Journal of Great Lakes Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Great Lakes Research

journal homepage: www.elsevier.com/locate/jglr

Evaluation and intercomparison of surface radiative and turbulent heat flux datasets over the Great Lakes region

Kyle Peco [a,*], Arka Mitra [b], Jiali Wang [a], T.C. Chakraborty [c], William J. Pringle [a], John D. Lenters [d], Peter D. Blanken [e], Christopher Spence [f], Pengfei Xue [a,g], Zhao Yang [c], Yun Qian [c], Robert Hetland [c]

[a] Environmental Science Division, Argonne National Laboratory, Lemont, IL 60439, United States
[b] Wageningen University & Research, Wageningen, the Netherlands
[c] Pacific Northwest National Laboratory, Richland, WA 99354, United States
[d] University of Michigan Biological Station, Pellston, MI 49769, United States
[e] University of Colorado, Boulder, CO 80309, United States
[f] Environment and Climate Change Canada, Gatineau K1A 0H3, Canada
[g] Michigan Technological University, Houghton, MI 49931, United States

ARTICLE INFO

Communicated by: Jay Austin

Keywords:
Great Lakes region
Surface radiative flux
Turbulent heat flux
Reanalysis
Satellites
Data validation

ABSTRACT

Surface radiative and turbulent heat fluxes over the Great Lakes strongly influence regional hydrological and meteorological processes. Their accurate representation is therefore critical for accurate numerical weather prediction and coupled atmosphere-lake models. However, direct flux observations remain sparse, limiting systematic evaluation of model performance. As a result, gridded reanalysis and satellite-based products are often used for downstream applications despite their uncertainties. Here we assess the ability of four widely used gridded datasets to represent surface heat fluxes over and around the Great Lakes: two reanalyses, ERA5 and MERRA-2, and two satellite-based products, CERES-EBAF and CLARA-A3. All four datasets provide surface downwelling shortwave and longwave radiation, while only the reanalyses provide turbulent latent and sensible heat fluxes. We evaluate these products against extensive in situ observations from the Great Lakes Evaporation Network (GLEN), AmeriFlux towers, and National Data Buoy Center (NDBC) buoys. Results show that CERES-EBAF most consistently reproduces the observed radiative flux patterns, with composite scores of 0.63 over lakes and 0.77 over land for shortwave, and 0.80 over lakes and 0.70 over land for longwave radiation, where 1 indicates perfect agreement. For turbulent fluxes over lakes, MERRA-2 performs better for both sensible and latent heat, with composite scores of 0.44 and 0.52, respectively. Over land, ERA5 better captures sensible and latent heat fluxes, with scores of 0.39 and 0.50. Overall, this study provides practical guidance for selecting gridded surface-flux products for Great Lakes model evaluation and regional surface energy-balance analysis, with implications for next-generation solar radiation and flux-product integration.

1. Introduction

Energy and water exchanges between the land (lake) surface and the atmosphere are governed by the surface radiative balance and turbulent heat fluxes, both of which exhibit strong spatial and temporal (e.g., seasonal and diurnal) variability. Absorbed solar radiation warms the land and lake surface; this energy may be conducted into the soil or lake water, emitted back to the atmosphere as longwave radiation, transferred upward as sensible heat, or used for evaporation as latent heat. Large water bodies such as the Great Lakes can strongly influence how available surface energy is partitioned between sensible heating and evaporation. The timing and magnitude of energy storage and release within the lakes influence key regional hydrometeorological phenomena, including lake-ice extent, lake-effect precipitation, and near-surface air temperatures (Wang et al., 2022, 2025), as well as lake thermal structure (e.g., stratification and turnover timing), which in turn affects aquatic ecosystems such as phytoplankton dynamics and fisheries productivity (Flood et al., 2021; Woolway, 2023). Given the complexity of

* Corresponding author.
E-mail address: kpeco@anl.gov (K. Peco).

https://doi.org/10.1016/j.jglr.2026.102994
Received 19 May 2026; Received in revised form 5 October 2026; Accepted 16 November 2026
0380-1330/© 2026 Published by Elsevier B.V. on behalf of International Association for Great Lakes Research. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

these interactions, accurately simulating surface energy exchange remains challenging for weather and climate models especially over the Great Lakes because they require a realistic representation of lake–atmosphere coupling (e.g., Xue et al., 2017, 2020). Therefore, surface radiative and turbulent heat fluxes are critical variables to evaluate in Numerical Weather Prediction (NWP) and coupled atmosphere–lake models (e.g., Kayastha et al., 2022).

Such evaluation, however, is hindered by limited availability of in situ observations, particularly over the lakes themselves (Lenters et al., 2013; Notaro et al., 2013). As an alternative, gridded satellite- and reanalysis-based datasets are widely used to evaluate modeled surface fluxes (Liu et al., 2005; Trenberth et al., 2009; Troy and Wood, 2009). Despite their widespread use, surface energy flux estimates from these products have not been comprehensively evaluated across the Great Lakes region. This evaluation is necessary because radiative and turbulent fluxes are not observed directly at comprehensive spatial scales; instead, they are derived from parameterizations that depend on observed or assimilated state variables such as cloud properties, surface albedo, vertical temperature and moisture profiles, and skin temperature (Gelaro et al., 2017; Hersbach et al., 2020; Kato et al., 2025). Differences in how these inputs are represented propagate into flux estimates and can produce distinct thermodynamic impacts on the lake–atmosphere system (e.g., differential surface heating, cooling, or moistening; Minallah and Steiner, 2021). Consequently, before gridded flux products are used for NWP or coupled model validation over the Great Lakes, it is essential to quantify their accuracy and characterize their limitations relative to available in situ observations.

Moukomla and Blanken (2017) conducted such evaluations using over-lake observations from the Stannard Rock and Spectacle Reef lighthouse sites for a lake-wide, pixel-based multi-year net all-wave radiation dataset spanning July 2001–December 2014. This net all-wave radiation dataset combines Moderate Resolution Imaging Spectroradiometer (MODIS) satellite remote sensing with the Modern-Era Retrospective Analysis (MERRA) reanalysis to represent spatial variability under all-sky conditions. Their analysis highlighted the dominant roles of cloud effects and surface albedo in modulating net shortwave radiation, underscoring the need for continued evaluation and validation of radiative flux products. Since then, satellite and reanalysis datasets have advanced substantially. State-of-the-science satellite climate data records, most notably the Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (CERES-EBAF; Kato et al., 2025) and the Clouds, Albedo and surface Radiation – AVHRR-based, Edition 3 (CLARA-A3; Karlsson et al., 2023), provide estimates of downwelling shortwave and longwave surface radiative fluxes. Widely used reanalyses, including the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis version 5 (ERA5) (Hersbach et al., 2020) and the Modern-Era Retrospective analysis for Research and Applications reanalysis version 2 (MERRA-2) (Gelaro et al., 2017; Rienecker et al., 2011), provide all-wave radiative and turbulent heat fluxes. While these datasets have been evaluated over global, oceanic, and terrestrial domains (e.g., Chakraborty and Lee, 2021; Josey et al., 2013; Karlsson et al., 2023; Kato et al., 2025; Loeb et al., 2018), they have not been systematically assessed with targeted focus on large freshwater systems such as the Great Lakes.

The present study addresses this gap by evaluating each gridded dataset against a comprehensive set of land- and lake-based observation networks, and by intercomparing surface fluxes from two reanalysis datasets (ERA5 and MERRA-2) and two satellite-based products (CERES-EBAF and CLARA-A3). Observational benchmarks are drawn from multiple networks, including the Great Lakes Evaporation Network (GLEN; Lenters et al., 2013), AmeriFlux, an international network of eddy covariance towers (Novick et al., 2018), and the National Data Buoy Center (NDBC; Meindl and Hamilton, 1992). Through these evaluations and intercomparisons, we identify the most reliable sources of surface radiative and turbulent heat flux for the Great Lakes region and provide practical guidance for selecting gridded flux products for evaluating NWP and coupled atmosphere-lake simulations of the Great Lakes surface energy budget. These findings also have implications for next generation solar radiation and flux-product integration. The remainder of the paper is organized as follows. Section 2 describes the gridded and observational datasets, data processing methods, and evaluation metrics. Section 3 presents the evaluation of radiative and turbulent flux variables against in situ observations, as well as an intercomparison between the gridded datasets, and discusses potential mechanisms underlying the differences between them. Section 4 summarizes the main findings and implications.

2. Datasets and methodology

2.1. Gridded datasets

Four gridded datasets were considered in this analysis. Two state-of-the-art satellite-based products: CERES-EBAF version 4.2 and CLARA-A3 represent the most recent releases of their respective climate data records and incorporate updated retrieval algorithms, homogenization procedures, and uncertainty estimates (Karlsson et al., 2023; Kato et al., 2025). CERES-EBAF was selected over CERES-SYN1deg because it provides monthly, energy-balanced, bias-adjusted radiative fluxes optimized for climate-scale accuracy, making it well suited for seasonal and climate-scale evaluation. Two atmospheric reanalysis products were evaluated: ERA5 and MERRA-2. ERA5 was selected for its relatively high spatial and temporal resolution and its widespread use in weather and climate model evaluation (Hersbach et al., 2020). MERRA-2 was included as a complementary reanalysis that employs different data assimilation strategies and model physics, providing an additional perspective on reanalysis-based estimates of surface radiative and turbulent heat fluxes (Gelaro et al., 2017). Detailed descriptions of each dataset are provided below.

CERES-EBAF v4.2 provides global, 1-degree monthly-mean top-of-atmosphere (TOA) and surface shortwave, longwave, and net radiative fluxes under clear-sky and all-sky conditions from March 2000 to the present. This product is designed to support studies of Earth's energy budget and climate model evaluations. CERES broadband radiometers carried on multiple polar-orbiting satellite platforms (including Terra, Aqua, Suomi-NPP, and NOAA-20), measure broadband radiances that are converted to TOA broadband upwelling radiative fluxes using angular distribution models (ADMs) (Kato et al., 2025). In the EBAF product, shortwave (SW) and longwave (LW) TOA fluxes are adjusted within their uncertainty ranges using an objective constraint algorithm to reduce the imbalance between global mean net TOA radiation and estimates of heat storage in the Earth–atmosphere system. Surface radiative fluxes are then derived from radiative transfer calculations using satellite-retrieved cloud and aerosol properties and reanalysis atmospheric state variables; these calculations are constrained to be consistent with the CERES EBAF TOA fluxes by adjusting input cloud, atmospheric, and surface properties within their uncertainties and recomputing the surface fluxes. For Edition 4.2, these calculations use cloud properties retrieved from co-located MODIS and Visible Infrared Imaging Radiometer Suite (VIIRS) imagers and MERRA-2 temperature and humidity profiles. Because MODIS and VIIRS provide cloud retrievals only at satellite overpass times, cloud properties are temporally interpolated between overpasses for the hourly flux calculations before aggregation to monthly means. Further details are provided in Kato et al. (2025). Given the climatological focus of this study, and that CERES-EBAF provides monthly means, we aggregate the other datasets to monthly averages for consistent comparison.

CLARA-A3 provides daily and monthly estimates of surface radiative fluxes and cloud properties from 1979 to the present at 0.25° horizontal spatial resolution, supporting regional-scale evaluation. CLARA-A3 is primarily derived from Advanced Very High-Resolution Radiometer (AVHRR) observations aboard National Oceanic and Atmospheric Administration (NOAA) and Meteorological Operational (MetOp) polar-

orbiting satellites (Cracknell, 1997; Karlsson et al., 2023). AVHRR narrowband radiances are used to retrieve cloud properties and surface albedo, which are combined with ancillary atmospheric and surface information in radiative transfer and empirical retrieval frameworks to estimate surface radiative fluxes. In contrast to CERES-EBAF, CLARA-A3 is not energy balanced using estimates of Earth heat storage, nor are its surface fluxes constrained to match broadband TOA fluxes. Instead, the CLARA-A3 surface radiation products are based on AVHRR-derived cloud and albedo information together with ancillary/reanalysis inputs; for example, surface incoming shortwave radiation uses AVHRR-derived cloud information with ERA5 water vapor, ozone, and surface albedo inputs, while surface downwelling longwave radiation is derived by adjusting ERA5 monthly downwelling longwave radiation using CLARA-A3 cloud fraction (Karlsson et al., 2023). This distinction is important because CERES-EBAF and CLARA-A3 represent different satellite-based retrieval approaches, with EBAF emphasizing global energy balance and TOA consistency, and CLARA-A3 emphasizing long-term, higher-resolution AVHRR-based regional surface radiation estimates.

