Evaluation of gridded precipitation datasets over international basins and large lakes

Yi Hong[1], Eric J Anderson[2], Lauren M Fry[3], Hong Xuan Do[4], James A Kessler[5], Laura K Read[6], Arezoo Rafieei-Nasab[7], Andrew D. Gronewold[8], and Lacey Mason[9]

[1]Cooperative Institute for Great Lakes Research
[2]National Oceanic and Atmospheric Administration (NOAA)
[3]University of Michigan–Ann Arbor
[4]University of Michigan
[5]NOAA Great Lakes Environmental Research Laboratory
[6]Tufts University
[7]NCAR
[8]University of Michigan-Ann Arbor
[9]NOAA

November 25, 2022

Abstract

Reliable precipitation estimates are a crucial backbone for supporting hydrologic modeling and other geophysical applications. However, watersheds that extend across international boundaries or those that contain large bodies of water pose particular challenges to acquiring consistent and accurate precipitation estimates. The North American Great Lakes basin is characterized by both of these features, which has led to long-standing challenges to water budget analysis and hydrologic prediction. In order to provide optimal conditions for hydrologic model calibration, retrospective analyses, and real-time forecasting, this study comprehensively evaluates four gridded datasets over the Great Lakes basin, including the Analysis of Record for Calibration (AORC), Canadian Precipitation Analysis (CaPA), Multi-sensor Precipitation Estimate (MPE), and a merged CaPA-MPE data set, in which these products are analyzed at multiple spatial and temporal scales using station observations and a statistical water balance model. In comparison with gauge observations from the Global Historical Climatology Network Daily (GHCN-D), gridded datasets generally agree with ground observations, however the international border clearly delineates a decrease in gridded precipitation accuracy over the Canadian portion of the basin. Analysis reveals that rank in gridded precipitation accuracy differs for overland and overlake regions, and between colder and warmer months. Overall, the AORC has the lowest variance compared to gauge observations and has greater performance over temporal and spatial scales. While CaPA and AORC may better capture atmospheric dynamics between land and lake regions, comparison with a statistical water balance model suggests that AORC and MPE provide the best estimates of monthly overlake precipitation.

Evaluation of gridded precipitation datasets over international basins and large lakes

Yi Hong[1], Eric J. Anderson[2], Lauren Fry[2], Hong Xuan Do[3,4], James Kesseler[2], Laura Read[5], Arezoo Rafieei Nasab[5], Andrew D. Gronewold[3], Lacey Mason[2]

[1] Cooperative Institute for Great Lakes Research, University of Michigan, Ann Arbor, MI, USA

[2] Great Lakes Environmental Research Laboratory, Ann Arbor, MI, USA

[3] School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA

[4] Faculty of Environment and Natural Resources, Nong Lam University, Ho Chi Minh City, Vietnam

[5] National Center for Atmospheric Research, Boulder, CO, USA

Corresponding author: Yi Hong (yhon@umich.edu)

Key Points:

- Comprehensively evaluated four gridded precipitation datasets over international basins and large bodies of water.
- Analyzed at multiple spatial and temporal scales using station observations and a statistical water balance model.
- Results reveal discrepancies between precipitation options and provide insight into potential downstream impacts to hydrologic modeling.

Abstract

Reliable precipitation estimates are a crucial backbone for supporting hydrologic modeling and other geophysical applications. However, watersheds that extend across international boundaries or those that contain large bodies of water pose particular challenges to acquiring consistent and accurate precipitation estimates. The North American Great Lakes basin is characterized by both of these features, which has led to long-standing challenges to water budget analysis and hydrologic prediction. In order to provide optimal conditions for hydrologic model calibration, retrospective analyses, and real-time forecasting, this study comprehensively evaluates four gridded datasets over the Great Lakes basin, including the Analysis of Record for Calibration (AORC), Canadian Precipitation Analysis (CaPA), Multi-sensor Precipitation Estimate (MPE), and a merged CaPA-MPE data set, in which these products are analyzed at multiple spatial and temporal scales using station observations and a statistical water balance model. In comparison with gauge observations from the Global Historical Climatology Network Daily (GHCN-D), gridded datasets generally agree with ground observations, however the international border clearly delineates a decrease in gridded precipitation accuracy over the Canadian portion of the basin. Analysis reveals that rank in gridded precipitation accuracy differs for overland and overlake regions, and between colder and warmer months. Overall, the AORC has the lowest variance compared to gauge observations and has greater performance over temporal and spatial scales. While CaPA and AORC may better capture atmospheric dynamics between land and lake regions, comparison with a statistical water balance model suggests that AORC and MPE provide the best estimates of monthly overlake precipitation.

1 Introduction

Precipitation is a vital component of the water cycle and is the variable most commonly associated with atmospheric circulation in weather and climate research. Accurate and reliable precipitation estimates are crucial for a comprehensive understanding of the climate and of the hydrological cycle, as well as the proper management of water resources, agriculture, and disaster management (C. Kidd et al., 2012; Sun et al., 2018).

Numerous precipitation datasets are accessible at regional and global scales, and each can be classified into one of the three categories: gauge-based datasets, satellite estimates or reanalysis products (Sun et al., 2018; Tapiador et al., 2012). Gauge-based datasets provide reliable precipitation estimates for specific locations, and provide ground-truth information to evaluate other precipitation products (Khandu et al., 2016; Salio et al., 2015). The shortcomings of gauge-based datasets are the poor spatial representation of precipitation patterns (owing to poor spatial coverage of observation stations, especially in sparsely populated and over large inland water bodies), and many gauges have not operated continuously or concurrently (Cole & Moore, 2008; Chris Kidd et al., 2017). Satellite estimates address these limitations, and can provide precipitation information at high spatial and temporal resolutions, moreover, precipitation information from different satellite sources (visible/infrared imagery, passive microwave) are often combined and blended with gauge-based data to improve accuracy (Duan et al., 2016; Maggioni & Massari, 2018). The final category, reanalysis products, includes datasets generated from numerical models that combine satellite and ground observations in order to generate a synthesized precipitation estimate which is consistent with the observations (Tapiador et al., 2012). Reanalysis products can provide spatially and temporally homogenous data that amalgamates all of the available high-quality observations, however, its reliability can vary substantially across different locations and time periods due to the quality and coverage of assimilated datasets (Sun et al., 2018). As a result, it is important to assess the accuracy of satellite and reanalysis datasets prior to any hydroclimatic applications.

Evaluation of precipitation estimates is particularly important in hydrologic modeling (Henn et al., 2018), as errors and biases in precipitation forcing can significantly impact model calibration and regionalization (Renard et al., 2010). The need for precipitation evaluation is even more crucial over watersheds that contain large inland lakes. Developing accurate precipitation estimates across these watersheds is difficult because simulation of physical processes across these massive freshwater surfaces is a challenging component in regional climate models (Xiao et al., 2016), and also because there is a lack of measurements over the lake (Holman et al., 2012). Moreover, many continental-scale precipitation datasets are susceptible to variations in monitoring infrastructure and data dissemination protocols when watershed, political, and jurisdictional boundaries do not align, that may cause unreliable and discontinuous precipitation data over international basins (Andrew D. Gronewold et al., 2018). In these regards, the Laurentian Great Lakes basin is a unique representation of the challenges facing precipitation data development: (i) four of the five Great Lakes and sub-basins are bisected by the international border between the United States and Canada; (ii) the vast surface waters represent 32% of the total basin area, equaling 245,310 km^2 (Hunter et al., 2015). No other basin in North America poses the same combination of precipitation data development challenges.

Recent studies by Great Lakes binational *ad hoc* federal partnerships indicate that the Canadian Precipitation Analysis (CaPA; Fortin et al., 2015) and Multi-sensor Precipitation Estimate (MPE) (D. Kitzmiller et al., 2013) data are the two most promising sources of precipitation for long term application to the region (Andrew D. Gronewold et al., 2018). Both products combine gauge and radar data to provide a best estimate of precipitation in near-real time, while CaPA also relies on the Global Environmental Multiscale (GEM) model (Côté, Desmarais, et al., 1998; Côté, Gravel, et al., 1998). In an effort to leverage the quality of these products, the Midwestern Regional Climate Center (MRCC) recently led the development of a binational precipitation product that merges CaPA and MPE data over the Great Lakes basin (referred to herein as “Merged”), relying on CaPA over Canadian land, MPE over land in the United States, and an arithmetic average of CaPA and MPE over the lake surface (Andrew D. Gronewold et al., 2018). These datasets are potentially beneficial for operational applications and water budget analysis.

Operational hydrologic prediction requires accurate precipitation data for model calibration and real-time forecasts. The National Oceanic and Atmospheric Administration (NOAA) National Water Model (NWM) has been deployed operationally as the framework for water prediction across the United States since 2016 (Alcantara et al., 2018). For version 2.1 of the NWM, calibration was performed using the Analysis of Record For Calibration (AORC), a reanalysis high-resolution dataset of near-surface weather based on surface, radar, and satellite observations (D. H. Kitzmiller et al., 2018). Nevertheless, no precipitation assessment of the AORC over the international Great Lakes basin exists. In fact, a comprehensive assessment of available precipitation data sets for the Great Lakes region has not been carried out despite the importance of overlake precipitation to provide reliable estimates for regional water accounting (Holman et al., 2012). This study fills in this gap by comprehensively evaluating four gridded datasets (AORC, MPE, CaPA, Merged MPE-CaPA), at multiple spatial (overland, overlake, sub-basin, country) and temporal (daily, monthly, annual) scales. The analysis aims to bring new insights into the performance of various precipitation products over large water bodies and across international boundaries, and contributes to the guidance of selecting precipitation products for operational development and for water practitioners across the regions.

In this paper, we first describe the study area and the precipitation datasets that we compare over the Great Lakes basin (Section 2). We then present the evaluation methods and statistical metrics used to assess the performance of different precipitation products (Section 3). Each precipitation product was then compared with gauge-based estimates, overland average and overlake averages across each of the sub-basins (Section 4). Furthermore, we discuss issues related to data consistency and accuracy, and make suggestions about data improvement and operational applications (Section 5). Finally, conclusions and perspectives are summarized in Section 6.

2 Study area and precipitation datasets

2.1 Study area

The North American Laurentian Great Lakes collectively constitute the largest freshwater surface (and second largest volume) on the planet (Andrew D. Gronewold et al., 2013). It contains nearly 20% of Earth's fresh unfrozen surface water (approximately 23,000 km3), and, with their surrounding basin, cover an area of about 766,000 km2 across the United States and Canada (Hunter et al., 2015) (Fig. 1). The Great Lakes basin forms a chain connecting the east-central interior of North America to the Atlantic Ocean. From the interior to the outlet at the Saint Lawrence River, water flows from Superior to Huron and Michigan, southward to Erie, and finally northward to Lake Ontario, which outflows to the St. Lawrence River.

Based on hydrological characteristics, the Great Lakes basin can be divided into four different sub-basins, including Superior, Michigan-Huron, Erie and Ontario (Fig. 1). Moreover, the sub-basins are divided among the jurisdictions of the Canadian province of Ontario and eight U.S. states (Michigan, Wisconsin, Minnesota, Illinois, Indiana, Ohio, Pennsylvania, and New York).

Table 1. Lake and land surface area for each hydrologic basin and the entire Great Lakes basin. The percent contribution of lake and land surfaces to each sub-basin is also indicated (Hunter et al., 2015).

