# Evaluating and improving modeled turbulent heat fluxes across the North American Great Lakes

Umarporn Charusombat[1], Ayumi Fujisaki-Manome[2,3], Andrew D. Gronewold[1], Brent M. Lofgren[1], Eric J. Anderson[1], Peter D. Blanken[4], Christopher Spence[5], John D. Lenters[6], Chuliang Xiao[2], Lindsay E. Fitzpatrick[2], and Gregory Cutrell[7]

[1]NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, Michigan, 48108 USA
[2]University of Michigan, Cooperative Institute for Great Lakes Research, Ann Arbor, Michigan, 48108, USA
[3]University of Michigan, Climate & Space Sciences and Engineering Department, Ann Arbor, Michigan, 48109, USA
[4]University of Colorado, Department of Geography, Boulder, Colorado, 80309, USA
[5]Environment and Climate Change Canada, Saskatoon, Saskatchewan, S7N 5C5, Canada
[6]University of Wisconsin-Madison, Center for Limnology, Boulder Junction, Wisconsin, 54512, USA
[7]LimnoTech, Ann Arbor, Michigan, 48108

*Correspondence to: Ayumi Fujisaki-Manome (ayumif@umich.edu)*

**Abstract** Turbulent fluxes of latent and sensible heat are important physical processes that influence the energy and water budgets of the North American Great Lakes. These fluxes can be measured *in situ* using eddy covariance techniques, and are regularly included as a component of lake-atmosphere models. To help ensure accurate projections of lake temperature and circulation and regional meteorology, here we validate the output of five algorithms used in three popular models to calculate surface heat fluxes; the Finite-Volume Community Ocean Model (FVCOM, with three different options for heat flux algorithm), the Weather Research and Forecasting (WRF) model, and the Large Lake Thermodynamics Model. These models are used in research and operational environments and concentrate on different aspects of the Great Lakes' physical system. We isolated only the code for the heat flux algorithms from each model and drove them using meteorological data from four over-lake stations within the Great Lakes Evaporation Network (GLEN), where eddy covariance measurements were also made, enabling co-located comparison. All algorithms reasonably reproduced the seasonal cycle of the turbulent heat fluxes, but all of the algorithms except for the Coupled Ocean Atmosphere Response Experiment (COARE) algorithm showed notable overestimation of the fluxes in fall and winter. Overall, COARE had the best agreement with eddy covariance measurements. The four algorithms other than COARE were altered by updating the parameterization of roughness length scales for air temperature and humidity to match those used in COARE, yielding sensible and latent heat fluxes with improved agreement with observations.

## 1. Introduction

Simulating physical processes within and across large freshwater water bodies is typically achieved using oceanographic-scale models representing heat and mass exchange below, above, and across the air-water interface. Verification and skill assessment of these models is limited, however, by the quality and spatial extent of observations and data. The datasets available for validation of ocean dynamical models, for example, include satellite-based surface water temperatures (Reynolds et al., 2007), sea surface height (Lambin et al., 2010), and when available, *in situ* measurements of sensible and latent heat fluxes (Edson et al., 1998). Dynamical and thermodynamic models for large lakes are often verified using similar measurements (Chu et al., 2011; Croley, 1989a, 1989b; Moukomla and Blanken, 2017; Xiao et al., 2016; Xue et al., 2017). However, the spatiotemporal resolution of *in situ* measurements for these variables in lakes is comparatively sparse (Gronewold and Stow, 2014), particularly for latent and sensible heat fluxes.

On the Laurentian Great Lakes (hereafter referred to as the Great Lakes), sensible and latent heat fluxes play an important role in the seasonal and interannual variability of critical physical processes including spring and fall lake evaporation (Spence et al., 2013), the onset, retreat, and spatial extent of winter ice cover ( Clites et al., 2014; Van Cleave et al., 2014;), and air mass modification including processes such as lake-effect snow (Wright et al., 2013). These phenomena, in turn, can impact lake water levels (Gronewold et al., 2013; Lenters, 2001), atmospheric and lake circulation patterns (Beletsky et al., 2006), and the fate and transport of watershed-borne pollutants (Michalak et al., 2013). For decades, Great Lakes dynamical and thermodynamic models simulating these processes have done so with minimal observations.

The Finite-Volume Community Ocean Model (FVCOM), for example, is a widely used hydrodynamic ocean model that has been found to provide accurate real-time nowcasts and forecasts of hydrodynamic conditions across the Great Lakes including currents, water temperature, and water level fluctuations across relatively fine spatiotemporal scales (Anderson et al., 2015; Anderson and Schwab, 2013; Bai et al., 2013; Xue et al., 2017). FVCOM is currently being developed, tested, and deployed across all of the Great Lakes as part of an ongoing update to the National Oceanic and Atmospheric Administration (NOAA) Great Lakes Operational Forecasting System (GLOFS). To date, however, there has been no direct verification of the turbulent heat flux algorithms intrinsic to FVCOM; this is an important step, in light of the fact that FVCOM flux algorithms were developed primarily for the open ocean and, until now, have

been assumed to provide reasonable turbulent heat flux simulations across broad freshwater surfaces as well.

The Large Lake Thermodynamic Model (Croley, 1989a, 1989b; Croley et al., 2002; Hunter et al., 2015), LLTM, is a conventional lumped conceptual lake model. It is employed in seasonal operational water supply and water level forecasting by water resource and hydropower management authorities (Gronewold et al., 2011) and is used as a basis for long-term historical monthly average evaporation records (Hunter et al., 2015). It has historically been calibrated and verified using observed ice cover and surface water temperature, but not turbulent flux data. Among more complex atmosphere-lake model systems, the Weather Research and Forecasting (WRF) system is increasingly used in Great Lakes applications (Xiao et al., 2016; Xue et al., 2015). However, only with the Global Environmental Multiscale model (GEM; Bélair et al., 2003a, 2003b; Deacu et al., 2012), have observed turbulent fluxes have been employed to assess predictive skill of net basin supply and regional climate conditions over the Great Lakes (Deacu et al., 2012).

To address this gap in the development and testing of lake-atmosphere exchange physically-based models for use on the Great Lakes, we employ data from a network of relatively novel year-round offshore eddy-covariance flux measurements collected over the past decade at lighthouse-based towers. Specific focus in this study are to determine: 1) the capability of the flux algorithms in reproducing inter-annual, seasonal, and daily latent and sensible heat fluxes, 2) how much variability occurs in the simulated latent and sensible heat fluxes from using different flux algorithms with common forcing data (e.g. meteorology and water surface temperature), and 3) the source of such variability and simulation errors. In particular, we will address how different parameterizations of roughness length scales affect simulations of turbulent latent and sensible heat fluxes over the water surface of the Great Lakes.

## 2. Methods

We begin by describing the measured meteorology and turbulent heat fluxes data used in this study, followed by the flux algorithms within larger modelling frameworks, and lastly the intercomparison methods used to evaluate the performance of the flux algorithms. We selected the time period from January 2012 through December 2014; this period is ideally suited for our study

since it allows for a comparison between two unusually warm (2012-2013) and unusually cold (2013-2014) winters (Clites et al., 2014).

### 2.1. Data

Meteorological and turbulent heat flux data were collected from four offshore, lighthouse-based monitoring platforms (Fig. 1): Stannard Rock (Lake Superior), White Shoal (Lake Michigan), Spectacle Reef (Lake Huron), and Long Point (Lake Erie). These observations are collected as part of a broader collection of fixed and mobile-based platforms collectively referred to as the Great Lakes Evaporation Network (GLEN, Lenters et al., 2013). These installations are referred to by the National Data Buoy Center (NDBC) as stations STDM4, WSLM4, and SRLM4 at Stannard Rock, White Shoal, and Spectacle Reef, respectively.

With the exception of Long Point, footprint analysis indicates each station is located sufficiently distant from shore so that there is no influence of the land surface on the turbulent flux measurements (flux footprint calculations show this to be predominantly the case during most times of the year; Blanken et al., 2011). Long Point, however, is located at the tip of a narrow, 40-km peninsula extending into Lake Erie. As a result, measured fluxes can be influenced by the upwind land surface when the wind direction is between $180^{o}$ and $315^{o}$, so these data were removed when measured wind directions were within this range.

#### 2.1.1. *Turbulent heat flux measurements*

All four eddy covariance systems follow conventional protocols for calculating turbulent fluxes, such as those established on Great Slave Lake (Northwest Territories, Canada) by Blanken et al. (2000). 30-minute mean turbulent fluxes of sensible and latent heat ($\lambda E$ and $H$, respectively; W $m^{-2}$; positive upward from the surface) were calculated from 10-Hz measurements of the vertical wind speed ($w$; $ms^{-1}$), air temperature ($T$; $^{\circ}C$), and water vapor density ($\rho_v$; g $m^{-3}$). Wind speed was measured using a 3D ultrasonic anemometer (Campbell Scientific CSAT-3), while water vapor density was measured using a krypton hygrometer (Campbell Scientific KH20). The statistics (means and covariances) of the high-frequency data were collected and processed at 30-minute intervals using a Campbell Scientific dataloggers. Corrections to the eddy covariance measurements included 2D coordinate rotation (Baldocchi et al., 1988), and corrections for air density fluctuations (Webb et al. 1980), sonic path length, high-frequency attenuation, and sensor

separation (Horst, 1997; Massman, 2000). Instrument heights above the mean water levels for meteorological and the eddy covariance measurements were 32.5 m at Stannard Rock, 29.5 m at Long Point, 30.0 m at Spectacle Reef, and 42.8 m at White Shoal.

As noted in section 2.1, the eddy covariance data at Long Point were filtered out when wind direction was between 180$^{o}$ S-315$^{o}$ NW to remove the land surface influence on the measured latent and sensible heat fluxes. We also applied cross-check filtering for the eddy covariance data at White Shoal and Spectacle Reef. The two stations were relatively close in distance and the measured latent and sensible heat fluxes at these stations were mostly similar, with daily averaged values differing by less than 100 W m$^{-2}$ (except during the ice-covered periods, which were not a focus of this study). There were outliers during July and August 2014 where the measured fluxes differed at the two stations by greater than 100 W m$^{-2}$. These data were removed, resulting in ~5 % loss of data points at White Shoal and Spectacle Reef. See Blanken et al. (2011) and Spence et al. (2011, 2013) for details of the measurements and flux corrections.

*2.1.2. Meteorological data and water surface temperature*

At the same heights as the turbulent flux instruments, half-hourly meteorological variables of wind speed, air temperature and relative humidity, and air pressure were obtained using RM Young wind sensors, Vaisala HMP45C thermohygrometers, and barometers (varied by site), respectively. Air pressure at Spectacle Reef was not measured and was approximated using data from the White Shoal station, a reasonable assumption given their close proximity. Water surface temperature for model input is taken from Great Lakes Surface Environmental Analysis (GLSEA, https://coastwatch.glerl.noaa.gov/glsea/doc/), which is a composite analysis based on NOAA Advanced Very High Resolution Radiometer (AVHRR) imagery. Lake surface temperatures are updated daily with an interpolation method using information from the cloud-free portions of the satellite imagery within ±10 days. The closest pixels to the observation sites were chosen to provide model inputs of water surface temperature. Ice concentration data provided by the National Ice Center (NIC) were used to decide whether eddy covariance measurements at each GLEN site were affected by ice cover. When ice concentration at the closest pixel to a GLEN station was greater than zero, we did not use any data for our comparison (i.e. the observed heat

fluxes, water surface temperature, and meteorological data). This was because the study focused on evaluating the turbulent heat fluxes over water during ice-free periods.

