie. Across the basin, annual rainfall increased by 102 mm from 1975 to 2017, with increases in intense rainfall occurring primarily during spring and summer seasons (Williams and King, 2020).

2.2. SWAT model characteristics

SWAT is a semi-distributed hydrologic and water quality model driven by daily time-scale inputs including precipitation, temperature, solar radiation, relative humidity and wind speed. This is a widely used model in the investigation of water resource issues associated with land use and management, especially in agricultural lands (Tan et al., 2020). The model used in this study was calibrated and validated for the Maumee River watershed considering field-scale best management practices (BMPs), and fertilizer applications (Apostel et al., in review). This SWAT model sub-divided the 358 sub-basins into 24,256 hydrologic response units (HRUs), each with ~70 acres area on average. The model accounted for unique combinations of crop rotations simplified to meet watershed crop characteristics; both inorganic and organic (including manure) fertilizer application; tile drainage, tillage management, and other management practices such as fertilizer incorporation and buffer strips; the increasing prevalence of soil stratification in the region; labile P calculation for each HRU; and new snow parameters based on data from NOAA’s Global Historical Climatology Network. The main objective of the calibration was to ensure the model simulates reasonable streamflow, nutrient and sediment load values. The calibration was performed for 2005–2015 period, with validation for 2000–2004. This model was run using the SWAT 2012 revision 635, which includes modification on the movement of soluble P through subsurface tile drains. The daily model calibration and validation metric results for TP, DRP and TN are in the supplementary information (SI Table 1).

2.3. Climate variables and climate models selection

Four global climate models from the Coupled Model Intercomparison Project Phase 5 (CMIP5) were used in this study: (1) CCSM4,
(2) MPI-ESM-MR (i.e. MPI), (3) CNRM-CM5 (i.e. CNRM), and (4) IPSL-CM5A-MR (i.e. IPSL). Though the models have distinct atmospheric grid resolutions, CCSM4 has the highest number of grid points within the study area compared to the other three models (Table 1). In this study, we used daily temporal resolution outputs with the Representative Concentration Pathway 8.5 (RCP8.5) as the future emissions scenario (Pachauri et al., 2014).

These models were selected from a set of 18 CMIP5 models, because they minimized error in historical seasonal-average temperature and precipitation over the WLEB (SI Table 2). Overall, the model that

---

**Table 1**
Description of the climate models used in this study.

<table>
<thead>
<tr>
<th>Model</th>
<th>ID in this study</th>
<th>Description</th>
<th>Atmospheric grid (lat × lon)</th>
<th>Grid points (lat × lon)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCSM4</td>
<td>CCSM4</td>
<td>National Center for Atmospheric Research (NCAR) - Community Earth System Model 4</td>
<td>0.94° × 1.25°</td>
<td>4 × 3 = 12</td>
</tr>
<tr>
<td>MPI-ESM-MR</td>
<td>MPI</td>
<td>Max Planck Institute for Meteorology - Earth System model running on mixed resolution grid</td>
<td>1.86° × 1.87°</td>
<td>2 × 2 = 4</td>
</tr>
<tr>
<td>CNRM-CM5</td>
<td>CNRM</td>
<td>Centre National de Recherches Météorologiques and Centre Européen de Recherche et de Formation Avancée</td>
<td>1.40° × 1.40°</td>
<td>2 × 2 = 4</td>
</tr>
<tr>
<td>IPSL-CM5A-MR</td>
<td>IPSL</td>
<td>Institut Pierre-Simon Laplace - Earth System Model for the 5th IPCC report - Mid resolution</td>
<td>1.27° × 2.5°</td>
<td>2 × 2 = 4</td>
</tr>
</tbody>
</table>

---

_L. Miralha, R.L. Muenich, D. Scavia et al._ Science of the Total Environment xxx (xxxx) xxx
provided the closest patterns to gridded observations across the temperature and precipitation metrics (the CCSM4 model) was also the one that had a realistic spatial representation of the Great Lakes. From a subset of modeled historical (1980–1999) and future (2046–2065, RCP 8.5 emission scenario) simulations, we used model precipitation (PCP), maximum temperature (TMAX), and minimum temperature (TMIN) as input to SWAT. We distinguish the SWAT run with climate model outputs from 1980 to 1999 as the “model-historical simulation”, and from 2046 to 2065 as the “future simulation”. A separate simulation using daily observations of PCP, TMAX and TMIN from 71 weather stations for 1980–1999 was also simulated for the historical time period, and this is referred to as the “model-observed simulation”. The “model-observed simulation” is treated as the baseline in this study.