Sub-basin	Sub-basin area (km²)	Lake surface area (km²)	Land surface area (km²)
Superior	210,000	82,100 (39%)	128,000 (61%)
Michigan-Huron	369,400	117,400 (32%)	252,000 (68%)
Erie	103,510	26,810 (26%)	76,700 (74%)
Ontario	83,000	19,000 (23%)	64,000 (77%)
Total	766,010	245,310 (32%)	520,700 (68%)

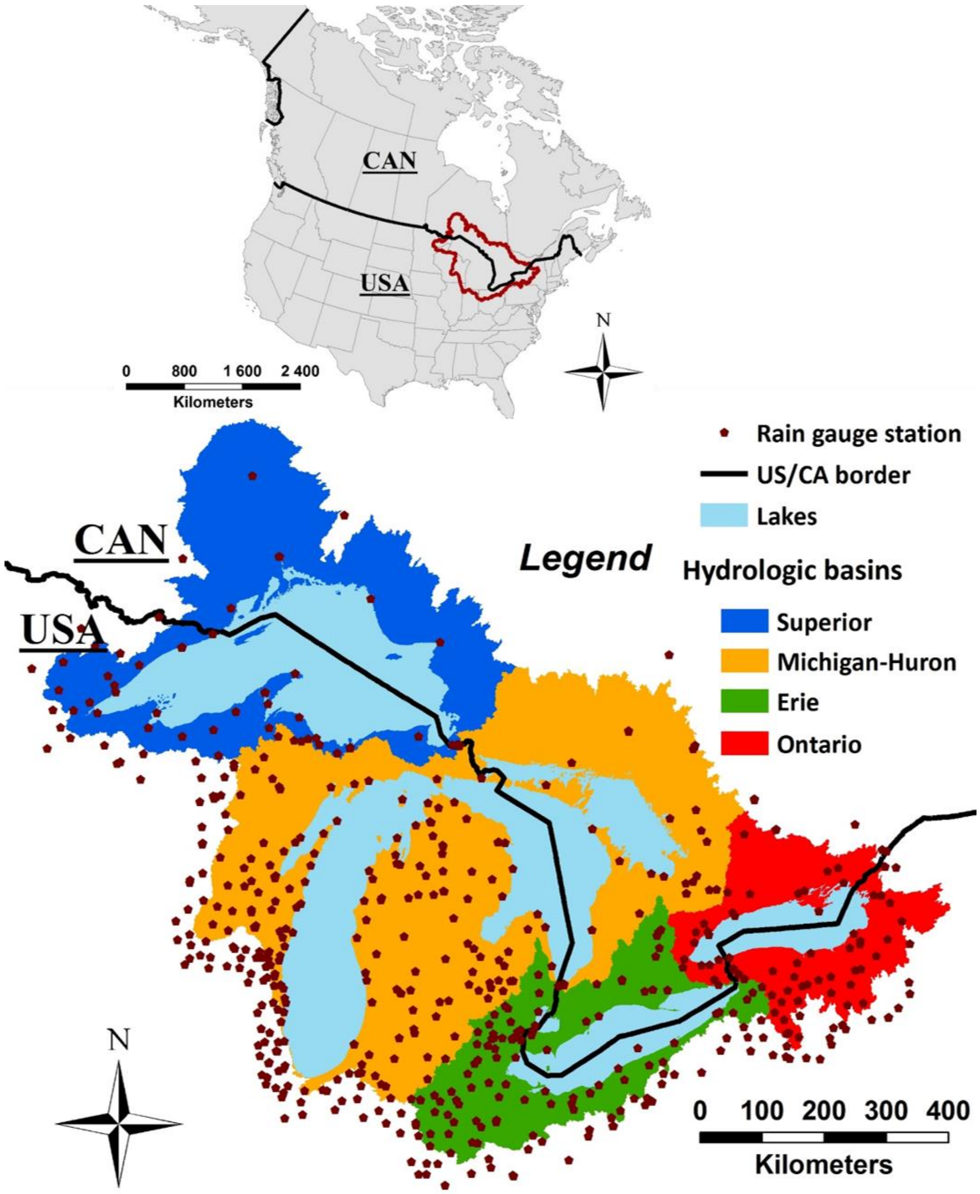

Fig. 1. The Great Lakes basin, 632 selected rain gauge stations, and delineation of four major hydrologic sub-basins: Superior (blue), Michigan-Huron (yellow), Erie (green), and Ontario (red). Lakes (light-blue). The solid black line indicates the USA/CAN border.

2.2. Datasets

This section briefly describes the four high resolution gridded precipitation products evaluated in this study (AORC, MPE, CaPA, Merged), and the reference data, which includes rain gauge observations from the Global Historical Climatology Network Daily database (hereafter referred to as GHCN-D) and sub-basin scale estimates of overlake and overland precipitation based on the gauge data (GLM-HMD). The analysis is performed from 2010 to 2019 for evaluating the data performance of recent 10 years. Table 2 summarizes the characteristics of the different precipitation datasets.

Table 2. Characteristics of the four gridded and two gauge based precipitation datasets over the Great Lakes region analyzed in this study.

Dataset	Citation	Spatial resolution	Finest temporal resolution	Coverage	Precipitation data sources[1]
AORC	(D. H. Kitzmiller et al., 2018)	1km	1-Hourly	Super-CONUS[2]	PRISM, (Livneh et al., 2015), NLDAS2, CFSR, MDR, WSI, NCEP (Stage II, IV) CMORPH
MPE	(D. Kitzmiller et al., 2013)	4 km	1-Hourly	Super-CONUS	HADS, MADIS, WSR-88D, space-based estimates from NESDIS
CaPA	(Lespinas et al., 2015)	10 km	6-Hourly	North America	SYNOP, METAR, RMCQ, SHEF, GOES, GEM
Merged CaPA-MPE	(Andrew D. Gronewold et al., 2018)	10 km	Daily	North America	CaPA, MPE
GHCN-D	(Menne et al., 2012)	Stations	Daily	Global	Gauge stations

[1]PRISM: Parameter elevation Regression on Independent Slopes Model;
NLDAS2: North American Land Data Assimilation System Forcing Fields Version 2;
CFSR: Climate Forecast System Reanalysis;
MDR: Manually-Digitized Radar;
WSI: Weather Services International;
NCEP: National Centers for Environmental Prediction;
CMORPH: Climate Prediction Center MORPHing technique;
HADS: Hydrometeorological Automated Data System;
MADIS: Meteorological Assimilation Data Ingest System;
WSR-88D: Weather Surveillance Radar-1988 Doppler;
NESDIS : National Environmental Satellite, Data, and Information Service;
SYNOP: Manual and automatic synoptic stations;
METAR: Aviation routine weather report;
RMCQ: Réseau Météorologique Coopératif du Québec;
SHEF: Standard Hydrometeorological Exchange Format;
GOES: Geostationary Operational Environmental Satellites;
GEM: Global Environmental Multiscale Model;
[2] "Super-CONUS" domain includes all contributing areas for contiguous U.S. surface waters.

GLM-HMD	(Hunter et al., 2015)	Interpolation of stations across basin-scale	Daily	Great Lakes basin	GHCN-D

2.2.1. AORC

Developed by the NOAA National Weather Service (NWS), the AORC surface precipitation is a reanalysis dataset that covers southern Canada, the contiguous United States, and northern Mexico. The domain includes all contributing areas for contiguous U.S. surface waters and is also referred to as the "Super-CONUS" region (D. H. Kitzmiller et al., 2018). It covers the period from 1979, at a time interval of 1 hour, with a grid resolution of approximately 1 km. The dataset was developed based on an approach similar to the North American Land Data Assimilation System Version 2 (NLDAS2; Xia et al. 2012), using multiple peer-reviewed and operational inputs to assimilate all weather information for forcing land-surface, snow, and hydrologic models. In order to decrease the uncertainty associated with national precipitation datasets, the AORC was officially used as the forcing data for NWM version 2.1 model calibration (Lahmers et al., 2019).

2.2.2. CaPA

The CaPA is a real-time gridded precipitation product provided by Environment and Climate Change Canada (ECCC). The grid has a resolution of approximately 10 km and the domain covers all of North America (Canada, USA and Mexico). It uses gauge data, radar reflectivity and the Geostationary Operational Environmental Satellite (GOES) imagery to modify a trial field provided by the GEM numerical weather prediction model using a statistical interpolation technique. The daily product has been operational since April 2011, but a hindcast starting in 2002 is available from ECCC.

2.2.3. MPE

The MPE (D. Kitzmiller et al., 2013) is currently used in the NWS to produce rainfall estimates that cover the 48 contiguous United States (CONUS) as well as portions of Canada and Mexico. MPE uses radar precipitation estimates from NWS and Department of Defense radars, hourly rain gage data and satellite precipitation estimates, and several other previously processed rainfall estimates, such as the NOAA National Severe Storms Laboratory (NSSL) Multi-Radar/Multi-Sensor (MRMS) data. These inputs are then manually analyzed by NWS to produce the daily best precipitation estimate on a 4 km grid on a 1 hour time step.

2.2.4. Merged CaPA-MPE

This "Merged" dataset (denoted as 'Mrg' in figures) is created using both CaPA and MPE datasets. Outside of CONUS, the CaPA data are applied at its native resolution. Within CONUS, the MPE data is resampled from its original 4 km cell size to a 10 km cell size. For cross-boundary areas and the Great Lakes, a 10-km buffer polygon was created on either side of the boundary of CONUS and was extended across the surface of the lakes. From both input datasets, point features intersecting this polygon are selected and appended into a single point feature class. An inverse distance weighted interpolation with a power setting of 0.5 and 10-

point variable search radius is used to create a new raster dataset with a 10-kilometer resolution. Finally, the interpolated raster data is mosaicked with the appropriate parts of the CaPA and resampled MPE data. The daily data is available from 2004.

2.2.5. GHCN-D

NOAA's GHCN-D data (Menne et al., 2012) are used as the reference station observations. GHCN-D is comprised of daily climate records from numerous sources that have been integrated and subjected to a quality control process. A dense network of rain gauge stations is recorded in the database. A total of 3,262 stations can be identified within and up to a range of 50 km outside the Great Lakes basin. Among which, 1,284 stations contain records after 2010 and we selected 632 stations that met our 90% temporal coverage threshold for the period of 2010 – 2019. The spatial distribution of these 632 stations are shown in Fig. 1. Rain gauges are particularly sparse over the northern part of the Superior and Michigan-Huron sub-basins, where the population density is low. Moreover, no station records can be found over the lakes themselves. This highlights the need for reliable alternative precipitation datasets to enhance understanding of water-related aspects of the whole basin, which is one motivation of this current study.

2.2.6. GLM-HMD

The NOAA Great Lakes Environmental Research Laboratory (GLERL) Hydrometeorological Database (GLM-HMD; Hunter et al., 2015) is used as the reference for evaluation of overlake and overland precipitation at the scale of each sub-basin (hereafter denoted as HMD in figures). GLM-HMD daily data are available at the site: https://www.glerl.noaa.gov/ftp/publications/tech_reports/glerl-083/UpdatedFiles/daily/. GLM-HMD uses GHCN-D and applies a modified version of conventional Thiessen weighting interpolation method (Croley & Hartmann, 1985) to calculate both daily overlake and overland precipitation estimates for each hydrologic sub-basin.

3 Evaluation methods

3.1. Comparisons with gauge observations

As described in the above sections, 632 GHCN-D gauge observations (Fig. 1) are used in this study to evaluate the AORC, MPE, CaPA and Merged products. Since these gauge stations are highly irregularly distributed over the Great Lakes basin, in order to avoid errors related to upscaling and interpolation methods (Hofstra et al., 2008), we directly carry out the grid-point comparisons of the gridded data and point observations. For each data product, we extracted precipitation from the grid cells that have the centroid closest to the rain gauge geographical coordinates. Together with gauge-based records, these time series form the product-gauge data pairs for evaluation. We analyzed biases, errors and correlations of each data product relative to gauge-based precipitation (from 2010 to 2019) using three experiment settings: (i) overall performance at the daily resolution across all stations; (ii) performance across different months of the year across all stations; and (iii) performance over the U.S. and Canadian portions of the basin over both daily and monthly resolutions.

Three commonly used statistical metrics were applied to analyze these product-gauge data pairs, including the Mean Absolute Error (MAE) to describe the discrepancies, the Percent Bias (PBias) showing the relative bias, and the coefficient of determination (R2) to represent the degree of collinearity. MAE, PBias, and R2 are defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(P_i - G_i)| \qquad (1)$$

$$PBias = 100 \times \frac{\sum_{i=1}^{n}(P_i - G_i)}{\sum_{i=1}^{n} G_i} \qquad (2)$$

$$R^2 = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(G_i - \bar{G})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2}\sqrt{\sum_{i=1}^{n}(G_i - \bar{G})^2}} \qquad (3)$$

Where n refers to the number of product-gauge data pairs; P_i and G_i represent the daily rainfall values of product and gauge, respectively; $\bar{P}$ and $\bar{G}$ are the mean value of product and gauge precipitation, respectively.

3.2. Comparing overland and overlake precipitation at sub-basin scale

As shown in Table. 1, lake surface area represents an important portion of the Great Lakes basin. We separately compute the daily mean of overland and overlake precipitation across each of the four hydrologic sub-basins by taking the averaged value of all grid cells within the polygon of the corresponding lake surface. This average is calculated using the command line tool (github.com/isciences/exactextract), which handles grid-cell size inconsistency and grid-cell/polygon intersection in a very precise manner. The GLM-HMD dataset is used as the reference data. Inter-comparisons of these datasets at sub-basin scales can provide information for calibration and uncertainty analysis of hydrological models. The same metrics of gauge-based evaluation (described in section 3.1) were then used to assess the performance of each gridded product over the land and lake portions of each hydrologic sub-basin.

3.3. Overlake precipitation analysis based on water balance closure

Since very few direct measurements of overlake precipitation are available, the accuracy of precipitation estimates over large lakes is not well understood (Holman et al., 2012; Xiao et al., 2016). In addition, previous studies also showed that overlake precipitation derived from existing precipitation datasets may be associated with high uncertainty and thus are not able to reconcile the water balance over multiple time period (A. D. Gronewold et al., 2016). In this study, we apply the Large Lake Statistical Water Balance Model (L2SWBM) (Do et al., 2020; Andrew D. Gronewold et al., 2020) to analyze the fidelity of different overlake precipitation estimates in the context of closing the water balance.

The L2SWBM assimilates multiple data sources into a Bayesian Marko chain Monte Carlo routine to infer feasible ranges of the major components (e.g. lake storage, overlake evaporation, inflow runoffs, overlake precipitation, etc.) of the water balance for each of the Great Lakes. To ensure the new estimates can reconcile the water balance over multiple periods, L2SWBM uses a conventional water balance equation to constrain the posterior inference. For this study, we run the L2SWBM for all of the Great Lakes from 2010 to 2019, while historical data from 1950 to 2009 were used to derive the "prior beliefs" of the possible ranges. In addition to the four datasets being evaluated, GLM-HMD precipitation was also assimilated in our simulations. For the other components of the water balance (i.e. lake levels, overlake evaporation, runoff and connecting channel flows), we used a database that was synthesized by the Coordinating Committee for Great Lakes Basic Hydrologic and Hydraulic Data (Do et al., 2020). The reliability of different gridded datasets (AORC, MPE, CaPA, Merged) can be assessed by comparing their overlake precipitation with the L2SWBM posterior inference of precipitation for each lake.