Infrared thermometers (IRTs, Apogee IRR-T) were also installed on the observation platforms to measure water surface temperature. However, test simulations showed that the flux values simulated using the water surface temperature from the IRTs were generally less reliable than when using the GLSEA data. Blanken et al. (2011) found that about 30% of the IRT-measured lake surface temperature observations were unreliable due to condensation, frost, and interference from other surfaces (e.g., the lighthouse or sky). It is likely that this issue affected the accuracy of IRT-measured water surface temperature during the period of our study. Therefore, we did not use the IRT-based measurements of water surface temperature as input to the simulations.

### 2.2. Flux algorithms

Five different flux algorithms that are incorporated into the three hydrodynamic/atmospheric/hydrologic models that are frequently used for Great Lakes operational and research applications were evaluated (Fig. 2).

In an early stage of its development, FVCOM required prescribed heat fluxes as forcing, not calculated, variables (Chen et al., 2006a, 2006b). In a subsequent version of FVCOM (Version 2.7), turbulent fluxes were calculated using the Coupled Ocean Atmosphere Response Experiment (COARE) Met Flux Algorithm version 2.6 (Fairall et al., 1996a,b) was first adopted in the official FVCOM (Chen et al., 2006a). The COARE Met Flux Algorithm is one of the most frequently used algorithms in the air–sea interaction community. It was subsequently modified and validated at higher winds in the version known as COARE 3.0 (Fairall et al., 2003) and the latest version COARE 3.5 (Edson et al., 2013) includes wave influences on the Charnock parameter (Charnock, 1955). FVCOM mostly incorporated these updates as the model was upgraded, including provision for freshwater implementation, except that the latest version of FVCOM (version 4.0) has not yet included wave influences on the Charnock parameter. Hereafter we refer to the COARE implementation in FVCOM as COARE, and it is equivalent to COARE 3.0. In FVCOM version 3 and later, two additional flux calculation algorithms were added (Chen et al., 2013): One was adapted from a flux coupler in the Community Earth System Model (CESM, Jordan et al. 1999; Kauffman and Large 2002) and also built into the code of the Los Alamos Sea Ice Model (CICE, Hunke et al., 2015). This algorithm will hereafter be referred to as J99 (i.e., Jordan et al., 1999).

The other algorithm, hereafter referred to as LS87 (Liu and Schwab, 1987) was originally developed at NOAA's Great Lakes Environmental Research Laboratory (GLERL) and subsequently used in a variety of Great Lakes research and operational applications (Anderson and Schwab 2013; Beletsky et al. 2003; Rowe et al. 2015; Wang et al. 2010; and many others). Inclusion of LS87 in FVCOM was tied to the fact that the algorithm was historically part of real-time nowcasts and forecasts of NOAA's GLOFS, which is based on the Princeton Ocean Model, and that GLOFS is transitioning its physical model to FVCOM.

The WRF model (Skamarock et al., 2008) is increasingly used for regional weather and climate model applications over the Great Lakes (Benjamin et al., 2016; Xiao et al., 2016; Xue et al., 2015). The WRF model includes a one-dimensional lake model that thermodynamically interacts with the overlaying atmosphere (WRF-lake, Bonan, 1995; Gu et al., 2015; Henderson-Sellers, 1986; Hostetler and Bartlein, 1990; Hostetler et al., 1993; Subin et al., 2012) and is adapted from the lake component within the Community Land Model version 4.5 (CLM 4.5, Oleson et al., 2013; Zeng et al., 1998). The algorithm for the turbulent heat flux calculation in WRF-lake is mainly based on Zeng et al., (1998) except that roughness length scales for temperature and humidity are constant for its WRF-lake application while they are updated dynamically in CLM 4.5. Hereafter, this algorithm in its WRF-lake application is referred to as Z98L.

Finally, we include the flux algorithm from the LLTM (Croley, 1989a,b; Croley et al., 2002; Hunter et al., 2015), which is a lumped conceptual lake model that was developed for hydrological research and forecasting for the Great Lakes. LLTM is developed to simulate evaporation and heat fluxes as a lake-wide average, rather than spatially distributed. This algorithm is based primarily on the work of Croley et al. (1989a,b) and is hereafter referred to as C89.

All of the above algorithms are based on applications of Monin-Obukhov similarity theory (Kantha and Clayson, 2000b; Obukhov, 1971), where the turbulent fluxes of sensible heat, latent heat, and momentum are expressed with state variable magnitudes associated with surface friction--$T^*$, $q^*$, $u^*$ for air temperature, specific humidity, and horizontal wind velocity, respectively. In each algorithm, the major differences are in the derivation of the bulk transfer coefficients, $C_H$ and $C_E$ for the sensible heat ($H$) and the latent heat ($\lambda E$), respectively, in the bulk expressions to calculate the sensible and latent heat fluxes:

$$H = \rho_a c_p C_H S(\theta_w - \theta_a), \qquad (1)$$

$$\lambda E = \rho_a \lambda C_E S(q_w - q_a), \quad (2)$$

where $\rho_a$ is the density of air; $c_p$ and $\boldsymbol{\lambda}$ are the specific heat of air and the latent heat of vaporization, respectively; $S$ is the average value of wind speed that includes the effect of the gustiness velocity in addition to horizontal wind speed $U$ (defined later); and $\theta_w$ and $\theta_a$ ($q_w$ and $q_a$) are potential temperature (specific humidity) of the water surface and of air at the measurement height, respectively.

The transfer coefficients have a dependence on atmospheric stability that can be expressed as:

$$C_D = \kappa \left[\ln\left(\frac{z}{z_0}\right) - \Psi_M(\zeta)\right]^{-2} \quad (3)$$

$$C_{H,E} = \kappa Pr_t^{-1} \left[\ln\left(\frac{z}{z_0}\right) - \Psi_M(\zeta)\right]^{-1} \left[\ln\left(\frac{z}{z_{0\theta,0q}}\right) - \Psi_{\theta,q}(\zeta)\right]^{-1} \quad (4)$$

$$u^* = S\sqrt{C_D} \quad (5)$$

$$T^* = \Delta\theta\sqrt{C_H} \quad (6)$$

$$q^* = \Delta q\sqrt{C_E}, \quad (7)$$

where $z_0$, $z_{0\theta}$, and $z_{0q}$ are roughness length scales for momentum, temperature, and humidity respectively; $C_D$ is the drag coefficient; $\kappa$ is von Kármán constant (0.40 for COARE, Z98L, and J99; 0.41 for C89, and 0.35 for LS87); $Pr_t$ is the turbulent Prandtl number (1.0 is used in all the algorithms); and $\Psi_{M,\theta,q}(\zeta)$ are the integrated forms of stability functions for momentum, temperature, and humidity, respectively. All algorithms assume that temperature and humidity have a common value of $\Psi$, i.e. $\Psi_\theta = \Psi_q = \Psi_M$. $\zeta = z/L$ is the stability factor, where $L$ is the Obukhov length and $z$ is the measurement height.

Differences among the algorithms are primarily in how they estimate $\Psi_{M,\theta,q}(\zeta)$ and $z_0$. The profile functions $\Psi_{M,\theta,q}(\zeta)$ are typically divided into three regimes, namely unstable, mildly stable, and strongly stable. All the algorithms use Businger-type parameterizations (Businger et al., 1971; Kraus and Businger, 1995) for the unstable regime (Table 1), except COARE which includes convective behaviour in highly unstable conditions by introducing a stability function for a

convective limit (Fairall et al., 1996a; Supplemental Tables 1 and 2). For stable conditions, Holtslag et al. (1990) is used in LS87, C89, and Z98L while Beljaars and Holtslag (1991) is used in J99 and COARE (Table 1). Note that there are minor differences in coefficients of $\Psi_{M,\theta,q}(\zeta)$ in the algorithms, which can be found in Supplemental Tables 1 and 2.

The roughness length scale for momentum, $z_0$, is often parameterized as a function of friction velocity $u^*$ ~~or the average value of wind speed $S$~~. The LS87, C89, and COARE algorithms apply Charnock's formula (Charnock, 1955; Smith, 1988):

$$z_0 = \frac{\alpha u^{*2}}{g} + \frac{0.11\nu}{u^*} \quad , \tag{8}$$

where $z_0$ is the roughness length scale of momentum, $\alpha$ is the Charnock parameter, $g$ is the acceleration due to gravity, and $\nu$ is kinematic viscosity. Because the value of $z_0$ feeds back into the value of $u^*$ via equations (3) and (5), equation (8) must be solved iteratively to arrive at final values of these variables. Here, COARE calculated the Charnock parameter $\alpha$ as a function of wind speed, while LS87 and C89 use a constant $\alpha$ (Table 1). In contrast to the Charnock formula (equation 8), J99 directly calculates $z_0$ as a function of wind speed based on Large and Pond (1981), while Z98L assumes $z_0$ to be a constant 0.001 m. In the original paper of Zeng et al. (1998), non-constant parameterizations for roughness length scales were used, namely Smith (1988) for momentum and Brutsaert (1982) for temperature and humidity. The constant value in Z98L is likely related to the fact that the implementation in WRF handles the lake surface as part of various land surface types, whose roughness lengths for momentum are often assumed to be constant (Mitchell et al., 2005; Oleson et al., 2013), while the original work of Zeng et al. (1998) assumed ocean surface applications.

Evidence suggests $z_0$ can be significantly larger than $z_{0\theta,q}$, because momentum is transported across the air-sea interface by pressure forces acting on roughness elements, while heat and water vapor must ultimately be transferred by molecular diffusion across the interfacial sublayer (Brutsaert, 1975; Garratt, 1992; Kantha and Clayson, 2000a). However, many land and lake models, including four of the five algorithms used in this study, assume the same roughness length for momentum and heat transfer; for example, Croley (1989b, C89); Liu and Schwab (1987, LS87); Oleson et al. (2013); Zeng et al. (1998, Z98L); the CICE application (J99), the previous

NCEP Eta model described in Chen et al., 1997; and the Canadian operational weather and hydrologic models described in Deacu et al. (2012). Deacu et al. (2012) showed that the same value for $z_0$ and $z_{0\theta,q}$ resulted in overestimation of turbulent heat fluxes over Lake Superior, and that the overestimation was reduced by using the smooth surface parameterization for $z_{0\theta,q}$, with an empirical coefficient based on Beljaars (1994).