2.4. Bias correction techniques

Bias correction (BC) was performed with daily climate model output (i.e. PCP, TMAX, TMIN) and equivalent daily observed data for the historical period (1980 to 1999). To bias correct, the observed values from the 71 weather stations were averaged to the nearest grid point of the climate model to ensure the spatial resolution of both the observed climate data and climate model output would be the same. To enable the comparison among BC techniques, we selected four well-established univariate BC methods and three recently developed multivariate BC methods. The four univariate methods were: (1) Delta; (2) Scaling; (3) Empirical Quantile Mapping (EQM); (4) Quantile Delta Mapping (QDM). The delta method adds the mean change signal between the simulated climate in the training period (p) and the test period (s) to the observations (o) (Eq. (1)), where the training period considers the data from 1980 to 1999 and the test period includes 2046–2065. The delta method is based on a difference for variables such as temperature, while for precipitation we apply a quotient approach.

\[
\Delta_{\text{TMAX/TMIN}} = o + (\text{mean} (p) - \text{mean} (s)) \\
\Delta_{\text{PCP}} = o \left( \frac{\text{mean}(p)}{\text{mean}(s)} \right)
\]

The scaling method scales the simulated variables based on the difference or quotient between the observed and simulated means (Eq. (2)). The difference is based on an additive equation that is usually applied to unbounded variables such as temperature, while the quotient or multiplicative is applied to variables such as precipitation, so frequency can be preserved.

\[
\text{Scaling}_{\text{TMAX/TMIN; additive}} = s - \text{mean} (p) + \text{mean} (o) \\
\text{Scaling}_{\text{PCP; multiplicative}} = \frac{s}{\text{mean} (p)} \times \text{mean} (o)
\]

EQM consists of calibrating the simulated cumulative distribution function (CDF) by adding the observed quantiles (i.e. mean delta change and the individual delta changes) in the corresponding simulated quantiles. This method adjusts 99 percentiles and linearly interpolates inside this range every two consecutive percentiles (Gutiérrez et al., 2019). EQM also has the option of extrapolation, which enables keeping the extreme values of the distribution; however, this option was not used in this study because of the potential risk of bias in the extremes of the distribution. Details of this method can be found in the documentation of the R package hyfo-biasCorrect (SantanderMetGroup, 2017). QDM is a quantile mapping method developed to avoid trend deterioration as in the traditional quantile mapping technique (Meyer et al., 2019). First, QDM extracts the climate change trend from the projected future quantiles. Then, the quantile mapping technique is applied to the detrended series. The quantile mapping is based on a transfer function that converts the CDF of the modeled data m, to match the CDF of the observed series o in a historical period h (Eq. (3)). The CDF function is denoted by F, and the bias correction of modeled PCP or TMP at time t within some projected period p is denoted by \(PCP_{h}^{m,p}(t)\) or \(TMP_{h}^{m,p}(t)\). This transfer function is only based on historical period information, excluding future model projection relationships (Cannon et al., 2015). Thus, QDM was developed as a solution to preserve model-projected relative changes in quantiles while correcting systematic biases in the quantiles of modeled series considering the observed values. The QDM transfer function is indicated by Eq. (4), where the additive sign changes to multiplicative if the variable to be corrected is precipitation. More details on QDM can be found in Cannon et al. (2015).

\[
\begin{align*}
PCP_{h}^{m} & = F_{su}^{-1}\left\{ F_{oh}^{m} \left[ PCP_{h}^{m,p}(t) \right] \right\} \\
 \text{PCP} & = \left\{ F_{su}^{-1}\left( F_{oh}^{m} [ PCP_{h}^{m,p}(t) ] \right) \right\ } \\
\text{or} & \left\{ F_{su}^{-1}\left( F_{oh}^{m} [ PCP_{h}^{m,p}(t) ] \right) \right\ } \\
\text{or} & \left\{ F_{su}^{-1}\left( F_{oh}^{m} [ PCP_{h}^{m,p}(t) ] \right) \right\ } \\
\text{EQM} & \text{ also has the option of extrapolation, which enables } \\
\text{keeping the extreme values of the distribution; however, this option was not used in this study because of the potential risk of bias in the extremes of the distribution.}
\end{align*}
\]

The multivariate bias correction (MBC) methods chosen were: (1) MBCp - Multivariate bias correction (Pearson correlation) that matches marginal distributions using QDM and the Pearson correlation dependence structure; (2) MBCr - Multivariate bias correction (Spearman rank correlation) that matches marginal distributions using QDM and the Spearman rank correlation dependence structure; and (3) MBC-N - Multivariate bias correction (N-pdf) that matches the multivariate distribution using QDM and the N-dimensional probability density function transform (Cannon, 2018; Cannon et al., 2015; Meyer et al., 2019). This method applies an orthogonal rotation to the data...
before QDM is executed. This rotation exposes QDM to a linear combination of the original variables, so QDM eventually corrects the probability distribution of the rotated data. For this method, the number of iterations used were the default (i.e., 30 iterations).

