The Effects of Lake Representation on the Regional Hydroclimate in the ECMWF Reanalyses

SAMAR MINALLAH and ALLISON L. STEINER

Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, Michigan

(Manuscript received 20 December 2020, in final form 15 February 2021)

ABSTRACT: Lakes are an integral part of the geosphere, but they are challenging to represent in Earth system models, which either exclude lakes or prescribe properties without simulating lake dynamics. In the ECMWF interim reanalysis (ERA-Interim), lakes are represented by prescribing lake surface water temperatures (LSWT) from external data sources, while the newer-generation ERA5 introduces the Freshwater Lake (FLake) parameterization scheme to the modeling system with different LSWT assimilation data sources. This study assesses the performance of these two reanalyses over three regions with the largest lakes in the world (Laurentian Great Lakes, African Great Lakes, and Lake Baikal) to understand the effects of their simulation differences on hydrometeorological variables. We find that differences in lake representation alter the associated hydrological and atmospheric processes and can affect regional hydroclimatic assessments. There are prominent differences in LSWT between the two datasets that influence the simulation of lake-effect snowstorms in the Laurentian winters and lake–land-breeze circulation patterns in the African region. Generally, ERA5 has warmer LSWT in all three regions for most months (by 2–12 K) and its evaporation rates are up to twice the magnitudes in ERA-Interim. In the Laurentian lakes, ERA5 has strong biases in LSWT and evaporation magnitudes. Over Lake Baikal and the African Great Lakes, ERA5 LSWT magnitudes are closer to satellite-based datasets, albeit with a warm bias (1–4 K), while ERA-Interim underestimates the magnitudes. ERA5 also simulates intense precipitation hot spots in lake proximity that are not visible in ERA-Interim and other observation-based datasets. Despite these limitations, ERA5 improves the simulation of lake–land circulation patterns across the African Great Lakes.

KEYWORDS: Inland seas/lakes; Lake effects; Precipitation; Climatology; Evaporation; Surface temperature; Reanalysis data

1. Introduction

Lakes are an important component of regional hydroclimates as they modulate the water and energy budgets through their markedly different heat capacities and albedo compared to terrestrial surfaces (Subin et al. 2012; Thiery et al. 2015). The presence of lakes alters micro- and mesoscale patterns of moisture budget components such as evaporation and precipitation (Sun et al. 2015; Wright et al. 2013), the surrounding air temperature profiles (Samuelsson et al. 2010), cloud formation (Holroyd 1971; Laird et al. 2017), and wind patterns (Crosman and Horel 2010; Gerken et al. 2014). They can produce localized extreme events in the form of wintertime lake-induced snowstorms in the northern midlatitudes (Notaro et al. 2013a,b; Zhao et al. 2012), summer convective storms due to lake-breeze fronts in the Laurentian Great Lakes region (Alexander et al. 2018; Boodoo et al. 2015; King et al. 2003), and severe thunderstorms over some lakes in Africa due to the diurnal cycle of lake–land-breeze circulation (Thiery et al. 2015, 2016). In the regions where the lakes cover considerable surface area (e.g., the Laurentian Great Lakes region), or where the concentration of small lakes is high (e.g., parts of Canada, Scandinavia, and the Tibetan Plateau), the diurnal and seasonal climatic patterns can be notably influenced by the presence of these water bodies (McDonald et al. 2012; Rouse et al. 2005; Samuelsson et al. 2010).

Despite their importance, lakes are difficult to simulate in Earth system models due to various challenges such as complex or unknown bathymetry, erratic changes in water levels and surface area, and varying geophysical and biochemical properties depending on the lake type (Pekel et al. 2016; Subin et al. 2012). Lake models are rarely coupled with climate and weather forecast models, which either exclude lake representation altogether (Mironov et al. 2010), prescribe sea surface temperatures to represent water temperatures (Fiedler et al. 2014), or introduce a shallow 1D lake model (Martynov et al. 2010, 2012; Samuelsson et al. 2010). The inclusion of a complex three-dimensional lake representation is atypical due to the computational costs, with only a few research applications for select large lakes that include dynamical lake processes (Mironov et al. 2010 and citations therein). Moreover, information on their physical characteristics and observation-based properties [e.g., lake surface water temperatures (LSWT)] are sparse and oftentimes subject to biases (Messager et al. 2016).

Reanalysis datasets provide spatiotemporally consistent information of the surface and atmospheric fields and are therefore used as model forcing data. This necessitates the assessment of lake behavior in these reanalyses to ensure they can capture the hydrological and atmospheric processes associated with lakes. European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim; Dee et al. 2014; Dee et al. 2011) has been used in past assessments of lake hydroclimates and as a forcing dataset (Dutra et al. 2010; and citations therein). Moreover, information on their physical characteristics and observation-based properties [e.g., lake surface water temperatures (LSWT)] are sparse and oftentimes subject to biases (Messager et al. 2016).
Frassl et al. 2018; Huziy and Sushama 2017). The fifth major global reanalysis produced by ECMWF (ERA5; Hersbach et al. 2020) has a higher spatiotemporal resolution (31-km hourly as compared to 79-km 6-hourly resolution of ERA-Interim) and can potentially provide improvements over ERA-Interim, which has been shown for some applications, for example, hydrological and agricultural modeling (Betts et al. 2019; Graham et al. 2019; Tarek et al. 2020).

Both ECMWF reanalyses use the Integrated Forecasting System (IFS) numerical model as their core, however, their representation of inland water bodies varies. Previously, the land surface heterogeneity in IFS was represented by the Hydrology Tiled ECMWF Scheme of Surface Exchanges over Land (HTESSEL) (ECMWF 2015) in which the lake tiles were prescribed a LSWT. In 2015, a lake model was introduced in IFS based on the Freshwater Lake (FLake; Mironov 2008; Mironov et al. 2010) parameterization scheme (Balsamo et al. 2012; ECMWF 2016). This relatively recent inclusion of a lake model in the IFS, together with changes in assimilation datasets used for LSWT, has altered the lake simulations in the newer ERA5.