4. Results

4.1. Comparison of precipitation products with rain gauge observations

4.1.1. Evaluation at daily time steps and transboundary impacts

Across all 632 rain gauges over the entire Great Lakes basin, boxplots of MAE, PBias, and R2 reveal discrepancies between the four gridded precipitation products and the observed precipitation for the period 2010 to 2019. The range of median values for these metrics across different datasets are: MAE around 2 mm/day, PBias in the range of ±5 %, and R2 between 0.6 and 0.8. Compared to existing studies (Duan et al., 2016; Sun et al., 2018), these results indicate a reasonable agreement between all evaluated products and the GHCN-D gauge observations. However, differences can be found between precipitation products as well. For instance, CaPA slightly underestimates the GHCN-D data, while the other datasets overestimate (Fig. 2b). Furthermore, although the AORC dataset seems to have the poorest performance according to these median values, its performance is less dispersed across different rain gauge stations (i.e. AORC has a smaller interquartile range), suggesting a better consistency across spatial scales for the entire basin.

Spatial variation of the performance of different gridded products is illustrated by maps of statistical metrics for the entire set of selected rain gauge stations over the period of evaluation, 2010 – 2019 (Fig. 3, S1, S2). Among these 632 stations, 529 stations are located in U.S., and 103 stations are situated in Canada. Using this division across the international boundary, MAE, PBias and R2 are calculated separately for stations located in the U.S. and Canada (Fig. 4). In general, metrics reveal the datasets perform better in the U.S. portion of the basin, with lower MAE and R2 relative to evaluation at Canadian stations. Similar to what was noted above, CaPA tends to underestimate on both sides of the border, whereas the other products tend to overestimate precipitation on the U.S. side and underestimate on the Canadian side (Fig. 3, 4b). However, these differences may derive from different protocols for generating and adjusting GHCN-D data for Canadian stations rather than an indication of poor skill of the various products; this issue is further discussed in Section 5.

The boxplots of MAE (Fig. 4a) and R2 (Fig. 4c) confirm the consistency of the AORC, with R2 values from 0.6 to 0.8 for U.S. stations and from 0.4 to 0.6 for Canadian stations. Again, the other datasets have larger interquartile ranges, varying from 0.4 to 0.8 for U.S. stations and from 0.2 to 0.4 for Canadian stations. However, the relative higher MAE and lower R2 for the AORC reported in Fig. 2a, c is driven by higher errors on the U.S. side as compared to the other products, whereas it actually outperforms the others on the Canadian side. In terms of PBias, Fig. 4b indicates that CaPA and Merged datasets have the best performance with no considerable difference found between U.S. and Canadian gauges. Yet, considering the MAE values for CaPA and Merged datasets, this suggests that large positive and negative biases exist for specific Canadian rain gauge stations.

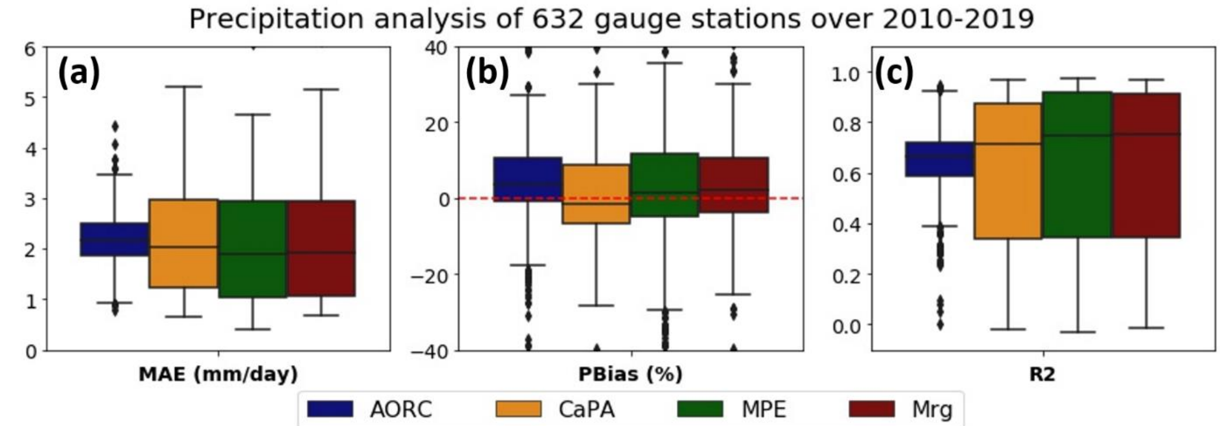

Fig. 2. Boxplots (median, 25th/75th percentiles, whiskers= represent the 1.5 standard deviation above and below the mean of the data) of MAE (**a**), PBias (**b**), and R2 (**c**) for daily precipitation at 632 rain gauge stations in the Great Lakes basin over the 2010 – 2019 period.

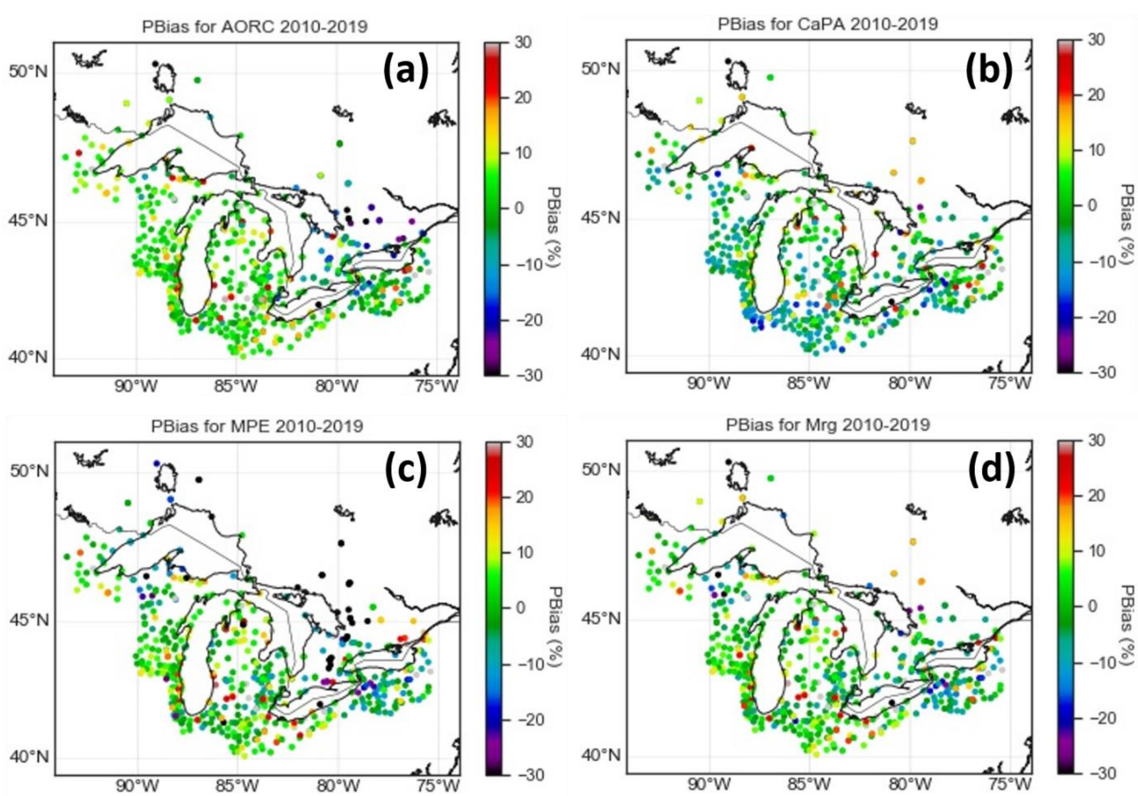

Fig. 3. Maps of PBias for AORC (a), CaPA (b), MPE (c), and Merged (d) at 632 rain gauge stations in the Great Lakes basin over 2010 – 2019.

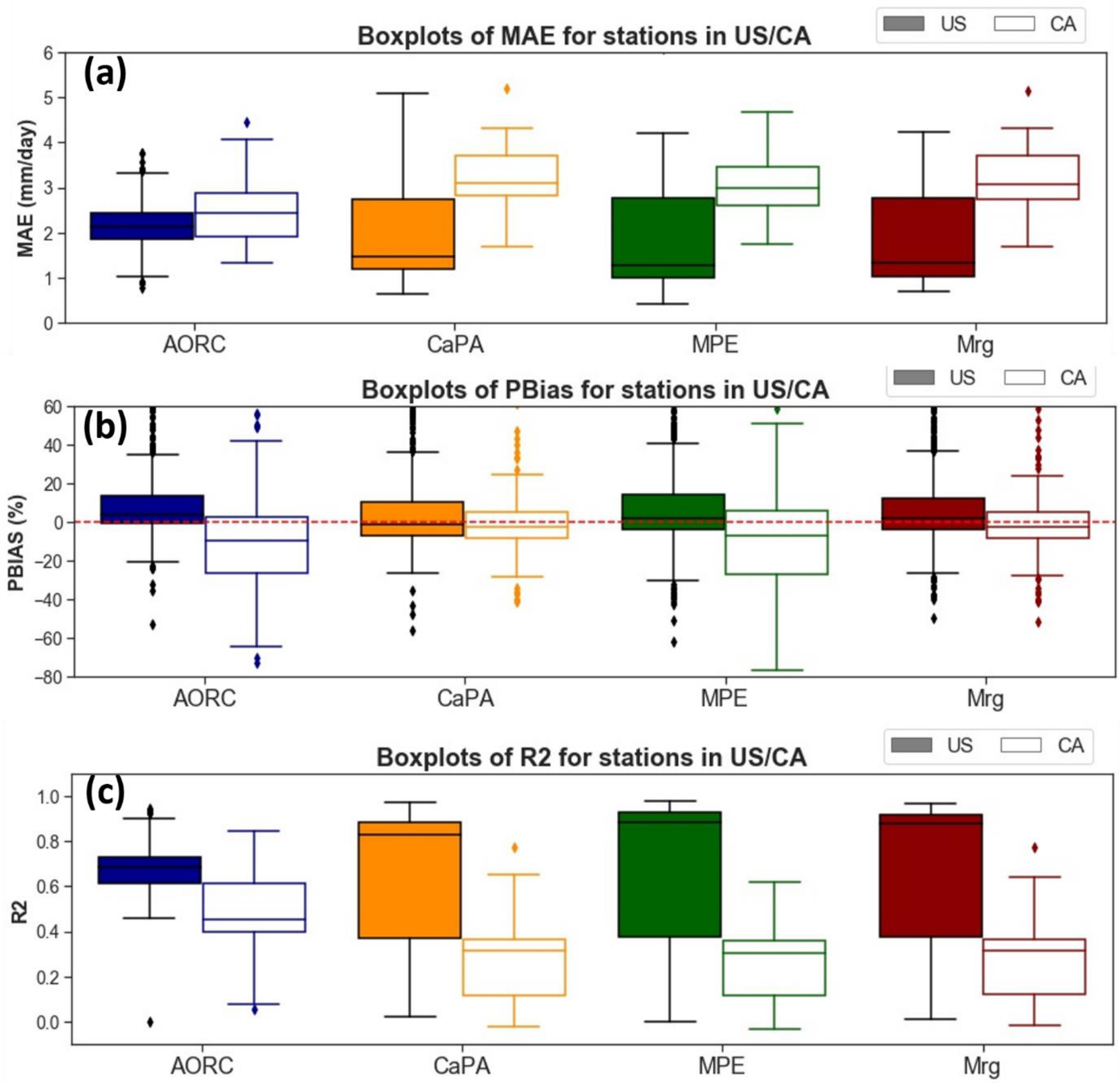

Fig. 4. Boxplots (median, 25th/75th percentiles, whiskers = represent the 1.5 standard deviation above and below the mean of the data) of MAE (**a**), PBias (**b**) and R^2 (**c**) of 2010 – 2019 daily precipitation at rain gauge stations located in U.S. (colored boxes) and the Canadian (white boxes) part of the Great Lakes basin.

4.1.2. Seasonal and monthly variation

In order to evaluate seasonal variations within gridded datasets, we first compare the accumulated monthly precipitation of each data product to the values aggregated from GHCN-D data. As shown in Fig. 5, all gridded datasets generally exhibit a seasonal pattern similar to that observed in the GHCN-D. Specifically, a lower magnitude of precipitation is observed from November to March relative to the April – October period. June is the wettest month of the year, regardless which dataset being considered, while other summer months (July to September) have a generally lower precipitation amount relative to other wet months (i.e. April, May and October).