While the univariate methods are based on mean deviations as well as quantile calculations, the multivariate techniques are based on the correlation or relationship among several variables to correct the variable set as target. MBC analyses typically require several inputs, but here we aimed to correct PCP based only on TMAX and TMIN inputs. The same was done to correct TMAX, using only TMIN and PCP inputs, while TMIN was corrected using PCP, and TMAX. We chose this simplified approach as these are typically the variables used in future climate studies in watershed modeling, and often the most readily available across climate models. Two main packages in R studio were used for these analyses: 1) hyfo- used to bias correct daily climate data, and 2) MBC. We did not use delta from the hyfo package, because it does not consider the quotient approach for precipitation. Plots of monthly precipitation and temperature variables from 1980 to 1999 per scenario (observed, no-bias corrected (historical), and bias corrected) can be found in the supplementary information (SI Fig. 2).

2.5. SWAT simulations

We ran several SWAT simulations to estimate TN, TP, and DRP loads. One simulation used the observed (i.e., 71 weather stations data) PCP, TMAX, and TMIN from 1980 to 1999 as inputs. We also ran SWAT simulations using PCP, TMAX, and TMIN from the climate models as inputs without bias correction for both the historical (1980–1999) and future (2051–2065) periods, and then with the 7 methods for bias correction for both the historical and future periods.

Fig. 4. Quantile-Quantile (Q-Q) plots between observed climate data (y-axis – solid red line) and each bias corrected method and climate model output (x-axis). PCP, TMAX, and TMIN are the overall average among the stations (for the observed) and the climate model grid (for the not bias-corrected and bias-corrected scenarios). The y-axis and solid red lines indicate the observed climate data from the 71 weather stations. For comparison, the model-historical (black dots) indicates the non-bias-corrected climate data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
This resulted in 69 SWAT simulations, 28 runs using the 4 climate models and 7 bias-correction methods for each time period, 1 run with observed climate data, 4 runs with observed climate data at each climate model resolution, and 8 model historical and future runs without bias correction. Because the first 5 years from the climate inputs were used as a spin-up period in SWAT, the historical and future comparisons were based on the period from 1985 to 1999 and from 2051 to 2065. Each historical run was plotted using Quantile-Quantile plots (Q-Q plots) of average monthly values of TN, TP and DRP per year from 1985 to 1999. The future Q-Q plots were based on the best historical scenarios and served to evaluate nutrient absolute change in the mid-century period from 2051 to 2065.

2.6. Evaluation of model fit, bias correction method, and future load impacts

We took a systematic approach to evaluating model fit, bias correction methods, and load impacts (Fig. 2). We first evaluated which climate model outputs without bias correction (i.e. model-historical) resulted in the best SWAT-modeled loads when compared to the model-observed loads from 1985 to 1999. Second, we evaluated which bias correction technique provided the best SWAT-generated loads using Q-Q plots and three metrics: (1) Percent Bias ($P_{BIAS}$), (2) Nash-Sutcliffe Efficiency (NSE), and (3) $R^2$. $P_{BIAS}$ is defined as the deviation between the simulated and observed values. Negative $P_{BIAS}$ indicates underestimation, and positive value indicates overestimation. NSE ranges from $-\infty$ to 1, with 1 reflecting a perfect match to observations. $R^2$ is the square of correlation coefficient $r$, which is a measure of the linear relationship between observed and simulated values. Average monthly values of TN, TP, and DRP were the inputs for $P_{BIAS}$, NSE, and $R^2$ calculations. We calculate these metrics based on SWAT simulations with the observed climate matched to each climate model resolution for optimal comparison with the climate-BC scenarios. To evaluate if direction and magnitude of change vary with bias correction, we calculated absolute change in nutrient loads for the mid-century (2051–2060) based on the SWAT outputs with and without bias correction. We acknowledge that the replication of this methodological approach in other regions will depend on the availability of observed water quality data and a calibrated watershed model.