Reanalysis datasets ERA5 and MERRA-2 provide model-based estimates of surface and TOA radiative fluxes (incoming, outgoing, and net shortwave and longwave fluxes; clear-sky and all-sky). ERA5 and MERRA-2 also offer turbulent heat flux variables (sensible and latent), which are not routinely available as operational products from the satellite datasets used here. ERA5 provides global hourly data from 1940 to the present at a 0.25° resolution (Hersbach et al., 2020), and MERRA-2 provides global hourly data from 1980 to the present at a 0.5° x 0.625° spatial resolution (Gelaro et al., 2017). In both reanalyses, radiative and turbulent fluxes are diagnostic model outputs constrained by assimilated atmospheric temperature and moisture profiles, cloud properties, and aerosol information, but they remain sensitive to the choice of parameterizations. In ERA5, surface radiative fluxes are computed using the ECMWF Integrated Forecasting System (IFS; ECMWF, 2016) radiation scheme, which uses the Rapid Radiative Transfer Model (RRTM; Mlawer et al., 1997) for longwave and shortwave radiative transfer and applies the Monte Carlo Independent Column Approximation (Marchuk et al., 1980) to represent subgrid cloud variability. Surface sensible and latent heat fluxes (and surface shear stress) are computed using a Monin–Obukhov similarity theory surface-layer formulation with stability-dependent transfer coefficients (Monin and Obukhov, 1954; Obukhov, 1971). Over the Great Lakes and other inland water bodies, prior to 2014, ERA5 represents lake thermodynamics using the Freshwater Lake model (FLake; Kirillin et al., 2011), which simulates lake temperature structure and ice evolution for inland water bodies. However, the reliability of the lake lower-boundary conditions is poor, and it contains substantial biases, including cold water temperature biases and excessive ice cover in winter and warm water temperature biases in summer (Simmons et al., 2021). From 2014 onward, ERA5 lake surface conditions over large lakes are constrained by external observational analyses, particularly the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA; Fiedler et al., 2014) which provides daily, gap-free sea/lake surface temperature and ice analyses based on satellite and in situ observations. These analyzed lake surface temperature and ice fields provide lower-boundary information for near-surface energy exchange, including bulk turbulent flux calculations based on wind speed and air–water temperature and humidity gradients. When direct observations are unavailable or discontinuous, these boundary fields are supplied through the external analysis product rather than through raw observations alone. Previous studies have documented biases in FLake-simulated lake surface temperature and lake ice cover, particularly for large and deep lakes (Minallah and Steiner, 2021; Wang et al., 2025). Therefore, ERA5 radiative and turbulent flux estimates over the Great Lakes, especially before 2014, may be affected by 1-D lake model errors in lake surface temperature and ice representation; we further evaluate and discuss these impacts in the Results section.

In MERRA-2, radiative fluxes are computed using the Goddard Earth Observing System Model, Version 5 (GEOS-5) radiation parameterizations (e.g., Chou and Suarez, 1994, 1999; Molod et al., 2015), and near-surface turbulent exchange is computed using a Monin–Obukhov similarity surface-layer scheme (Helfand and Schubert, 1995). Importantly, MERRA-2 also assimilates satellite-derived aerosols, which constrain their impact on the surface radiative fluxes. The GEOS-5 implementation includes a viscous sublayer treatment for heat and moisture exchange over non-land surfaces, and ocean roughness formulations that vary with wind regimes with additional modifications (e.g., Kondo, 1975; Large and Pond, 1981; Molod et al., 2015). In MERRA-2, inland water bodies are represented within the GEOS-5 land–atmosphere modeling framework using a fractional tiling approach, in which grid cells may contain both land and water surface types. Surface fluxes over water are computed using Monin–Obukhov similarity theory with stability-dependent transfer coefficients and wind-dependent roughness formulations, consistent with the model's treatment of ocean surfaces (Molod et al., 2015; Reichle et al., 2017). Unlike ERA5, which employs a dedicated lake model (FLake), MERRA-2 does not explicitly resolve lake thermodynamics (e.g., vertical temperature structure or heat storage) and instead represents inland water surfaces through bulk exchange parameterizations within the coupled surface layer scheme. While this approach may be adequate for smaller and shallower lakes, it can underrepresent key vertical and horizontal lake processes in large, deep systems such as the Great Lakes.

2.2. In situ datasets

To evaluate the radiative and turbulent heat flux gridded datasets described in Section 2.1, we leveraged three observational networks, namely GLEN (Lenters et al., 2013), NDBC (Meindl and Hamilton, 1992), and AmeriFlux (Novick et al., 2018) across the Great Lakes Basin. The locations and data availability for all three in situ datasets (16 sites in total) are shown in Table 1 and Fig. 1. Specifically, we used five GLEN lighthouse stations: Long Point in Lake Erie, Granite Island and Stannard Rock in Lake Superior, Spectacle Reef in Lake Huron, and White Shoal in Lake Michigan (Fig. 1). These stations report standard meteorological variables and surface flux measurements, including

Table 1
Relevant metadata for all in situ observations taken across the Great Lakes region. SW = Shortwave radiative flux, LW = longwave, LH = latent heat, SH = sensible heat.

Station	Network	Latitude, Longitude	Temporal Coverage	Variables
Long Point	GLEN	42.56, −80.05	2012–2018	SW, LW, LH, SH
Granite Island	GLEN	46.82, −87.41	2010–2018	LH, SH
Spectacle Reef	GLEN	45.77, −84.00	2009–2018	SW, LW, LH, SH
Stannard Rock	GLEN	47.38, −87.22	2008–2018	SW, LW, LH, SH
White Shoal	GLEN	45.77, −85.53	2012–2018	SW, LH, SH
Buoy 45,013	NDBC	43.10, −87.85	2015–present	SW
Buoy 45,028	NDBC	46.81, −91.82	2014–present	SW
Buoy 45,170	NDBC	41.75, −86.96	2014–present	SW
Buoy 45,175	NDBC	45.87, −84.77	2015–present	SW
CA-Gro	AmeriFlux	48.21, −82.15	2003–2014	SW, LW, LH, SH
CA-TPD	AmeriFlux	42.63, −80.55	2012–2017	SW, LW, LH, SH
US-CRT	AmeriFlux	41.62, −83.34	2011–2013	SW, LW, LH, SH
US-MMS	AmeriFlux	39.32, −86.41	1999–2020	SW, LW, LH, SH
US-ORv	AmeriFlux	40.02, −83.01	2011–2016	SW, LH, SH
US-Syv	AmeriFlux	46.24, −89.34	2001–2021	SW, LH, SH
US-UMB	AmeriFlux	45.36, −84.69	2007–2021	SW, LW, LH, SH

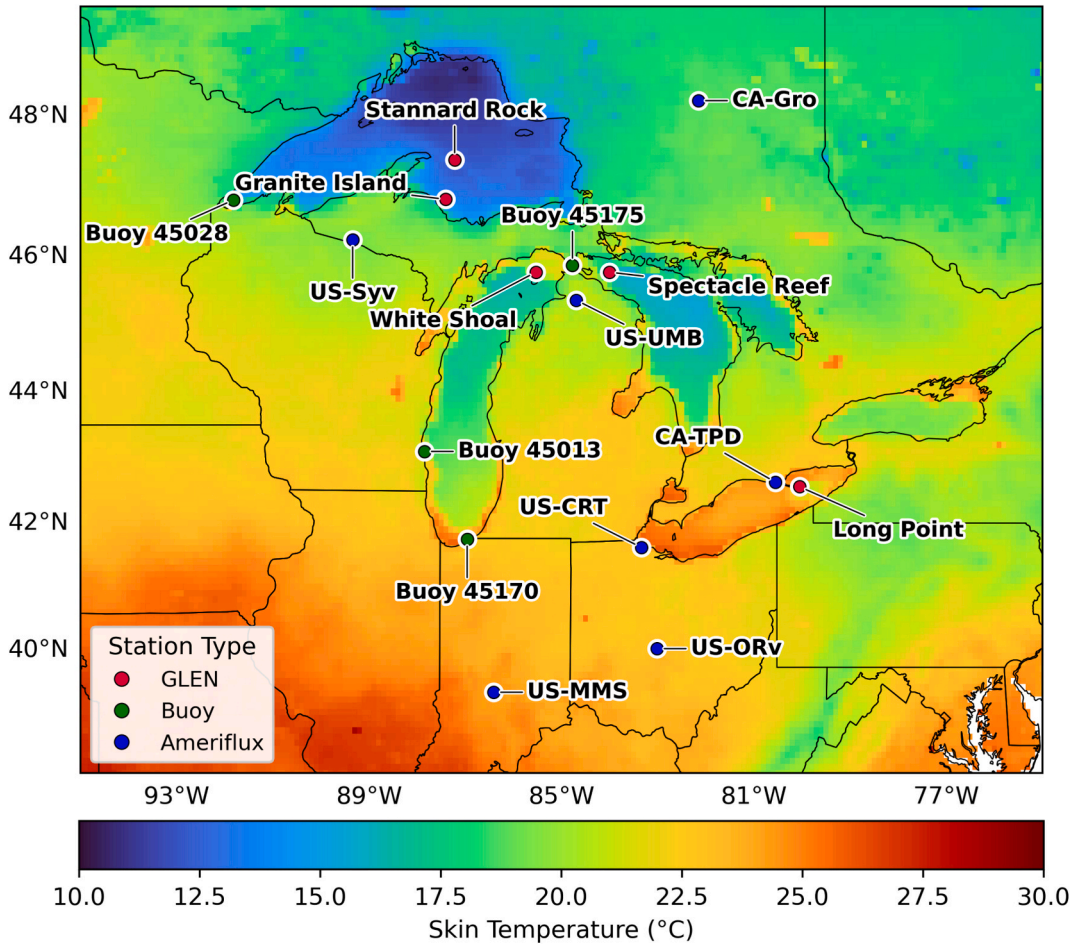

Fig. 1. Locations of all in situ observation sites used in this study plotted alongside average skin temperature sourced from ERA5-Land (9 km spatial resolution) for summer (June–August) 2018. More specific station information can be found in Table 1, and colors differentiate the observing network/organization. Spectacle Reef, Stannard Rock, and White Shoal represent open-water locations, with water extending at least 10 km in all directions. Granite Island is also surrounded by open water but may be marginally influenced by the island itself. Long Point in Lake Erie is more strongly influenced by nearby land; however, wind-direction filtering was applied to minimize land-contaminated conditions (Lenters et al., 2013). NDBC buoys 45,013 and 45,175 in Lake Michigan are closer to shore (~2–3 km offshore), whereas buoys 45,028 in Lake Superior and 45,170 in Lake Michigan are in more open water (~6–8 km offshore).

temperature, wind speed, humidity, downwelling shortwave and longwave radiation, and sensible and latent heat fluxes. Measurements are collected atop lighthouses approximately 30 m above the lake surface at 30-min intervals. Data availability begins between 2008 and 2010, with records extending through 2018, depending on the site. Some sites are also now part of AmeriFlux and report more recent data from 2019 to the present, such as Stannard Rock (Nicholls et al., 2024). NDBC provides downwelling shortwave radiation along with standard meteorological and oceanic variables from nearshore buoy sites on Lakes Superior and Michigan. In total, downwelling shortwave radiation from four NDBC buoys is used in this study (Fig. 1), with data available from 2015 to the present at 30-min resolution. AmeriFlux (Novick et al., 2018) provides land-based measurements of incoming shortwave and longwave radiation and latent and sensible heat fluxes at sites around the Great Lakes Basin (Fig. 1). AmeriFlux sites also report a broad suite of meteorological and ecosystem flux variables designed to quantify exchanges between terrestrial ecosystems and the atmosphere. We selected seven AmeriFlux sites for this analysis, including near-lake locations (CA-TPD, US-CRT, and US-UMB) and farther inland sites (CA-Gro, US-MMS, US-ORv, and US-Syv; Fig. 1). These sites complement the over-lake observations and allow us to assess whether spatial patterns emerge in gridded datasets across lake-influenced versus inland environments.

2.3. Data processing

For the point validations against in situ observations, each gridded dataset was evaluated on its native grid. We used nearest-neighbor extraction rather than regridding all products to a common grid to avoid introducing interpolation artifacts across land–water boundaries, which are especially important in the Great Lakes region. This choice preserves each product's native representation of surface type and lake geometry. For radiative fluxes, the analysis focused on monthly climatological variability, for which exact point collocation is less critical than for turbulent fluxes. For turbulent heat fluxes, nearest-neighbor cells were manually checked to ensure that lake-based observations were paired with lake grid cells and land-based observations with land grid cells. It is important to note that, despite these careful methodological considerations, the varying native spatial resolutions between datasets, especially in geographic regions with heterogeneous land types, may still impart some influence on overall dataset performance. Furthermore, varying temporal availability of the corresponding in situ data naturally has additional effects on dataset performance as well.

In situ data were quality controlled at their native temporal resolution prior to aggregation to monthly means (to match the temporal resolution of CERES-EBAF). Specifically, negative values of downwelling radiative fluxes were removed from each observational dataset. For the half-hourly GLEN records, spurious spikes in downwelling

shortwave and longwave radiation were filtered by retaining values between 0 and 1200 W m^{-2}; the same thresholding was applied to the 30-min buoy shortwave measurements. In the specific case of nighttime hours, GLEN shortwave pyranometers frequently produced small negative offsets; these overnight negative values were effectively converted to 0 W m^{-2} before analysis. Turbulent fluxes were filtered using analogous outlier screening, but physically plausible negative values were retained. Overall, only 0.1–0.2% of radiative flux data and 2.0–2.5% of turbulent heat flux data was discarded via these filtering methodologies. After quality control, monthly means were computed only for months with at least 90% data availability (~27 days of valid observations).

Lastly, for the intercomparison between the four gridded datasets, without considering observational data, we bilinearly interpolated all fields to a common 0.5° × 0.5° latitude–longitude grid to be able to compute statistical metrics consistently across products. Seasonal climatological means were then computed over the common period 2001–2020 to represent typical surface flux conditions for winter (December–February), spring (March–May), summer (June–August), and fall (September–November). For each season, inter-dataset variance was calculated at every grid cell using the seasonal mean flux values, providing a spatial measure of dataset spread for surface downwelling shortwave and longwave radiation. This analysis highlights where disagreements are largest (e.g., over land versus lakes) and how the magnitude and location of discrepancies vary by season.

2.4. Metrics for reanalysis and satellite data evaluation

To assess agreement between the gridded datasets and in situ observations, we employed multiple statistical and distributional metrics. Error and association metrics included root mean square error (RMSE), mean bias error (MBE), and the Pearson correlation coefficient (r), which together evaluate differences in magnitude, systematic bias, and temporal covariation. To compare flux distributions, emphasizing magnitude across the full range of values rather than time alignment, we also computed an overlap ratio (OR).