To explore the seasonal variation of the performance of each gridded product, we also analyze MAE and PBias relative to the GHCN-D for each month separately (Fig. 6). As shown in Fig. 6a, the first and third quartiles of MAE across different datasets and gauge stations are approximately ranged from 1 to 2 mm/day for “dry” months (Nov. – Mar.); and 1 to 4 mm/day for “wet” months (Apr. – Oct.). According to the results of PBias (Fig. 6b), all of the analyzed products tend to get larger positive bias in cold months (Nov. – Mar.), with higher variations among stations. Conversely, all products are more likely to have negative bias in warmer months (Apr. – Oct.), with lower spatial variations. Among these products, CaPA reveals the most significant seasonal variation. CaPA largely overestimates the GHCN-D data during Nov. – Mar.; the first and third quartiles of PBias during these months can be up to +10 % and +50 %, respectively. While CaPA significantly underestimates in summer (Jun. – Aug.), the first and third quartiles of PBias are approximately at 0 and -20 %, respectively. July is the month that CaPA records the largest negative bias compared to GHCN-D. Furthermore, the seasonal variation of AORC is less significant than other gridded products, with the median values of PBias varied from 0 to +10 % over different months, implying a better consistency at temporal scales.

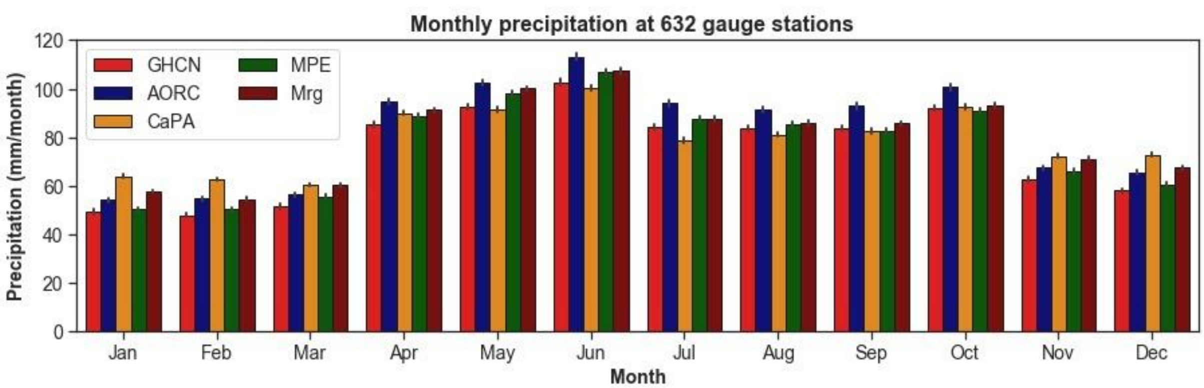

Fig. 5. Barplots of monthly mean precipitation at 632 rain gauge stations in the Great Lakes basin over 2010 – 2019.

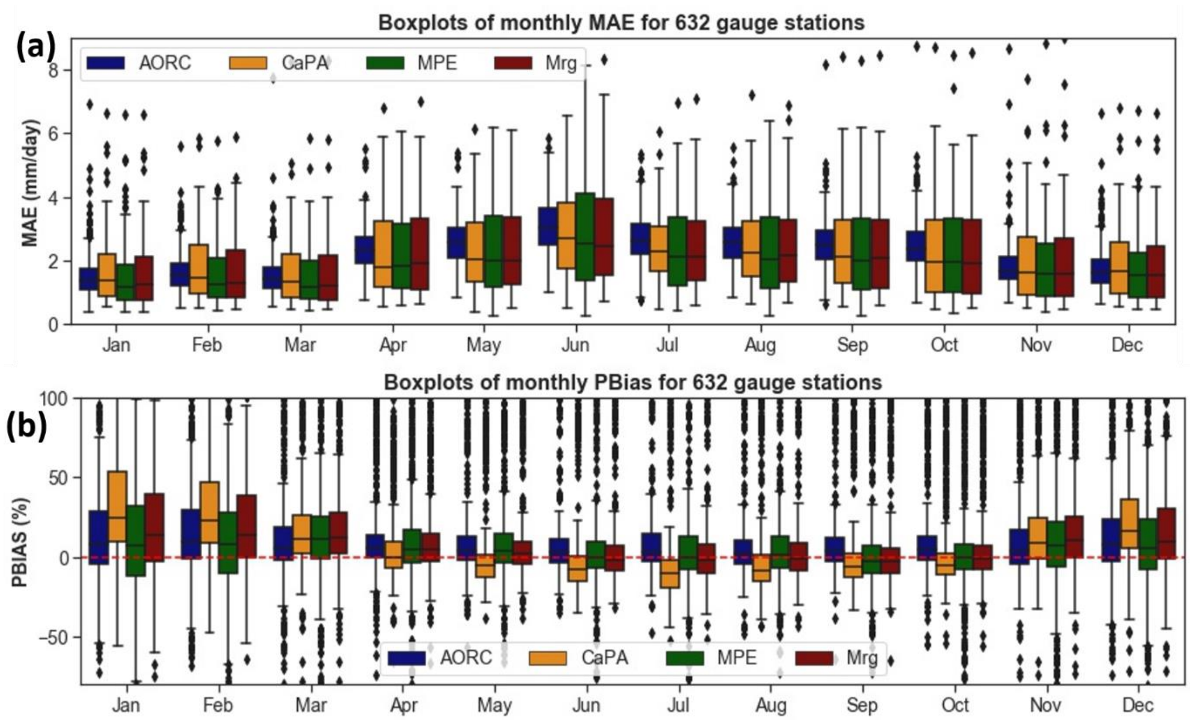

Fig. 6. Boxplots (median, 25th/75th percentiles, whiskers= represent the 1.5 standard deviation above and below the mean of the data) of MAE (**a**), and PBias (**b**) for each month at 632 rain gauge stations in the Great Lakes basin over 2010 – 2019.

4.2. Comparison of averaged overland/overlake precipitation at sub-basin scales

To explore differences in product performance for overland and overlake regions, we compare annual, monthly, and daily accumulated precipitation for each sub-basin. Annual accumulated precipitation for overland and overlake areas of Michigan-Huron are described in Fig. 7 as an example, results for other sub-basins are presented in Fig. S3-5. In general, an increasing trend of annual precipitation from 2010 to 2019 is observed for all five precipitation datasets. This increasing trend is more pronounced in the MPE product, resulting in the highest values in 2018 and 2019. Conversely, the AORC contains the least rise in precipitation overland and CaPA the least amount of rise overlake for the Michigan-Huron sub-basin (Fig. 7), though trends are similar across other sub-basins (Fig. S3-5). In general, the interannual change in precipitation is consistent across products with the exception of the AORC, which has divergent years in Michigan-Huron, Erie, and Ontario sub-basins. Overall, the annual dynamic patterns of different datasets are similar for overland and overlake precipitation, suggesting that there is no substantial difference between overland and overlake precipitation estimates for all the evaluated products on an annual basis.

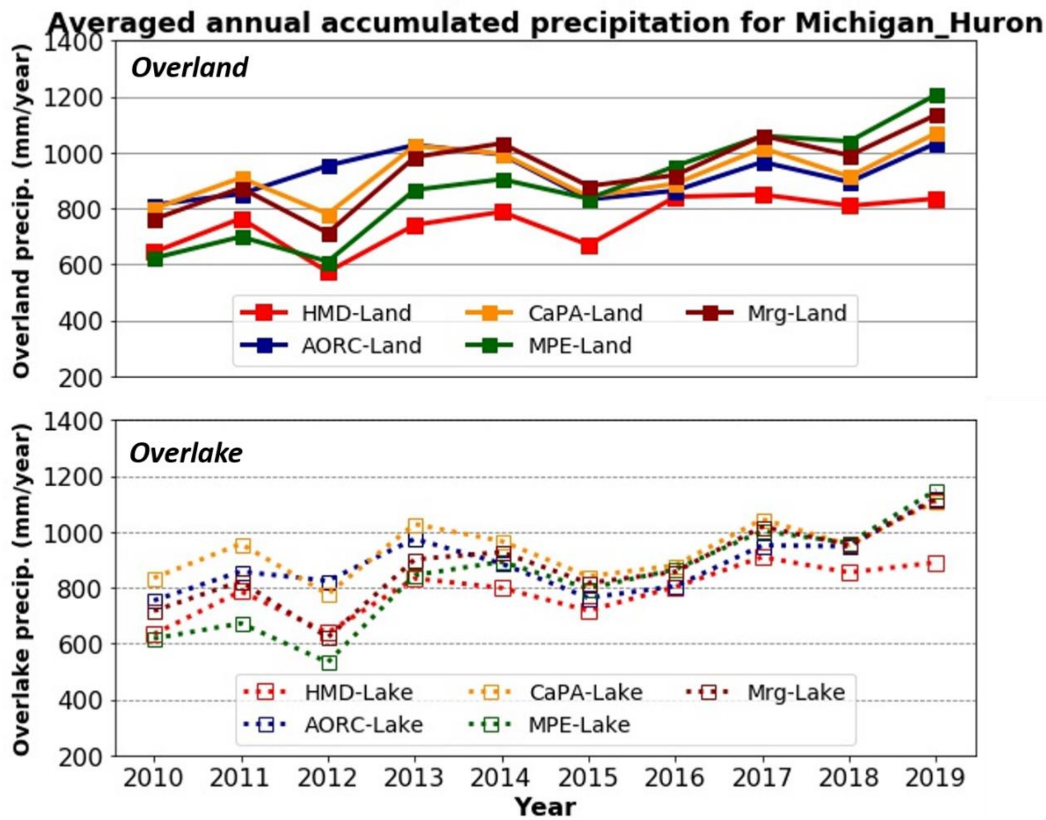

Fig. 7. Annual averaged overland/overlake precipitation for Michigan-Huron sub-basin. Solid lines represent overland precipitation (upper subplot), and dashed lines indicate overlake precipitation (lower subplot).

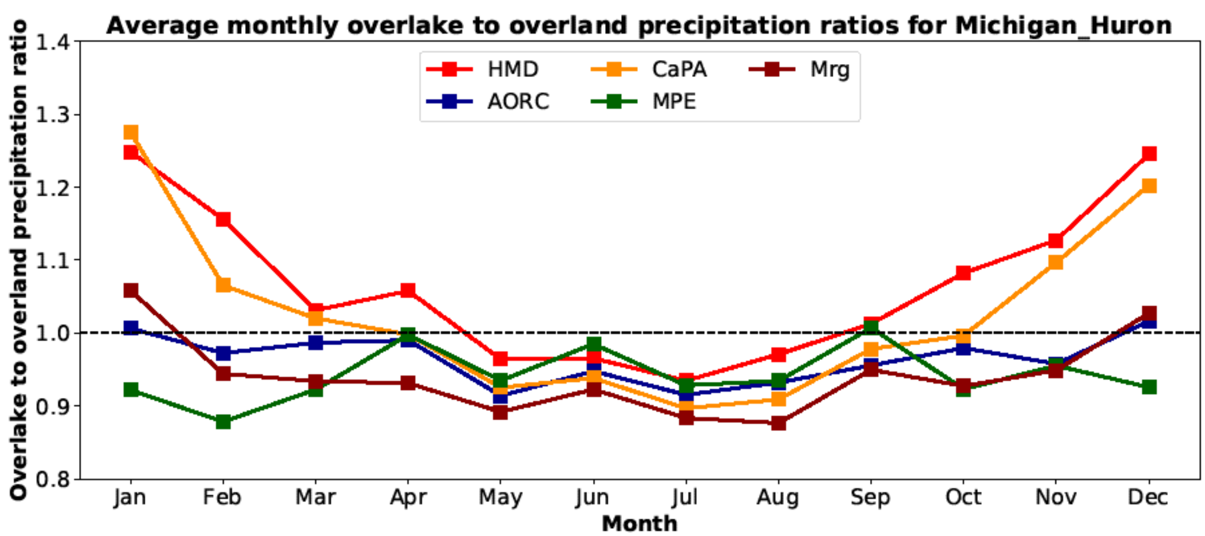

Fig. 8. Averaged monthly accumulated overlake-to-overland precipitation ratio for Michigan-Huron sub-basin.

Due to lake-atmospheric feedbacks, overlake precipitation across each of the Great Lakes is generally higher in cold months relative to the that of warm months (Holman et al., 2012). For testing the suitability of different datasets for this assumption, averaged monthly accumulated overlake-to-overland precipitation ratio (R_p) is compared in Fig. 8. The Michigan-Huron sub-basin is presented as an example, and results from the other sub-basins are shown in Fig. S6-8. Comparisons of R_p values from different datasets indicate that CaPA reflect seasonal dynamics that are not only stronger than any other gridded datasets, but that are also much closer to the gauge-based estimations and what we might expect from a large lake-dominated system (Holman et al., 2012). More specifically, we find that the R_p values for Michigan-Huron sub-basin from CaPA and GLM-HMD range from a maximum of 1.3 in January to a minimum of 0.9 in July, and are less than 1.0 from April to October. Whereas, the difference between winter and summer R_p values from MPE, AORC and Merged datasets are less than 0.2, and the seasonal dynamics cannot be clearly observed especially for MPE and AORC products.