Dutra et al. (2010) conducted a global offline assessment of FLake coupled with HTESSEL and found that the model simulations with lakes increased evaporation rates as compared to evaporation over resolved lakes in ERA-Interim. They concluded that on annual time scales, the differences were small and localized in nature but “the impact can be significant in some specific areas and seasons.” Balsamo et al. (2012), in their assessment of the IFS implementation of FLake, found that this inclusion resulted in a “non-negligible signal in near surface temperature as a consequence of the lake thermal inertia” with a positive bias over parts of Laurentian Great Lakes and negative bias over Norwegian lakes when compared to satellite-based MODIS lake temperature estimates. Further, they found improved model performance in predicting summertime LSWT and poorer results in the winter season due to errors in lake-ice initial conditions, highlighting the sensitivity to this forcing. Thiery et al. (2014) tested FLake for two African lakes, Kivu and Tanganyika, and found the model reproduced the lake thermal structure, but stated that deep warm lakes required further consideration, especially when coupled with weather prediction or climate models where the forcing data may be biased. They concluded that accurate representation of the lake thermal structure is sensitive to meteorological forcing and input data such as lake depth.

This study aims to assess the performance of the two generations of ECMWF reanalyses (ERA-Interim and ERA5) in their representation of inland water bodies and simulation of LSWT for three regions with substantial lake presence (the Laurentian Great Lakes, the African Great Lakes, and Lake Baikal). We investigate the role of lakes as an important component of the regional hydroclimate, and their influence in generating the lake-effect processes (circulation patterns and local precipitation) by analyzing hydroclimatic variables (vapor pressure gradient, lake evaporation, surface winds, cloud water content, and moisture flux convergence) on long-term climatological (2001–18) and meteorological (daily, subdaily) time scales. In all three regions, we observe considerable differences in LSWT, lake evaporation rates, and precipitation patterns around the lakes between the two reanalyses. Over the Laurentian Great Lakes region, the failure of LSWT representation in ERA5 was mentioned by Hersbach et al. (2020), and we further discuss the consequences of these biases in LSWT on the lake-effect processes.

2. Study domain

We investigate three regions containing large lakes by volume and surface area: the Laurentian Great Lakes in North America, including Lakes Superior, Michigan, Huron, Erie, and Ontario (Fig. 1a); Lake Baikal in Russia, which is the largest freshwater lake in the world by volume (Fig. 1b); and the African Great Lakes, including Lakes Victoria, Tanganyika, and Malawi, which form part of the East African Rift System (Fig. 1c). Lakes Tanganyika, Malawi, and Baikal are narrow and deep rift lakes, while the Laurentian lakes are prehistoric glacial lakes with large surface area (see Table S1 in the online supplemental material). The large volume and surface area of these lakes modulate the local temperatures, wind patterns, and moisture content both at meteorological (e.g., lake-effect weather phenomenon) and long-term climatic time scales (Notaro et al. 2013b; Scott and Huff 1996; Thiery et al. 2015; Williams et al. 2014).

3. Datasets

ERA5 is the newest generation reanalysis from ECMWF with approximately 31-km global resolution and hourly analysis and forecast fields (Hersbach et al. 2020). The previous-generation ERA-Interim has a coarser ~79-km global spatial resolution with 3-hourly estimates of the surface and two-dimensional parameters, and 6-hourly estimates of the three-dimensional fields (Dee et al. 2011). For consistency in analysis, we regridded both datasets to 0.5° × 0.5° resolution and averaged monthly fields over the 2001–18 period for long-term climatologies. To assess the effects of LSWT on other meteorological fields, we further conduct case studies at daily and subdaily time scales using the native spatial resolutions.

ECMWF uses the IFS as its atmospheric model and data assimilation system. ERA-Interim is based on the older version of IFS (Cycle 31R2 released in 2006), where the inland water bodies are represented in the land surface hydrology model HTESSEL, that includes various “tiles” to represent surface types (e.g., open water, frozen water, bare ground, vegetation; Balsamo et al. 2009) and only gridscale lakes (lake cover > 50%) are considered. In this scheme, the LSWT is prescribed as follows:

- The surface temperatures for the Laurentian Great Lakes are integrated from external sea surface temperature (SST) sources including National Centers for Environmental Prediction (NCEP) Real-Time Global sea surface temperature (from 2002 to 2009) and Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA; from 2009 onward) (Dee et al. 2011).
For all other lakes not resolved in OSTIA or NCEP (including the African Lakes and Lake Baikal), one month lagged 2-m temperature from ERA-40 is used for LSWT, which does not simulate diurnal changes and interannual variability. In the newer IFS versions, starting from Cycle 41R1 released in 2015 and Cycle 41R2 released in 2016 (which forms the basis of the ERA5) the LSWT representation differs from that of ERA-Interim:

- The surface temperature for the Laurentian Great Lakes is extracted from external SST data from the Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST2) and OSTIA. However, from 1979 to 2013, a failure to utilize these observationally based SST datasets resulted in errors over the Great Lakes (Hersbach et al. 2020), and the FLake model is used instead (H. Hersbach 2020, personal communication).

- For all other inland water bodies, FLake model is modified and coupled with the HTESSEL surface scheme. FLake is a one-dimensional, two-layer scheme that represents an evolving temperature profile of the lake and the integral budgets of the heat and kinetic energy of the two lake layers (upper mixed layer and lake thermocline). This model has two time-independent inputs, lake cover and lake depth, that allows inclusion of subgrid lakes (lake cover < 50%) (ECMWF 2016).