For each pairing between the gridded dataset (‘*model*’) and observations (‘*obs*’), the RMSE and MBE were computed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_i^{model} - X_i^{obs}\right)^2} \tag{1}$$

$$MBE = \frac{1}{N}\sum_{i=1}^{N}\left(X_i^{model} - X_i^{obs}\right) \tag{2}$$

where X_i^{model} and X_i^{obs} represent model (gridded dataset) and observed values at the i-th monthly timestep, and N is the sample size (number of months). RMSE reflects the overall accuracy of the model and is particularly sensitive to outliers but does not indicate the direction of model bias. MBE complements RMSE by capturing the model's systematic bias in magnitude or direction but can incorrectly indicate good model performance if there are equal magnitudes of positive and negative errors.

The correlation coefficient (r) was used to assess the linear association between the gridded dataset and observations in time, with an r value of 1 indicating a perfect positive correlation. In the case of radiative and turbulent heat fluxes, there are strong seasonal signals, which can produce a correlation coefficient very close to 1. To reduce this influence of the shared annual cycle on the correlation coefficient, Pearson's r was calculated using de-seasonalized flux anomalies. Specifically, the monthly climatological mean was removed in both observations and gridded datasets before calculating their correlations:

$$O_t' = O_t - \overline{O}_{m(t)} \tag{3}$$

$$M_t' = M_t - \overline{M}_{m(t)} \tag{4}$$

The reported correlation coefficient was then calculated as the Pearson correlation between the observed and modeled anomalies:

$$r_{\text{anom}} = \frac{\sum_{t=1}^{N}\left(O_t' - \overline{O'}\right)\left(M_t'\overline{M'}\right)}{\sqrt{\sum_{t=1}^{N}\left(O_t'\overline{O'}\right)^2}\sqrt{\sum_{t=1}^{N}\left(M_t'\overline{M'}\right)^2}} \tag{5}$$

where O_t' and M_t' are the de-seasonalized observed and modeled flux anomalies, respectively, and N is the number of paired valid time steps. All Pearson correlation presented in Results are based on de-seasonalized time series.

Overlap ratio (OR) is defined as the fractional overlap between two normalized frequency distributions. Initially, probability density functions ($(f(X))$) were approximated using kernel density estimation, specifying Scott's Rule for bandwidth smoothing. ORs were then calculated between the density functions using the following formula:

$$OR = \int_{-\infty}^{\infty}\left(f_{model}(X)\,, f_{obs}(X)\right)\,dX \tag{6}$$

The OR ranges from 0 (no overlap) to 1 (identical distributions), providing a simple measure of similarity and demonstrating directional differences when visualizing the two density functions.

We also constructed a composite score by considering all the metrics we examined. Because RMSE and MBE are unbound and not on a common scale, these metrics were normalized prior to aggregation. In contrast, r and OR were retained in their native form because both range from 0 to 1, given that we do not see negative correlations between the observations and gridded datasets. Each metric was assigned a predefined weight reflecting its relative importance: RMSE and MBE were weighted more heavily (0.35 and 0.25, respectively) because they account for magnitude errors across the time series, whereas r and OR were each weighted at 0.20 because they capture more limited aspects of agreement. By perturbing these weights, we found that while alternate weighting schemes generated slightly different composite score values, overall dataset rankings remained the same.

To provide a documented, scalar measure of dataset performance, we calculated two additional complementary skill scores for each gridded dataset relative to the in situ observations: the refined Willmott index of model performance (Willmott et al., 2012), d_r, and the modified coefficient of efficiency (Legates and McCabe Jr., 1999), E_1. These metrics were selected because they are dimensionless and can be applied consistently across radiative and turbulent heat fluxes. Both scores were computed using paired monthly modeled and observed values after temporal matching and removal of missing values.

The refined Willmott index, d_r, evaluates model agreement using the summed absolute error relative to the potential error associated with deviations from the observed mean. It is defined as:

$$d_r = \begin{cases} 1 - \dfrac{\sum_{i=1}^{n}\left|P_i - O_i\right|}{2\sum_{i=1}^{n}\left|O_i - \bar{O}\right|}, & \text{if } \sum_{i=1}^{n}|P_i - O_i| \le 2\sum_{i=1}^{n}|O_i - \bar{O}| \\ \dfrac{2\sum_{i=1}^{n}\left|O_i - \bar{O}\right|}{\sum_{i=1}^{n}\left|P_i - O_i\right|} - 1, & \text{if } \sum_{i=1}^{n}|P_i - O_i| > 2\sum_{i=1}^{n}|O_i - \bar{O}| \end{cases} \tag{7}$$

where P_i is the gridded dataset estimate, O_i is the observed value, $\bar{O}$ is the mean of the observations, and n is the number of paired samples. The refined Willmott index ranges from −1 to 1, where $d_r = 1$ indicates perfect agreement. Values closer to zero or below zero indicate decreasing model performance.

We also calculated the modified coefficient of efficiency, E_1, which compares the absolute error of the gridded dataset against the error that would result from using the observed mean as a baseline predictor:

$$E_1 = 1 - \frac{\sum_{i=1}^{n} \left| P_i - O_i \right|}{\sum_{i=1}^{n} \left| O_i - \bar{O} \right|} \quad (8)$$

The E_1 score has a direct baseline interpretation: $E_1 = 1$ indicates perfect agreement, $E_1 = 0$ indicates performance equivalent to using the observed mean, and $E_1 < 0$ indicates performance worse than the observed-mean baseline. Because E_1 is based on absolute deviations rather than squared errors, it is less sensitive to isolated large errors than traditional squared-error efficiency metrics. This property is useful for turbulent heat fluxes, which can exhibit episodic variability and larger short-term deviations than radiative fluxes.

Together, d_r and E_1 provide complementary measures of model performance. The refined Willmott index quantifies overall agreement relative to the potential observational variability, while the modified coefficient of efficiency evaluates whether a gridded dataset provides predictive skill beyond the observed climatological mean. These scores were used alongside RMSE, mean bias error, Pearson correlation, and distributional overlap to assess both the magnitude and structure of dataset errors.

3. Results

3.1. Evaluation against observations

This section validates each gridded dataset against in situ observations. Although these observations are spatially sparse relative to the coverage of gridded products, they span both lake and land locations over one to two decades and provide a robust benchmark for evaluating gridded surface flux estimates. The specific in situ stations chosen for the

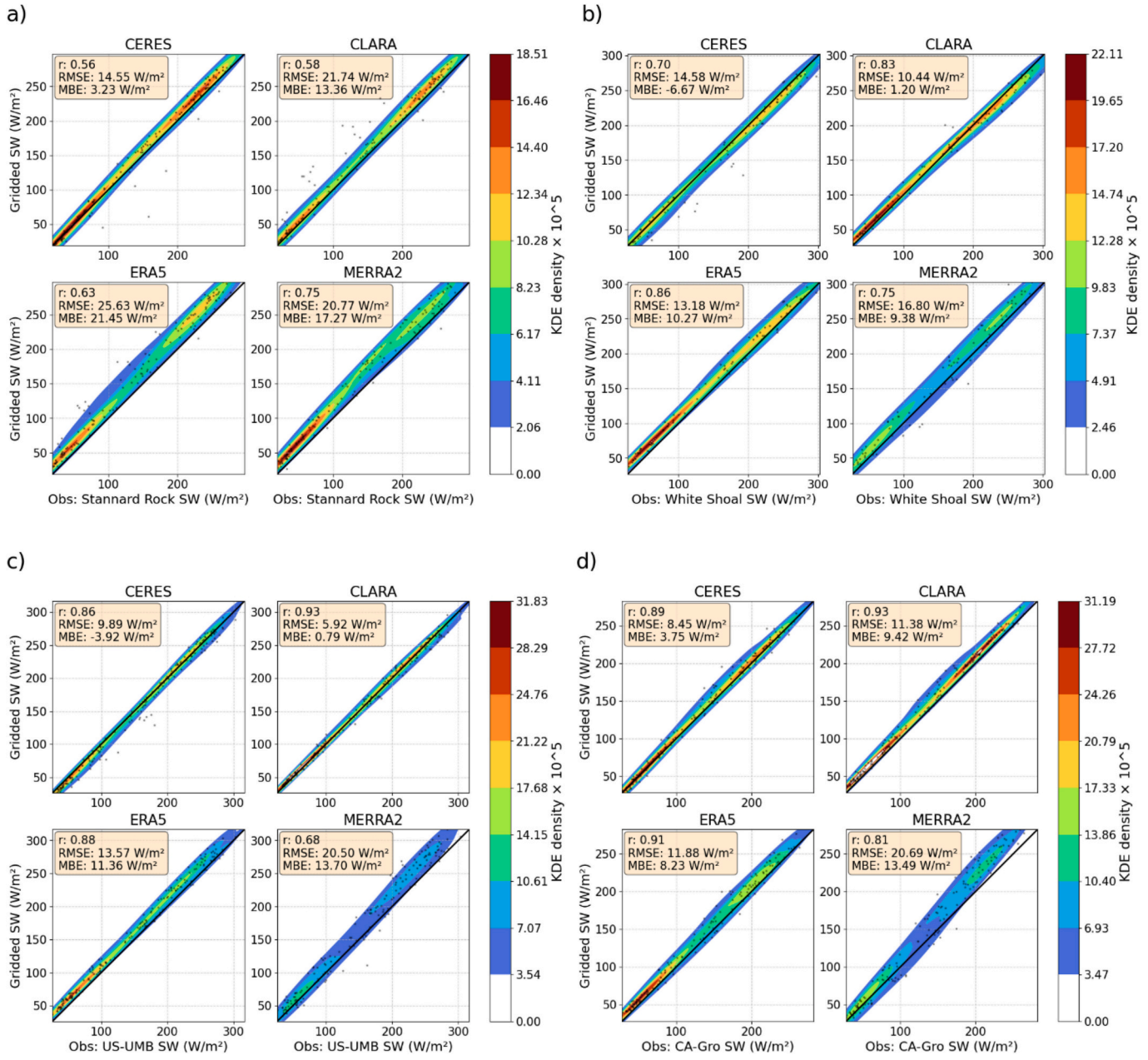

Fig. 2. Contour density functions plotted between observational data and the four gridded datasets (CERES-EBAF, CLARA, ERA5 and MERRA-2) for two lake locations, Stannard Rock (a) and White Shoal (b), and two land locations, US-UMB (c) and CA-Gro (d) for downwelling shortwave radiation. Darker/warmer colors indicate a higher density of scatter points, with a 1:1 line drawn to visualize each dataset's systematic bias.

plots in this section were selected for their representativeness in demonstrating the magnitudes and patterns of radiative and turbulent heat fluxes over different surface types.

3.1.1. Radiative flux

We first compared monthly mean downwelling shortwave and longwave flux distributions from each gridded dataset with observations at the nearest locations. Probability density functions and two-dimensional kernel-density contour plots were used to assess distributional similarity, and overlap ratios were computed to quantify agreement. As shown in Fig. 2, all four gridded datasets capture the overall distribution of observed downwelling shortwave radiation over two lake and two land sites. The satellite products generally agree more closely with observations than the reanalyses, indicated by tighter concentrations of scatter points around the 1:1 line (Fig. 2), and all datasets perform better over land than over lakes. CERES-EBAF and CLARA-A3 yield mean overlap ratios of 0.94 and 0.93 over lakes, and 0.97 and 0.97 over land, respectively (Fig. 6a). ERA5 and MERRA-2 show mean overlap ratios of 0.91 and 0.89 over lakes and 0.96 and 0.94 over land, respectively (Fig. 6a).