As part of the current study, we intend to conduct a similar experiment to Deacu et al. (2012), namely, updating the original $z_{0\theta,q}$ parameterization in the LS87, C89, Z98L, and J99 algorithms to a more realistic parameterization. We conduct this experiment to identify errors in $\lambda E$ and $H$ simulations with these algorithms' original $z_{0\theta,q}$ formulation and to evaluate how much the errors could be reduced in this way. We use an alternative $z_{0\theta,q}$ formulation that is based on Fairall et al. (2003), which is used in COARE. The formulation utilizes the Liu–Katsaros–Businger model (LKB; Liu et al., 1980), with updates described in Fairall et al. (2003), where a simpler empirical relationship was formulated to represent the LKB model, based on a fit to observational data:

$$z_{0\theta,q} = \min\left(1.6\times10^{-4}, 5.8\times10^{-5}Rr^{-0.72}\right), \quad (9)$$

where $Rr=u^{*}z_0/\nu$ is the roughness Reynolds number, which is also updated throughout the iterations. We test both of the original and updated parameterizations for $z_{0\theta,q}$ in the heat flux simulations.

"Gustiness" velocity $w_g$ is included in Z89L and COARE to account for the additional flux induced by the convective boundary layer in low wind speed regimes. The average value of wind speed $S$ is defined as,

$$S = \sqrt{U^2 + {w_g}^2}, \quad (10)$$

where $U$ is the mean horizontal wind speed. $w_g$ is defined as

$$w_g = \beta \frac{g}{\rho_a}\left[\frac{H}{c_p T} + 0.61E\right]., \quad (11)$$

where $\beta$ is an empirical constant set to $\beta$ =1.2 in COARE and $\beta$ =1.0 in Z89L. Further details of the gustiness velocity formulations are described by Fairall et al. (1996a). In LS87, C89, and J99, $S$ is assumed to be identical to $U$.

All algorithms require meteorological inputs of horizontal wind speed $U$, potential air temperature $\theta_a$, potential temperature at the water surface $\theta_w$, a humidity-related variable (dew point for LS87, relative humidity for C89, Z98L, COARE, and specific humidity for J99), air pressure, and sensor height. These meteorological inputs should represent a temporal mean field over the corresponding eddy covariance measurement. $U$, $\theta_a$, and $\theta_w$ can be directly used in eqs. (1) and (2), while $q_w$ and $q_a$ need to be derived from relative humidity, water surface (or air) temperature, and air pressure.

### 2.3. Intercomparison methods

We take the following steps to compare and verify simulated sensible and latent heat fluxes against observed fluxes:

1) The five algorithms were forced by half-hourly meteorological data ($U$, $\theta_a$, $\theta_w$, relative humidity, air pressure). Missing values were assigned for simulated heat fluxes when any observed values of $U$, $\theta_a$, $\theta_w$, and relative humidity were not available or when lake ice was present at a site.
2) Temporal averaging was applied to simulated and observed fluxes. We first calculated daily averaged $\lambda E$ and $H$. Gap-matching was applied to the simulated and observed fluxes. If either of the simulated or observed $\lambda E$ ($H$) were missing values at a half-hourly time step, both the simulated and observed $\lambda E$ ($H$) at this time step were not used for daily averaging. This was conducted so that daily averages from the simulation (roughly continuous in time) were adequately compared with those from the observations, which had more frequent gaps. When more than 24 out of 48 data points were missing in a day, a missing value (-9999) was assigned. For time series comparison, a 10-day moving average was applied to simulated and observed fluxes in order to smooth the synoptic variability and highlight comparison of the respective seasonal cycles. Daily averaging was used for one-to-one comparisons (i.e. scatter plots).
3) Root-mean-square errors (RMSEs) and mean bias were calculated for daily sensible and latent heat fluxes.

4) Errors of daily $\lambda E$ and $H$ were calculated as functions of $\theta_w$-$\theta_a$, $q_w$-$q_a$, $U$, $C_{H,E}$, and $\zeta$.

## 3. Results

### 3.1. Observed and modeled seasonal cycles

Figure 3 shows the time series of air temperature, water surface temperature from GLSEA, relative humidity, and wind speed at the four stations. The time series for Stannard Rock are relatively gap-free throughout the three years, while there are some data gaps in the time series for the other stations. The air temperature time series are characterized by a typical seasonal cycle, along with relatively warm and cold winters in 2012-2013 and 2013-2014, respectively (Fig. 3a). This is also reflected in the water surface temperature time series (Fig. 3b), where only White Shoal and Long Point were affected by ice cover in the winter (January-March) of 2012-2013, shown as gaps in the time series, whereas all four stations were affected by ice cover in the winter of 2013-2014. In the spring and summer of 2012 (April-September), the water surface temperature was anomalously warm compared with the same periods of 2013 and 2014 (particularly at Stannard Rock). Relative humidity generally fluctuated between 50 % and 90 % (Fig. 3c), while wind speed (Fig. 3d) shows a weak seasonal cycle of relatively high wind speeds during fall and winter (October-March) and lower wind speeds during spring and summer (April-September).

Figures 4-7 show visual comparisons of 10-day moving average time series of $\lambda E$ and $H$ at each of the four stations. Overall, all five algorithms simulated the general seasonal cycles of $\lambda E$ and $H$, including the observed high fluxes during fall and winter and low fluxes during summer and spring that is typical for large North American lakes (Blanken et al., 1997, 2000, 2011; Spence et al., 2011). On the other hand, there are notable overestimations of $\lambda E$ and $H$ by the original algorithms, particularly at Stannard Rock (Fig. 4) in the fall ($\lambda E$) and winter ($H$).

The observed $\lambda E$ and $H$ at Stannard Rock are largely gap-free (Fig. 4), showing most continuous timeseries of seasonal cycles, aside from periods of high ice coverage during the cold winter of 2013-2014. A few additional data gaps also occurred, including late summer of 2012, a longer data gap during January-May 2014, and very short data gap during December 2013.

At Stannard Rock, Late-fall (October-December) $\lambda E$ and $H$ were relatively low in 2012 (3-month averages 84 W m$^{-2}$ for $\lambda E$ and 55 W m$^{-2}$ for $H$) and high in 2013 (119 W m$^{-2}$ in $\lambda E$, 85 W m$^{-2}$ for $H$), indicating preconditioning of the following mild and severe winters, respectively. During spring and summer of both years (April-September), the observed $\lambda E$ and $H$ were much lower due to the cool lake surface relative to the overlying air. The simulated $\lambda E$ and/or $H$ mostly reproduced these lower values, but also showed occasionally negative values (Fig. 4), such as during May 2012 and July 2014. During these periods, the air was predominantly warmer than the water surface (i.e $T_w$-$T_a$ < 0, Fig. 1), and specific humidity gradients were near zero during May 2012 and reversed (i.e., air-to-water) during July 2014, forcing the algorithms to simulate near-zero and negative (i.e. downward) fluxes, respectively. However, the observed $\lambda E$ and $H$ fluxes remained close to zero, but slightly positive.

The forcing dataset for White Shoal (Fig. 3) is relatively gap-free as well, but there was an missing data period before October 2013 for $\lambda E$, and data gaps in $H$ due to ice cover (Fig. 5). White Shoal tends to be influenced by ice cover even in mild winters, since typical south-westerly winds push ice in Lake Michigan downwind, causing ice accumulation in northern parts of the lake near White Shoal. As such, the exclusion of turbulent flux data for this analysis during the mild winter of 2012-2013 at White Shoal was due to ice cover. These observations also showed contrasting late-fall heat fluxes during the two years: three-month average $H$ was 40 W m$^{-2}$ during October-December 2012 and 61 W m$^{-2}$ during October-December 2013. Some model underestimation of the sensible heat flux ($H$) occurred during July-September 2013 and June-October 2014.

The Spectacle Reef forcing dataset (Fig. 3) and flux dataset (Fig. 6) both contained a long gap from March 2012 to September 2013 due to electrical problems from lightning strikes. A data gap in $\lambda E$ and $H$ during January-March 2014 was due to ice cover, but unlike White Shoal, Spectacle Reef is less affected by ice cover. This is because winds carry ice that forms nearshore toward the east and offshore in Lake Huron, keeping the area around the flux tower largely in open water. Indeed, the station was not affected by ice cover in the winter of 2012-2013 (based on the NIC data), but this period was included in the above-referenced long data gap due to lighting strikes.

The dataset at Long Point (Fig. 7) shows the largest number of data gaps due to the additional filtering according to wind direction of 180$^o$ (S)-315$^o$ NW, which included typical south-westerly winds in this region. The significant data gaps at Spectacle Reef and Long Point, therefore, do not allow us to compare the late-fall fluxes between the anomalous two years. However, for the

purpose of the algorithm verification, the data at the two stations were still valuable, and forcing datasets were largely continuous (Fig. 3).

Also shown in Figs. 4-7 are model results using both the original and updated $z_{0\theta,q}$ parameterizations (eq. 9). The original results of LS87, C89, Z98L, and J99 showed overestimation of $\lambda E$ and $H$ at Stannard Rock by anywhere from 33-50 % for most of the algorithms to ~80 % overestimation for Z98L (both $\lambda E$ and $H$) and LS87 ($\lambda E$) (Fig. 4, Table 2). These overestimations were particularly obvious during high flux events in fall and winter (October-March). The overestimation at Stannard Rock was significantly lessened to roughly 24-33 % error by using the updated $z_{0\theta,q}$ formula (eq. 9). This is consistent with the findings of Deacu et al. (2012), who showed the improvements of simulated latent and sensible heat flux simulation by updating the roughness length scale parameterization at Stannard Rock for the December 2008 simulation period. Similar improvements are noted at White Shoal (Fig. 5), Spectacle Reef (Fig. 6), and Long Point (Fig. 7).

### 3.2. Comparison of daily mean fluxes

While the 10-day running mean time series of $\lambda E$ and $H$ provided an effective way to illustrate the overall cycle (Figs. 4-7), abrupt changes in $\lambda E$ and $H$ often occur on daily timescales, caused by the passage of frontal systems and cold air outbreaks (Blanken et al., 2008). Thus, we further evaluated the performance of the various algorithms at daily timescales by means of scatter plots of observed and modeled daily mean heat fluxes (Figs. 8 and 9). Data points of $\lambda E$ (Fig. 8) diverged more from the 1:1 line than $H$ (Fig. 9), showing both overestimated fluxes (at Stannard Rock and Long Point with Z98L) and underestimated fluxes (at Spectacle Reef). Overall, the updated $z_{0\theta,q}$ formula reduced simulated $\lambda E$, generally bringing the fluxes into better agreement with observations. An exception to this occurred for $\lambda E$ at Spectacle Reef, where the agreement became slightly worse with the updated formulation. The error reduction ratio was negative and the mean bias was more negative with the updated formulation at this station (Table 2). For $H$ (Fig. 9, Table 3), notable overestimation was seen in the original J99, LS87, and Z98L, particularly at relatively large heat loss values (> ~300 W $m^{-2}$). At Stannard Rock, Spectacle Reef, and Long Point, this overestimation was significantly improved with the updated $z_{0\theta,q}$ formula. At White Shoal, however, the improvement was not as

significant, 28 % compared to 57 % for Stannard Rock, 69 % for Spectacle Reef, and 50 % for Long Point according to error reduction ratios (Table 3).