In the following assessment, this change in simulation of the LSWT (i.e., inclusion of HadISST2 and the FLake model)
between ERA-Interim and ERA5 affects the water temperature seasonality and other related meteorological fields. To study these differences, we analyze the following fields: 1) skin/surface temperature (K), which is defined as “the temperature of the surface at radiative equilibrium” and is identical to SST for inland water bodies; 2) evaporation rate (mm day$^{-1}$) where negative values show condensation; 3) total precipitation (mm), which is the sum of convective and large-scale precipitation in both solid and liquid form; 4) horizontal and vertical wind velocities (m s$^{-1}$) at pressure levels, where the vertical wind velocity is converted from omega (Pa s$^{-1}$) to meters per second and scaled by a factor of 5; 5) 10-m u- and v-wind components (m s$^{-1}$); and 6) vertically integrated moisture flux divergence (mm day$^{-1}$). In IFS, the evaporation flux over water bodies is computed as

$$E = \rho_a |U_L| C_H [q_L - q_{sat}(T_{sk})],$$

where $\rho_a$ is the air density, $|U_L|$ is the wind speed at the lowest atmospheric model level ($\sim 1013.25$ hPa), $C_H$ is the turbulent exchange coefficient, $q_L$ is the humidity at the lowest atmospheric model level, and $q_{sat}(T_{sk})$ is the saturated specific humidity at the skin temperature in each model grid cell (ECMWF 2016). We calculate the vapor pressure gradient (hPa) as the difference between the saturation vapor pressure at the surface (at skin temperature of the lake) and the actual vapor pressure at the 2-m height (at 2-m dewpoint temperature) using the August–Roche–Magnus approximation (Lawrence 2005), with positive vapor pressure gradient indicating higher values at the surface. We also compute the specific cloud water content (kg kg$^{-1}$) as the sum of specific cloud liquid and ice contents.

Two observation-based gridded precipitation products are used for comparison: monthly 0.5° × 0.5° Climatic Research Unit Time series, version 4.04 (CRU TS4.04), averaged over 2001–18 (Harris et al. 2020), and monthly 0.5° × 0.5° Global Precipitation Climatology Centre (GPCC) dataset averaged over 2001–16 (Schneider et al. 2014; Schneider et al. 2017). For LSWT, we use the Along-Track Scanning Radiometer (ATSR) Reprocessing for Climate: Lake Surface Water Temperature and Ice Cover (ARC-Lake) and Global Observatory of Lake Responses to Environmental Change (GloboLakes) daily time series that provides satellite-derived LSWT at high spatial resolutions (0.05° × 0.05°) from 1996 to 2016 (Carrea and Merchant 2019). The period from 2001 to 2016 is used for comparison with the reanalyses. We note that for the lakes in our study domains, ARC-Lake and GloboLakes data are not spatiotemporally consistent. Specifically, the number of days with available data varies from month to month and year to year. For example, over Lake Baikal, the measurements for the month of February are missing, except for the year 2015 with only 12 days of data. When LSWT measurements are available for any given day, they cover only a fraction of the lake surface area (e.g., Fig. 9) and, therefore, may not represent the mean LSWT. For GloboLakes, we compute the percentage of data available (Fig. 4) as the product of two metrics: 1) percentage of days with available data for each month, accumulated over the 16-yr daily time series, and 2) fraction of lake grid cells with available data, averaged for each month over the 16-yr period. For the Laurentian Great Lakes, we also use Great Lakes Surface Environmental Analysis (GLSEA) Surface Water Temperature, produced by NOAA Great Lakes Environmental Research Laboratory, for the 2001–18 period.

4. Results and discussion

a. Laurentian Great Lakes

1) LAKE CLIMATOLOGY

In the Laurentian Great Lakes region, the warmest months are from June to September (mean seasonal temperature of 291 K), the air temperature is below freezing from December to March (≈267 K), while the spring and autumn seasons from April to May (≈282 K) and October to November (≈278 K), respectively, are relatively short. The domain-averaged 2-m air temperature seasonality is identical for ERA-Interim and ERA5 (Fig. 1d). This regional seasonal cycle, together with lake dynamic and thermodynamic processes such as surface ice cover, convective overturning, and lake stratification (Kravtsov et al. 2018; Sugiyama et al. 2018), contributes to differential surface temperature patterns where the lakes show a delayed heating and cooling response in the summer and winter seasons, respectively, as compared to the surrounding land (Scott and Huff 1996). The spatial patterns of skin temperature in ERA-Interim show this contrasting behavior of lakes (Fig. 2 for select months). In warm months, for example, the larger northern lakes (Superior, Michigan, and Huron) are much cooler than their surroundings (by 9–10 K in June and July, Fig. 2b), while in the cold season they have much higher skin temperatures (on average an 8–10-K difference from December to February, Figs. 2a,d). The seasonal transition months (e.g., March and September; Fig. 2c) display negligible contrast in the lake–land skin temperatures.

The ERA5 simulation of LSWT is noticeably different than ERA-Interim. This issue is identified by Hersbach et al. (2020) and the ECMWF climate bulletin, stating that “there was an inadvertent failure in production to utilize observationally based analyses of the water temperatures of the Great Lakes” over the 1979–2013 period that resulted in erroneous values of LSWT in ERA5. The skin temperature seasonality for the two reanalyses (Fig. 4a) shows that ERA5 has higher Lake Superior LSWT from June through December when compared to ERA-Interim, with a maximum difference of 4.7 K in July, while in February, ERA5 is 4.8 K colder than ERA-Interim. Overall, the ERA5 and ERA-Interim have similar seasonal cycles, with an amplification of the ERA5 LSWT seasonality. ERA5 also simulates higher LSWT than ERA-Interim over Lake Erie from May to September, with the maximum difference of 3.2 K in June (Fig. 4b). The two satellite-based LSWT datasets, GLSEA and GloboLakes, have more data gaps in winter that may arise due to cloud-cover detection errors for the infrared measurements (Bulgin et al. 2018; Wentz et al. 2000). Our assessment of GloboLakes data availability indicates that, throughout the year, LSWT measurements are sparse and may not be representative of the mean monthly climatology for the full lake surface. The percentage of missing
data is even higher from November to March for both Lakes Erie and Superior (Figs. 4a,b). Therefore, caution must be exercised when comparing reanalyses and model outputs with satellite-based datasets that can contain biases in the seasonal temperature profiles.