For downwelling longwave radiation, performance is weaker (Figs. 3 and 6a). The satellite products again show the best agreement, with overlap ratios exceeding 0.91 over both land and lakes (Figs. 3 and 6a). ERA5 also performs reasonably well, with mean overlap ratios at or above 0.90 for both lake and land locations. In contrast, MERRA-2 shows the lowest agreement, with data points clustering below the 1:1 line (Fig. 3), with overlap ratios of 0.84 over lakes and 0.86 over land. This overall reduced performance is consistent with the strong sensitivity of downwelling longwave radiation to skin temperature, surface emissivity, boundary-layer temperature and humidity, and cloud regimes, all of which can vary substantially over heterogeneous terrain and near coastlines (Tian et al., 2023). Overlap ratios for all additional station

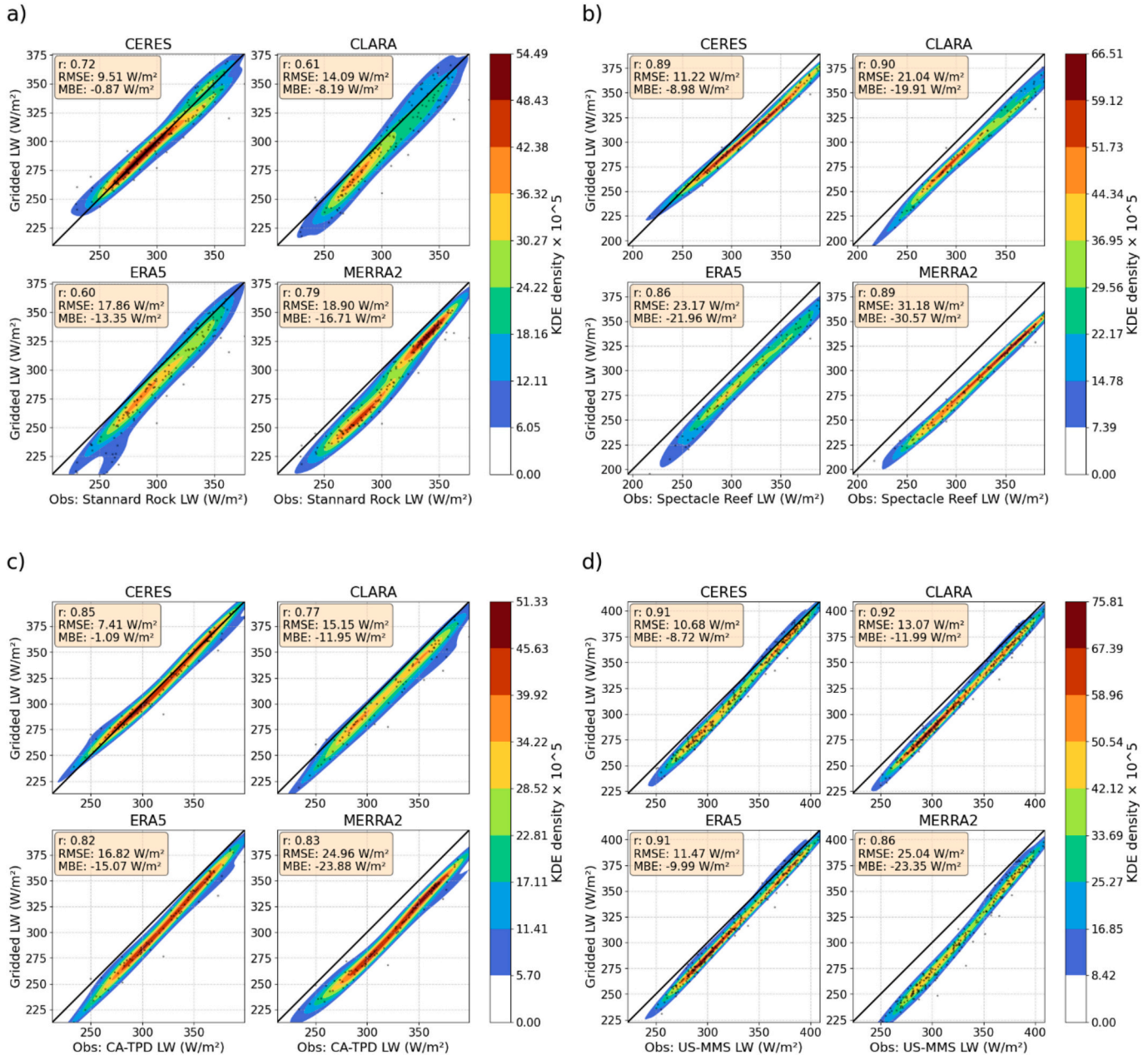

Fig. 3. Same as Fig. 2, but for downwelling longwave radiation. Two lake locations are Stannard Rock (a) and Spectacle Reef (b), and two land locations are CA-TPD (c) and US-MMS (d).

evaluations are provided in the Supplementary Tables for both shortwave and longwave fluxes (Electronic Supplementary Material (ESM) Tables S1 and S2).

Next, we evaluated the temporal agreement between the gridded datasets and observations by plotting the multi-year average monthly time series (Figs. 4 and 5). We also calculated RMSE, MBE, and deseasonalized Pearson's r on the monthly average radiative fluxes (summarized in Fig. 6b–d); all datasets capture the seasonal cycle of both downwelling shortwave and longwave radiation very well, including the timing of seasonal minima and maxima (Figs. 4 and 5). Interannual variability was also well captured, with Pearson's r ~ 0.63–0.76 for shortwave radiation over the lakes and 0.79–0.86 for longwave radiation over the lakes, as well as shortwave and longwave radiation over land (0.75–0.90, Fig. 6b). In terms of errors, the satellite products (CERES-EBAF and CLARA-A3) performed best, with smaller RMSEs and MBEs than the reanalyses (Fig. 6c and d). ERA5 generally follows behind the satellite products, whereas MERRA-2 consistently shows larger RMSEs and MBEs over both land and lakes for both shortwave and longwave radiation (Fig. 6c and d).

Interestingly, all datasets demonstrated positive biases in downwelling shortwave fluxes and negative biases in downwelling longwave fluxes over both lake and land surface classes (Fig. 4, Fig. 5, Fig. 6c). These positive biases in downwelling shortwave fluxes were strongest in the reanalysis datasets, especially during the shoulder seasons (e.g., spring and fall) over the lakes, where monthly mean overestimations reached as high as ~40 W m^{-2}. Negative biases in longwave fluxes were also strongest from fall to spring over both land and lakes (Fig. 5). These biases were particularly evident over large and deep lakes, as shown at Stannard Rock (Figs. 4b and 5b). This opposite bias pattern in shortwave and longwave fluxes is consistent with systematic errors in low-level cloud representation: when there are fewer clouds, downwelling shortwave radiation increases while downwelling longwave radiation decreases (Kato et al., 2018; Mitra et al., 2025).

The key question, therefore, becomes whether satellite and reanalysis products can represent cloud occurrence and cloud radiative properties reasonably well. During late fall and early winter, enhanced lake–atmosphere heat and moisture fluxes promote frequent cloud formation, particularly during cold-air outbreaks over the Great Lakes (e.g., Kristovich and Laird, 1998; Notaro et al., 2013). These cold-season boundary-layer cloud regimes are often shallow, spatially heterogeneous, and mixed-phase, making them difficult to represent accurately in numerical model parameterizations (e.g., Bodas-Salcedo et al., 2014; Cesana and Chepfer, 2012). Although low-level clouds are often readily detected by satellite imagers, retrievals of cloud fraction, cloud-top height, phase, optical thickness, and liquid water path can be sensitive to the sensor, viewing geometry, surface background, multilayer cloud structure, and retrieval assumptions (e.g., Karlsson et al., 2023; Stubenrauch et al., 2013). These uncertainties are important because errors in cloud microphysics and microphysical properties directly affect retrieved or modeled shortwave and longwave surface radiative fluxes.

Consistent with the relatively large biases seen for shortwave and longwave fluxes during the spring and fall seasons, we find relatively greater inter-dataset discrepancies in cloud cover among the four datasets (ESM Fig. S1). During late winter and early spring, cloud representation becomes even more difficult because of the presence of ice and its highly variable spatial patterns. In winter, excessive lake ice suppresses moisture release and limits cloud formation, leading to more downwelling shortwave and less downwelling longwave radiation. This is evident in MERRA-2, where the lakes are unrealistically represented as inland-water bodies, producing the least cloud cover in winter and fall (ESM Fig. S1), and also in ERA5, which overestimates winter lake ice cover (Wang et al., 2025). In addition, we found that the reanalyses exhibit lower near-surface humidity over water surfaces during stable conditions (e.g., spring), further reducing the cloud fraction and modeled longwave emission. Together these factors produce a consistent pattern of positive shortwave and negative longwave biases across datasets.

3.1.2. *Turbulent flux*

Turbulent heat fluxes govern the exchange of heat and moisture across the land–atmosphere and lake–atmosphere interfaces. Radiative fluxes are strongly influenced by atmospheric column properties, including clouds, aerosols, water vapor, and temperature and humidity profiles. Turbulent fluxes, in contrast, depend more directly on near-surface gradients and exchange processes, including wind speed, air–surface temperature and humidity differences, boundary-layer stability, and surface roughness. Their estimation also requires accurate representation of land and lake surface states, including skin temperature, soil moisture, vegetation, lake water temperature, and ice cover. Because the two satellite-based datasets used here do not provide turbulent heat fluxes, this evaluation is limited to the two reanalyses (ERA5 and MERRA-2). Although turbulent fluxes can be derived from near-surface meteorological measurements (e.g., Charusombat et al., 2018), differences among algorithms introduce additional uncertainty that

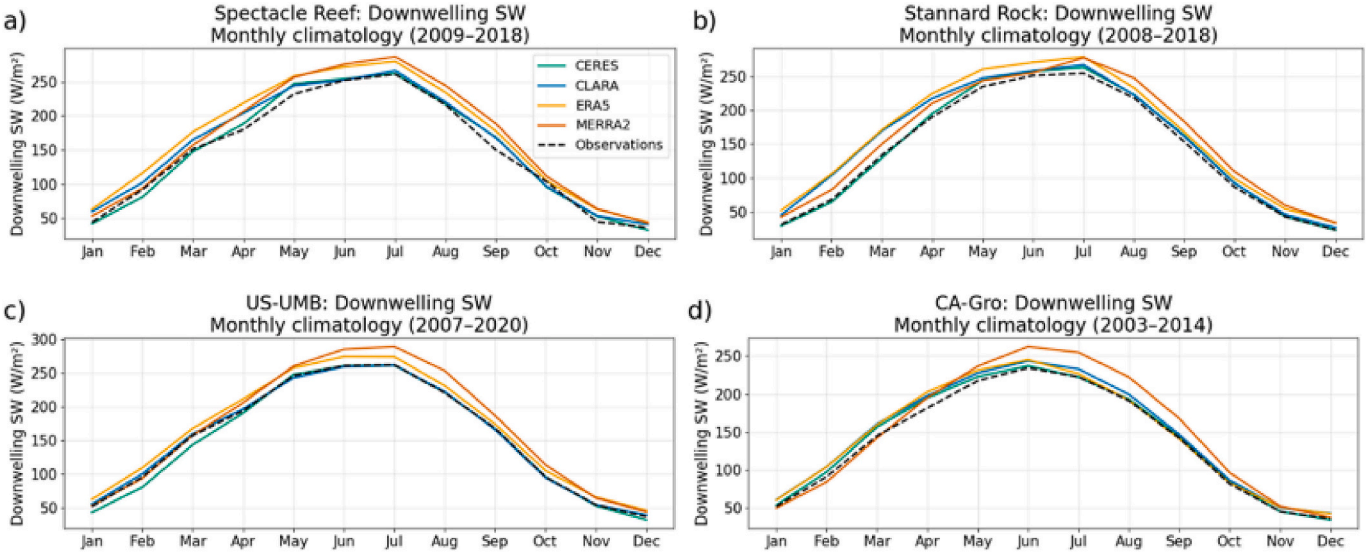

Fig. 4. Multiyear average monthly time series for downwelling shortwave radiation for lake locations Spectacle Reef (a) and Stannard Rock (b), and land locations US-UMB (c) and CA-Gro (d).

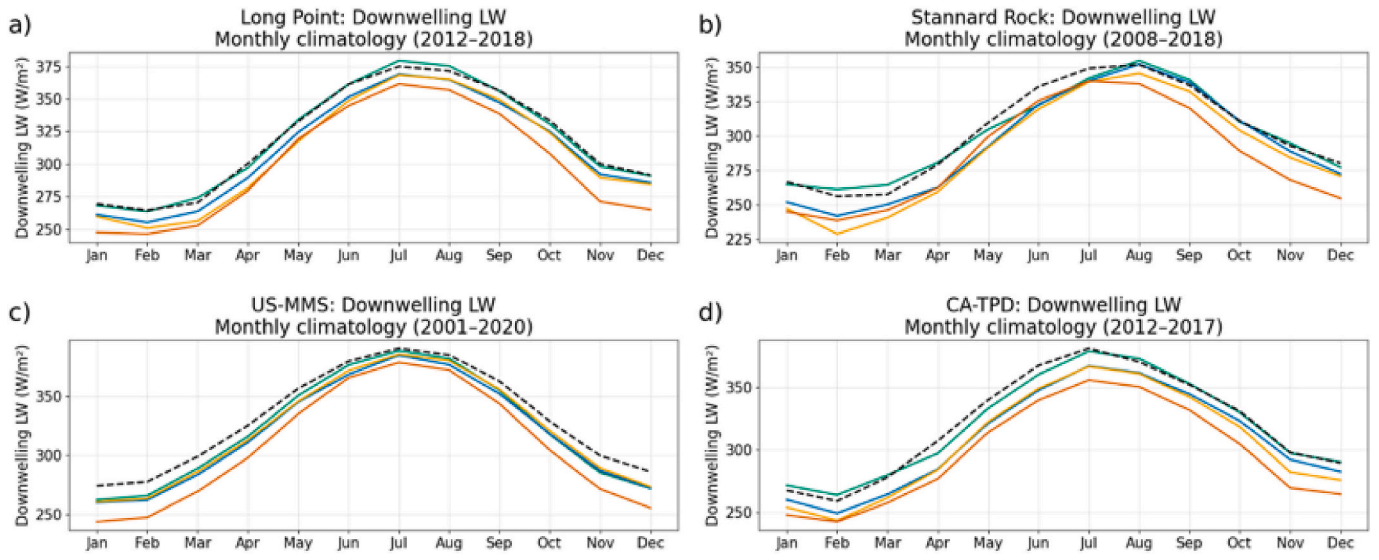

Fig. 5. Same as Fig. 4 but for downwelling longwave radiation for lake locations Long Point (a) and Stannard Rock (b), and land locations US-MMS (c) and CA-TPD (d).