Table. 3 shows the degree of agreement of daily precipitation between gridded products and GLM-HMD at overland and overlake portions of each sub-basin. From the MAE and R2 values, AORC fits better with GLM-HMD for most overlake and overland areas of different sub-basins. This finding is consistent to the previous findings (section 4.1). Because fewer Canadian rain gauge stations are available, Canadian stations tend to carry much more weight in the interpolation procedures used by the GLM-HMD data at sub-basin scale. Moreover, as AORC agrees better with Canadian observations (as presented in Fig. 4), a better performance is observed at the sub-basin scale. In addition, the MAE values show that better agreements between gridded products and GLM-HMD can be found for Superior and Michigan-Huron than Erie and Ontario. Nevertheless, in spite of the modest MAE values (1 – 2 mm/day), the PBias values reveal considerable bias between products and GLM-HMD (up to 70%), particularly for the overland areas. These results suggest that systematic positive bias may exist between gridded products and GLM-HMD.

Table 3. Summary of MAE (a), PBias (b), and R^2 (c) for comparing daily averaged precipitation at overland/overlake areas of each sub-basin over 2010 - 2019. Bolded values indicate the best performance for each area. Gridded products are evaluated against GLM-HMD dataset.

(a) MAE	Overland				Overlake			
(mm/day)	AORC	CaPA	MPE	Merged	AORC	CaPA	MPE	Merged
Superior	**1.65**	1.88	1.78	1.85	**1.70**	2.27	1.90	2.04
Michigan-Huron	1.81	1.97	**1.79**	1.98	**1.71**	2.06	1.77	1.89
Erie	**2.03**	2.29	2.29	2.32	**2.49**	3.09	3.09	3.06
Ontario	**2.07**	2.59	2.57	2.44	**1.93**	2.58	2.42	2.46

(b) PBIAS	Overland				Overlake			
(%)	AORC	CaPA	MPE	Merged	AORC	CaPA	MPE	Merged
Superior	17.88	17.24	**9.42**	16.02	11.71	24.83	**-1.80**	12.69
Michigan-Huron	70.28	65.53	**57.04**	70.12	41.99	45.59	**28.84**	37.86
Erie	68.77	**56.44**	62.91	64.54	-3.22	**0.28**	2.91	1.33
Ontario	**26.49**	34.54	30.69	28.21	**6.84**	18.90	8.95	14.74

(c) R^2	Overland				Overlake			
(-)	AORC	CaPA	MPE	Merged	AORC	CaPA	MPE	Merged
Superior	**0.57**	0.51	0.54	0.52	**0.60**	0.43	0.48	0.46
Michigan-Huron	**0.62**	0.48	0.57	0.50	**0.64**	0.51	0.57	0.54
Erie	**0.63**	0.44	0.47	0.46	**0.43**	0.29	0.31	0.30
Ontario	**0.54**	0.41	0.40	0.43	**0.57**	0.43	0.45	0.45

4.3. Assessing the reliability of overlake precipitation with L2SWBM

In order to test the reliability of different overlake precipitation data to reconcile lake water balance dynamics over long time periods, we then compared overlake precipitation estimates of different products with the posterior distributions of L2SWBM (Fig. 9). To explicitly demonstrate the accuracy of each data product in the context of lake water balance, the percentage of monthly precipitation records falling within the 95% credible intervals of L2SWBM outputs are illustrated in Fig. 9 as well.

The results clearly indicate that across all lakes, AORC and MPE are the most reliable overlake precipitation datasets for reconciling lake water balance, with up to 80 % of monthly precipitation records falling within the 95% credible intervals of L2SWBM posterior distributions. While AORC is slightly better for Lake Superior and Michigan-Huron, and MPE is lightly better for Lake Erie and Ontario. Overall, the performance for Lake Ontario is inferior to other lakes, as less than of 60% of all monthly precipitation records is within the 95% credible intervals. Although GLM-HMD is used as the input for L2SWBM for both prior and analysis periods, AORC and MPE fit better with the L2SWBM outputs, suggesting closer estimates to the actual overlake precipitation. Whereas no significant enhancement can be observed with CaPA and Merged, meaning these datasets are not better than gauge-based overlake precipitation estimates. According to the timeseries plots, we can find that MPE is below the 95% credible intervals for 2010 – 2012, particularly for Lake Superior and Michigan-Huron; and AORC is above the 95% credible intervals for 2012 – 2013, more obviously for Lake Michigan-Huron and Erie. These findings are in accordance with our previous results shown in Fig. 7.

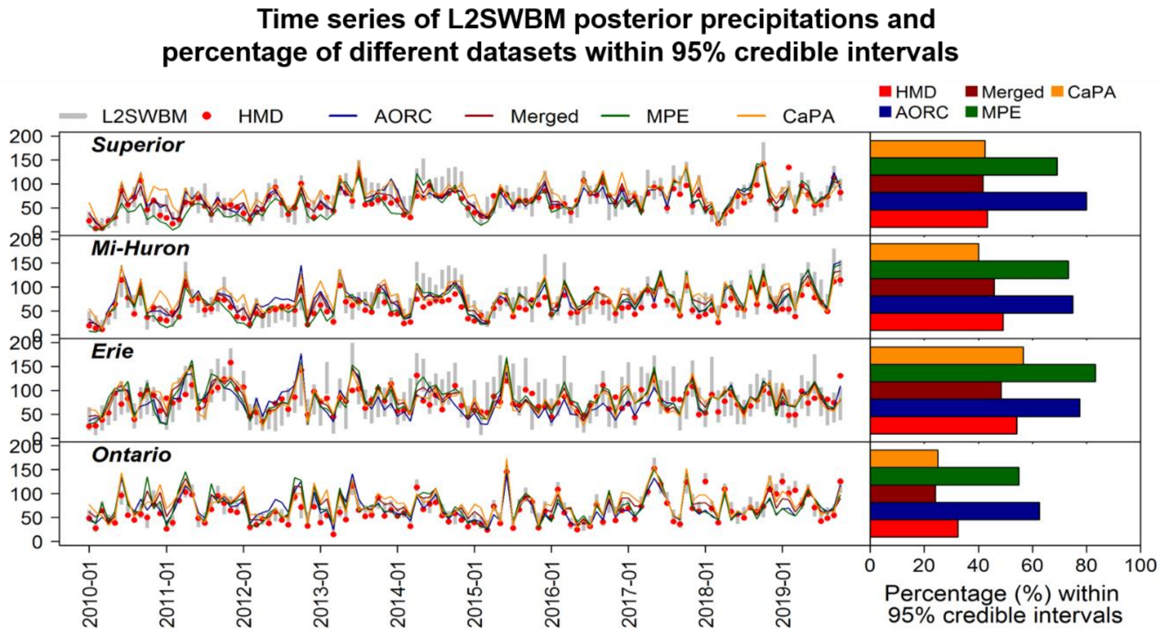

Fig. 9. Time series of monthly precipitation at the four lakes of L2SWBM posterior distributions (as 95% credible intervals, grey bars), GLM-HMD (red points), AORC (blue lines), CaPA (orange lines), MPE (green lines), and Merged (brown lines). Barplots represent the percentage of monthly precipitation from each product that fall within the 95% credible intervals of L2SWBM posterior distributions.

5. Summary and discussion

Accurate precipitation estimates over international basins and large bodies of water can be a challenge to weather and hydrologic forecasting. The North American Great Lakes basin combines these issues and poses a unique challenge for hydrologic science. In this study, we evaluate four leading gridded precipitation data sets used for hydrologic model calibration and simulation to understand differences across many spatial and temporal scales. For the period 2010-2019, the AORC, MPE, CaPA, and Merged MPE-CaPA products are compared with 632 gauged observations (GHCN-D), analyzed for overland and overlake sub-basins, and evaluated against a Bayesian statistical analysis framework (L2SWBM) used for water budget accounting.

Comparisons with gauge observations indicate a generally good agreement between all evaluated products and the GHCN-D dataset, however, a large difference is found between U.S. and Canadian precipitation accuracy. In addition, based on comparisons at sub-basin scales and the analysis with L2SWBM model, the ability of different gridded precipitation products to capture overlake precipitation, and the effects of large lakes on overlying atmospheric stability, varies dramatically between products. In all, these results reveal the discrepancies between precipitation options for international basins and those containing large bodies of water, and provide insight into potential downstream impacts to hydrologic model development and prediction.

5.1. Poor agreement with GHCN-D data of Canadian stations

From Fig. 5-6, a sudden change of performance of gridded datasets can be observed between U.S. and Canadian regions of the Great Lakes basin. In contrast to the relatively higher performance at U.S. stations, gridded datasets poorly agreed with precipitation observations at Canadian stations. This phenomenon may be because of the differences in data collection and adjustment protocols of GHCN-D for rain gauge stations between the two countries.

In GHCN-D, the U.S. collection contains daily data from a dozens of separate datasets archived by NOAA. Precipitation data for U.S. rain gauge stations are corrected for the resulting archive-quality data. However, for Canadian rain gauge stations, GHCN-D does not contain adjustments for biases resulting from historical changes in instrumentation and observing practices, thus systematic bias for certain stations might be important for applications of GHCN-D. For example, GHCN-D Canadian stations record daily total precipitation as the sum of daily rainfall and daily Snow Water Equivalent (SWE) from snowfall. Daily ruler measurements of snowfall are automatically converted to daily SWE using a constant conversion ratio that assumes 10 mm of snowfall to be equivalent to 1 mm of SWE. On the contrary, SWE conversion ratio varies substantially (ranges from AAA to BBB) across all U.S. stations archived within the GHCN-D. The constant SWE conversion ratio across Canada was found to potentially underestimate up to 15 % of the total precipitation for most stations of Southern Canada (Wang et al., 2017), leading to a remarkably low values of GLM-HMD (derived from GHCN-D precipitation) at sub-basin scales relative to other gridded products as shown in Fig. 7.

Another reason underlying the poor agreement between gridded precipitation and GHCN-D records is the potentially inaccurate geographical positions (longitudes, latitudes) of rain gauges reported in their metadata. Meta-data of U.S. stations in GHCN-D are regularly updated and users have access to the latest information. Whereas, this information should be further analyzed for GHCN-D Canadian stations. For example, in Great Lakes basin, a total of 373 GHCN-D Canadian stations contain data after 2010, 61 duplicated Lat-Long coordinates can be identified for these stations. Moreover, we compared the location of these 373 stations with 3346 ECCC rain gauge stations extracted from the latest version (2016) of the Adjusted Daily Rainfall and Snowfall (AdjDlyRS) datasets (https://open.canada.ca/data/en/dataset/d8616c52-a812-44ad-8754-7bcc0d8de305), only 4 among them have identical Lat-Long coordinates. This phenomenon suggests that the location of Canadian stations in GHCN-D might should be checked and updated.

5.2. Precipitation estimates over large lakes

According to historical measurements and studies (Holman et al., 2012), realistic basin-wide precipitation estimates could have higher relative overlake precipitation in cold months, and lower overlake precipitation in warm months. A common hypothesis for this phenomenon is that the relatively cool air over the lakes (water temperature lower than air temperature, resulting relatively lower near surface air temperature) during warm season will inhibit the growth of convective storms resulting in less rain over the lakes; conversely, the relatively warm lake during the winter (water temperature higher than air temperature, resulting relatively higher near surface air temperature) will initiate convective instability through the flux of heat and moisture into the cold air advecting over the lakes.

However, this phenomenon is rarely considered by existing hydro-meteorological studies at lake and sub-basin scales. For sub-basins which feature large freshwater surfaces, these processes may significantly impact the prediction of lake water budget and water balance at sub-basin scales. Since one-third of the Great Lakes basin water budget is derived from precipitation falling directly on the lake surface (Table 1), particularly for Lake Superior (39 %) and Lake Michigan-Huron (32 %), this study preliminarily addresses this issue by analyzing the overland-overlake precipitation patterns of different products.

As illustrated in Fig. 8, and S6-8, the overland-overlake seasonal variation is more significant for Lake Superior and Michigan-Huron than Lake Erie and Ontario. This order follows the rank of lake surface areas of sub-basin, indicating that stronger seasonal variations can be observed for sub-basins with larger lake surface proportions. On the other hand, it can be noted that CaPA and AORC reflect stronger seasonal dynamics than MPE, particularly for Lake Superior and Michigan-Huron. Since no rain gauge measurement is available over the water surfaces, there is no “ground truth” for calibrating satellite images on these overlake areas. Products only derived from rain gauge observations and satellite data might not well capture these overland/overlake seasonal variations. Results in section 4.2 reaffirm this assumption that CaPA and AORC, which are reanalysis products relying on meteorological models, can better represent the seasonal variations of the ratio of overland to overlake precipitation. Whereas, these variations are less represented by MPE and GLM-HMD, which are satellite-gauge blended and gauge based datasets.

On the other hand, results from analysis with L2SWBM suggest that AORC and MPE could have closer estimates to the realistic overlake precipitation based on a long-term water balance aspect. That underscores AORC for both correctly estimating overlake precipitation, and properly representing differences between overland and overlake precipitation.