3. Results

3.1. Assessment of climate data without bias-correction

Comparing averages of observed PCP, TMAX, and TMIN values with the climate model output (i.e. historical with no BC) revealed that bias differs among climate models (Fig. 3). PCP bias generally increased with the magnitude of the observed values, where MPI and IPSL overestimated observed PCP by 38% and 20%, respectively. Due to better alignment with observed values, bias in CNRM PCP was the lowest among the model ensemble (7%, deviation from the observed climate ($\sigma$) = 5.03) while its standard deviation was higher than CCSM4 ($P_{BIAS}$ = 11%; $\sigma$ = 3.98), which performed better for larger values. The MPI model had the highest degree of deviation from the observed values ($\sigma$ = 6.2), followed by IPSL ($\sigma$ = 5.38). For temperature, on average, the climate models appeared to predict the historical TMAX well, with deviation from the observed values occurring mostly in the extremes of the distribution. Overall, TMAX and TMIN were underestimated in all models except for CCSM4 TMAX (Fig. 3). CCSM4 TMIN also show strong underestimation with $P_{BIAS}$ = $-61.1%$. IPSL ranked first ($P_{BIAS}$ = $-1.5%$) when predicting historical TMAX, followed by MPI ($-4%$), CNRM ($-5%$), and CCSM4 ($5%$). Prior to bias correction, CCSM4 and CNRM historical PCP outputs fit the observed PCP better, while MPI and IPSL were best in fitting observed temperature. Underestimation in CNRM was the most severe, with the model ranking last in performance ($-74%$ bias for TMIN). MPI and IPSL had the lowest bias (both $-57%$), but this was still high compared to PCP and TMAX. CNRM had the overall lowest bias for PCP, while MPI and IPSL were best for TMIN and TMAX, respectively.
3.2. Bias correction of climate output

The application of BC generally moved the historical no-BC values closer to observed values (Fig. 4). The Delta method was able to correct most of the bias between the historical and the observed PCP. In general, overestimates increased with increasing PCP values as well as with extremes for TMAX and TMIN. For TMAX and TMIN, the mismatch mostly occurred when using the scaling approaches. Results varied for each BC technique by model and climate output. For example, while EQM performed well when correcting PCP outputs from CNRM, IPSL, and MPI, the technique’s fit worsened when correcting CCSM4 PCP.

3.3. SWAT - scenarios without climate bias correction

Monthly SWAT TP and DRP loads, driven by historical daily precipitation and temperature from the climate models without bias correction, were over-predicted compared to observed loads (Fig. 5). While the overestimates increased with increasing loads in all scenarios, performance also varied based on climate model and nutrient. DRP and TP loads based on CCSM4 inputs were closer to observed compared to other models (Fig. 5). Modeled TN showed better fits despite the excess for other loading variables within the ensemble, and IPSL-CMSA-MR inputs resulted in the best overall TN fit (Fig. 5).

![Fig. 7. Quantile-Quantile (Q-Q) plots showing comparison between the DRP, TP, and TN loads from the SWAT driven by the 71 climate stations and the SWAT driven by bias-corrected precipitation (PCP), minimum temperature (TMIN), and maximum temperature (TMAX) climate model outputs.](image-url)
For seasonal loads averaged over the time period of the analysis, all models overestimated DRP and TP loads from November to March (Fig. 6). IPSL-CM5A-MR underestimated TN loads most of the spring months, although IPSL-CM5A-MR TN results were a closer fit to model-observed values. Interestingly, CCSM4 simulation results appeared to follow closely March and April model-observed TN average load values.

3.4. SWAT output with climate bias correction

After bias correcting PCP, TMAX, and TMIN and re-running the SWAT model, there were clear differences in load estimates among bias correction methods (Fig. 7). Among the bias correction methods, Delta resulted in loads closest to observations, followed by MBCn and QDM, respectively. MBCn and the other multivariate BC techniques (i.e. MBCp and MBCr), however, performed similarly to QDM, with MBCn outperforming QDM in most cases. In general, while the majority of the BC scenarios showed improvements over the simulations without bias correction, EQM resulted in larger deviation from observed loads except for TN.

Comparing the performance among climate models using the best BC methods (i.e. Delta, QDM, and MBCn), results showed that CCSM4 performed best for loads below 100 and 500 ton for DRP and TP, respectively (Fig. 8). For TP, CCSM4 usually underestimated smaller loads and overestimated higher ones. When applying Delta, all models seem to perform similarly, with all models performing well when predicting smaller loads for DRP and TP while larger loads were best predicted by MPI and IPSL scenarios. When applying MBCn and QDM, CCSM4 and CNRM performed best principally when simulating smaller loads. CNRM-CM5 appeared to perform best in simulating higher DRP loads, but is poor in simulating TP and TN. MPI-ESM-MR simulated higher TP loads better than the other climate model scenarios, but it does not match as well as CCSM4 for the mid-range of observed loads.