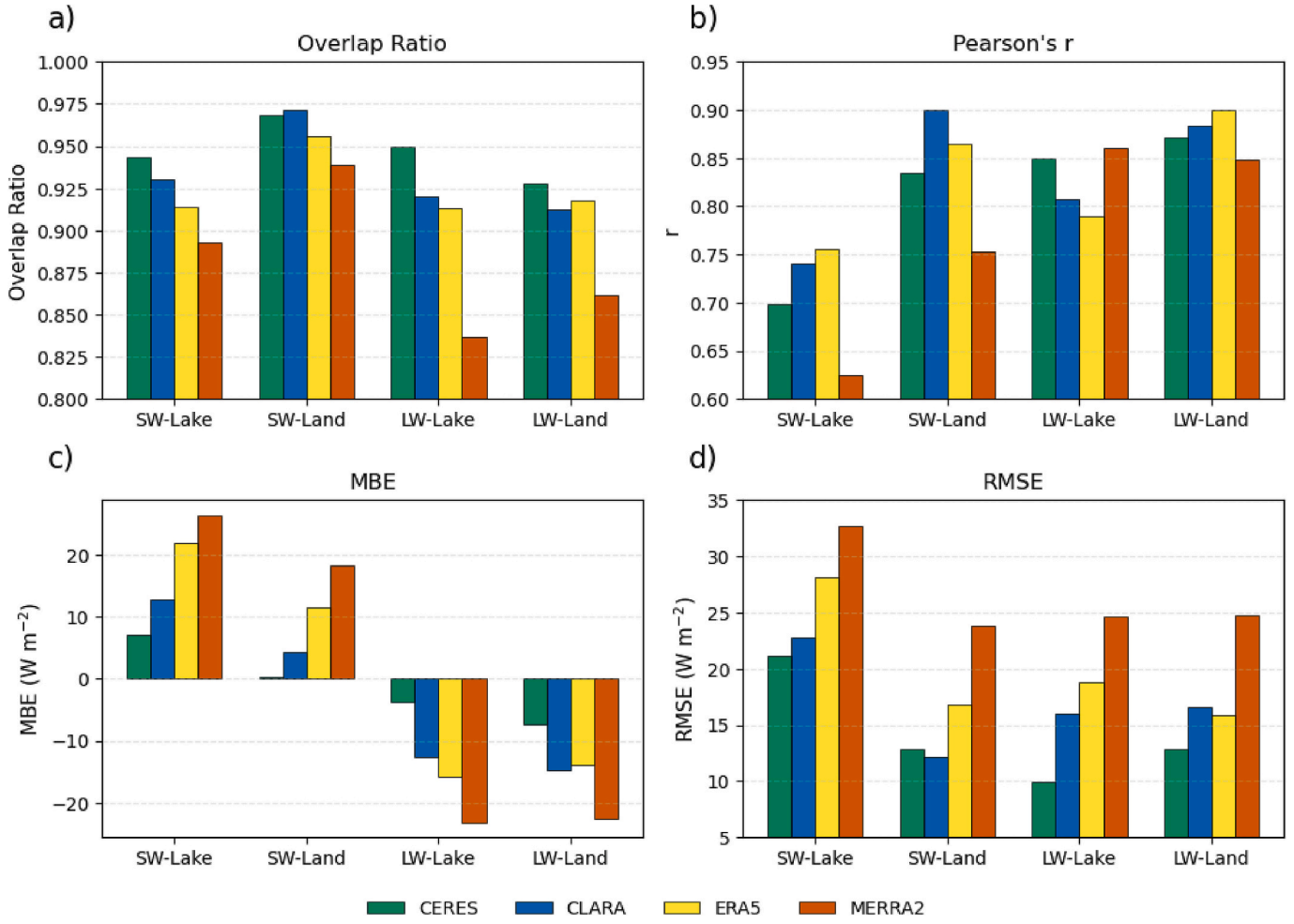

Fig. 6. Average statistical metrics, OR (a), de-seasonalized Pearson's r (b), MBE (c), and RMSE (d) for surface downwelling shortwave and longwave fluxes for both lake and land observational locations (Table 1).

could affect the evaluation. Therefore, the turbulent flux evaluation is constrained to the observational measurements that directly provide these flux variables.

Given the sensitivity of turbulent fluxes to both surface condition and

near-surface atmospheric states, alongside ERA5's known lake surface temperature and ice cover biases prior to 2014 (Minallah and Steiner, 2021; Simmons et al., 2021; Wang et al., 2025), we first examine the seasonal cycle of latent and sensible heat fluxes before and after 2014 (Fig. 7). We found that prior to 2014, ERA5 underestimates turbulent heat fluxes at the GLEN sites during winter, consistent with cold lake-water biases and excessive ice cover, both of which suppress air-water exchange through the insulation effect of ice. In summer, warm lake-water biases are associated with enhanced latent heat flux and increased evaporation. These issues are most apparent over large and deep lake regions (demonstrated by Granite Island in Fig. 7a and Stannard Rock in Lake Superior). Accordingly, the subsequent evaluation, regarding lake locations, focuses on the period after 2014, when the known ERA5 lake water temperature and ice issues over the Great Lakes are substantially reduced (Simmons et al., 2021; Wang et al., 2022).

ERA5 and MERRA-2 demonstrate moderate skill in capturing the observed distributions and temporal behavior of turbulent heat fluxes, with neither dataset showing a clear overall advantage across variables, surface types, and evaluation metrics (ESM Fig. S2; ESM Tables S3 and S4). ERA5 performs slightly better over land, while MERRA-2 tends to perform somewhat better over shallow lakes, although this pattern is not consistent across all four metrics. Both ERA5 and MERRA-2 also capture the contrasting thermal response of land and lakes. Over land, turbulent fluxes respond closely to air temperature and solar radiation, increasing rapidly in spring as air temperature rises and solar radiation increases (Fig. 8). Over lakes, however, turbulent fluxes increase later, typically in summer for shallow lakes and in fall for large and deep lakes (Fig. 7). These evolving thermal gradients between land, lakes, and the atmosphere help modulate regional weather and climate and contribute to distinctive local phenomena such as lake breezes in summer and lake-effect snow in winter. Overall, both ERA5 and MERRA-2 demonstrate moderate errors relative to the observed latent and sensible heat fluxes, especially over lake locations, but do represent the overarching temporal trends based on seasonal warming/cooling.

Finally, we assessed average diurnal cycles using the hourly observations and reanalysis data. This analysis is especially useful because turbulent heat fluxes show strong intraday variability over land, while diurnal variations over lakes are much weaker (Fig. 9) because the high heat capacity of deep water suppresses the large temperature fluctuations that drive land-based diurnal flux cycles. Both ERA5 and MERRA-2 capture this basic land-lake contrast, although important biases remain. Over most lake sites, both reanalyses show a notable negative bias in latent and sensible heat fluxes (Fig. 9a, c, e, g). This is especially evident in summer for sensible heat flux, when the datasets show little to no heat release from the lakes, consistent with Fig. 8. This suggests that the modeled lake surface water temperature is too cool relative to observations and remains cooler than the overlying atmosphere, as shown in Fig. 2 of Wang et al. (2022) comparing ERA5 lake surface temperature with NOAA Great Lakes Surface Environmental Analysis (GLSEA) dataset.

Land locations show a markedly different behavior, with pronounced diurnal cycles tied to strong day-night temperature swings. These diurnal swings are larger in summer, when solar radiation and air-surface temperature gradients are stronger. For both sensible and latent heat fluxes, ERA5 and MERRA-2 overestimate the magnitude of these afternoon peaks, with biases reaching ~50–100 W m^{-2} at the diurnal maxima (Fig. 9f and h). This suggests that the reanalyses overestimate land-surface temperature and moisture relative to observations. One likely reason is the coarse land-surface representation in reanalyses. Many land sites reside in mixed land-use settings, including grass, trees, and roads, with a range of surface roughness and thermal properties that are not resolved in these coarse datasets. Another

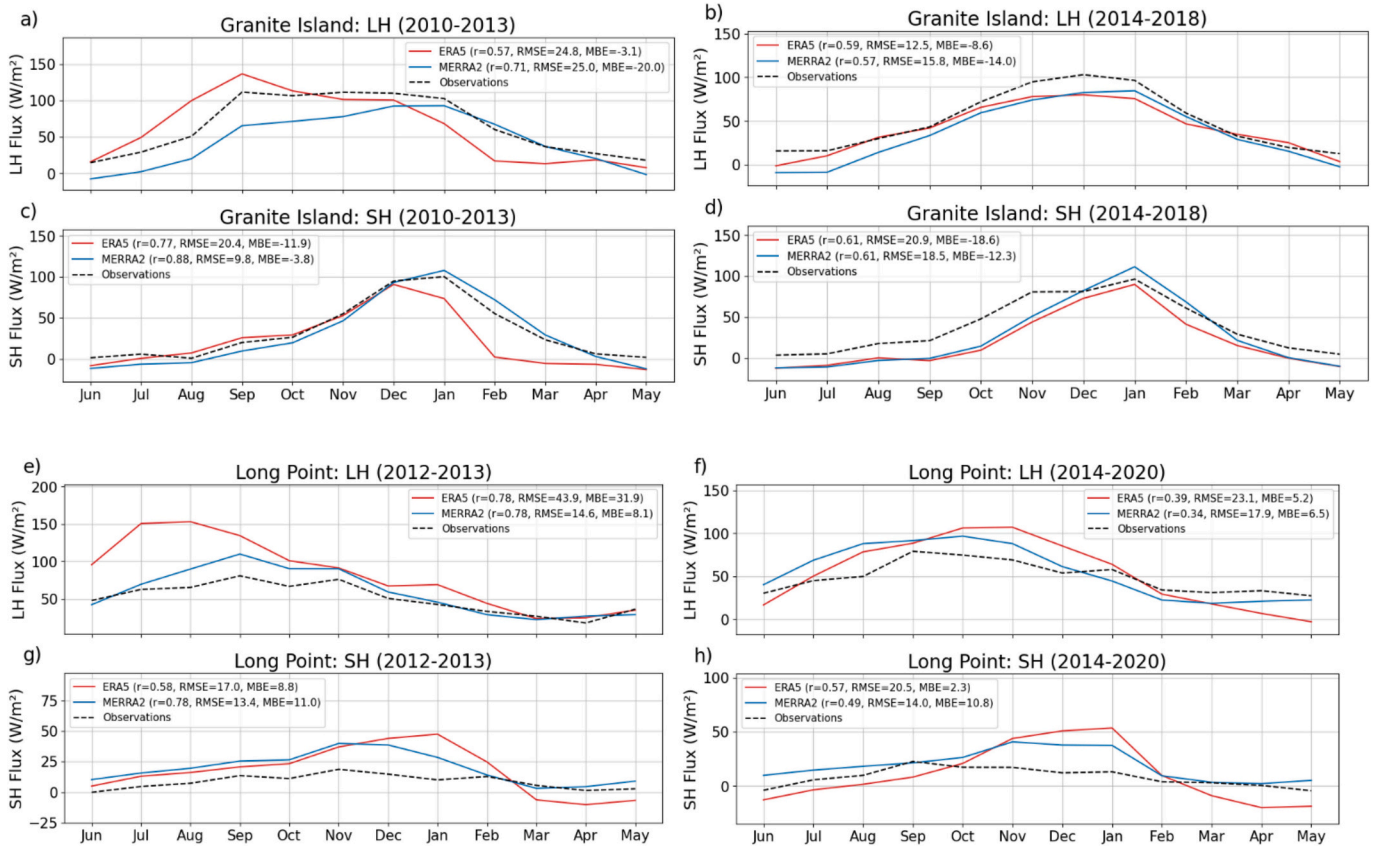

Fig. 7. Average monthly time series of sensible and latent heat flux from ERA5 and MERRA-2 over lake locations Granite Island (a–d), Lake Superior, and Long Point (e–h), Lake Erie prior to 2014 (left) and 2014 onward (right).

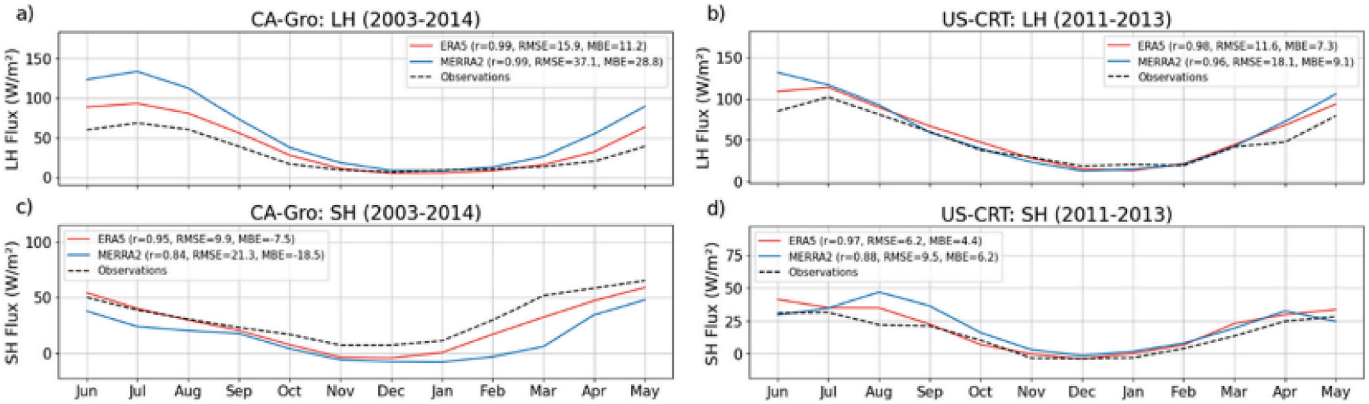

Fig. 8. Same as Fig. 7 but for land locations CA-Gro (a, c) and US-CRT (b, d). These land sites show all years available, as ERA5 lake ice and temperature errors prior to 2014 had negligible influence on the turbulent heat fluxes for land locations.

potential explanation for these latent and sensible heat overestimations in reanalysis datasets lies in the overestimation in downwelling shortwave fluxes, as shown in Section 3.1.1. Overall, these results suggest that the diurnal cycle is not reliably represented in the reanalyses and needs to be further evaluated by considering other atmospheric factors such as winds. The diurnal cycle of these reanalysis turbulent fluxes therefore should be used with caution for model evaluation.