Lake surface evaporation rates have a thermal dependency that can be affected by biases in skin temperature profiles. The vapor pressure gradient between the surface and the overlying air indicates the evaporation patterns over the lakes, while wind velocity near the surface also affects the magnitudes [Eq. (1)]. In the evaporation climatological mean for ERA-Interim (Fig. S1a), condensation occurs over parts of the larger northern lakes (Superior, Michigan, and Huron) from May to July driven by the negative vapor pressure gradient (Fig. 3b) due to low LSWT (Fig. 2b). At the end of summer (September), the two southern lakes (Erie and Ontario) have comparable evaporation rates to the surrounding land while Lake Superior has relatively lower magnitude (Fig. 2k). From October to February, a positive vapor pressure gradient (Figs. 3a,d), caused by the higher LSWT (Figs. 2a,d), results in larger evaporation rates over the water bodies (Figs. 2i,l). This pattern is similar to that observed in other reanalyses, including NASA MERRA-2, NCEP CFSR, and NCEP NARR (Minallah and Steiner 2021).

The evaporation magnitudes and seasonality of the two ECMWF reanalyses are remarkably different. Over Lake Superior, there is a sharp increase in the evaporative flux from August to December in ERA5, and its evaporation is more than twice the magnitude of ERA-Interim in September (3.53 ± 0.22 mm day$^{-1}$ and 1.36 ± 0.70 mm day$^{-1}$, respectively; Fig. 4a). Over Lake Erie as well, the evaporation rates increase rapidly from May onward and are approximately 1.8 times higher than ERA-Interim in June and July (Fig. 4b). In spatial patterns, these differences are more prominent and ERA5 estimates evaporation fluxes that are uncharacteristic of the lake behavior (Figs. 2m–p and Fig. S1b). In June, the evaporation rate over western edge of Lake Erie is the highest in the region and this maximum spreads across all the lakes by September, which is opposite of ERA-Interim. ERA5 also shows higher evaporation over the lakes from October to December (Fig. 2p) and lower magnitudes in February over Lake Superior (Fig. 2m). In terms of vapor pressure gradient, the differences between the two datasets are evident (Figs. 3e–h). For example, in June, the ERA5 gradient is weaker, resulting

---

**Fig. 2.** Spatial patterns of the 2001–18 mean skin temperature for (a)–(d) ERA-Interim, (e)–(h) ERA5 minus ERA-Interim skin temperature difference, (i)–(l) ERA-Interim evaporation, and (m)–(p) ERA5 minus ERA-Interim evaporation difference over the Laurentian Great Lakes region. Negative values for evaporation show condensation. Four months are selected (February, June, September, and December) to show the seasonal variations. ERA-Interim and ERA5 evaporation magnitudes for all months are shown in Fig. S1.
in condensation only over Lake Superior (Fig. S1), whereas
in September, ERA5 has higher vapor pressure gradient
(Fig. 3g), which results in the anomalously high evaporation
rates (Fig. 2o).

2) LAKE-EFFECT PRECIPITATION CASE STUDY

To understand how water bodies alter the regional meteorology in the Great Lakes region, we conduct a case study to
assess the link between LSWT and lake-effect precipitation in
the winter months and look at the atmospheric profiles for two
cross sections: 43.75°N crossing Lakes Michigan, Huron, and
Ontario, and 47.5°N crossing Lake Superior (dashed lines in
Fig. 1a) over a 6-day period (1–6 February 2007).

During the February 2007 event, ERA5 has higher LSWT at
the 43.75°N cross section (Fig. 5d) and lower Lake Superior
temperature (Fig. 5j) than ERA-Interim, similar to what is
observed in the climatology (Figs. 2a,e). We correspondingly
observe larger evaporation rates over the lakes with higher
temperatures (Figs. 5e,h), which provides the moisture to feed
cloud formation. The warm surface also leads to stronger

FIG. 3. Spatial patterns of the surface minus 2-m vapor pressure gradient (2001–18 mean)
over the Laurentian lakes for (a)–(d) ERA-Interim and (e)–(h) ERA5 for the same 4 months
in Fig. 2.
updrafts, which is especially pronounced over Lake Ontario in ERA5 from 3 February onward (Figs. 6g–l) and Lake Superior in ERA-Interim (Figs. 6m–r). In both cases, the specific cloud water content (CWC) overhead and downwind of the lakes is high, which can identify regions of precipitation. In ERA-Interim, 1–3 February simulates higher CWC over Lake Superior, extending further downwind of the lake (Figs. 6m–o). Since the atmosphere is quite cold, most of the moisture precipitates out (Fig. 5i). In ERA5, the high-intensity precipitation zones over Lake Ontario on 3 and 4 February (Fig. 5f) similarly correspond to region with high CWC from 850- to 700-hPa height (Figs. 6i,j), whereas in ERA-Interim the updrafts are relatively weaker, CWC is concentrated lower in the atmosphere (around 850 hPa; Figs. 6c,d), and the precipitation magnitudes are less intense (Fig. 5c). In ERA5, there are downdraft regions over the western shore of Lake Superior (Figs. 6s–x), which are absent in ERA-Interim when its LSWT is warmer (Fig. 5g). We note that the actual process of precipitation formation is more complex and not all CWC will precipitate out. For example, in ERA-Interim, Lake Ontario has high CWC content on 2 February (Fig. 6b), which does not correspond to high evaporation or precipitation magnitudes (Figs. 5b,c). In general, we see strong links between warmer LSWT, higher evaporation rates, stronger updrafts, higher CWC, and subsequently high precipitation magnitudes, which highlights that LSWT play a role in precipitation formation processes within this region in cold months, and this manifests in the two reanalyses depending on their treatment of the lakes.