3.5. Model metrics comparison

Based on fit statistics (Tables 2–4), Delta outperformed all the other methods, with MBCn ranking second most of the time. The choice of BC method also depends on the metric applied. MBCn ranked best after Delta when considering NSE and $R^2$, but based on $PBIAS$, Scaling and QDM rank best after Delta, with QDM ranking best at least once among all metrics when compared to Scaling. Most of the $PBIAS$ results were positive, indicating consistent overestimation of nutrient loads. DRP and TP loads were highly and consistently overestimated among all models and BC methods. However, DRP overestimates were generally less than TP. $PBIAS$ varied significantly among bias correction methods. For example, for CCSM4, EQM had $PBIAS$ of 173 for TP, while Delta and QDM values were 0 and 12.9, respectively. TN was underestimated only for IPSL-CM5A-MR scenarios. However, it was overestimated by the other models (i.e. CCSM4, CNRM, and MPI). Based on NSE and $R^2$, MPI performed best for DRP and TP among the BC methods (Table 3–4). However, for TN, IPSL had the most optimal
months (~5 mm/month), except in the end of summer and beginning of fall season, and flow overall decreased (~50 mm/month), except in the winter, and July. In all BC methods, however, snowfall was consistently predicted to decrease in the future. Rainfall is expected to increase in the winter and in the first and third months of spring, while in other months are projected to decrease, which agrees with the increasing flow and decreasing snowfall patterns of QDM and MBCn scenarios. For temperature, we observed that both TMIN and TMAX are projected to increase in the future, which explains the increasing ET patterns in QDM and MBCn scenarios. These temperature patterns also explain the consistent decrease in snowfall for mid-century, because SWAT calculates snowfall by partitioning rainfall based on temperature values.

Combining both assessments (Figs. 9 and 10), March, April, and May DRP, TP and TN loads are projected to decrease as flow decreases and ET increases with potentially higher temperatures in the future. These results not only show how bias correction impacts the nutrient modeling results, but also how it impacts the hydrological processes that can drive the nutrient load results.

### 4. Discussion

We investigated the influence of bias correcting climate model precipitation and temperature output to be used as input to a watershed nutrient model. We compared the ability of the watershed model to simulate historical nutrient loads when driven by outputs (i.e. PCP, TMAX, and TMIN) from 4 climate models (CCSM4, CNRM-CMS, IPSL-CM5A-MR, and MPI-ESM-MR) without bias correction. We showed that model performance varied by nutrient type and

### Table 2

Percent Bias (PBIAS) result for model, bias correction method, and nutrient type. Calculations were based on monthly-averaged loads between loads from the SWAT driven by the observed climate data at the climate model resolution and the SWAT loads driven by the climate outputs used for each bias correction scenario. Best metric other than Delta is highlighted in each row.

<table>
<thead>
<tr>
<th>Model</th>
<th>Nutrient</th>
<th>No BC</th>
<th>Delta</th>
<th>Scaling</th>
<th>EQM</th>
<th>QDM</th>
<th>Mbc-N</th>
<th>Mbc-p</th>
<th>Mbc-r</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCSM4</td>
<td>TP</td>
<td>0.00</td>
<td>31.10</td>
<td>173.40</td>
<td>12.90</td>
<td>16.80</td>
<td>12.90</td>
<td>12.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>0.00</td>
<td>2.90</td>
<td>53.50</td>
<td>5.10</td>
<td>4.70</td>
<td>5.10</td>
<td>5.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DRP</td>
<td>0.00</td>
<td>25.50</td>
<td>110.60</td>
<td>12.40</td>
<td>14.60</td>
<td>12.40</td>
<td>12.40</td>
<td></td>
</tr>
<tr>
<td>CNRM-CMS</td>
<td>TP</td>
<td>0.00</td>
<td>23.90</td>
<td>72.80</td>
<td>16.30</td>
<td>19.20</td>
<td>18.80</td>
<td>15.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>0.00</td>
<td>23.20</td>
<td>57.00</td>
<td>7.20</td>
<td>7.20</td>
<td>6.50</td>
<td>4.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DRP</td>
<td>0.00</td>
<td>14.60</td>
<td>45.70</td>
<td>7.20</td>
<td>9.10</td>
<td>7.90</td>
<td>7.10</td>
<td></td>
</tr>
<tr>
<td>MPI-ESM-MR</td>
<td>TP</td>
<td>0.00</td>
<td>27.10</td>
<td>129.70</td>
<td>42.10</td>
<td>28.50</td>
<td>43.50</td>
<td>40.60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>0.00</td>
<td>14.30</td>
<td>31.00</td>
<td>10.80</td>
<td>14.40</td>
<td>11.30</td>
<td>11.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DRP</td>
<td>0.00</td>
<td>22.80</td>
<td>88.90</td>
<td>35.10</td>
<td>24.70</td>
<td>32.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IPSL-CM5A-MR</td>
<td>TP</td>
<td>0.00</td>
<td>80.10</td>
<td>49.00</td>
<td>61.00</td>
<td>34.20</td>
<td>62.20</td>
<td>64.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>0.00</td>
<td>54.00</td>
<td>31.90</td>
<td>43.40</td>
<td>24.40</td>
<td>43.60</td>
<td>45.40</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3