3.2. Composite scores

Considering the radiative flux variables (downwelling shortwave and longwave), CERES-EBAF shows the strongest overall performance. Across each flux type and surface category, CERES-EBAF consistently achieves the highest composite score (Table 2). For shortwave fluxes, CLARA-A3 also performs well and ranks just behind CERES-EBAF, with composite scores of 0.57 at lake locations and 0.74 at land locations. The reanalysis datasets perform weaker overall, although ERA5 modestly outperforms MERRA-2 for all flux categories except for the turbulent heat fluxes (latent, sensible) over the lakes. In general, all datasets perform better over land than over lake locations for shortwave radiation, with land-based composite scores typically ~0.10–0.15 higher. For example, CERES-EBAF achieves a shortwave composite score of 0.77 over land compared with 0.63 over lakes (Table 2). For longwave radiation, CERES-EBAF again leads, with a composite score of 0.80 over lakes and 0.70 over land; in contrast to shortwave, differences between land and lake performance are smaller. CLARA-A3 does not show the same relative skill for longwave as for shortwave and is slightly outperformed by ERA5 over land. MERRA-2 shows the lowest performance for longwave radiation across both surface types.

For the turbulent heat fluxes, ERA5 performance drops, especially over the lake locations. Interestingly, despite only considering the temporal period outside of ERA5's known lake temperature and ice cover biases (after 2014), the dataset struggles to capture the latent and sensible heat fluxes over the lakes themselves. MERRA-2 does relatively better over water (mostly shallow lakes), with composite scores of 0.44 and 0.52 for sensible and latent heat, respectively. Alternatively, ERA5 dominates over land, with much higher composite scores for both latent and sensible heat fluxes, compared to MERRA-2.

We calculated two additional complementary skill scores for each gridded dataset relative to the in situ observations: the refined Willmott index of model performance (Willmott et al., 2012), d_r, and the modified coefficient of efficiency (Legates and McCabe Jr., 1999), E_1, which can also be found in Table 2. All skill scores agree with each other in terms of relative data performance. For clarification, the originally developed composite score was normalized within each surface–flux comparison group. As a result, the score represented a relative ranking among datasets rather than an absolute measure of model skill.

Specifically, RMSE and absolute MBE were scaled by the maximum RMSE and maximum absolute MBE among the datasets being compared, which forced the poorest-performing dataset within each group to receive a normalized value of zero for that metric. This approach amplified differences among datasets, even in cases where all datasets showed similar absolute agreement with the observations. For example, over lake locations for downwelling shortwave radiation, the composite scores varied from 0.30 to 0.63, whereas the refined Willmott scores ranged only from 0.80 to 0.83. The latter result is more consistent with the time series and distributional comparisons, which showed broadly similar performance among the datasets (Figs. 3–5). Therefore, the Willmott index and coefficient of efficiency skill scores were used to characterize absolute dataset performance and complement the original composite score which is to be interpreted as a relative ranking metric.

3.3. Regional intercomparison across four gridded datasets

Building on the site-based validation against in situ observations in previous sections, we next zoom out to the basin scale and examine how similar the four gridded datasets are in their spatial representation of Great Lakes surface radiative and turbulent heat-flux climatology. This regional intercomparison serves two purposes. First, it highlights where flux estimates are most robust across datasets and where they diverge, providing information useful for evaluation of conventional numerical models and emerging artificial intelligence (AI)-based emulators, as well as for surface energy balance analyses. Second, it reinforces that the Great Lakes region, especially the open-water regions, remain observationally under-sampled, which limits calibration of retrieval algorithms, parameterizations, and lake-process treatment in reanalyses. While this spatial analysis is not intended to be definitive, it provides a first step toward identifying where additional analysis and targeted process studies are most needed, and where future data-integration efforts could be most effective.

To quantify dataset similarity, we calculate the inter-dataset variance of seasonal mean fluxes. For downwelling shortwave radiation, datasets show relatively high consistency in winter and fall, with low variances (0–30 $W^2\,m^{-4}$) across much of the region (Fig. 10a). In spring, however, variance increases over the lakes, most prominently over Lake Superior, reaching ~120–160 $W^2\,m^{-4}$ (Fig. 10a). The springtime disagreement over the lakes likely reflects differences in lake surface state (e.g., temperature and partial ice cover) across datasets. During spring, Lake Superior often remains partially ice covered, and differences in how datasets detect or represent ice and surface temperature can produce strong lake–air thermal contrasts that influence low-cloud amounts (Ackerman et al., 2013; Wang et al., 2022).

Because downwelling shortwave radiation is highly sensitive to cloud fraction and optical thickness, even modest differences in surface

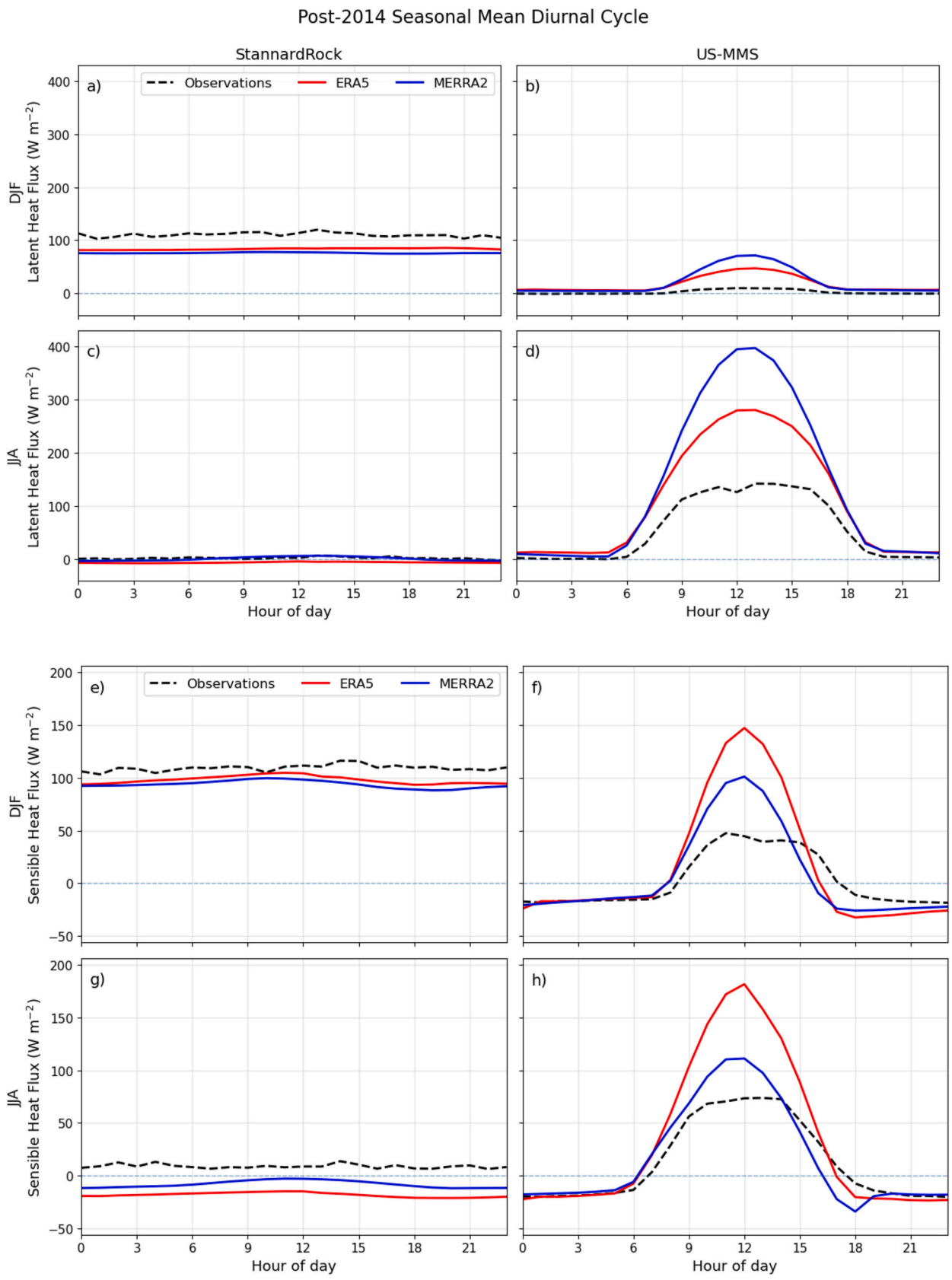

Fig. 9. Average diurnal cycles for winter and summer for one lake location, Stannard Rock (a, c, e, g) and one land location, US-MMS (b, d, f, h) for latent heat flux (a–d) and sensible heat flux (e–h).

Table 2
Composite scores for the gridded datasets for radiative and turbulent fluxes, considering all locations available in lake and over land. The closer the score is to 1, the better the performance is. Bold indicates best datasets for the particular variable over that particular region (lake or land).

Variable	Location	Dataset	Composite Score	Willmott Index (d_r)	Coefficient of Efficiency (E_1)
Shortwave	Lake	**CERES-EBAF**	**0.63**	**0.83**	**0.65**
		CLARA-A3	0.57	0.82	0.64
		ERA5	0.42	0.80	0.60
		MERRA-2	0.30	0.80	0.60
	Land	**CERES-EBAF**	**0.77**	**0.92**	**0.84**
		CLARA-A3	0.74	0.91	0.83
		ERA5	0.56	0.89	0.78
		MERRA-2	0.35	0.84	0.67
Longwave	Lake	**CERES-EBAF**	**0.80**	**0.90**	**0.81**
		CLARA-A3	0.58	0.82	0.63
		ERA5	0.50	0.78	0.57
		MERRA-2	0.34	0.68	0.36
	Land	**CERES-EBAF**	**0.70**	**0.87**	**0.74**
		CLARA-A3	0.57	0.84	0.69
		ERA5	0.59	0.85	0.70
		MERRA-2	0.35	0.72	0.43
Sensible Heat	Lake	ERA5	0.29	0.55	0.09
		MERRA-2	**0.44**	**0.65**	**0.30**
	Land	**ERA5**	**0.53**	**0.61**	**0.22**
		MERRA-2	0.22	0.48	−0.04
Latent Heat	Lake	ERA5	0.25	0.46	−0.07
		MERRA-2	**0.52**	**0.59**	**0.18**
	Land	**ERA5**	**0.39**	**0.63**	**0.26**
		MERRA-2	0.20	0.50	0.01

state or boundary-layer structure can translate into large differences in incoming solar radiation. In summer, larger variances ($> 160\ W^2\ m^{-4}$) appear mainly over land surrounding the lakes, particularly west of Lakes Michigan and Superior and southeast of Lake Erie. This enhanced summer land variance is driven largely by the difference between MERRA-2 and the other three datasets. When MERRA-2 is excluded (Fig. 10b), variance over land decreases in all seasons, especially summer (down to less than $20\ W^2\ m^{-4}$), suggesting that land-surface conditions (and associated boundary-layer and cloud responses) differ systematically in MERRA-2 relative to the other three datasets and may influence turbulent mixing and convective cloud development (Ackerman et al., 2013). One of the reasons could be the coarse resolution of MERRA-2 which limits its capability to capture many of the atmospheric properties that are important for deriving the shortwave radiation.

For downwelling longwave radiation, inter-dataset variance is generally smaller than for shortwave, with typical values of $\sim 50–90\ W^2\ m^{-4}$, even though the longwave flux magnitude itself is larger year-round ($\sim 250–400\ W\ m^{-2}$ over the Great Lakes region) compared to shortwave flux. Variance peaks in fall over the larger lakes (Superior, Michigan, and Huron), reaching $\sim 80–100\ W^2\ m^{-4}$, and then decreases into winter ($\sim 60–80\ W^2\ m^{-4}$). The lower variance is consistent with the fact that downwelling longwave is governed primarily by near-surface temperature, lower-tropospheric humidity, and cloud-base emissivity, which tend to vary more smoothly and are better constrained in reanalysis products (Mitra et al., 2025).

As with shortwave radiation, a substantial fraction of the land-based longwave variance is associated with MERRA-2: excluding MERRA-2 reduces the land variance by $30\ W^2\ m^{-4}$, leaving the largest remaining discrepancies concentrated over the lakes, with the greatest spread again occurring in spring ($\sim 70–80\ W^2\ m^{-4}$; Fig. 10b). This pattern suggests that lake surface conditions during the winter-to-spring ice-transition season remain a key source of uncertainty, potentially through their impacts on evaporation, near-surface temperature and moisture, and thus the thermodynamic profiles that control longwave radiation. The stronger land-based differences involving MERRA-2 may also reflect differences in aerosol treatment and assimilation within the GEOS system, which can alter surface shortwave radiation relative to products that use different aerosol assumptions or constraints (Buchard et al., 2017; Chakraborty and Lee, 2021; Randles et al., 2017).

Intercomparison of turbulent heat fluxes is limited to the two reanalysis datasets (ERA5 and MERRA-2) and reveals substantial differences (Fig. 10e–f). For the latent heat flux, inter-dataset variance is small ($< 20\ W^2\ m^{-4}$) during winter and early spring, when land surfaces are cold and occasional lake-ice formation suppresses evaporation. In summer, discrepancies increase substantially over land, when land surface temperatures are highest, with variances reaching $\sim 400\ W^2\ m^{-4}$ across the northern Great Lakes region (Fig. 10c). Higher variance also appears over Lake Erie in summer ($\sim 375\ W^2\ m^{-4}$), consistent with its small size and shallow depth relative to the other Great Lakes, which allow lake temperature (and evaporation) to respond rapidly to atmospheric warming. In fall, land-based differences diminish, while variance over the lakes increases and becomes the largest over four of the five Great Lakes ($240–400\ W^2\ m^{-4}$), when lake surface water temperatures remain warm and evaporation intensifies as the lakes release moisture to the atmosphere.