**Fig. 4.** 2001–18 mean seasonality of the LSWT (K; solid line) and evaporation (mm day$^{-1}$; dashed line) along with the 95% confidence intervals for ERA-Interim (blue) and ERA5 (red) averaged over four lake surfaces: (a) Superior, (b) Erie, (c) Baikal, and (d) Victoria. GloboLakes LSWT (2001–16 mean) is shown in green line for each lake, while GLSEA LSWT is shown in dotted line for lakes Superior and Erie. The green bars show percentage data available for each month over the 16-yr period for the GloboLakes dataset.
b. Lake Baikal

Lake Baikal is located in the south Siberian climate characterized by cold, dry winters and cool summers (Köppen-Geiger climate classification; Kottek et al. 2006). This region has a Gaussian seasonality with the air temperature maximum in July (~290 K), while the temperature is below freezing from October until April and gradually warms in May (~280 K; Fig. 1e). The regional temperature profile, together with the narrow and deep lake bathymetry, plays a role in modulation of the lake-effect climate. Differential heating patterns also exist between the north and south basins due to the different air temperatures and lake depths, where the north basin is shallower (Todd and Mackay 2003).

In this region during the cold period (January–April), the land–lake skin temperature contrast is not as pronounced (Fig. S4). From May/June, the lake surface warms, but is still relatively colder than the surrounding land till August. By September/October, as the land cools, the delayed heating of the lake results in higher LSWT than the surrounding land. Both ERA-Interim and ERA5 follow similar skin temperature patterns and magnitudes from January through July (Fig. 4c). However, they diverge from August onward, where ERA5 LSWT is warmer than ERA-Interim by 2.7 K in August to 12.4 K in November. The ERA5 lake surface warms more and retains this heat for a longer duration than simulated in ERA-Interim.

The stark difference in the LSWT between ERA-Interim and ERA5 affects the lake evaporation patterns and magnitudes, which also diverge in August and subsequent months. For example, in November, the mean lake evaporation rate is 0.1 ± 0.01 mm day⁻¹ in ERA-Interim as compared to 2.38 ± 0.08 mm day⁻¹ for ERA5 (Fig. 4c). The higher LSWT in ERA5 (Fig. 7b) induces a positive air-to-surface vapor pressure gradient (Fig. 7d), which is a contributing factor to higher evaporation rates (Fig. 7f). For the same month, ERA5 has high-precipitation zones eastward of the lake that are absent in ERA-Interim (Figs. 8a,b). Comparison with two gridded observation-based precipitation datasets (CRU TS4.04 and GPCC; Figs. 8c,d) indicates that ERA5 overestimates while ERA-Interim underestimates precipitation magnitudes near both the north and south basins. ERA5 also has higher precipitation directly north of Lake Baikal over the Baikal mountains, which is likely due to the topography of the region and not associated with the lake simulation.

**COMPARISON WITH LSWT OBSERVATIONS**

Spatiotemporally consistent LSWT buoy or satellite measurements are lacking for this region, thus climatological comparison with observations is subject to inconsistencies. For example, the GloboLakes seasonality is strikingly different than the two reanalyses in December–April, when clouds likely interrupt retrieval of surface measurements and make the comparison unreliable (Fig. 4c). We use merged ATSR-1, ATSR-2, and Advanced ATSR (AATSR) daytime measurements provided by the ARC-Lake project (MacCallum and Merchant 2014) for Lake Baikal at 0.05° resolution for select days in October and November, when adequate spatial coverage is available to compare with the reanalyses (Fig. 9). The comparison suggests that ERA-Interim is much colder, while ERA5 LSWT is closer to the merged ATSR measurements albeit with somewhat higher temperatures along the lake center (by 1–4 K). On the other hand, lake evaporation requires further corroboration with observations as it is sensitive
to LSWT and small positive biases in temperature can result in anomalously high evaporation rates. The extra moisture provided by lake evaporation from September to December is likely a source for the higher-intensity precipitation observed in ERA5 (Fig. 8b for November).

c. African Great Lakes

1) Lake Climatology

Given the proximity to the equator, temperature seasonality in the African Great Lakes region is only 3–4 K in amplitude (Fig. 1c), where the months of September–November are slightly warmer than the remaining months (Fig. 1f). The African region does not demonstrate a sharp land–lake skin temperature contrast in ERA-Interim, and the LSWT is similar in magnitude to the surrounding land (Fig. S5). However, ERA5 exhibits higher LSWT throughout the year; for example, Lake Victoria’s skin temperature is 3.1–4.7 K greater than ERA-Interim (Fig. 4d). We also note that not only are the LSWT much greater in ERA5, the land skin temperature is generally higher than ERA-Interim, as shown in February when the difference between the two reanalyses is the highest (Figs. 10a,b; Fig. S5). The GloboLakes LSWT is slightly closer to ERA5 in magnitude; however, ERA5 seems to overestimate the LSWT magnitudes whereas ERA-Interim underestimates it (Fig. 4d).