5.3. Applications of precipitation products for hydrological modeling of the Great Lakes basin

For the four evaluated gridded datasets, CaPA and MPE are available for nowcast and hindcast simulations, while AORC and Merged can be only used for hindcast simulations. For nowcasting applications, according to results shown in Table 3, MPE is better than CaPA for sub-basins of Lake Superior, Lake Michigan-Huron and Lake Ontario, while CaPA performs better for the sub-basin of Lake Erie. As for hindcasting applications, since the performance of AORC is less dispersed at spatial and temporal scales (Fig. 2, Fig. 6), and AORC fits better with GLM-HMD for most sub-basins (Table 3), AORC could be the appropriate choice of climate forcing for hydrological modeling for the entire Great Lakes watershed. Whereas, for some specific sub-basins (i.e. overland areas of Michigan-Huron), other products (i.e. MPE) might be preferred. In addition, it is informative to notice that in Table 3, overland and overlake consistently favors AORC for MAE (under 2 mm/day) and R2 (above 0.6), however, important PBias values can be noted (20 % - 70%). This result implies that consistent bias may exist with AORC dataset and should be considered for model calibration and uncertainty analysis.

On the other hand, hydrological processes across overland areas and in the lake are often simulated by different models. Land surface hydrological models commonly use simple routing schemes for computing in-lake processes (i.e. NWM (Lahmers et al., 2019), SWAT (Arnold et al., 1998)); while lake models usually do not include land surface processes (i.e. FVCOM (Chen et al., 2003), Delft3D (Deltares, 2016)). Therefore, when performing land surface modeling, the correctness of precipitation products on overland areas is important; where gridded products can be evaluated by comparing with the ground observations. On the contrary, for lake water modeling, appropriate estimates of overlake precipitation are crucial. Considering seasonal variations of the ratio of overland to overlake precipitation, and using L2SWBM to infer reliable overlake precipitation range, can help modeler to select proper products for overlake precipitation. For large lake basins like the Great Lakes, a reasonable approach would be to couple land surface hydrology models (e.g. WRF-Hydro), with sophisticated lake hydrodynamic models (i.e. FVCOM), in which each approach could benefit from the most appropriate precipitation forcing.

Acknowledgments

The study was supported by funding awarded to the Cooperative Institute for Great Lakes Research (CIGLR) through the U.S. National Oceanic and Atmospheric Administration (NOAA) Cooperative Agreement with the University of Michigan. The CIGLR contribution number is 1150. The authors would like to thank Tim Hunter of NOAA for providing GLM-HMD data and valuable suggestions.

Data availability

All the datasets that support the analysis of this study are publically available:

- AORC is provided by the National Weather Service of NOAA, archived online at https://hydrology.nws.noaa.gov/aorc-historic/.
- CaPA, MPE, and Merged CaPA-MPE can be accessed from the website of Midwestern Regional Climate Center: https://mrcc.illinois.edu/cliwatch/northAmerPcpn/getArchive.jsp.
- GLM-HMD is provided by the Office of Water Prediction of NOAA, datasets are described in this paper: (Hunter et al., 2015), and can be obtained from: https://www.glerl.noaa.gov/ftp/publications/tech_reports/glerl-083/UpdatedFiles/daily/.
- GHCN daily data is available online: https://www.ncdc.noaa.gov/ghcnd-data-access.
- The reference for AdjDlyRS is: (Wang et al., 2017), the data can be achieved from: https://open.canada.ca/data/en/dataset/d8616c52-a812-44ad-8754-7bcc0d8de305.

References

Alcantara, M. A. S., Kesler, C., Stealey, M. J., Nelson, E. J., Ames, D. P., & Jones, N. L. (2018). Cyberinfrastructure and Web Apps for Managing and Disseminating the National Water Model.

JAWRA Journal of the American Water Resources Association, *54*(4), 859–871. https://doi.org/10.1111/1752-1688.12608

Arnold, J. G., Srinivasan, R., Muttiah, R. S., & Williams, J. R. (1998). Large area hydrologic modeling and assessment Part I: Model development. *Journal of the American Water Resources Association*, *34*(1), 73–89. https://doi.org/10.1111/j.1752-1688.1998.tb05961.x

Chen, C., Liu, H., & Beardsley, R. C. (2003). An Unstructured Grid, Finite-Volume, Three-Dimensional, Primitive Equations Ocean Model: Application to Coastal Ocean and Estuaries. *Journal of Atmospheric and Oceanic Technology*, *20*(1), 159–186. https://doi.org/10.1175/1520-0426(2003)020<0159:AUGFVT>2.0.CO;2

Cole, S. J., & Moore, R. J. (2008). Hydrological modelling using raingauge- and radar-based estimators of areal rainfall. *Journal of Hydrology*, *358*(3), 159–181. https://doi.org/10.1016/j.jhydrol.2008.05.025

Côté, J., Gravel, S., Méthot, A., Patoine, A., Roch, M., & Staniforth, A. (1998). The Operational CMC–MRB Global Environmental Multiscale (GEM) Model. Part I: Design Considerations and Formulation. *Monthly Weather Review*, *126*(6), 1373–1395. https://doi.org/10.1175/1520-0493(1998)126<1373:TOCMGE>2.0.CO;2

Côté, J., Desmarais, J.-G., Gravel, S., Méthot, A., Patoine, A., Roch, M., & Staniforth, A. (1998). The Operational CMC–MRB Global Environmental Multiscale (GEM) Model. Part II: Results. *Monthly Weather Review*, *126*(6), 1397–1418. https://doi.org/10.1175/1520-0493(1998)126<1397:TOCMGE>2.0.CO;2

Croley, T. E., & Hartmann, H. C. (1985). Resolving Thiessen polygons. *Journal of Hydrology*, *76*(3), 363–379. https://doi.org/10.1016/0022-1694(85)90143-X

Deltares. (2016). 3D/2D modelling suite for integral water solutions - Delft3D User Manual. Retrieved from https://www.deltares.nl

Do, H. X., Smith, J. P., Fry, L. M., & Gronewold, A. D. (2020). Seventy-year long record of monthly water balance estimates for Earth's largest lake system. *Scientific Data*, *7*(1), 276. https://doi.org/10.1038/s41597-020-00613-z

Duan, Z., Liu, J., Tuo, Y., Chiogna, G., & Disse, M. (2016). Evaluation of eight high spatial resolution gridded precipitation products in Adige Basin (Italy) at multiple temporal and spatial scales. *Science of The Total Environment*, *573*, 1536–1553. https://doi.org/10.1016/j.scitotenv.2016.08.213

Fortin, V., Roy, G., Donaldson, N., & Mahidjiba, A. (2015). Assimilation of radar quantitative precipitation estimations in the Canadian Precipitation Analysis (CaPA). *Journal of Hydrology*, *531*, 296–307. https://doi.org/10.1016/j.jhydrol.2015.08.003

Gronewold, A. D., Bruxer, J., Durnford, D., Smith, J. P., Clites, A. H., Seglenieks, F., et al. (2016). Hydrological drivers of record-setting water level rise on Earth's largest lake system. *Water Resources Research*, *52*(5), 4026–4042. https://doi.org/10.1002/2015WR018209

Gronewold, Andrew D., Fortin, V., Lofgren, B., Clites, A., Stow, C. A., & Quinn, F. (2013). Coasts, water levels, and climate change: A Great Lakes perspective. *Climatic Change*, *120*(4), 697–711. https://doi.org/10.1007/s10584-013-0840-2

Gronewold, Andrew D., Fortin, V., Caldwell, R., & Noel, J. (2018). Resolving Hydrometeorological Data Discontinuities along an International Border. *Bulletin of the American Meteorological Society*, *99*(5), 899–910. https://doi.org/10.1175/BAMS-D-16-0060.1

Gronewold, Andrew D., Smith, J. P., Read, L. K., & Crooks, J. L. (2020). Reconciling the water balance of large lake systems. *Advances in Water Resources*, *137*, 103505. https://doi.org/10.1016/j.advwatres.2020.103505

Henn, B., Newman, A. J., Livneh, B., Daly, C., & Lundquist, J. D. (2018). An assessment of differences in gridded precipitation datasets in complex terrain. *Journal of Hydrology*, *556*, 1205–1219. https://doi.org/10.1016/j.jhydrol.2017.03.008

Hofstra, N., Haylock, M., New, M., Jones, P., & Frei, C. (2008). Comparison of six methods for the interpolation of daily, European climate data. *Journal of Geophysical Research: Atmospheres*, *113*(D21). https://doi.org/10.1029/2008JD010100

Holman, K. D., Gronewold, A., Notaro, M., & Zarrin, A. (2012). Improving historical precipitation estimates over the Lake Superior basin. *Geophysical Research Letters*, *39*(3). https://doi.org/10.1029/2011GL050468

Hunter, T. S., Clites, A. H., Campbell, K. B., & Gronewold, A. D. (2015). Development and application of a North American Great Lakes hydrometeorological database — Part I: Precipitation, evaporation, runoff, and air temperature. *Journal of Great Lakes Research*, *41*(1), 65–77. https://doi.org/10.1016/j.jglr.2014.12.006

Khandu, Awange, J. L., & Forootan, E. (2016). An evaluation of high-resolution gridded precipitation products over Bhutan (1998–2012). *International Journal of Climatology*, *36*(3), 1067–1087. https://doi.org/10.1002/joc.4402

Kidd, C., Bauer, P., Turk, J., Huffman, G. J., Joyce, R., Hsu, K.-L., & Braithwaite, D. (2012). Intercomparison of High-Resolution Precipitation Products over Northwest Europe. *Journal of Hydrometeorology*, *13*(1), 67–83. https://doi.org/10.1175/JHM-D-11-042.1

Kidd, Chris, Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society*, *98*(1), 69–78. https://doi.org/10.1175/BAMS-D-14-00283.1

Kitzmiller, D., Miller, D., Fulton, R., & Ding, F. (2013). Radar and Multisensor Precipitation Estimation Techniques in National Weather Service Hydrologic Operations. *Journal of Hydrologic Engineering*, *18*(2), 133–142. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000523

Kitzmiller, D. H., Wu, W., Zhang, Z., Patrick, N., & Tan, X. (2018). The Analysis of Record for Calibration: A High-Resolution Precipitation and Surface Weather Dataset for the United States. *AGU Fall Meeting Abstracts*, *41*. Retrieved from http://adsabs.harvard.edu/abs/2018AGUFM.H41H..06K

Lahmers, T. M., Gupta, H., Castro, C. L., Gochis, D. J., Yates, D., Dugger, A., et al. (2019). Enhancing the Structure of the WRF-Hydro Hydrologic Model for Semiarid Environments. *Journal of Hydrometeorology*, *20*(4), 691–714. https://doi.org/10.1175/JHM-D-18-0064.1

Lespinas, F., Fortin, V., Roy, G., Rasmussen, P., & Stadnyk, T. (2015). Performance Evaluation of the Canadian Precipitation Analysis (CaPA). *Journal of Hydrometeorology*, *16*(5), 2045–2064. https://doi.org/10.1175/JHM-D-14-0191.1

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., et al. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Scientific Data*, *2*(1), 150042. https://doi.org/10.1038/sdata.2015.42

Maggioni, V., & Massari, C. (2018). On the performance of satellite precipitation products in riverine flood modeling: A review. *Journal of Hydrology*, *558*, 214–224. https://doi.org/10.1016/j.jhydrol.2018.01.039

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., & Houston, T. G. (2012). An Overview of the Global Historical Climatology Network-Daily Database. *Journal of Atmospheric and Oceanic Technology*, *29*(7), 897–910. https://doi.org/10.1175/JTECH-D-11-00103.1

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*(5). https://doi.org/10.1029/2009WR008328

Salio, P., Hobouchian, M. P., García Skabar, Y., & Vila, D. (2015). Evaluation of high-resolution satellite precipitation estimates over southern South America using a dense rain gauge network. *Atmospheric Research*, *163*, 146–161. https://doi.org/10.1016/j.atmosres.2014.11.017

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K.-L. (2018). A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Reviews of Geophysics*, *56*(1), 79–107. https://doi.org/10.1002/2017RG000574

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A. T., et al. (2012). Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, *104–105*, 70–97. https://doi.org/10.1016/j.atmosres.2011.10.021

Wang, X. L., Xu, H., Qian, B., Feng, Y., & Mekis, E. (2017). Adjusted Daily Rainfall and Snowfall Data for Canada. *Atmosphere-Ocean*, *55*(3), 155–168. https://doi.org/10.1080/07055900.2017.1342163

Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., et al. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, *117*(D3). https://doi.org/10.1029/2011JD016048

Xiao, C., Lofgren, B. M., Wang, J., & Chu, P. Y. (2016). Improving the lake scheme within a coupled WRF-lake model in the Laurentian Great Lakes. *Journal of Advances in Modeling Earth Systems*, *8*(4), 1969–1985. https://doi.org/10.1002/2016MS000717

[Water Resources Research]

Supporting Information for

[Evaluation of gridded precipitation datasets over international basins and large lakes]

[Yi Hong[1], Eric J. Anderson[2], Lauren Fry[2], Hong Xuan Do[3,4], James Kesseler[2], Laura Read[5], Arezoo Rafieei Nasab[5], Andrew D. Gronewold[3], Lacey Mason[2]]

[[1] Cooperative Institute for Great Lakes Research, University of Michigan, Ann Arbor, MI, USA

[2] Great Lakes Environmental Research Laboratory, Ann Arbor, MI, USA

[3] School for Environment and Sustainability, University of Michigan, Ann Arbor, MI, USA

[4] Faculty of Environment and Natural Resources, Nong Lam University, Ho Chi Minh City, Vietnam

[5] National Center for Atmospheric Research, Boulder, CO, USAd]

Contents of this file

Figures S1 to S8

Introduction

In addition to Fig. 3, Fig. S1-2 represent maps of MAE and R^2 for different gridded products.