Because latent heat flux is highly sensitive to air-water humidity gradients, small differences between the two reanalyses can amplify into large inter-dataset variance. For sensible heat flux, elevated variance is concentrated near shorelines during winter and summer ($\sim 150\ W^2\ m^{-4}$), when lake–air temperature contrasts are strongest. Under these conditions, even modest differences in lake water or air temperatures between ERA5 and MERRA-2 can lead to pronounced differences in the sensible heat flux. Some shoreline variances may also be influenced by regridding, because nearshore grid cells can contain heterogeneous land–water fractions that differ between the native resolutions of the reanalyses. During spring, variances become more widespread within and around the lakes ($90–150\ W^2\ m^{-4}$), consistent with the strong sensitivity to differences in lake surface-state representation in reanalyses during the transition season, as discussed earlier.

4. Conclusions

Gridded datasets are widely used to characterize surface energy balance across many regions of the globe, yet they have not been systematically evaluated over the Great Lakes region. Here we evaluated four state-of-the-science gridded datasets: two satellite-based products (CERES-EBAF and CLARA-A3) and two reanalyses (ERA5 and MERRA-2), against available in situ observations to assess their ability to represent surface energy fluxes over and around the Great Lakes. We identified substantial inter-dataset differences in both surface downwelling shortwave and longwave radiation. Overall, the satellite-based products consistently outperformed the reanalyses, with CERES-EBAF demonstrating the highest overall skill and MERRA-2 the lowest for both shortwave and longwave radiation. CERES-EBAF most accurately reproduced monthly-mean radiative flux magnitudes and variability and best matched observed radiative flux distributions across diverse land- and lake-based measurement sites. For turbulent heat fluxes, uncertainties and inter-dataset differences were larger than for radiative

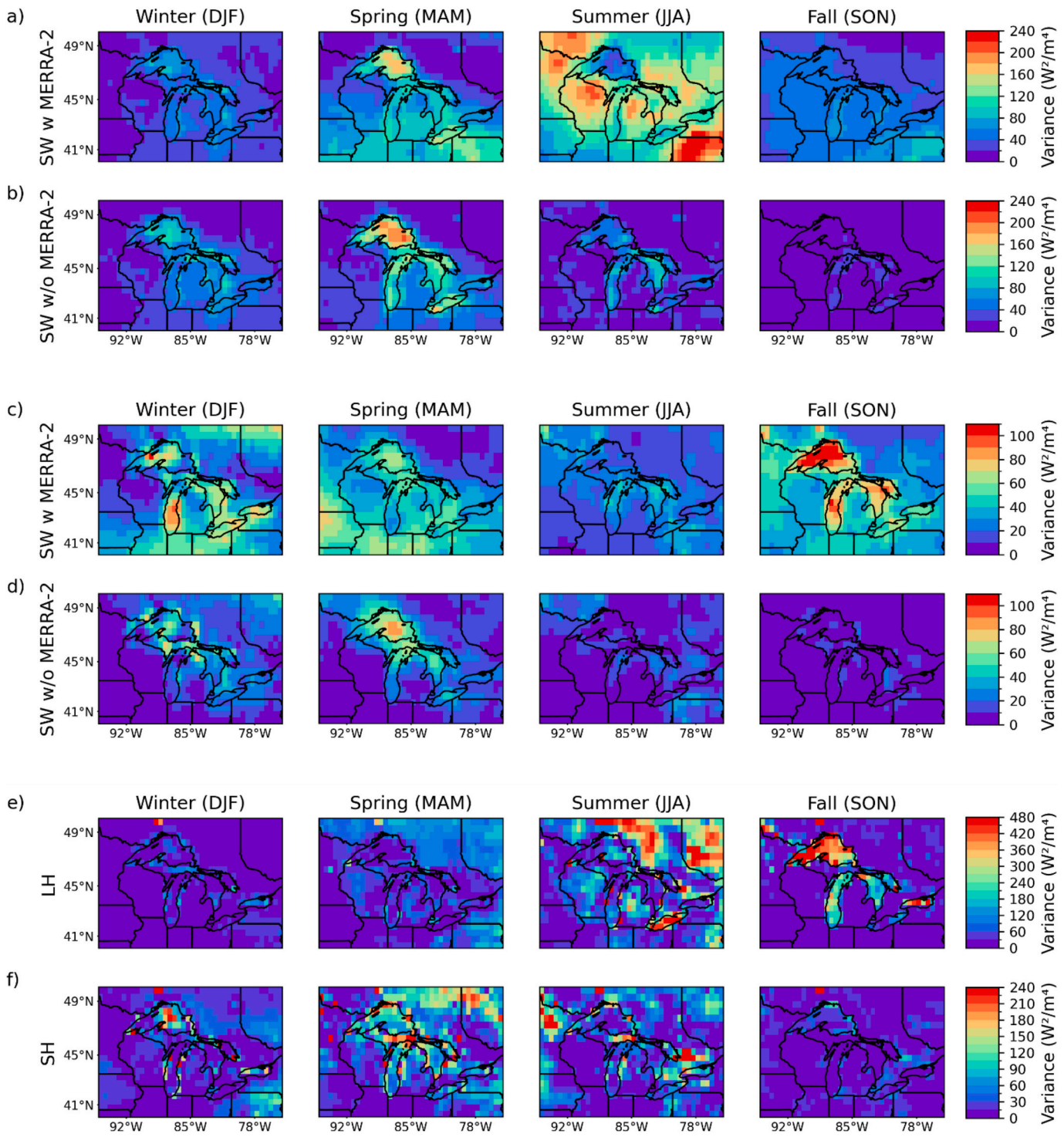

Fig. 10. Variances calculated for downwelling shortwave radiation (2001−2020) across the four gridded datasets, CERES, CLARA, ERA5, and MERRA-2 (a) and then across the datasets excluding MERRA-2 (b). Panels (c) and (d) are the same as panels (a) and (b) but for downwelling longwave radiation. Variances for latent heat (e) and sensible heat (f) fluxes calculated between reanalysis datasets ERA5 and MERRA-2.

fluxes. Both reanalyses show only moderate skill in capturing latent and sensible heat flux variability, underscoring the greater challenges in simulating surface–atmosphere exchange processes relative to radiative transfer, particularly over large freshwater systems where lake state (temperature and ice) strongly modulates fluxes.

This analysis motivates several next steps. First, targeted process diagnostics are needed to attribute discrepancies to specific drivers (e.g., cloud regime differences, aerosol assumptions, lake surface temperature/ice treatment, or boundary-layer humidity biases). Second, future work could evaluate whether dataset differences depend on specific meteorological and lake-state regimes, such as cold-air outbreaks, lake-effect periods, and spring ice-transition weeks, rather than reflecting only long-term climatological biases. Third, additional over-lake radiative flux, turbulent flux, and other meteorological observations, especially during shoulder seasons and high-evaporation periods, would substantially improve calibration and evaluation of both satellite

retrieval algorithms and reanalysis surface layer parameterizations. Finally, the insights from this intercomparison can inform strategies for integrated flux products over the Great Lakes, including dataset blending or bias correction conditioned on surface type, season, and lake state. As an example, the National Solar Radiation Data Base (Sengupta et al., 2018) relies heavily on MERRA-2 reanalysis for key atmospheric inputs (e.g., aerosol properties, atmospheric profiles, wind direction and speed, snow depth, surface temperature and pressure). However, it is important to note that these results are geographically specific to the Great Lakes region, and the findings of this investigation should not be extrapolated to other large lake systems, which may have different climate regimes, seasonal ice trends, lake bathymetry, and surrounding topography.

Our results and a previous study (Sheridan et al., 2025) on wind evaluation suggest that alternative reanalysis inputs (e.g., ERA5) may improve certain atmospheric properties used in solar radiation estimation over the Great Lakes region. More broadly, several machine-learning approaches could support flux integration at monthly scales by leveraging sparse in situ observations and complementary strengths across gridded datasets. A straightforward option is conditional bias correction (e.g., Gneiting et al., 2005), in which a flexible regression model (e.g., gradient-boosted trees or generalized additive models) learns systematic flux errors relative to observations as a function of surface type (land versus lake), season, and lake state (e.g., ice fraction, lake surface temperature), and applies these corrections across the basin. A second option is regime-dependent dataset blending (e.g., a mixture-of-experts framework; Chevuturi et al., 2023) that learns spatially and seasonally varying weights for each product, allowing the integrated estimate to draw more heavily on the most skillful dataset under specific conditions (e.g., satellite products for shortwave radiation, and reanalyses where turbulent flux information is required).

As of writing, continuing advancements are being made in the development and improvement of flux products, with efforts focusing on increasing spatial resolution and developing more realistic representations of lake-atmosphere exchanges. Specifically, ECMWF's forthcoming ERA6 reanalysis increases horizontal resolution to 14 km and improves the embedded FLake lake model, which may lead to more realistic lake-atmosphere interactions and may consequentially improve the associated surface fluxes (Hersbach et al., 2026). Likewise, satellite-based datasets are undergoing enhancements. Notably, CERES Edition 5 is well underway and incorporates updated cloud retrievals, radiative-transfer calculations, aerosol information, and more consistent MODIS and VIIRS observations, which are all enhancements that will likely improve surface radiative flux representations within the Great Lakes region (NASA CERES Data Management Team, 2025). These developments suggest that the biases documented in the present benchmarking study should be periodically reassessed as new product generations become available.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

Portions of the manuscript text were edited for clarity and grammar with assistance from ChatGPT (GPT-5.2, OpenAI). Additionally, Python scripts for processing and analyzing data were refined using ChatGPT. After using this tool, the authors have reviewed and edited the content as needed and take full responsibility for all content in the published article.

CRediT authorship contribution statement

Kyle Peco: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Arka Mitra:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **Jiali Wang:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Conceptualization. **T.C. Chakraborty:** Writing – review & editing, Supervision. **William J. Pringle:** Writing – review & editing, Supervision. **John D. Lenters:** Writing – review & editing, Supervision, Resources, Data curation. **Peter D. Blanken:** Writing – review & editing, Supervision, Resources, Data curation. **Christopher Spence:** Writing – review & editing, Supervision, Resources, Data curation. **Pengfei Xue:** Writing – review & editing, Supervision. **Zhao Yang:** Writing – review & editing, Supervision. **Yun Qian:** Writing – review & editing, Supervision. **Robert Hetland:** Writing – review & editing, Supervision, Project administration.

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.

Acknowledgements

This study is primarily supported by COMPASS-GLM, a multi-institutional project supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research as part of the Regional and Global Modeling and Analysis (RGMA) program, Multi-sector Dynamics Modeling (MSD) program, and Earth System Model Development (ESMD) program. This study is also Contribution No. 143 of the Great Lakes Research Center at Michigan Technological University. Computational resources are provided by the DOE-supported National Energy Research Scientific Computing Center and Argonne Leadership Computing Facility. The Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830.

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.jglr.2026.102994.

Data availability

All datasets, including the gridded datasets and the in situ data used in this study are freely available. ERA5 reanalysis data is accessible through Climate Data Store: https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels. CLARA-A3 satellite data is also available through the climate data store: https://cds.climate.copernicus.eu/datasets/satellite-surface-radiation-budget. MERRA-2 reanalysis data is available through NASA Earthdata: https://disc.gsfc.nasa.gov/datasets?project=MERRA-2. CERES-EBAF data is also accessible through NASA Earthdata: https://asdc.larc.nasa.gov/project/CERES/CERES_EBAF_Edition4.2

Observational radiative flux and turbulent heat flux data were obtained from the Great Lakes Evaporation Network (GLEN). Data for Stannard Rock, Long Point, and Granite Island are publicly available through the AmeriFlux data portal: https://ameriflux.lbl.gov/. Additional GLEN sites used in this study (White Shoal and Spectacle Reef) are described in the GLEN dataset publication (Lenters et al., 2013) and are available upon request or through associated project repositories. The seven land-based AmeriFlux sites used in this study (CA-Gro, CA-TPD, US-CRT, US-MMS, US-ORv, US-Syv, and US-UMB) can also be accessed and downloaded from the AmeriFlux data portal: https://ameriflux.lbl.gov/. The four buoys used for the shortwave radiative flux evaluation can be downloaded from the National Data Buoy Center portal: National Data Buoy Center.

Data processing scripts were written in Python and processed GLEN data are being reviewed and will be published on ESS-Dive.

References

Ackerman, S.A., Heidinger, A., Foster, M.J., Maddux, B., 2013. Satellite regional cloud climatology over the Great Lakes. Remote Sens. 5, 6223–6240.

Bodas-Salcedo, A., Williams, K.D., Ringer, M.A., Beau, I., Cole, J.N.S., Dufresne, J.-L., Koshiro, T., Stevens, B., Wang, Z., Yokohata, T., 2014. Origins of the solar radiation biases over the Southern Ocean in CFMIP2 models. J. Clim. 27, 41–56.

Buchard, V., Randles, C.A., da Silva, A.M., Darmenov, A., Colarco, P.R., Govindaraju, R., Ferrare, R., Hair, J., Beyersdorf, A.J., Ziemba, L.D., Yu, H., 2017. The MERRA-2 aerosol reanalysis, 1980 onward. Part II: evaluation and case studies. J. Clim. 30, 6851–6872.

Cesana, G., Chepfer, H., 2012. How well do climate models simulate cloud vertical structure? A comparison between CALIPSO-GOCCP satellite observations and CMIP5 models. Geophys. Res. Lett. 39, L20803.