Nash-Sutcliffe Efficiency (NSE) results for model, bias correction method, and nutrient type. Calculations were based on monthly-averaged loads between loads from the SWAT driven by the observed climate data at the climate model resolution and the SWAT loads driven by the climate outputs used for each bias correction scenario. Best metric other than Delta is highlighted in each row.

<table>
<thead>
<tr>
<th>Model</th>
<th>Nutrient</th>
<th>No BC</th>
<th>Delta</th>
<th>Scaling</th>
<th>EQM</th>
<th>QDM</th>
<th>Mbc-N</th>
<th>Mbc-p</th>
<th>Mbc-r</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCSM4</td>
<td>TP</td>
<td>0.00</td>
<td>1.93</td>
<td>7.96</td>
<td>1.07</td>
<td>0.99</td>
<td>1.07</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>0.00</td>
<td>0.67</td>
<td>1.13</td>
<td>0.82</td>
<td>0.73</td>
<td>0.82</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DRP</td>
<td>0.00</td>
<td>1.62</td>
<td>3.91</td>
<td>1.07</td>
<td>0.81</td>
<td>1.07</td>
<td>1.07</td>
<td></td>
</tr>
<tr>
<td>CNRM-CMS</td>
<td>TP</td>
<td>0.00</td>
<td>2.04</td>
<td>3.53</td>
<td>2.29</td>
<td>1.73</td>
<td>2.27</td>
<td>2.45</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>0.00</td>
<td>0.82</td>
<td>0.64</td>
<td>0.63</td>
<td>0.61</td>
<td>0.64</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DRP</td>
<td>0.00</td>
<td>1.08</td>
<td>1.58</td>
<td>0.99</td>
<td>0.72</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>MPI-ESM-MR</td>
<td>TP</td>
<td>0.00</td>
<td>0.19</td>
<td>3.11</td>
<td>0.45</td>
<td>0.24</td>
<td>0.50</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>0.00</td>
<td>0.72</td>
<td>0.60</td>
<td>0.41</td>
<td>0.44</td>
<td>0.45</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DRP</td>
<td>0.00</td>
<td>0.41</td>
<td>2.39</td>
<td>0.76</td>
<td>0.45</td>
<td>0.76</td>
<td>0.69</td>
<td></td>
</tr>
<tr>
<td>IPSL-CM5A-MR</td>
<td>TP</td>
<td>0.00</td>
<td>3.36</td>
<td>1.99</td>
<td>2.43</td>
<td>1.34</td>
<td>2.49</td>
<td>2.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>0.00</td>
<td>0.31</td>
<td>0.20</td>
<td>0.29</td>
<td>0.25</td>
<td>0.31</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DRP</td>
<td>0.00</td>
<td>2.76</td>
<td>1.84</td>
<td>2.24</td>
<td>1.33</td>
<td>2.26</td>
<td>2.32</td>
<td></td>
</tr>
</tbody>
</table>
magnitude, but that driving SWAT with CCSM4 led to the best fit to observations. Like other recent studies (Ehret et al., 2012; Maraun et al., 2010; Teutschbein and Seibert, 2013) we found that bias correcting climate model outputs, to correct for systematic and random errors in global and regional climate model outputs, can alter the findings of a study. Nutrient load studies that consider future climate should evaluate the impact of any bias correction methods applied across temporal scales (annual to monthly) and depending on nutrient of interest.

Our results reveal several inherent uncertainties persisted within simulations for the western Lake Erie basin, confirming behavior found in previous studies for regions outside of the Great Lakes. First, using climate model outputs as input to SWAT can increase uncertainty within loading predictions (Mehdi et al., 2015). A study addressing SWAT climate data input issues in the northeast of Brazil suggested that the choice of climatic inputs is critical for better representation of watershed processes, directly influencing nutrient load forecasts (de Almeida Bressiani et al., 2015). In our study, for example, we found that, in general, CCSM4 climate outputs performed best among the other models, but that MPI-ESM-MR performed better at higher DRP and TP loads, and IPSL-CM5A-MR performed best for TN. However, we also found that applying bias correction was necessary to improve the fit to model-observed loads.