For all three lakes in this region, the warmer lake temperatures in ERA5 drive higher vapor pressure gradients (Fig. 10d) and lake evaporation rates (Fig. 10f). It is important to note that the 3–5-K difference in LSWT between the ERA5 and

---

**Fig. 6.** Vertical profile of the specific cloud water content and winds (m s$^{-1}$), from 1000 to 700 hPa, for the same cross sections in Fig. 5. The vertical winds are scaled by a factor of 5.
ERA-Interim drive a twofold difference in evaporation magnitudes, and this can result in vastly different regional hydroclimatic simulations. For example, in February, Lake Victoria’s evaporation rate for ERA5 is $4.91 \pm 0.1 \text{ mm day}^{-1}$, as compared to $2.44 \pm 0.05 \text{ mm day}^{-1}$ for ERA-Interim (Fig. 4d). The regional precipitation magnitudes and spatial patterns also have pronounced differences between the two datasets, especially with the simulation of very intense precipitation over Lake Malawi and other spurious precipitation hot spots around the three larger lakes in ERA5 (e.g., shown for February in Fig. 13b). Specific differences between ERA5 and ERA-Interim can be attributed to the finer spatiotemporal model resolution of the former, which may result in the differences in precipitation processes, specifically the partitioning of convective and stratiform precipitation. However, the distinct precipitation hot spots (in Fig. 13b) indicate that lake representation likely plays a role in producing these patterns (discussed in section 4d).

2) CASE STUDY

The African Great Lakes exert a strong influence on the atmospheric dynamics and localized circulation patterns (Thiery et al. 2015). Precipitation processes in the lake vicinity are driven by diurnal variations in the land surface temperature.
relative to the LSWT, which induce a land–lake-breeze circulation that can frequently result in intense thunderstorms (Thiery et al. 2016; Virts and Goodman 2020; Woodhams et al. 2019). We analyze the precipitation patterns and local atmospheric dynamics for one such precipitation event (7–8 April 2017) over Lake Victoria to assess how the reanalyses simulate their diurnal cycles. There are large-scale dynamic controls that also influence precipitation, including the ITCZ location, prevailing easterly winds, moisture influx from outside the region (Indian Ocean), and topographic effects (Anyah et al. 2006; Woodhams et al. 2019); however, here we focus on the diurnal circulation strength that depends on land–lake temperature gradients.

Similar to the observed LSWT April climatology where ERA5 is 4.2 K warmer than ERA-Interim (Fig. 4d), the subdaily lake temperatures from 7 to 8 April are also higher in ERA5 (Fig. 11b). Due to the simplified lake representation in the older-generation ERA-Interim model core, the diurnal cycle of LSWT is not properly captured (Fig. 11a), and the lake temperatures from 0300 to 0900, 0900 to 1500, 1500 to 2100, and 2100 to 0300 East Africa time (EAT) remain nearly identical (~1–2-K difference). Despite the newer-generation model core using the FLake model, the LSWT diurnal cycle in ERA5 is also minimal (Fig. 11b). Nonetheless, the ERA5 10-m wind directions reverse over the lake from southwesterly over daytime (0900–2100 EAT) to easterly over nighttime (2100–0900 EAT), where the lake–land temperature contrast plays a role in the lake-effect circulation. From 0300 to 0900 EAT, the winds over the lake surface are directed toward the northwest in both reanalyses. As the day heats up from 0900 to 1500 EAT, the lake breeze is evident in ERA-Interim on both the eastern and western shores; however, in ERA5 this pattern is reversed on the western edge due to the colder land temperature around 31°E. In ERA5, this wind pattern continues until the land temperature drops from 2100 EAT onward. For ERA-Interim, the temperature decline over the land surface is higher than ERA5 (1500–2100 EAT), yet the overlake wind directions remain westerly. At 2100–0300 EAT the overlake wind patterns have different directions in ERA-Interim and ERA5, where the latter simulates easterly winds while ERA-Interim simulates southerly winds. These differences between the two reanalyses can arise due to 1) nuanced variations in the land–lake temperature contrast, 2) different spatiotemporal resolutions, and 3) the topography representation of the Eastern and Western Rift Valleys surrounding Lake Victoria, the effect of which is visible in ERA5 along 30° and 36°E.

The vertically integrated moisture flux divergence (VIMFD) patterns during night/early morning (0300–0900 EAT) for both ERA-Interim and ERA5 show convergence over most of Lake Victoria and particularly on the western shore, with stronger convergence in ERA5 (Figs. 11c,d). From 0900 to 1500 EAT,
there is divergence over the lake and convergence over the surrounding land, where the terrestrial region has a weaker signal in ERA-Interim. In ERA5, this convergence zone propagates over the eastern shore of the lake by evening (1500–2100 EAT) and directly over lake by night (2100–0300 EAT). Prior work (Minallah and Steiner 2021; Seager and Henderson 2013) shows that the VIMFD patterns are influenced by both the spatial and temporal resolutions of the

Fig. 9. Lake Baikal LSWT on select days for (a) ERA-Interim, (b) ERA5, and (c) ARC-Lake. The days are shown on the left.
variables, and some differences between the two can be attributed to the coarser resolution of ERA-Interim.

The vertical profile of winds at 1S cross section across the lake (Figs. 12a,b) show a counterclockwise circulation at 850–800 hPa along the eastern shore of Lake Victoria at 0900–1500 EAT in both ERA-Interim and ERA5. This circulation pattern arises as the afternoon lake breeze in the planetary boundary layer (1000–850 hPa) brings cool air over the warm land surface resulting in updrafts at the eastern shore of the lake, driving reverse land to lake flow aloft (~700 hPa) and subsidence over the lake. In ERA5, the counterclockwise circulation on the western shore at 0900–1500 EAT is opposite of the expected circulation pattern (i.e., lake to land flow at the surface driving overland updrafts, leading to a clockwise circulation as the air returns aloft and descends over the cooler lake). This is because the land skin temperature on the western shore at 1S is cooler than the lake and 10-m winds are directed toward the warmer lake (Fig. 11b), which is opposite of ERA-Interim that has warmer land
temperatures than lake, driving surface winds moving from the lake to land (Fig. 11a). By 1500–2100 EAT, the circulation pattern on the eastern shore shifts over the lake in ERA5, while its strength is weakened in ERA-Interim. At night/early morning (2100–0300 and 0300–0900 EAT), ERA5 has westward flow in PBL over the lake, with strong updrafts in the 850–700-hPa layer at the western shore. For ERA-Interim, there are some differences in the circulation pattern, for example, at 2100–0300 EAT, the eastern shore in the PBL has eastward winds and updrafts over land, opposite to the simulated patterns in ERA5.