Chakraborty, T., Lee, X., 2021. Large differences in diffuse solar radiation among current-generation reanalysis and satellite-derived products. J. Clim. 34, 6635–6650.

Charusombat, U., Fujisaki-Manome, A., Gronewold, A.D., Lofgren, B.M., Anderson, E.J., Blanken, P.D., Spence, C., Lenters, J.D., Xiao, C., Fitzpatrick, L.E., Cutrell, G., 2018. Evaluating and improving modeled turbulent heat fluxes across the north American Great Lakes. Hydrol. Earth Syst. Sci. 22, 5559–5578.

Chevuturi, A., Tanguy, M., Facer-Childs, K., Martínez-de la Torre, A., Sarkar, S., Thober, S., Samaniego, L., Rakovec, O., Kelbling, M., Sutanudjaja, E.H., Wanders, N., Blyth, E., 2023. Improving global hydrological simulations through bias-correction and multi-model blending. J. Hydrol. 621, 129607.

Chou, M., Suarez, M.J., 1994. Technical Report Series on Global Modeling and Data Assimilation: An Efficient Thermal Infrared Radiation Parameterization for Use in General Circulation Models. NASA Technical Memorandum 104606, Vol. 3. https://ntrs.nasa.gov/api/citations/19950009331/downloads/19950009331.pdf [accessed 30 September 2026].

Chou, M., Suarez, M.J., 1999. Technical Report Series on Global Modeling and Data Assimilation 15: A Solar Radiation Parameterization for Atmospheric Studies. NASA Technical Memorandum 104606, Vol. 15. https://ntrs.nasa.gov/api/citations/19990060930/downloads/19990060930.pdf [accessed 30 September 2026].

Cracknell, A.P., 1997. The Advanced Very High Resolution Radiometer AVHRR. CRC Press, London.

ECMWF, 2016. IFS Documentation CY41R2. Operational Implementation 8 March 2016. Part IV: Physical Processes. European Centre for Medium-Range Weather Forecasts, Reading, England.

Fiedler, E.K., Martin, M.J., Roberts-Jones, J., 2014. An operational analysis of lake surface water temperature. Tellus 66, 21247.

Flood, B., Wells, M., Dunlop, E., Young, J., 2021. Vertical oscillations of the thermocline caused by internal waves modify Coldwater pelagic fish distribution: results from a large stratified lake. J. Great Lakes Res. 47, 1386–1399.

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J.E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S.D., Sienkiewicz, M., Zhao, B., 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J. Clim. 30, 5419–5454.

Gneiting, T., Raftery, A.E., Westveld, A.H., Goldman, T., 2005. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon. Weather Rev. 130, 1098–1118.

Helfand, H.M., Schubert, S.D., 1995. Climatology of the simulated Great Plains low-level jet and its contribution to the continental moisture budget of the United States. J. Clim. 8, 784–806.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., Thépaut, J.-N., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146, 1999–2049.

Hersbach, H., Bell, B., Herbert, C., Kaandorp, M., Poli, P., Radu, R., Schepers, D., Simmons, A., Soci, C., Steele, L., Strube, C., Buontempo, C., McNally, A., 2026. From ERA5 to ERA6. In: EMS Annual Meeting 2026, Utrecht, Netherlands, 6–11 September 2026, EMS2026–251. https://doi.org/10.5194/ems2026-251.

Josey, S.A., Gulev, S., Yu, L., 2013. Exchanges through the ocean surface. In: Siedler, G., Griffies, S., Gould, J., Church, J. (Eds.), Ocean Circulation and Climate: A 21st Century Perspective, 2nd ed. Academic Press, pp. 115–140.

Karlsson, K.-G., Stengel, M., Meirink, J.F., Riihelä, A., Trentmann, J., Akkermans, T., Stein, D., Devasthale, A., Eliasson, S., Johansson, E., Håkansson, N., Solodovnik, I., Benas, N., Clerbaux, N., Selbach, N., Schröder, M., Hollmann, R., 2023. CLARA-A3: the third edition of the AVHRR-based CM SAF climate data record on clouds, radiation and surface albedo covering the period 1979 to 2023. Earth Syst. Sci. Data 15, 4901–4926.

Kato, S., Loeb, N.G., Rose, F.G., Ham, S.-H., Thorsen, T.J., Rutan, D.A., Miller, W.F., Caldwell, T.E., Doelling, D.R., Sun-Mack, S., 2025. Seamless continuity in CERES energy balanced and filled (EBAF) surface radiation budget across multiple satellites. J. Clim. 26, 2719–2740.

Kato, S., Rose, F.G., Rutan, D.A., Thorsen, T.J., Loeb, N.G., Doelling, D.R., Huang, X., Smith, W.L., Su, W., Ham, S.-H., 2018. Surface irradiances of edition 4.0 clouds and the earth's radiant energy system (CERES) energy balanced and filled (EBAF) data product. J. Clim. 31, 4501–4527.

Kayastha, M.B., Ye, X., Huang, C., Xue, P., 2022. Future rise of the Great Lakes water levels under climate change. J. Hydrol. 612, 128205.

Kirillin, G., Hochschild, J., Mironov, D., Terzhevik, A., Golosov, S., Nützmann, G., 2011. FLake-global: online lake model with worldwide coverage. Environ. Model. Softw. 26, 683–684.

Kondo, J., 1975. Air-sea bulk transfer coefficients in diabatic conditions. Bound.-Layer Meteorol. 9, 91–112.

Kristovich, D.A.R., Laird, N.F., 1998. Observations of widespread lake-effect cloudiness: influences of lake surface temperature and upwind conditions. Weather Forecast. 13, 811–821.

Large, W.G., Pond, S., 1981. Open Ocean momentum flux measurements in moderate to strong winds. J. Phys. Oceanogr. 11, 324–336.

Legates, D.R., McCabe Jr., G.J., 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res. 35, 233–241.

Lenters, J.D., Anderton, J.B., Blanken, P., Spence, C., Suyker, A.E., 2013. Assessing the impacts of climate variability and change on Great Lakes evaporation. In: Brown, D., Bidwell, D., Briley, L. (Eds.), 2011 Project Reports. Great Lakes Integrated Sciences and Assessments (GLISA) Center. http://glisaclimate.org/media/GLISA_Lake_Evaporation.pdf [accessed September 30, 2026].

Liu, J., Curry, J.A., Rossow, W.B., Key, J.R., Wang, X., 2005. Comparison of surface radiative flux data sets over the Arctic Ocean. J. Geophys. Res. Oceans 110, 1–13.

Loeb, N.G., Doelling, D.R., Wang, H., Su, W., Nguyen, C., Corbett, J.G., Liang, L., Mitrescu, C., Rose, F.G., Kato, S., 2018. Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) top-of-atmosphere (TOA) edition-4.0 data product. J. Clim. 31, 895–918.

Marchuk, G.I., Mikhailov, G.A., Nazaraliev, M.A., Darbinjan, R.A., Kargin, B.A., Elepov, B.S., 1980. Monte Carlo Methods in Atmospheric Optics. In: Springer Series in Optical Sciences, vol. 12. Springer-Verlag, Berlin.

Meindl, E.A., Hamilton, G.D., 1992. Programs of the National Data Buoy Center. Bull. Am. Meteorol. Soc. 73, 985–993.

Minallah, S., Steiner, A.L., 2021. The effects of lake representation on the regional hydroclimate in the ECMWF reanalyses. Mon. Weather Rev. 149, 1747–1766.

Mitra, A., Ghate, V., Rutan, D., 2025. Comparison of CERES SYN1deg radiative fluxes with those derived from observations at the ARM ENA site. J. Clim. 38, 6475–6491.

Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J., Clough, S.A., 1997. Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. J. Geophys. Res. Atmos. 102, 16663–16682.

Molod, A., Takacs, L., Suarez, M., Bacmeister, J., 2015. Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2. Geosci. Model Dev. 8, 1339–1356.

Monin, A.S., Obukhov, A.M., 1954. Basic laws of turbulent mixing in the atmosphere near the ground. Tr. Akad. Nauk SSSR Geofiz. Inst. 24, 163–187.

Moukomla, S., Blanken, P.D., 2017. Estimating the Great Lakes net radiation using satellite remote sensing and MERRA reanalysis. Int. J. Digit. Earth 10, 764–784.

NASA CERES Data Management Team, 2025. CERES Data Management Team spring 2025 Meeting: CERES Edition 5. NASA Langley Research Center.

Nicholls, E.M., Spence, C., Hedstrom, N., Lenters, J.D., Blanken, P.D., 2024. Lake Superior evaporation: a long-term eddy covariance dataset at Stannard Rock Lighthouse (2008–2022). Sci. Data 11, 1152.

Notaro, M., Holman, K., Zarrin, A., Fluck, E., Vavrus, S., Bennington, V., 2013. Influence of the Laurentian Great Lakes on regional climate. J. Clim. 26, 789–804.

Novick, K.A., Biederman, J.A., Desai, A.R., Litvak, M.E., Moore, D.J.P., Scott, R.L., Torn, M.S., 2018. The AmeriFlux network: a coalition of the willing. Agric. For. Meteorol. 249, 444–456.

Obukhov, A.M., 1971. Turbulence in an atmosphere with a non-uniform temperature. Bound.-Layer Meteorol. 2, 7–29.

Randles, C.A., da Silva, A.M., Buchard, V., Colarco, P.R., Darmenov, A., Govindaraju, R., Smirnov, A., Holben, B., Ferrare, R., Hair, J., Shinozuka, Y., Flynn, C.J., 2017. The MERRA-2 aerosol reanalysis, 1980 onward. Part I: system description and data assimilation evaluation. J. Clim. 30, 6823–6850.

Reichle, R.H., Draper, C.S., Liu, Q., Girotto, M., Mahanama, S.P.P., Koster, R.D., De Lannoy, G.J.M., 2017. Assessment of MERRA-2 land surface hydrology estimates. J. Clim. 30, 2937–2960.

Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R.D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C.R., Reichle, R., Robertson, F. R., Ruddick, A.G., Sienkiewicz, M., Woollen, J., 2011. MERRA: NASA'S modern-era retrospective analysis for research and applications. J. Clim. 24, 3624–3648.

Sengupta, M., Xie, Y., Lopez, A., Habte, A., Maclaurin, G., Shelby, J., 2018. The National Solar Radiation Data Base (NSRDB). Renew. Sust. Energ. Rev. 89, 51–60.

Sheridan, L.M., Wang, J., Draxl, C., Bodini, N., Phillips, C., Duplyakin, D., Tinnesand, H., Rai, R.K., Flaherty, J.E., Berg, L.K., Jung, C., Young, E., Kotamarthi, R., 2025. Performance of wind assessment datasets in United States coastal areas. Wind Energy Sci. 10, 1551–1574.

Simmons, A., Hersbach, H., Muñoz-Sabater, J., Nicolas, J., Vamborg, F., Berrisford, P., de Rosnay, P., Willett, K., Woollen, J., 2021. Low Frequency Variability and Trends in Surface Air Temperature and Humidity from ERA5 and Other Datasets. ECMWF Technical Memorandum 881. https://doi.org/10.21957/ly5vbtbfd.

Stubenrauch, C.J., Rossow, W.B., Kinne, S., Ackerman, S., Cesana, G., Chepfer, H., Di Girolamo, L., Getzewich, B., Guignard, A., Heidinger, A., Maddux, B.C., Menzel, W. P., Minnis, P., Pearl, C., Platnick, S., Poulsen, C., Riedi, J., Sun-Mack, S., Walther, A., Winker, D., Zeng, S., Zhao, G., 2013. Assessment of global cloud datasets from satellites: project and database initiated by the GEWEX Radiation Panel. Bull. Am. Meteorol. Soc. 94, 1031–1049.

Tian, Y., Zhong, D., Ghausi, S.A., Wang, G., Kleidon, A., 2023. Understanding variations in downwelling longwave radiation using Brutsaert's equation. Earth Syst. Dyn. 14, 1363–1374.

Trenberth, K.E., Fasullo, J.T., Kiehl, J., 2009. Earth's global energy budget. Bull. Am. Meteorol. Soc. 90, 311–324.

Troy, T.J., Wood, E.F., 2009. Comparison and evaluation of gridded radiation products across northern Eurasia. Environ. Res. Lett. 4, 045008.

Wang, J., Xue, P., Peco, K., Huang, C., Pringle, W.J., Yang, Z., Chakraborty, T.C., Qian, Y., Hetland, R.D., 2025. Influence of Lake Ice Biases in Reanalysis Data on Downscaled Climate Simulations over the Great Lakes Region. ESS Open Archive. Preprint. https://doi.org/10.22541/essoar.175916613.30361389/v1.

Wang, J., Xue, P., Pringle, W., Yang, Z., Qian, Y., 2022. Impacts of lake surface temperature on the summer climate over the Great Lakes region. J. Geophys. Res. Atmos. 127, e2021JD036231.

Willmott, C.J., Robeson, S.M., Matsuura, K., 2012. A refined index of model performance. Int. J. Climatol. 32, 2088–2094.

Woolway, R.I., 2023. The pace of shifting seasons in lakes. Nat. Commun. 14, 2101.

Xue, P., Pal, J.S., Ye, X., Lenters, J.D., Huang, C., Chu, P.Y., 2017. Improving the simulation of large lakes in regional climate modeling: two-way lake–atmosphere coupling with a 3D hydrodynamic model of the Great Lakes. J. Clim. 30, 1605–1627.

Xue, P., Rizzoli-Malanette, P., Wei, J., Eltahir, E., 2020. Coupled ocean-atmosphere modeling over the maritime continent: a review. J. Geophys. Res. Oceans 125, e2019JC014978.