Stannard Rock showed small groups of $\lambda E$ and $H$ around the origin, where the simulated fluxes underestimated the observed fluxes (i.e. below the 1:1 line). These data represented two summer periods when the observed fluxes were near zero, but the simulated fluxes were negative (see the discussion in section 3.1). At White Shoal, there was a population of $H$ values below the 1:1 line, representing periods when the simulation results underestimated the observations during July-September 2013 and June-October 2014 (see the discussion in section 3.1).

### 3.3. Error dependence on meteorological conditions and transfer coefficients

Figures 10 and 11 show the magnitude of error in simulated daily $\lambda E$ and $H$ (i.e., difference from observations) as functions of $\theta_w$-$\theta_a$, $q_w$-$q_a$, $C_H$, $C_E$, $U$, and $\zeta$=$z/L$ for the five algorithms at Stannard Rock. Similar results were observed in the error and bias analyses at the other sites (supplementary Figures S1-S6). There are several features common in all the algorithms: The $H$ ($\lambda E$) errors were positively correlated with $\theta_w$-$\theta_a$ ($q_w$-$q_a$) for negative values of $\theta_w$-$\theta_a$ ($q_w$-$q_a$); the amplitudes of the errors become large (both positive and negative) as wind speed increases; and the majority of data are in the range $-2<\zeta<0$ (unstable). Most notably, the transfer coefficients $C_H$ and $C_E$ were significantly reduced with the updated $z_{0\theta,q}$ formula, which also reduced the error in the $\lambda E$ and $H$ simulations. This was to be expected, since $z_{0\theta}$ and $z_{0q}$ are directly translated into $C_H$ and $C_E$ respectively. The study period did not include the occurrence of highly unstable conditions ($\zeta<<-1$, Figs. 10 and 11, Supplemental figures S1-S6). Therefore, the period was not sufficient to evaluate the convective behaviour treatment in COARE. Also, the study period did not include sustained low wind speeds. Fairall et al. (1996a) note that the gustiness parameterization has only a modest effect until the wind speed becomes less than 2-3 m $s^{-1}$. The wind speeds during our study period were mostly greater than 3 m $s^{-2}$ (Figs. 10 and 11, Supplemental figures S1-S6) and therefore did not allow us to evaluate the influence of the gustiness parameterizations in COARE and Z98L.

## 4. Discussion

The simulation results of four of the five algorithms investigated here (J99, C89, LS87, Z98L) were overall improved by the updated $z_{0\theta,q}$ formula (Eq. 9), bringing the simulation results into closer correspondence with the COARE simulations. In our study period, we did not see clear advantages in the simulation results with the other differences among the algorithms. For example, we did not observe a clear difference in the results when using the various stability functions (i.e. $\Psi_{M,\theta,q}$) in the algorithms. Evaluations of the convective behaviour treatment in COARE and the gustiness effect in COARE and Z98L were not possible as our study period did not have appropriate conditions, as mentioned in section 3.3. The notably smaller value of von Kármán constant used in LS87 (0.35) also contributes to the simulated $\lambda E$ and $H$. Indeed, a test simulation with $\kappa$=0.41 with LS87 resulted in ~30% larger values of $\lambda E$ and $H$ (not shown). However, this makes the algorithm's overestimation of $\lambda E$ and $H$ even worse. Thus, the most important factor in our analyses to improve the $\lambda E$ and $H$ simulations was the parameterization of roughness length scales for temperature and humidity ($z_{0\theta}$ and $z_{0q}$). Formulae for $z_{0\theta,q}$ with smooth surface parameterization (such as Eq. 9) have been widely used for air-sea interaction modeling (e.g. Beljaars, 1994; Fairall et al., 2003) and have been also verified in lake applications (Deacu et al., 2012). It is reasonable that future updates of the four algorithms should include the updated or similar formulation of $z_{0\theta,q}$

The inclusion of the updated $z_{0\theta,q}$ formula will not guarantee immediate improvement of the parent model systems. This is because each of the model systems is complex and must embrace uncertainties from all aspects, including forcing, dynamics, and boundary conditions. Typically, such a system is calibrated to provide best estimates of certain variables for its own purpose (e.g. water temperature for the implementation of FVCOM in GLOFS), and a sudden change to a single aspect of the system would lose a balance that has been achieved by extensive calibration. An ideal approach to improve model systems would have to be more comprehensive in terms of model variables of which a system is expected to provide best estimates. For example, in FVCOM, it may be a combination of improvements to a meteorological data set that drives the hydrodynamic model and to a bulk flux algorithm in it.

Simulated negative values of $\lambda E$ and $H$ contrast with near zero, but positive, observed values during summer at Stannard Rock (Fig. 4, around May 2012 and July 2014). Although the

magnitude of these negative values is much smaller than the positive values in other seasons which are more influential on the annual energy budget, there is still an issue that the reasons behind this discrepancy that are not fully understood. A similar discrepancy was found at Long Point (Fig. 7, around April 2013) and at White Shoal (Fig. 5, around June 2014) although the discrepancy was only for $H$ and the magnitude of the discrepancy was smaller than at Stannard Rock. The discrepancy remained even after updating the $z_{0\theta,q}$ formula. During these periods, the temperature gradients between the air (at sensor heights) and at the water surface were commonly negative (the air was warmer) and wind speeds ranged from 6-12 m $s^{-1}$, resulting in the negative fluxes (i.e., downward) simulated by the bulk flux algorithms. One possible reason is that the sensors were above the constant flux layer during these periods, and therefore, the similarity theory on which the bulk flux algorithms are based on was not applicable. However, given the sensor heights at these stations, such a condition are likely rare.

Other possible sources of the discrepancy could be in the forcing data, particularly, uncertainties in the GLSEA water surface temperature data. As described in section 2.1.2, the information for cloudy areas is created using an interpolation method from the satellite imageries within ±10 days. Therefore, the GLSEA data tends to have lower accuracy and could miss abrupt changes in water surface temperature for cloudy days. The IRT-measured water surface temperature showed somewhat warmer water surface temperature than the GLSEA data during these discrepancy periods (Supplemental figure S7), indicating possible underestimation of water surface temperature in the GLSEA data and resulting in false negative $\lambda E$ and $H$. However, we concluded earlier that the accuracy of the IRT-measured water surface temperature was limited (see section 2.1.2). An ideal way to confirm the GLSEA accuracy for such analyses would be an *in situ* water surface temperature measurements at the flux tower sites using buoys, for example (which is now being done at Stannard Rock and planned for White Shoal). Also, a recent experimental method by Moukomla and Blanken (2016), who derived water surface temperature from MODIS (or Moderate Resolution Imaging Spectroradiometer) for all-sky conditions, may be tested in the future.

The normalized RMSEs at Long Point were worse than those at the other stations even though we filtered out the data with wind directions in the range of 180° S -315° NW. which should be sufficiently large to remove any possible land surface contamination. We again suspect the water surface temperature data could be an error source. As noted in section 2.1, the station is on the

shore of a narrow peninsula extending into Lake Erie. The satellite-based observations of water surface temperature tend to lose their accuracy near the coast due to pixel contamination, and thus the GLSEA accuracy at this station could be lower. For such a location, FVCOM, a full hydrodynamic model, may be appropriate to reproduce the observed fluxes, but it should have sufficient horizontal resolution to represent the complex bathymetry around the peninsula, which is essential to reproduce the spatial pattern of heat capacity in the water column correctly, and therefore, the water surface temperature.

## 5. Summary and conclusions

This study focused on the validation of surface latent and sensible heat fluxes ($\lambda E$ and $H$, respectively) from the surface of the Great Lakes. We isolated the surface flux algorithms commonly used in Great Lakes physical modeling and tested each algorithm using observed meteorology and lake surface temperatures by comparing their output to several eddy covariance stations within the GLEN that measure *in situ* lake surface fluxes. All algorithms reproduced the seasonal cycle of $\lambda E$ and $H$ reasonably well during a warm period (2012-mid 2013) and cold period (late 2013-2014). However, four of the original algorithms (i.e., except for COARE) presented notable disagreement with the observations under certain conditions; significant positive biases in $H$ were found under high upward (cooling of water surface) heat flux conditions for the algorithms other than COARE; the errors in $H$ were also positively correlated with the temperature difference between air and water.

These errors were significantly improved by introducing the updated $z_{0\theta,q}$ formula based on Fairall et al., (2003), which is well supported by the air-sea interaction modeling community. The update led to reduced transfer coefficients $C_H$ and $C_E$, reducing the overestimation of the simulated heat fluxes. With the updated formula for $z_{0\theta,q}$, the four models (LS87, C89, J99, Z98L) simulated heat fluxes similar to COARE's. While it is reasonable to adopt the updated formula in the parent model systems where these algorithms are included, this does not guarantee immediate improvement of simulations by the parent model systems, since these model systems are calibrated to provide best simulations for certain variables by embracing uncertainties in all aspects. We used *in situ* meteorological forcing to drive the algorithms, which is generally ideal, but in operational

practice, it is not possible to use *in situ* data over the entire lake surface. For example, GLOFS uses interpolated and/or model-forecasted meteorological forcing, which inevitably includes additional sources of error.

It should not be a great surprise that adjustment of roughness length $z_{0\theta,q}$ is a primary factor in correcting turbulent fluxes. In eqs. (3) and (4), $z_{0\theta,q}$ and $\Psi_{M,\theta,q}(\zeta)$ are the only terms for which some discretion is left for the algorithm to specify a value. One anchoring point for $\Psi_{M,\theta,q}(\zeta)$ is that it must be zero for neutral stability conditions. As long as the algorithms' values of $\Psi_{M,\theta,q}(\zeta)$ do not disagree strongly, the value of $z_{0\theta,q}$ will primarily control the turbulent fluxes.

We successfully evaluated the flux algorithms, which are an important aspect of Great Lakes water and energy balance modeling, and we identified and reduced errors in simulated heat fluxes from these algorithms. We recommend that bulk flux algorithms use an appropriate parameterization for $z_{0\theta}$ and $z_{0q}$ instead of assuming them equal to $z_0$, or simply employ the COARE algorithm, which presented the best agreement with the eddy covariance measurements in this study. We also recommend simultaneous *in situ* measurement of water surface temperature at the flux tower locations in order to allow more robust comparison between the eddy covariance measurements and simulated $\lambda E$ and $H$ by a column model (e.g., the five algorithms independently driven by the forcing data in this study).

Accurate simulation of the turbulent heat fluxes from the lake surface is important to a wide range of lake-atmosphere and earth system applications, from long-term water balance estimates to numerical prediction of lake levels, weather, lake ice, and regional climate. Communities are increasingly dependent on numerical geophysical models for these types of applications. The continued monitoring of turbulent heat fluxes at the offshore GLEN sites is critical for such models to be improved in future studies.