The mechanisms generating the CWC can be separated by the altitude: the CWC at higher altitudes (>700 hPa) in both

FIG. 11. Lake Victoria LSWT (K) together with 10-m wind vectors (m s⁻¹) for (a) ERA-Interim and (b) ERA5 and vertically integrated moisture flux divergence (mm day⁻¹) for (c) ERA-Interim and (d) ERA5, where the blue (red) colors indicate convergence (divergence). (left to right) The four columns are for 7–8 Apr 2017 covering 0300–0900, 0900–1500, 1500–2100, and 2100–0300 EAT (UTC + 3 h).
products is likely associated with synoptic-scale moisture transport, whereas content at lower heights (850–700 hPa) can be influenced by moisture from the surface and the land–lake circulation effects. In the 850–700-hPa zone, ERA-Interim has negligible CWC ($<0.5 \times 10^{-3}$ kg kg$^{-1}$), whereas ERA5 has higher CWC, especially over the lake surface (Figs. 12a,b). We hypothesize that this is due to the higher lake surface evaporation in ERA5, which provides extra moisture to feed the CWC. In case of ERA-Interim, the evaporation amount is relatively smaller, which is also observed in the long-term monthly climatology of evaporation (Fig. 4d).

In the diurnal patterns of convective precipitation, ERA5 captures the known (Thiery et al. 2016, 2017) storm development...
In the afternoon and evening (1500–2100 EAT), a precipitation storm develops over land on the eastern edge of Lake Victoria (Fig. 12d) due to the lake-breeze effect (Fig. 11b). The storm propagates over lake (2100–0300 EAT), and from 0300 to 0900 EAT, we observe a strong precipitation zone over Lake Victoria due to the cool-land–warm-lake convective circulation. The precipitation band moves northwest over the land by noon (0900–1500 EAT). In ERA5, there is correspondence between the CWC (Fig. 12b), convergence zones (Fig. 11d), and the convective precipitation zones (Fig. 12d), that is, generally, high cloud water content together with the convergence zones are collocated with convective precipitation zones throughout the diurnal cycle. In ERA-Interim, the cloud water content is generally restricted to the upper atmosphere, and the regions of convergence and CWC appear decoupled with convective precipitation along the 1°S transect. We note that while ERA5 captures the land–lake-breeze dynamics, its LSWT estimates and precipitation magnitudes for all months are higher than what is observed (Figs. 4d and 13b, respectively).

d. Effects of lake representation on precipitation patterns

The ERA5 FLake scheme utilizes two input fields, lake cover and lake depth, to drive the physical calculations of LWST. Here, we evaluate these input fields to understand how...
the lake representation may influence the LWST and resulting simulation of precipitation. The lake cover input variable is a fractional value ranging from 0 to 1. When evaluating over central Africa, some rivers and intermittent water bodies are identified as lakes (Fig. 13d); for example, the Congo River and several of its tributaries are visible in the lake cover fraction west of Lake Victoria and Tanganyika, as well as the River Zambezi south of Lake Malawi and the presence of its two large lakes, Kariba (around 17°S, 28°E) and Cabora Bassa (16°S, 32°E). Other nonlake features are also identified in the lake cover, including the Sua Pan salt plains at 21°S, 26°E. Dutra et al. (2010) state that the categorization in the FLake-HTESSEL model does not distinguish between lakes and other inland open water surfaces, and this is evident in the lake cover fraction in the African Great Lakes region.

The lake depth parameter also shows some potential errors, as it provides a depth for various regions and geographic features without an underlying water body, for example, the mountains around Lake Baikal, Putorana Plateau in Russia, and parts of the Great Plains in North and South Dakota of the United States (not shown). In the regions around the Congo River west of Lake Tanganyika and around Lake Malawi (Fig. 13e), lake depth is present when lake cover is not. Together, these two terms provide the lake categorization for FLake, and we show the product of lake cover and depth (Fig. 13f) as a proxy to assess the locations where the model will simulate lakes. While the main Great Lakes are well represented, the proxy suggests that other grid cells in the region are also simulated as lakes. Further, regions of high-intensity precipitation in ERA5 (Fig. 13b) frequently correspond to the cells categorized as lakes. For example, the precipitation band at approximately 12°S, 32°E corresponds to “lake cells” visible in Fig. 13f next to Lake Malawi. Other precipitation features around Lakes Tanganyika and Victoria also correspond with the presence of a modeled lake.

Comparing the precipitation patterns of the two reanalyses with observation-based CRU TS4.04 dataset (Fig. 13c) and satellite-based PERSIANN-CDR dataset (not shown), we note that ERA5 captures the synoptic-scale patterns of precipitation, and therefore the large-scale systems are likely adequately represented in ERA5. The presence of lakes will influence the local convective systems, and we note that the spurious ERA5 convective precipitation hot spots are collocated with features simulated as lakes in ERA5, which are absent in the ERA-Interim, likely due to its coarser spatial resolution and absence of subgrid lake representation. This indicates that both erroneous categorization of lake surface in input parameters and simulation of LSWT can influence the simulation of convective precipitation magnitudes in proximity to the simulated lakes. The higher resolution of ERA5 may affect topography-induced precipitation as well, and further analysis is required to isolate the causes of intense precipitation in ERA5.