Fig. S3-5 illustrate annual accumulated precipitation for overland and overlake areas of Superior, Erie and Ontario, which complement the Fig. 7 (Michigan-Huron).

Fig. S6-8 describe averaged monthly accumulated overlake-to-overland precipitation ratio ($\boldsymbol{R_p}$) for Superior, Erie, and Ontario. Results for Michigan-Huron are presented in Fig. 8.

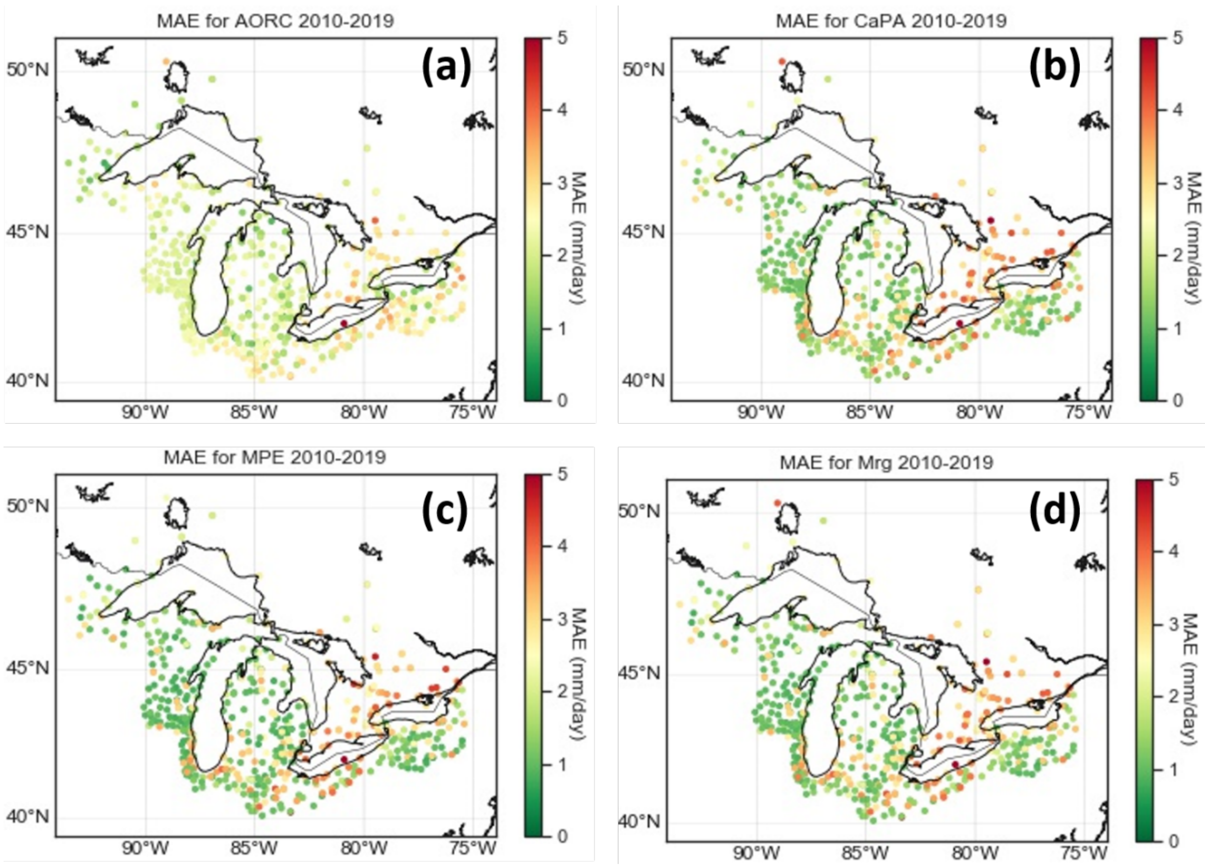

Fig. S1. Maps of MAE for AORC (a), CaPA (b), MPE (c), and Merged (d) at 632 rain gauge stations in the Great Lakes basin over 2010 – 2019.

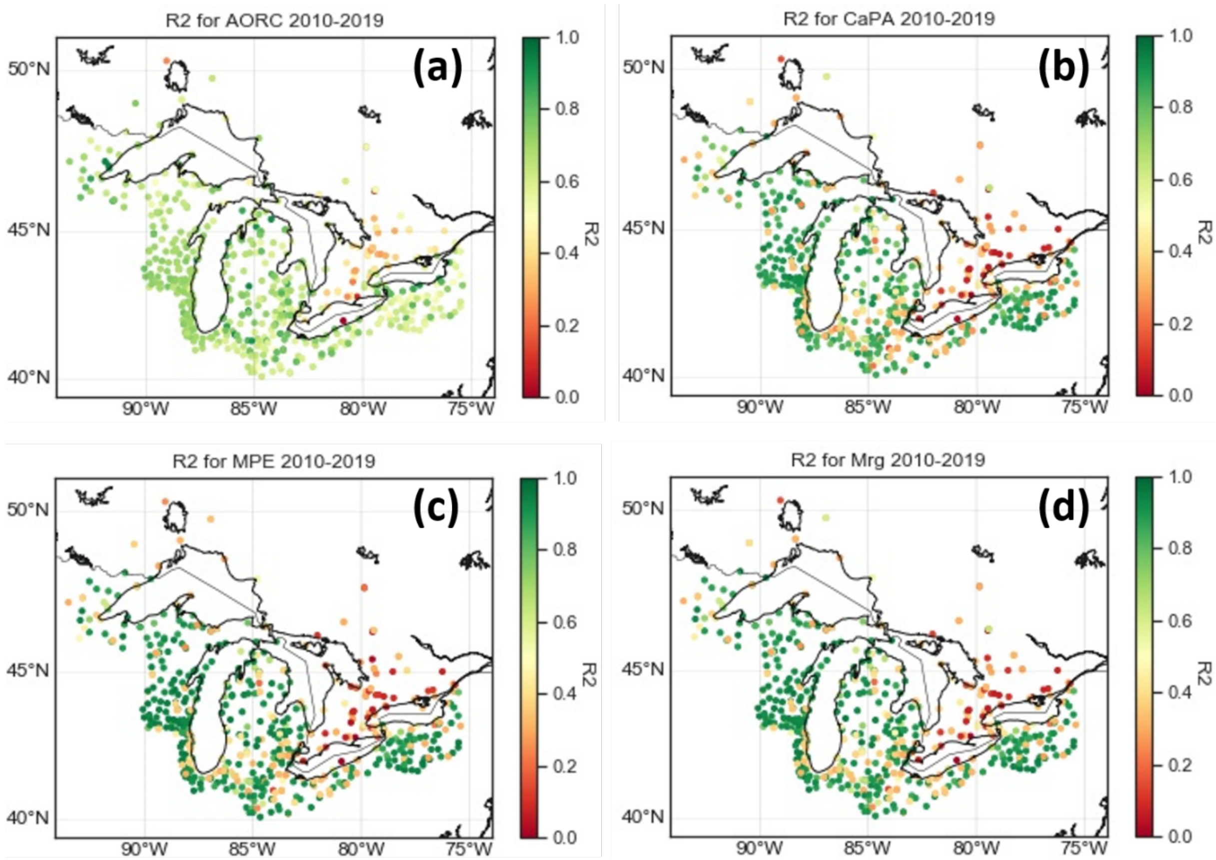

Fig. S2. Maps of R2 for AORC (a), CaPA (b), MPE (c), and Merged (d) at 632 rain gauge stations in the Great Lakes basin over 2010 – 2019.

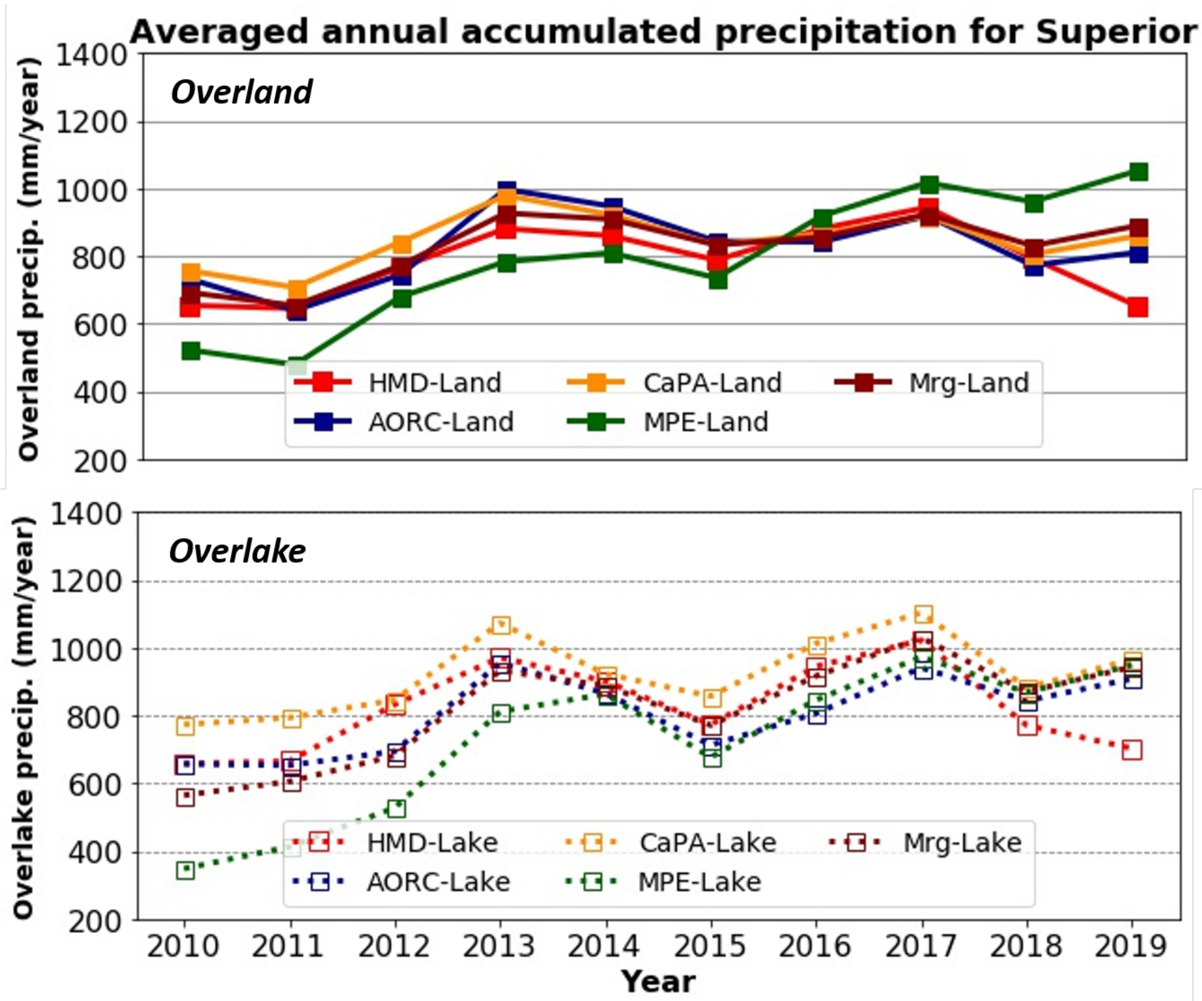

Fig. S3. Annual averaged overland/overlake precipitation for Superior sub-basin. Solid lines represent overland precipitation (upper subplot), and dashed lines indicate overlake precipitation (lower subplot).