**Acknowledgments**

This research is funded by the US Environmental Protection Agency's Great Lakes Restoration Initiative (GLRI) and the NOAA' Coastal Storms Program awarded to the Cooperative Institute for Great Lakes Research (CIGLR) through the NOAA Cooperative Agreement with the University of Michigan (NA12OAR4320071). Dr. Umarporn Charusombat was supported by a National Research Council Research Associateship award at the NOAA Great Lakes

Environmental Research Laboratory. The data used in this research were kindly provided by the Great Lakes Evaporation Network (GLEN). GLEN data compilation and publication were provided by LimnoTech, the University of Colorado at Boulder, and Environment and Climate Change Canada under Award/Contract No. 10042-400759 from the International Joint Commission (IJC) through a subcontract with the Great Lakes Observing System (GLOS). The statements, findings, conclusions, and recommendations are those of the authors and do not reflect the views of GLEN, LimnoTech, the University of Colorado at Boulder, Environment and Climate Change Canada, the IJC, or GLOS. The authors thank Drs. Chris Fairall and Dev Niyogi for comments that improved the quality of this manuscript. This is GLERL Contribution Number XXXX and CIGLR contribution XXXX.

**References**

Anderson, E. J. and Schwab, D. J.: Predicting the oscillating bi-directional exchange flow in the Straits of Mackinac, J. Great Lakes Res., 39(4), 663–671, doi:10.1016/j.jglr.2013.09.001, 2013.

Anderson, E. J., Bechle, A. J., Wu, C. H., Schwab, D. J., Mann, G. E. and Lombardy, K. A.: Reconstruction of a meteotsunami in Lake Erie on 27 May 2012; Roles of atmospheric conditions on hydrodynamic response in enclosed basins, J. Geophys. Res., 120, 1–16, doi:10.1002/2014JC010564., 2015.

Andreas, E. L.: A theory for the scalar roughness and the scalar transfer coefficients over snow and sea ice, Boundary-Layer Meteorol., 38(1–2), 159–184, doi:10.1007/BF00121562, 1987.

Bai, X., Wang, J., Schwab, D. J., Yang, Y., Luo, L., Leshkevich, G. A. and Liu, S.: Modeling 1993-2008 climatology of seasonal general circulation and thermal structure in the Great Lakes using FVCOM, Ocean Model., 65, 40–63, doi:10.1016/j.ocemod.2013.02.003, 2013.

Baldocchi, D. D., Hicks, B. B. and Meyers, T. P.: Measuring biosphere–atmosphere exchanges of biologically related gases with micrometeorological methods, Ecology, 69, 1331–1340, 1988.

Bélair, S., Brown, R., Mailhot, J., Bilodeau, B. and Crevier, L.-P.: Operational implementation of the ISBA land surface scheme in the Canadian Regional Weather Forecast Model. Part I: Warm season results, J. Hydrometeorol., 4(2), 352–370, doi:10.1175/1525-

7541(2003)4<352:OIOTIL>2.0.CO;2, 2003a.

Bélair, S., Brown, R., Mailhot, J., Bilodeau, B. and Crevier, L.-P.: Operational implementation of the ISBA land surface scheme in the Canadian Regional Weather Forecast Model. Part II: Cold season results, J. Hydrometeorol., 4(2), 371–386, doi:10.1175/1525-7541(2003)4<371:OIOTIL>2.0.CO;2, 2003b.

Beletsky, D., Schwab, D. J., Roebber, P. J., McCormick, M. J., Miller, G. S. and Saylor, J. H.: Modeling wind-driven circulation during the March 1998 sediment resuspension event in Lake Michigan, J. Geophys. Res., 108(C2), 3038, doi:10.1029/2001JC001159, 2003.

Beletsky, D., Schwab, D. and McCormick, M.: Modeling the 1998-2003 summer circulation and thermal structure in Lake Michigan, J. Geophys. Res. Ocean., 111(10), 1–18, doi:10.1029/2005JC003222, 2006.

Beljaars, A. C. M.: The parametrization of surface fluxes in large-scale models under free convection, Q. J. R. Meteorol. Soc., 121(522), 255–270, doi:10.1002/qj.49712152203, 1994a.

Beljaars, A. C. M.: The parametrization of surface fluxes in large-scale models under free convection, Q. J. R. Meteorol. Soc., 121(522), 255–270, doi:10.1002/qj.49712152203, 1994b.

Beljaars, A. C. M. and Holtslag, A. A. M.: Flux parameterization over land surfaces for atmospheric models, J. Appl. Meteorol., 30(3), 327–341, doi:10.1175/1520-0450(1991)030<0327:FPOLSF>2.0.CO;2, 1991.

Benjamin, S. G., Weygandt, S. S., Brown, J. M., Hu, M., Alexander, C., Smirnova, T. G., Olson, J. B., James, E., Dowell, D. C., Grell, G. A., Lin, H., Peckham, S. E., Smith, T. L., Moninger, W. R., Kenyon, J. and Manikin, G. S.: A North American hourly assimilation and model forecast cycle: The rapid refresh, Mon. Weather Rev., 144(4), 1669–1694, doi:10.1175/MWR-D-15-0242.1, 2016.

Blanken, P. D., Black, T. a., Yang, P. C., Neumann, H. H., Nesic, Z., Staebler, R., den Hartog, G., Novak, M. D. and Lee, X.: Energy balance and canopy conductance of a boreal aspen forest: Partitioning overstory and understory components, J. Geophys. Res., 102(D24), 28915, doi:10.1029/97JD00193, 1997.

Blanken, P. D., Rouse, W. R., Culf, A. D., Spence, C., Boudreau, L. D., Jasper, J. N., Kochtubajda, B., Schertzer, W. M., Marsh, P. and Verseghy, D.: Eddy covariance measurements of evaporation from Great Slave Lake, Northwest Territories, Canada, Water Resour. Res., 36(4), 1069–1077, doi:10.1029/1999WR900338, 2000.

Blanken, P. D., Rouse, W. R. and Schertzer, W. .: The time scales of evaporation from Great Slave Lake, in *Atmospheric Dynamics of a Cold Region: The Mackenzie GEWEX Study Experience, vol. II: Hydrol*, edited by M.-K. Woo, pp. 181–196, Springer, New York., 2008.

Blanken, P. D., Spence, C., Hedstrom, N. and Lenters, J. D.: Evaporation from Lake Superior: 1. Physical controls and processes, J. Great Lakes Res., 37(4), 707–716, doi:10.1016/j.jglr.2011.08.009, 2011.

Bonan, G. B.: Sensitivity of a GCM simulation to inclusion of inland water surfaces, J. Clim., 8(11), 2691–2704, doi:10.1175/1520-0442(1995)008<2691:SOAGST>2.0.CO;2, 1995.

Brutsaert, W.: A theory for local evaporation (or heat transfer) from rough and smooth surfaces at ground level, Water Resour. Res., 11(4), 543–550, doi:10.1029/WR011i004p00543, 1975.

Brutsaert, W.: Evaporation into the Atmosphere - Theory, History and Applications, Springer., 1982.

Businger, J. A., Wyngaard, J. C., Izumi, Y. and Bradley, E. F.: Flux-profile relationships in the atmospheric surface layer, J. Atmos. Sci., 28(2), 181–189, doi:10.1175/1520-0469(1971)028<0181:FPRITA>2.0.CO;2, 1971.

Charnock, H.: Wind stress on a water surface, Q. J. R. Meteorol. Soc., 81(350), 639–640, doi:10.1002/qj.49708135027, 1955.

Chen, C., Beardsley, R. C. and Cowles, G.: An unstructured grid, finite-volume coastal ocean model FVCOM user manual, Oceanography, 19(1), 78–89, doi:10.5670/oceanog.2006.92, 2006a.

Chen, C., Beardsley, R. C. and Cowles, G.: An unstructured grid, finite volume coastal ocean model (FVCOM) system, Oceanography, 19(1), 78–89, doi:10.5670/oceanog.2006.92, 2006b.

Chen, C., Beardsley, R., Cowles, G., Qi, J., Lai, Z., Gao, G., Stuebe, D., Xu, Q., Xue, P., Ge, J., Hu, S. and Ji: An unstructured grid, Finite-Volume Coastal Ocean Model FVCOM -- User Manual, Tech. Rep., SMAST/UMASSD-13-0701, Sch. Mar. Sci. Technol., Univ. Mass. Dartmouth, New Bedford., 416 pp, 2013.

Chen, F., Janjić, Z. and Mitchell, K.: Impact of atmospheric surface-layer parameterizations in the new land-surface scheme of the NCEP Mesoscale Eta Model, Boundary-Layer Meteorol., 85(3), 391–421, doi:10.1023/A:1000531001463, 1997.

Chu, P. Y., Kelley, J. G. W., Mott, G. V., Zhang, A. and Lang, G. A.: Development, implementation, and skill assessment of the NOAA/NOS Great Lakes Operational Forecast System, Ocean Dyn., 61(9), 1305–1316, doi:10.1007/s10236-011-0424-5, 2011.

Clites, A. H., Wang, J., Campbell, K. B., Gronewold, A. D., Assel, R., Bai, X. and Leshkevich, G. A.: Cold Water and High Ice Cover on Great Lakes in Spring 2014, Earth Obs. Syst., 95(34), 305–312, 2014.

Croley, T. E.: Verifiable evaporation modeling on the Laurentian Great Lakes, Water Resour. Res., 25(5), 781–792, doi:10.1029/WR025i005p00781, 1989a.

Croley, T. E. .: Lumped Modeling of Laurentian Great Lakes Evaporation , Heat Storage , and Energy Fluxes for Forecasting and Simulation, NOAA Tech. Memo., ERL GLERL-, 48, 1989b.

Croley, T. E., Assel, R. A. and Arbor, A.: Great Lakes Evaporation Model Sensitivities and Errors, Second Fed. Interag. Hydrol. Model. Conf. Subcomm. Hydrol. Interag. Advis. Comm. Water Data, (9), 1–12, 2002.

Deacu, D., Fortin, V., Klyszejko, E., Spence, C. and Blanken, P. D.: Predicting the net basin supply to the Great Lakes with a hydrometeorological model, J. Hydrometeorol., 13(6), 1739–1759, doi:10.1175/JHM-D-11-0151.1, 2012.

Edson, J. B., Hinton, A. A., Prada, K. E., Hare, J. E. and Fairall, C. W.: Direct covariance flux estimates from mobile platforms at sea, J. Atmos. Ocean. Technol., 15(2), 547–562, doi:10.1175/1520-0426(1998)015<0547:DCFEFM>2.0.CO;2, 1998.

Edson, J. B., Jampana, V., Weller, R. a., Bigorre, S. P., Plueddemann, A. J., Fairall, C. W., Miller, S. D., Mahrt, L., Vickers, D. and Hersbach, H.: On the Exchange of Momentum over the Open Ocean, J. Phys. Oceanogr., 43(8), 1589–1610, doi:10.1175/JPO-D-12-0173.1, 2013.

Fairall, C. W., Bradley, E. F., Rogers, D. P., Edson, J. B. and Young, G. S.: Bulk parameterization of air-sea fluxes for Tropical Ocean-Global Atmosphere Coupled-Ocean Atmosohere Response Experiment, J. Geophys. Res., 101, 3747–3764, 1996a.