Second, the way N and P processes are represented in the watershed model can influence how they respond to climate. For example, SWAT has been shown to routinely underestimate soil solution P, which likely leads to the underestimation of dissolved P (Vadas et al., 2013; Vadas and White, 2010), and climate change would impact particulate and dissolved P differently. While modeling P in SWAT will improve over time, studies generally report satisfactory performance in modeling N (Cakir et al., 2020; Logsdon and Chaubey, 2013). This may explain why in our study the metrics were better for N when compared to TP and DRP. Additionally, Kujawa et al. (2020) showed that the main source of uncertainty when predicting N loads is related to the climate uncertainty, whereas uncertainty in P loads is mostly linked to uncertainty in the hydrologic model (SWAT).

While several studies have shown that climate bias correction impacts streamflow, precipitation, and snowfall models (Bhowmik et al., 2017; Meyer et al., 2019; Teutschbein and Seibert, 2012; Wörner et al., 2019), we showed that BC selection also impacts watershed nutrient model performance. We also demonstrated that the impact of BC choice varies by type of nutrient modeled, climate model used, and the fit metric of interest. For example, we found that CCSM4 outputs led to the best watershed model performance for DRP and TP, principally after Delta, QDM, and MBCn bias correction. However, with the exception of Delta, based on NSE, MBCn is the most optimal BC method. On the contrary, if PBIAS is the metric of interest, QDM is the most proper BC choice, after Delta. Interestingly, for TN, the no-BC scenario appears to perform reasonably well in terms of PBIAS and NSE for the IPSL-CM5A-MR model. This variation among model metrics and nutrient load per climate model was also observed by Yuan et al. (2020), although they did not compare among bias correcting methods.

Table 4

<table>
<thead>
<tr>
<th>Model</th>
<th>Nutrient</th>
<th>No BC</th>
<th>Delta</th>
<th>Scaling</th>
<th>EQM</th>
<th>QDM</th>
<th>Mbc-N</th>
<th>Mbc-p</th>
<th>Mbc-r</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCSM4</td>
<td>TP</td>
<td>0.10</td>
<td>1.00</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.11</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>0.08</td>
<td>1.00</td>
<td>0.08</td>
<td>0.12</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>DRP</td>
<td>0.08</td>
<td>1.00</td>
<td>0.09</td>
<td>0.09</td>
<td>0.09</td>
<td>0.10</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>CNRM-CMS</td>
<td>TP</td>
<td>0.22</td>
<td>1.00</td>
<td>0.21</td>
<td>0.21</td>
<td>0.19</td>
<td>0.23</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>0.14</td>
<td>1.00</td>
<td>0.13</td>
<td>0.18</td>
<td>0.15</td>
<td>0.16</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>DRP</td>
<td>0.22</td>
<td>1.00</td>
<td>0.21</td>
<td>0.22</td>
<td>0.20</td>
<td>0.25</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>MPI-ESM-MR</td>
<td>TP</td>
<td>0.29</td>
<td>1.00</td>
<td>0.34</td>
<td>0.33</td>
<td>0.36</td>
<td>0.30</td>
<td>0.35</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>0.17</td>
<td>1.00</td>
<td>0.16</td>
<td>0.20</td>
<td>0.19</td>
<td>0.19</td>
<td>0.18</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>DRP</td>
<td>0.27</td>
<td>1.00</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
<td>0.36</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>IPSL-CM5A-MR</td>
<td>TP</td>
<td>0.12</td>
<td>1.00</td>
<td>0.10</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.13</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>0.12</td>
<td>1.00</td>
<td>0.13</td>
<td>0.12</td>
<td>0.12</td>
<td>0.14</td>
<td>0.12</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>DRP</td>
<td>0.11</td>
<td>1.00</td>
<td>0.12</td>
<td>0.14</td>
<td>0.13</td>
<td>0.14</td>
<td>0.12</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Fig. 9. Monthly average absolute load change for the midcentury (2051–2065) period based on the overall best SWAT results driven by no bias-corrected (no BC) and best bias corrected CCSM4 outputs for Dissolved Reactive Phosphorus (DRP), Total Phosphorus (TP), and Total Nitrogen (TN).
Although we found Delta to be the optimal BC method, hydrological modeling studies have suggested Delta has both advantages and disadvantages (Cannon et al., 2020). One advantage is that it bases future predictions on historical reality, which is already understood, but it can be useful if the study seeks to understand the impact of the current climate regime on a future scenario. However, future interactions among climate, flow, or nutrient dynamics might differ from the historical period under non-stationary conditions (Lenderink et al., 2007). We showed that Delta bias corrections led to opposite directions of change when compared to other relevant univariate (QDM) and multivariate (MBCn) BC techniques, although QDM and MBCn reflected different magnitude changes. Delta has the potential to preserve the mean of the distribution in case of variables such as temperature as well as the mean and variance of bounded variables distribution such as precipitation. In the case of precipitation, Delta could improve the mean statistic but deteriorate the simulated variance principally when considering the extremes of a distribution (Cannon et al., 2020). This deterioration may explain the opposite direction of change when predicting monthly DRP, TP, and TN using Delta in comparison to QDM and MBCn. Therefore, Delta should be used with caution depending on the objective of the study, and because it assumes future climate, flow, and nutrient dynamics will function similarly to the present.