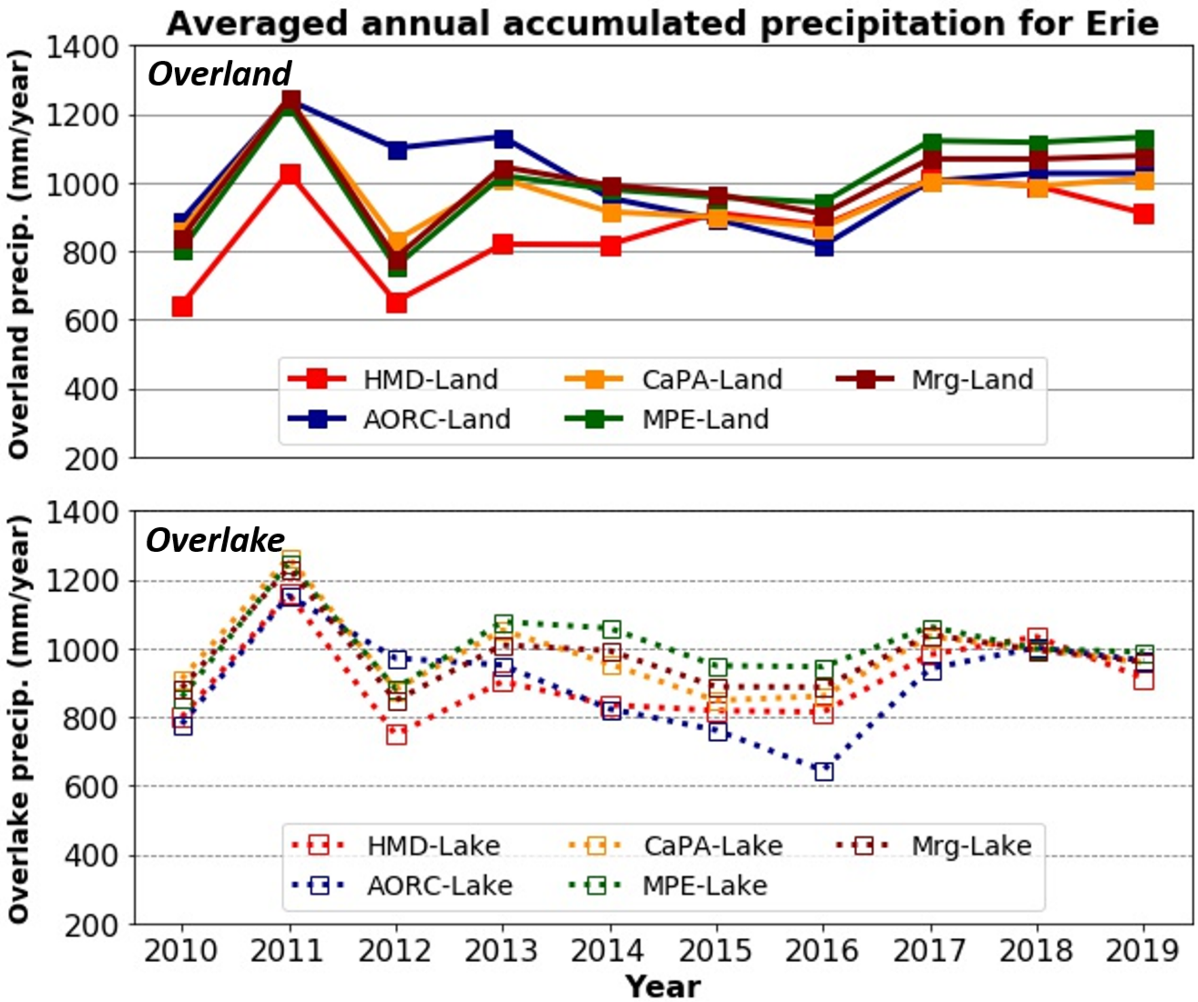

Fig. S4. Annual averaged overland/overlake precipitation for Erie sub-basin. Solid lines represent overland precipitation (upper subplot), and dashed lines indicate overlake precipitation (lower subplot).

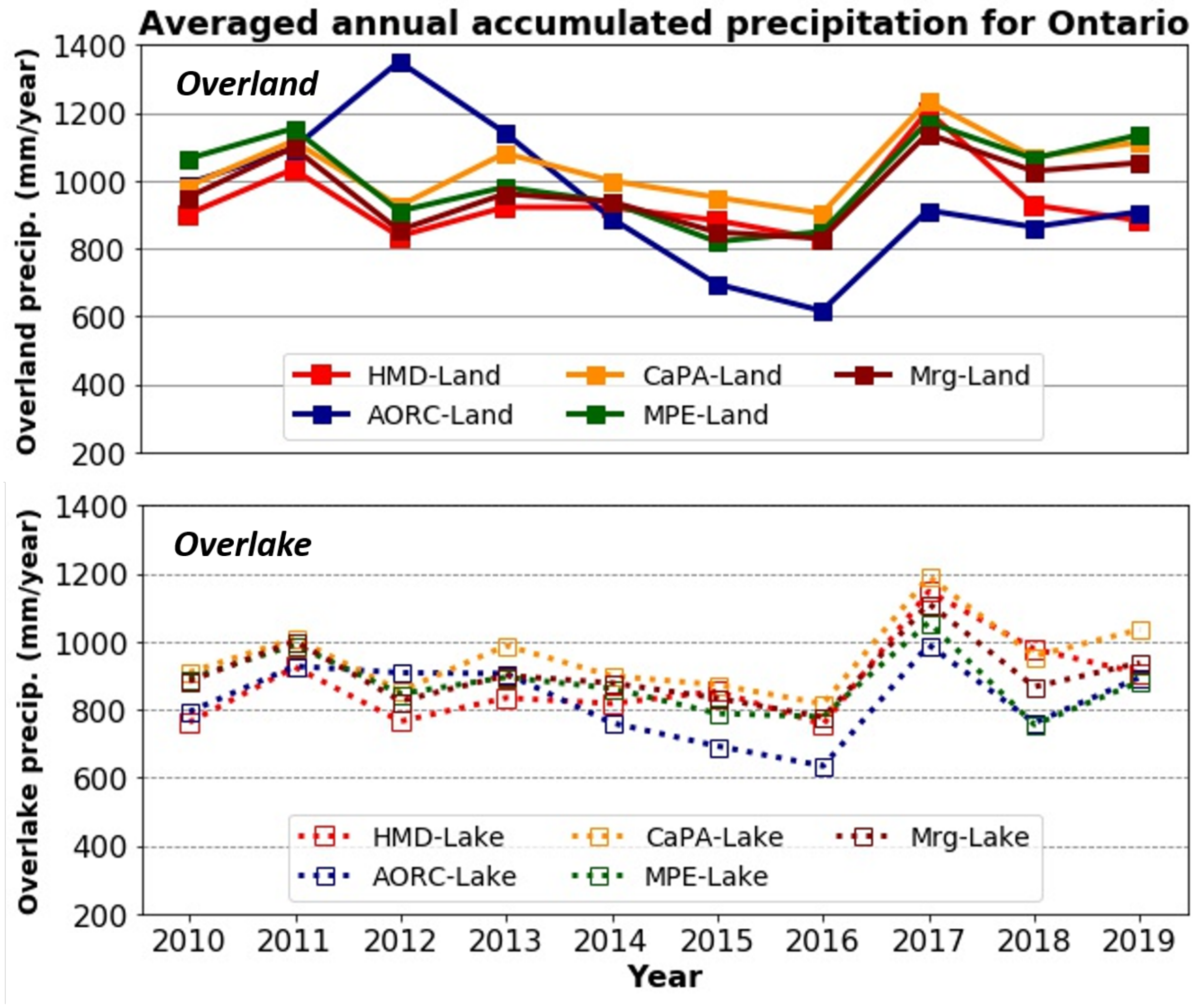

Fig. S5. Annual averaged overland/overlake precipitation for Ontario sub-basin. Solid lines represent overland precipitation (upper subplot), and dashed lines indicate overlake precipitation (lower subplot).

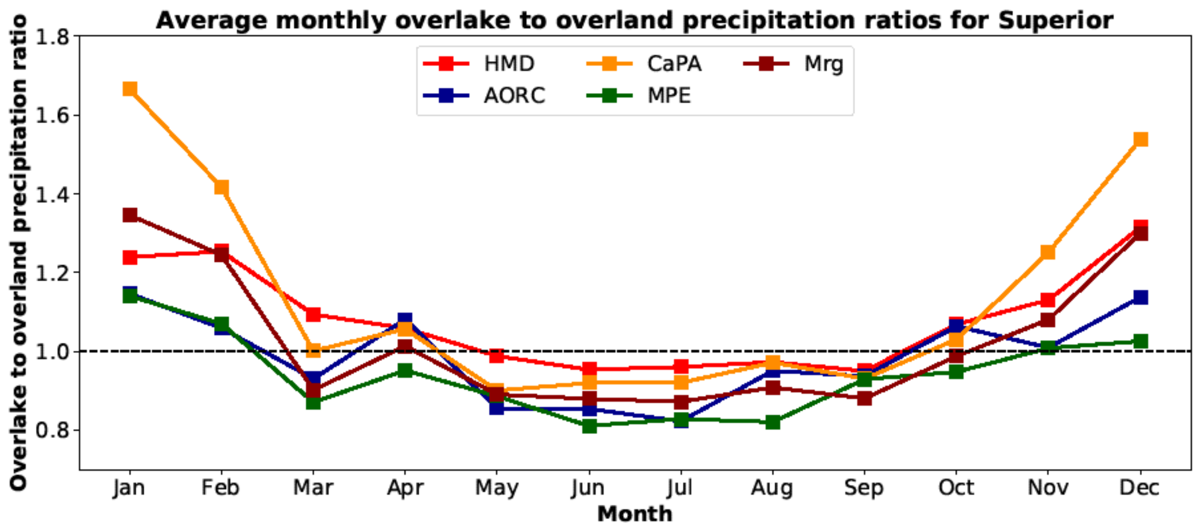

Fig. S6. Averaged monthly accumulated overland-to-overlake precipitation ratio for Superior sub-basin.

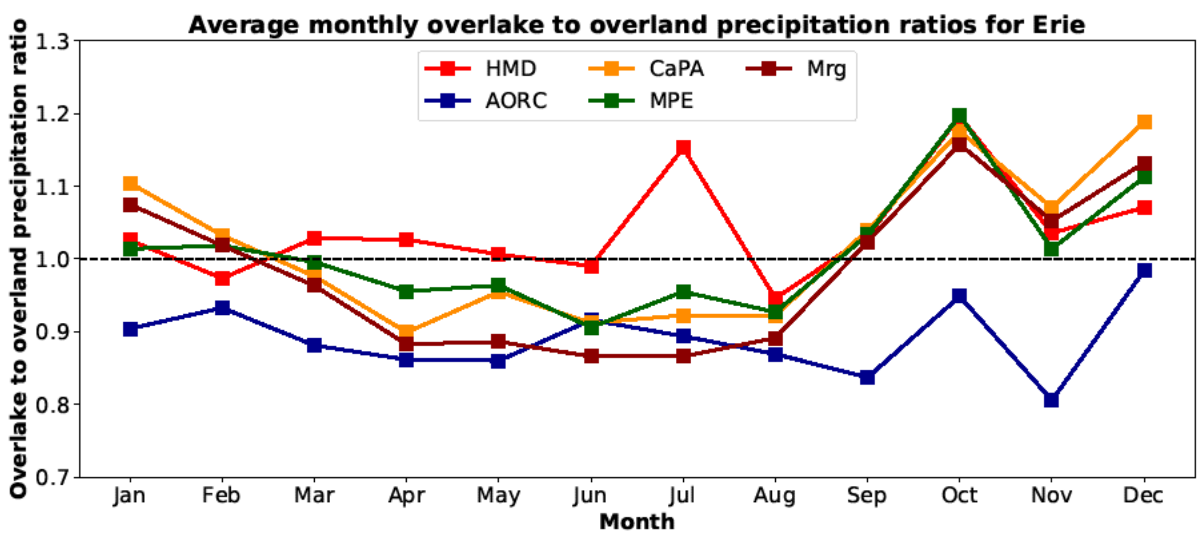

Fig. S7. Averaged monthly accumulated overland-to-overlake precipitation ratio for Lake Erie sub-basin.

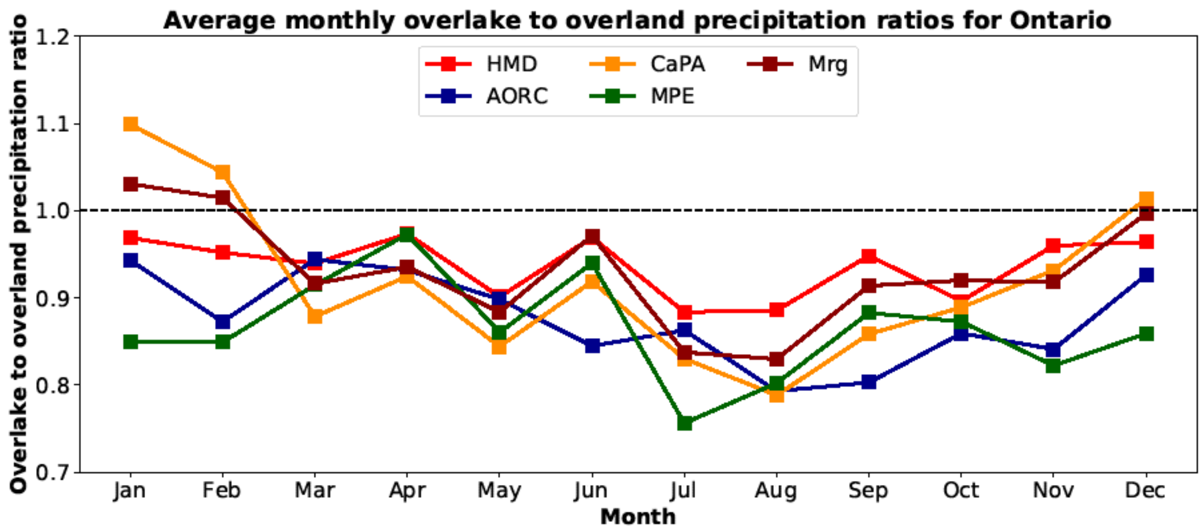

Fig. S8. Averaged monthly accumulated overland-to-overlake precipitation ratio for Lake Ontario sub-basin.