5. Conclusions

In our assessment of the three regions with the largest freshwater volume in the world (the Laurentian Great Lakes, the African Great Lakes, and Lake Baikal), we found considerable differences in the lake representation between the new-generation ERA5 and its predecessor ERA-Interim. Depending on the lake region, the LSWT are prescribed and/or simulated with different methods. In ERA-Interim, the Laurentian Great Lakes LSWT are prescribed from external SST datasets (specifically, OSTIA and NCEP), whereas for the other two lake systems, a simpler approach of using 1-month-lagged 2-m temperature from ERA-40 is employed to represent the delayed heating/cooling response of LSWT. In ERA5, the land surface model incorporates a new parameterization scheme, the FLake model, that uses lake cover and lake depth to identify regions with inland water bodies and resolves both gridscale and subgrid lakes. For the Laurentian system, ERA5 failed to accurately simulate the LSWT, which introduced considerable errors in the representation of lakes and related hydrometeorological fields. There are also prominent differences in the LSWT between the two reanalyses in the other two regions.

Specifically, ERA5 generally simulates much warmer LSWT that can be present throughout the year (e.g., Lake Victoria, which on average is 4.1 K warmer than ERA-Interim) or in specific months and seasons only (e.g., Lake Baikal from August to January with the highest difference of 12.4 K in November, and Lake Erie in late spring and summer months with an average difference of 2.3 K). In some cases, the lakes in ERA5 tend to heat quickly upon onset of the summer season and retain the heat for longer durations (Lake Baikal), while other locations show a sharp rise and fall in temperature with higher seasonal maxima/minima (Lake Superior). In all three regions, the LSWT differences are substantial enough to simulate different regional moisture budget components and water cycle. LSWT can also impact evaporative fluxes and precipitation, which we evaluate with both climatological averages and short-term meteorological case studies. Lakesurface evaporation is affected by the vapor pressure gradient and LSWT, and ERA5 has considerably higher evaporation rates that can be more than twice the magnitude in ERA-Interim for some months and regions.

The effects of LSWT on the regional hydroclimate can also propagate to the mesoscale weather patterns. The turbulent and radiative exchanges over water bodies and lake thermodynamics alter the atmospheric boundary layer structure and produce important meteorological phenomenon such as lake-effect precipitation, lake-breeze circulations, and storms in lake proximity depending on the seasonal and diurnal cycles (Alexander et al. 2018; Notaro et al. 2013b; Thiery et al. 2015; Zhao et al. 2012). We see these lake–atmosphere feedbacks in the lake proximity where ERA5 shows high-intensity precipitation areas not present in ERA-Interim and observation-based datasets (CRU and GPCP). While some differences between the two reanalyses arise due to the finer spatiotemporal model resolution of ERA5, the precipitation formation processes indicate that in the Laurentian winters, high LSWT correspond with higher vapor pressure gradient and higher lake evaporation, stronger updrafts, and more cloud water content that consequently generates more precipitation. In the African regions, the lake–land-breeze circulation is better
captured in ERA5, due to its finer spatiotemporal resolution and more comprehensive lake representation, despite an overestimation of the LSWT when compared to the GloboLakes dataset. However, the LSWT in ERA5 does not exhibit a diurnal cycle and its precipitation simulation is more intense compared to ERA-Interim, the observation-based CRU, and the satellite-based PERSIANN-CDR datasets. For Lake Baikal, ERA5 LSWT is closer but somewhat warmer than the GloboLakes and ARC-Lake datasets, whereas ERA-Interim underestimates the LSWT. ERA5 again shows zones of high-intensity precipitation around the lake that are not observed in CRU or GPCC precipitation.

We also analyzed two ancillary input parameters into the FLake scheme of ERA5, lake depth and lake cover, and found that they often erroneously represent rivers, other intermittent water bodies, and some topographic features as lakes in the parameterization scheme (Dutra et al. 2010). In ERA5, many of the intense precipitation zones correspond to the presence of an underlying lake grid cell. While the synoptic scale often dominates the total precipitation magnitude, inland water bodies influence convective precipitation mechanisms, and therefore they must be correctly simulated in reanalyses and models to understand the regional hydroclimates, weather systems, and extreme events.

The aim of this analysis is to assess the performance of two generations of ECMWF reanalyses in regions with large water bodies and to determine their feasibility for regional hydroclimatic assessments. The differences in how lakes are simulated in the two versions cause variations in hydroclimatic parameters at climatological and meteorological time scales. The ECMWF’s IFS modeling system is also used for numerical weather prediction and operational forecasting, therefore any bias in lake simulation may extend beyond the reanalyses. Moreover, many researchers use ERA-Interim and, now, the newer-generation ERA5 for hydrological modeling and related applications due to its higher spatiotemporal resolution and improved performance in other cases and regions (Betts et al. 2019; Dullaart et al. 2019; Graham et al. 2019; Tarek et al. 2020), therefore it is important to identify these differences that may impact the simulation of the regional climate. Coupling a lake-parameterization scheme with the land model is a step in the right direction for improved representation of lake–atmosphere interactions, and this study can assist in resolving differences in future model and reanalysis versions.

Acknowledgments. This study is supported by National Science Foundation Grant OCE-1600012 and University of Michigan MCubed Funding. We thank Dr. Hans Hershbach from ECMWF and Dr. Yiwen Mei from the University of Michigan Ann Arbor for their feedback on the initial draft, and the three anonymous reviewers for their detailed review and valuable feedback.

REFERENCES


ing seasonal to sub-daily water temperature dynamics in a large shallow lake. Water, 10, 594, https://doi.org/10.3390/w10050594.