Fairall, C. W., Bradley, E. F., Godfrey, J. S., Wick, G. A., Edson, J. B. and Young, G. S.: Cool-skin and warm-layer effects on sea surface temperature, J. Geophys. Res., 101(C1), 1295–1308, doi:10.1029/95JC03190, 1996b.

Fairall, C. W., Bradley, E. F., Hare, J. E., Grachev, A. A. and Edson, J. B.: Bulk parameterization of air-sea fluxes: Updates and verification for the COARE algorithm, J. Clim., 16(4), 571–591, doi:10.1175/1520-0442(2003)016<0571:BPOASF>2.0.CO;2, 2003.

Garratt, J. R.: The Atmospheric Boundary Layer, Cambridge University Press, Cambridge., 1992.

Gronewold, A. D., Clites, A. H., Hunter, T. S. and Stow, C. A.: An appraisal of the Great Lakes advanced hydrologic prediction system, J. Great Lakes Res., 37(3), 577–583,

doi:10.1016/j.jglr.2011.06.010, 2011.

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

Gronewold, A. D. and Stow, C. A.: Water loss from the Great Lakes, Science, 343(6175), 1084–1085, doi:10.1126/science.1249978, 2014.

Gu, H., Jin, J., Wu, Y., Ek, M. B. and Subin, Z. M.: Calibration and validation of lake surface temperature simulations with the coupled WRF-lake model, Clim. Change, 129(3–4), 471–483, doi:10.1007/s10584-013-0978-y, 2015.

Henderson-Sellers, B.: Calculating the surface energy balance for lake and reservoir modeling: A review, Rev. Geophys., 24(3), 625–649, doi:10.1029/RG024i003p00625, 1986.

Holtslag, A. A. M., De Bruijn, E. I. F. and Pan, H.-L.: A High Resolution Air Mass Transformation Model for Short-Range Weather Forecasting, Mon. Weather Rev., 118(8), 1561–1575, doi:10.1175/1520-0493(1990)118<1561:AHRAMT>2.0.CO;2, 1990.

Horst, T. W.: A simple formula for attenuation of eddy fluxes measured with first-order-response scalar sensors, Boundary-Layer Meteorol., 82(2), 219–233, doi:10.1023/A:1000229130034, 1997.

Hostetler, S. and Bartlein, P.: Simulation of Lake Evaporation with Application to Modeling Lake Level Variations, Oregon, Water Resour. Res., 26(10), 2603–2612, 1990.

Hostetler, S. W., Bates, G. T. and Giorgi, F.: Interactive coupling of a lake thermal model with a regional climate model, J. Geophys. Res., 98(D3), 5045, doi:10.1029/92JD02843, 1993.

Hunke, E. C., Lipscomb, W. H., Turner, A. K., Jeffery, N. and Elliott, S.: CICE : the Los Alamos Sea Ice Model documentation and software user's manual LA-CC-06-012, , 115, 2015.

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

Jordan, R. E., Andreas, E. L. and Makshtas, A. P.: Heat budget of snow-covered sea ice at North Pole 4, J. Geophys. Res., 104(C4), 7785, doi:10.1029/1999JC900011, 1999.

Kantha, L. H. and Clayson, C. A.: Surface exchange processes, in Small Scale Processes in Geophysical Fluid Flows, pp. 417–509., 2000a.

Kantha, L. H. and Clayson, C. A.: Surface Exchange Processes, in Small Scale Processes in

Geophysical Fluid Flows, pp. 417–509., 2000b.

Kauffman, B. G. and Large, W. G.: The CCSM coupler, version 5.0.1. Technical Note, Boulder, CO USA. [online] Available from: http://www.cesm.ucar.edu/models/ccsm2.0.1/cpl5/users_guide.pdf, 2002.

Kraus, E. B. and Businger, J. A.: Atmosphere-Ocean Interaction, 2nd ed., Oxford University Press., 1995.

Lambin, J., Morrow, R., L.-L., F., Willis, J. K., Bonekamp, H., Lillibrige, J., Perbos, J., Zaouche, G., Vaze, P., Bannoura, W., Parisot, F., Thouvenot, E., Coutin-Faye, S., Lindstorm, E. and Mingogno, M.: The OSTM/Jason-2 Mission, Mar. Geod., 33, 4–25, doi:http://dx.doi.org/10.1080/01490419.2010.491030, 2010.

Large, W. G. and Pond, S.: Open ocean momentum flux measurements in moderate to strong winds, J. Phys. Oceanogr., 11, 324–336, doi:10.1175/1520-0485(1981)011<0324:OOMFMI>2.0.CO;2, 1981.

Large, W. G., McWilliams, J. C. and Doney, S. C.: Oceanic Vertical Mixing - a Review and a Model with a Nonlocal Boundary-Layer Parameterization, Rev. Geophys., 32(94), 363–403, doi:10.1029/94rg01872, 1994.

Lenters, J. D.: Long-term Trends in the Seasonal Cycle of Great Lakes Water Levels, J. Great Lakes Res., 27(3), 342–353, doi:10.1016/S0380-1330(01)70650-8, 2001.

Lenters, J. D., Anderton, J. B., Blanken, P. D., Spence, C. and Suyker, A. E.: Assessing the impacts of climate variability and change on Great Lakes evaporation. 2011 Proj. Reports. D. Brown, D. Bidwell, L. Briley, eds. Available from Gt. Lakes Integr. Sci. Assessments Cent., 2013.

Liu, P. C. and Schwab, D. J.: A comparison of methods for estimating u*, from given uz and air-sea temperature differences, J. Geophys. Res., 92(C6), 6488–6494, 1987.

Liu, W. T., Katosaros, K. B. and Businger, J. A.: Bulk Parameterization of Air-Sea Exchanges of Heat and Water Vapor Including the Molecular Constraints at the Interface, J. Atmos. Sci., 37(12), 2798–2800, doi:10.1175/1520-0469(1980)037<2798:COPOAS>2.0.CO;2, 1980.

Massman, W. J.: A simple method for estimating frequency response corrections for eddy covariance systems, Agric. For. Meteorol., 104(3), 185–198, doi:10.1016/S0168-1923(00)00164-7, 2000.

Michalak, A. M., Anderson, E. J., Beletsky, D., Boland, S., Bosch, N. S., Bridgeman, T. B., Chaffin, J. D., Cho, K., Confesor, R., Daloglu, I., DePinto, J. V., Evans, M. A., Fahnenstiel, G. L.,

He, L., Ho, J. C., Jenkins, L., Johengen, T. H., Kuo, K. C., LaPorte, E., Liu, X., McWilliams, M. R., Moore, M. R., Posselt, D. J., Richards, R. P., Scavia, D., Steiner, A. L., Verhamme, E., Wright, D. M. and Zagorski, M. A.: Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions, Proc. Natl. Acad. Sci., 110(16), 6448–6452, doi:10.1073/pnas.1216006110, 2013.

Mitchell, K., Ek, M., Wong, V., Lohmann, D., Koren, V., Schaake, J., Duan, Q., Gayno, G., Moore, B., Grunmann, P., Tarpley, D., Ramsay, B., Chen, F., Kim, J., Pan, H., Lin, Y., Marshall, C., Mahrt, L., Meyers, T. and Ruscher, P.: Noah Land Surface Model (LSM) User's Guide., 2005.

Moukomla, S. and Blanken, P. D.: Remote sensing of the North American Laurentian Great Lakes' surface temperature, Remote Sens., 8(4), 1–15, doi:10.3390/rs8040286, 2016.

Moukomla, S. and Blanken, P. D.: The Estimation of the North American Great Lakes Turbulent Fluxes Using Satellite Remote Sensing and MERRA Reanalysis Data, Remote Sens., 9(141), 14 pages, doi:10.3390/rs9020141, 2017.

Obukhov, A. M.: Turbulence in an atmosphere with a non-uniform temperature, Boundary-Layer Meteorol., 2(1), 7–29, doi:10.1007/BF00718085, 1971.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, W. J., Subin, Z. M., Swenson, S. C., Thornton, P. E., Bozbiyik, A., Fisher, R., Heald, C. L., Kluzek, E., Lamarque, J.-F., Lawrence, P. J., Leung, L. R., Lipscomb, W., Muszala, S., Ricciuto, D. M., Sacks, W., Sun, Y., Tang, J. and Yang, Z.-L.: Technical description of version 4.5 of the Community Land Model (CLM), Boulder, CO USA., 2013.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S. and Schlax, M. G.: Daily high-resolution-blended analyses for sea surface temperature, J. Clim., 20(22), 5473–5496, doi:10.1175/2007JCLI1824.1, 2007.

Rowe, M. D., Anderson, E. J., Wang, J. and Vanderploeg, H. A.: Modeling the effect of invasive quagga mussels on the spring phytoplankton bloom in Lake Michigan, J. Great Lakes Res., 41, 49–65, doi:10.1016/j.jglr.2014.12.018, 2015.

Skamarock, W. C., Klemp, J. B., Dudhi, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W. and Powers, J. G.: A Description of the Advanced Research WRF Version 3, Tech. Rep., (June), 113, doi:10.5065/D6DZ069T, 2008.

Smith, S. D.: Coefficients for Sea Surface Wind Stress, Heat Flux, and Wind Profiles as a Function of Wind Speed and Temperature, J. Geophys. Res., 93(C12), 15467–15472, 1988.

Smith, S. D.: Water vapor flux at the sea surface, Boundary-Layer Meteorol., 47, 277–293, doi:10.1007/BF00122334, 1989.

Spence, C., Blanken, P. D., Hedstrom, N., Fortin, V. and Wilson, H.: Evaporation from Lake Superior: 2. Spatial distribution and variability, J. Great Lakes Res., 37(4), 717–724, doi:10.1016/j.jglr.2011.08.013, 2011.

Spence, C., Blanken, P. D., Lenters, J. D. and Hedstrom, N.: The importance of spring and autumn atmospheric conditions for the evaporation regime of Lake Superior, J. Hydrometeorol., 14(5), 1647–1658, doi:10.1175/JHM-D-12-0170.1, 2013.

Subin, Z. M., Riley, W. J. and Mironov, D.: An improved lake model for climate simulations: Model structure, evaluation, and sensitivity analyses in CESM1, J. Adv. Model. Earth Syst., 4(2), 1–27, doi:10.1029/2011MS000072, 2012.

Van Cleave, K., Lenters, J. D., Wang, J. and Verhamme, E. M.: A regime shift in Lake Superior ice cover, evaporation, and water temperature following the warm El Niño winter of 1997-98, Limnol. Oceanogr., 59(6), 1889–1898, doi:10.4319/lo.2014.59.6.1889, 2014.

Wang, J., Hu, H., Schwab, D., Leshkevich, G., Beletsky, D., Hawley, N. and Clites, A.: Development of the Great Lakes Ice-circulation Model (GLIM): Application to Lake Erie in 2003-2004, J. Great Lakes Res., 36(3), 425–436, doi:10.1016/j.jglr.2010.04.002, 2010.