Studies that applied MBC to predict snow precipitation found MBC techniques, that account for the interdependencies among variables used as inputs, are superior to the univariate techniques such as QDM (Meyer et al., 2019). In this study, MBCn outperformed QDM in most of the cases, likely because in addition to correcting quantile dependent biases including frequencies and intensities of wet days, MBCn includes multivariate dependence structure among variables in the correction. However, this performance varied by climate model, metric of interest, and nutrient type. For CNRM-CMS, QDM outperformed both MBCn and MBCp but not MBCr when considering PBIAS. In the MPI-ESM-MR model scenarios, QDM performed poorly in comparison to MBCs. However, the difference among metrics for QDM and MBCs is not high, which may be because we only used temperature data to correct precipitation and vice-versa. It was interesting that MBCs performed well even though only temperature and precipitation were used as inputs. If more variables such as solar radiation, wind speed, and relative humidity were used the MBC performance would probably improve, but this finding is helpful for data limited study areas.

Fig. 10. Monthly hydrological process changes for each bias correction method: rainfall (mm), maximum (TMAX) and minimum (TMIN) temperature (°C), evapotranspiration (ET, mm), snowfall (mm), and flow (cms).
Analyzing the impact of BC choice in future nutrient load changes using the overall best model-BC combinations (i.e. CCSM4 no BC; CCSM4 Delta; CCSM4 QDM; CCSM4 MBCn), we found that BC choice impacts the direction of load. Kalcic et al. (2019) compared their no-BC results with a Delta BC scenario and suggested that bias correcting can significantly impact nutrient load predictions. Our results expand and emphasize the importance of addressing uncertainty in the choice of climate models and BC methods when modeling nutrients. Furthermore, we found that loads, based on no-BC, QDM, and MBCn, are likely to decrease in the future spring period due to decreases in flow and snowfall, as well as increases in temperature and evapotranspiration. However, these hydrological changes were the opposite when considering Delta as the BC method, further illustrating the importance of this choice for future simulations. Future studies should consider a combination of bias correction methods among precipitation and temperature for modeling purposes.

Bias correction of climate data is a common practice among watershed modelers across the globe, yet the impacts of different correction methods are rarely evaluated. We acknowledge that applying BC does not create climate variability, and so there is a stationarity of bias in simulations. However, BC is usually necessary (depending on the climate model choice and the objective of the study) to correct systematic errors of climate model outputs principally in the calibrating period (Hakala et al., 2019). Without this correction, the error propagation when forecasting could be substantial. The benefits of this study go beyond the Great Lake region in the US, and our approach can be applied to any location in the world where watershed models have been developed and calibrated as well as where data exists. In places where there is limited water quality data to investigate the specific influence of BC on water quality predictions, this work can serve as a guide for which methods might be most appropriate.

CRediT authorship contribution statement

Miralha completed the formal analysis, data curation, writing of original draft, visualization and contributed to the methodology development. Muenich was responsible for conceptualization, methodology development, review and editing of the original draft, supervision, and project administration. Scavia, Kalcic, Steiner were involved in the conceptualization, methodology development, review and editing of final draft. Apostel, Wells and Basile contributed resources and data curation in addition to reviewing and editing the final draft. Kirchhoff was involved in conceptualization and review and editing of final draft. Muenich, Scavia, Kalcic, and Kirchhoff were involved with funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was funded by the National Science Foundation Coastal SEES grant OCE-1600012. We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. A special thanks to Harry for helping to “push” the final version of this manuscript to the finish line.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2020.143039.

References

Kalcic, M.M., Muenich, R.L., Basile, S., Steiner, A.L., Kirchhoff, C., Scavia, D., 2019. Climate change and nutrient loading in the Western Lake Erie Basin: warming can counteract...