Webb, E.K., Pearman, G.I, and Leuning, R.: Correction of flux measurements for density effects due to heat and water vapour transfer, Q. J. R. Meteorol. Soc., 106, 85–100, doi:10.1002/qj.49710644707, 1980.

Wright, D. M., Posselt, D. J. and Steiner, A. L.: Sensitivity of lake-effect snowfall to lake ice vover and temperature in the Great Lakes region, Mon. Weather Rev., 141(2), 670–689, doi:10.1175/MWR-D-12-00038.1, 2013.

Xiao, C., Lofgren, B. M., Wang, J. and Chu, P. Y.: Improving the lake scheme within a coupled WRF-lake model in the Laurentian Great Lakes, J. Adv. Model. Earth Syst., 8(4), 1969–1985, doi:10.1002/2016MS000717, 2016.

Xue, P., Schwab, D. J. and Hu, S.: An investigation of the thermal response to meteorological forcing in a hydrodynamic model of Lake Superior, J. Geophys. Res. Ocean., 120, 5233–5253, doi:10.1002/jgrc.20224, 2015.

Xue, P., Pal, J. S., Ye, X., Lenters, J. D., Huang, C., and Chu, P. Y.: Improving the simulation of large lakes in regional climate modeling: Two-way lake-atmosphere coupling with a 3-D

hydrodynamic model of the Great Lakes, J. Clim., 30, 1605-1627, doi:10.1175/JCLI-D-16-0225.1, 2017.

Zeng, X., Zhao, M. and Dickinson, R. E.: Intercomparison of bulk aerodynamic algorithms for the computation of sea surface fluxes using TOGA COARE and TAO data, J. Clim., 11(10), 2628–2644, doi:10.1175/1520-0442(1998)011<2628:IOBAAF>2.0.CO;2, 1998.

Zeng, X., Shajkh, M., Dai, Y., Dickinson, R. E. and Myneni, R.: Coupling of the Common Land Model to the NCAR Community Climate Model, J. Clim., 15(14), 1832–1854, doi:10.1175/1520-0442(2002)015<1832:COTCLM>2.0.CO;2, 2002.

**Table 1 – Summary of flux algorithm specifications.**

| Algorithm name | Parent Model | Stability | | Parameterization of roughness length scales for | | Gustiness | References |
|---|---|---|---|---|---|---|---|
| | | | | momentum $z_0$ | temperature and humidity $z_{0\theta,q}$ | | |
| | | Unstable | Stable | | | No | |
| LS87 | FVCOM | Similar to Businger et al. (1971) | Holtslag et al. (1990) | $z_0 = \alpha \frac{u^{*2}}{g} + 0.11 \frac{\nu}{u^*}$ $\alpha$=0.011 | $z_0 = z_{\theta,q}$ | No | Liu and Schwab (1987) |
| C89 | LLTM | Businger et al. (1971) | Holtslag et al. (1990) | $z_0 = \alpha \frac{u^{*2}}{g}$ $\alpha$=0.0101 | $z_0 = z_{\theta,q}$ | No | Croley, 1989a,b) |
| Z98L | WRF-Lake | Businger et al. (1971) | Holtslag et al. (1990) | $z_0$ =0.001 m (Smith, 1988 for ocean) | $z_0 = z_{\theta,q}$ (Brutsaert, 1975 for ocean) | Fairall et al. (1996), $\beta$=1.0 | Zeng (1998) |
| J99 | FVCOM, UG-CICE | Businger et al. (1971) | Beljaars and Holtslag, (1991) | $z_0 = z\exp\left[-\kappa\left(\frac{2.7\times10^{-3}}{U} + 1.42\times10^{-4} + 7.64\times10^{-5}U\right)^{-1}\right]$ (Large et al., 1994) | $z_0 = z_{\theta,q}$ (Jordan (1999) used Andreas, (1987) for ice surface. | No | Jordan (1999), Hunke et al. (2015) |
| COARE | FVCOM | Businger et al. (1971), Convective behavior: Fairall et al.,(1996) | Beljaars and Holtslag, (1991) | $z_0 = \alpha \frac{u^{*2}}{g} + 0.11 \frac{\nu}{u^*}$ $\alpha$: function of wind speed | $z_{\theta,q} = \min(1.6\times10^{-4}, 5.8\times10^{-5}Rr^{-0.72})$ | Fairall et al. (1996), $\beta$=1.2 | Fairall et al., (1996ab), Edson et al. (2013) |

**Table 2** Statistics of simulated latent heat flux $\lambda E$ for 2012-2014. For J99, LS87, Z98L, and C89, RMSEs with the updated $z_{0\theta,q}$ formulation are shown. Numbers in parentheses denote RMSEs with the original $z_{0T,q}$ formulation. An error reduction ratio (%) is calculated for mean RMSEs of J99, LS87, Z98L, and C89. A mean flux (W $m^{-2}$) and mean normalized RMSE are calculated for all the five algorithms.

| | RMSE [$Wm^{-2}$] | | | | | Error reduction ratio [%] | Mean flux [W $m^{-2}$] | Mean Normalized RMSE | Mean bias [%] |
|---|---|---|---|---|---|---|---|---|---|
| | COARE | J99 | LS87 | Z98L | C89 | | | | |
| Stannard Rock | 27.2 | 33.1 (32.6) | 29.3 (39.9) | 29.1 (79.9) | 29.1 (38.9) | 37.0 | 56.9 | 0.53 (0.84) | 29.6 (38.3) |
| White Shoal | 25.2 | 36. (25.3) | 28.3 (25.4) | 27.8 (68.0) | 27.6 (25.8) | 17.0 | 61.1 | 0.49 (0.59) | 1.4 (24.0) |
| Spectacle Reef | 70.4 | 83.8 (66.8) | 68.5 (61.9) | 67.4 (72.6) | 71.3 (62.5) | -10.3 | 116.1 | 0.63 (0.57) | -27.8 (-3.2) |
| Long Point | 42.9 | 40.1 (42.1) | 47.9 (46.5) | 49.1 (104.3) | 45.8(47.8) | 24.1 | 50.7 | 0.90 (1.19) | 27.4 (49.6) |
| Mean RMSE [$Wm^{-2}$] | 41.4 | 48.3 (41.7) | 43.5 (43.5) | 43.3 (81.2) | 43.5 (43.8) | 15.0 | 81.5 | 0.52 (0.64) | - |
| Mean bias [%] | 4.3 | -9.1 (8.7) | 15.7 (18.9) | 16.5 (88.5) | 10.9 (19.6) | - | - | - | 7.7 (27.1) |

**Table 3** Same as Table 2, but for sensible heat flux $H$.

| | RMSE [$Wm^{-2}$] | | | | | Error reduction ratio [%] | Mean flux [W $m^{-2}$] | Normalized RMSE | Mean bias [%] |
|---|---|---|---|---|---|---|---|---|---|
| | COARE | J99 | LS87 | Z98L | C89 | | | | |
| Stannard Rock | 27.4 | 29.1 (50.2) | 26.3 (84.7) | 26.5 (77.4) | 23.6 (32.1) | 56.8 | 39.1 | 0.68 (1.56) | 26.6 (48.9) |
| White Shoal | 32.3 | 31.4 (37.9) | 31.8 (50.8) | 31.9 (52.8) | 31.0 (32.9) | 27.7 | 40.7 | 0.78 (1.07) | -24.9 (7.8) |
| Spectacle Reef | 11.4 | 13.2 (27.2) | 13.9 (60.4) | 11.9 (65.3) | 13.3 (13.8) | 68.6 | 46.1 | 0.28 (0.90) | 6.3 (44.8) |
| Long Point | 27.2 | 26.7 (45.5) | 28.5 (65.6) | 27.6 (63.2) | 21.5 (32.9) | 49.7 | 11.7 | 2.2 (4.4) | 18.5 (31.4) |
| Mean RMSE | 24.6 | 25.1 (40.2) | 25.1 (65.4) | 24.5 (64.7) | 22.4 (28.0) | 51.0 | 38.0 | 0.64 (1.30) | - |
| Mean bias [%] | 4.2 | 7.5 (31.4) | 13.6 (61.7) | 6.3 (62.0) | 1.5 (11.0) | - | - | - | 6.6 (33.2) |

**Captions of Figures**

**Figure 1. Map of the Laurentian Great Lakes including the locations of offshore lighthouse-based monitoring stations used in this study. Adapted from Lenters et al. (2013). Instrument heights above the mean water level are 32.5 m at Stannard Rock, 29.5 m at Long Point, 30.0 m at Spectacle Reef, and 42.8 m at White Shoal.**

**Figure 2. Schematic diagram showing the relationship between the parent model systems (FVCOM, WRF-Lake, and LLTM) and the flux algorithms used in the parent model systems. Detail description of each flux algorithm is listed in Table 1.**

**Figure 3. 10-day running mean time series of meteorological variables at the four stations. Air temperature and relative humidity were measured with Vaisala HMP45C thermohygrometers and wind speed were measured with the CSAT-3 (See section 2.1.1 or Figure 1 for the sensor heights). Water surface temperature is taken from GLSEA. Data at pixels closest to the stations are used. The data gaps in water surface temperature from January to April denote periods during which the site was affected by lake ice cover. Measurements at Long Point and White Shoal started in May and June of 2012. There is also a long data gap between February 2012 and June 2013 at Spectacle Reef.**

**Figure 4. 10-day running mean time series of latent ($\lambda E$) and sensible ($H$) heat fluxes at Stannard Rock. Black lines denote observed $\lambda E$ and $H$ and the same for (a), (b) and (c), (d), respectively. The $\lambda E$ and $H$ simulations employ the original $z_{0\theta,q}$ formula in (a), (c) and with the updated $z_{0\theta,q}$ formula in (b) and (d). The COARE simulation results are unchanged from (a) to (b) or from (c) to (d).**

**Figure 5. The same as Figure 4, but at White Shoal.**

**Figure 7. The same as Figure 4, but at Long Point.**

**Figure 6. The same as Figure 4, but at Spectacle Reef.**

**Figure 8. Scatter plots of latent heat flux ($\lambda E$) comparing the observed ($x$-axis) and the simulated ($y$-axis) daily mean fluxes. Each row shows comparisons with a specific**

**algorithm at the four stations, while each column shows comparisons with the five algorithms at a specific station. Grey and blue dots indicate the results with the original and updated $z_{0\theta,q}$ formulae, respectively.**

**Figure 9. The same as Figure 8, but for sensible heat flux ($H$).**

**Figure 10. Errors in daily mean latent heat flux ($y$-axis) versus specific humidity difference between the water surface and air at the sensor height $q_w$-$q_a$ [kg kg$^{-1}$], transfer coefficient $C_E$ [-], wind speed $U$ [m s$^{-1}$], and stability factor $z/L$ ($x$-axis) for the five algorithms at Stannard Rock. Grey and blue dots indicate the results using the original and updated $z_{0\theta,q}$ formulae, respectively.**

**Figure 11. Errors in daily mean sensible heat flux ($y$-axis) versus potential temperature difference between the water surface and air at the sensor height $\theta_w$-$\theta_a$ [$^o$C], transfer coefficient $C_H$ [-], wind speed $U$ [m s$^{-1}$], and stability factor $z/L$ ($x$-axis) for the five algorithms at Stannard Rock. Grey and blue dots indicate the results with the original and updated $z_{0\theta,q}$ formulae, respectively.**

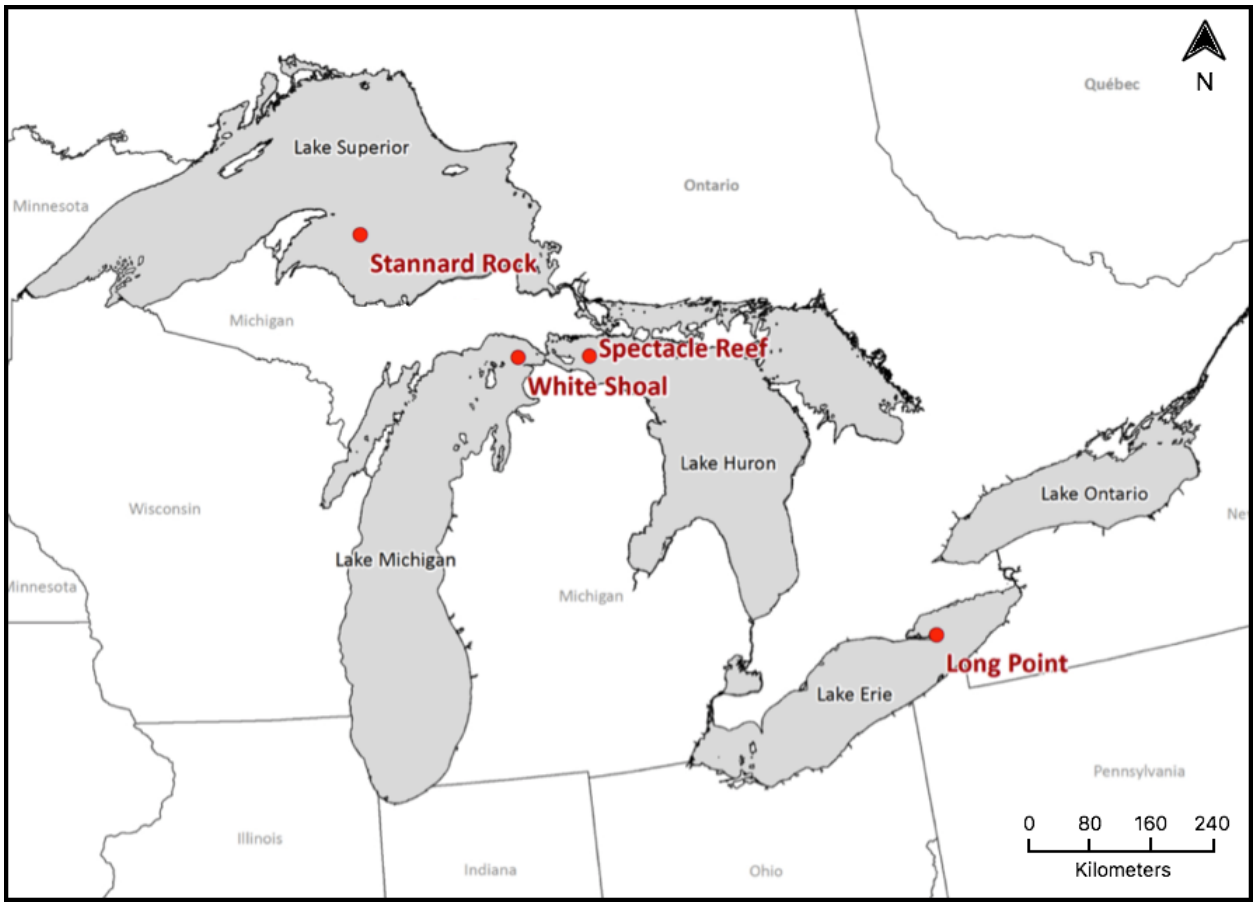


**Figure 1. Map of the Laurentian Great Lakes including the locations of offshore lighthouse-based monitoring stations used in this study. Adapted from Lenters et al. (2013). Instrument heights above the mean water level are 32.5 m at Stannard Rock, 29.5 m at Long Point, 30.0 m at Spectacle Reef, and 42.8 m at White Shoal.**

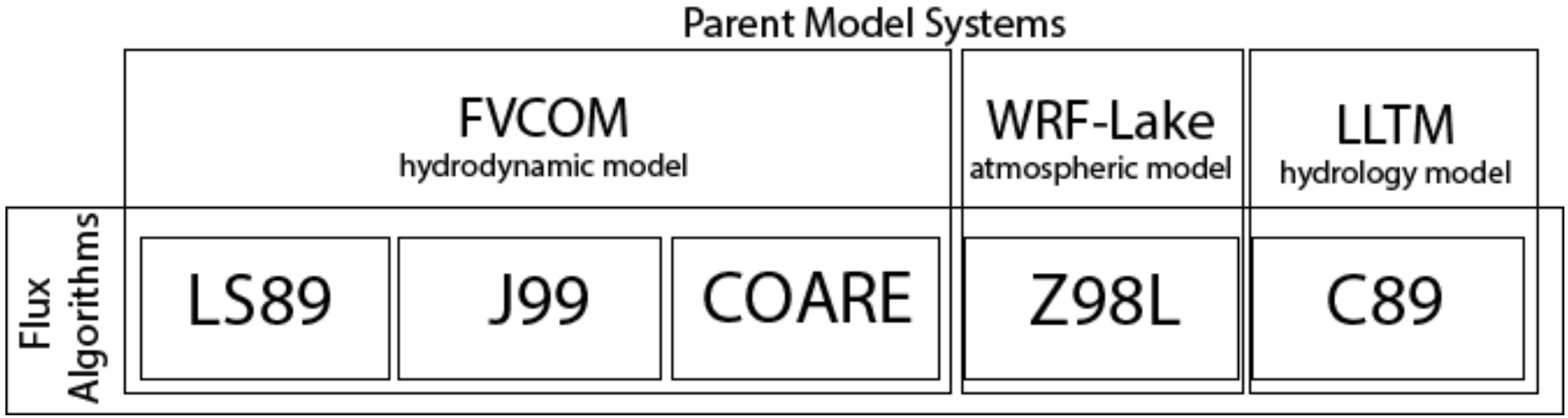


**Figure 2. Schematic diagram showing the relationship between the parent model systems (FVCOM, WRF-Lake, and LLTM) and the flux algorithms used in the parent model systems. Detail description of each flux algorithm is listed in Table 1.**

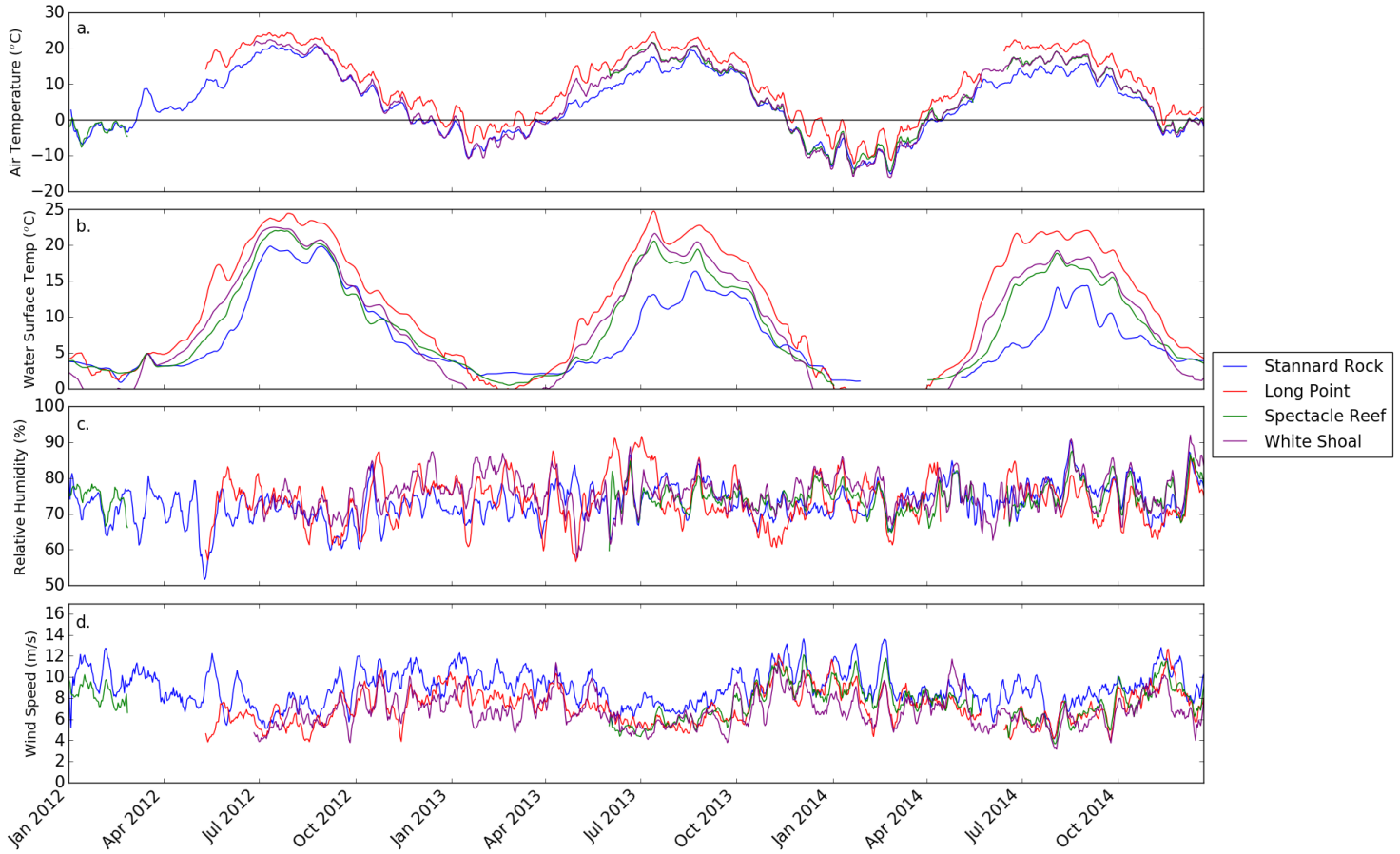


**Figure 3. 10-day running mean time series of meteorological variables at the four stations. Air temperature and relative humidity were measured with Vaisala HMP45C thermohygrometers and wind speed were measured with the CSAT-3 (See section 2.1.1 or Figure 1 for the sensor heights). Water surface temperature is taken from GLSEA. Data at pixels closest to the stations are used. The data gaps in water surface temperature from January to April denote periods during which the site was affected by lake ice cover. Measurements at Long Point and White Shoal started in May and June of 2012. There is also a long data gap between February 2012 and June 2013 at Spectacle Reef.**

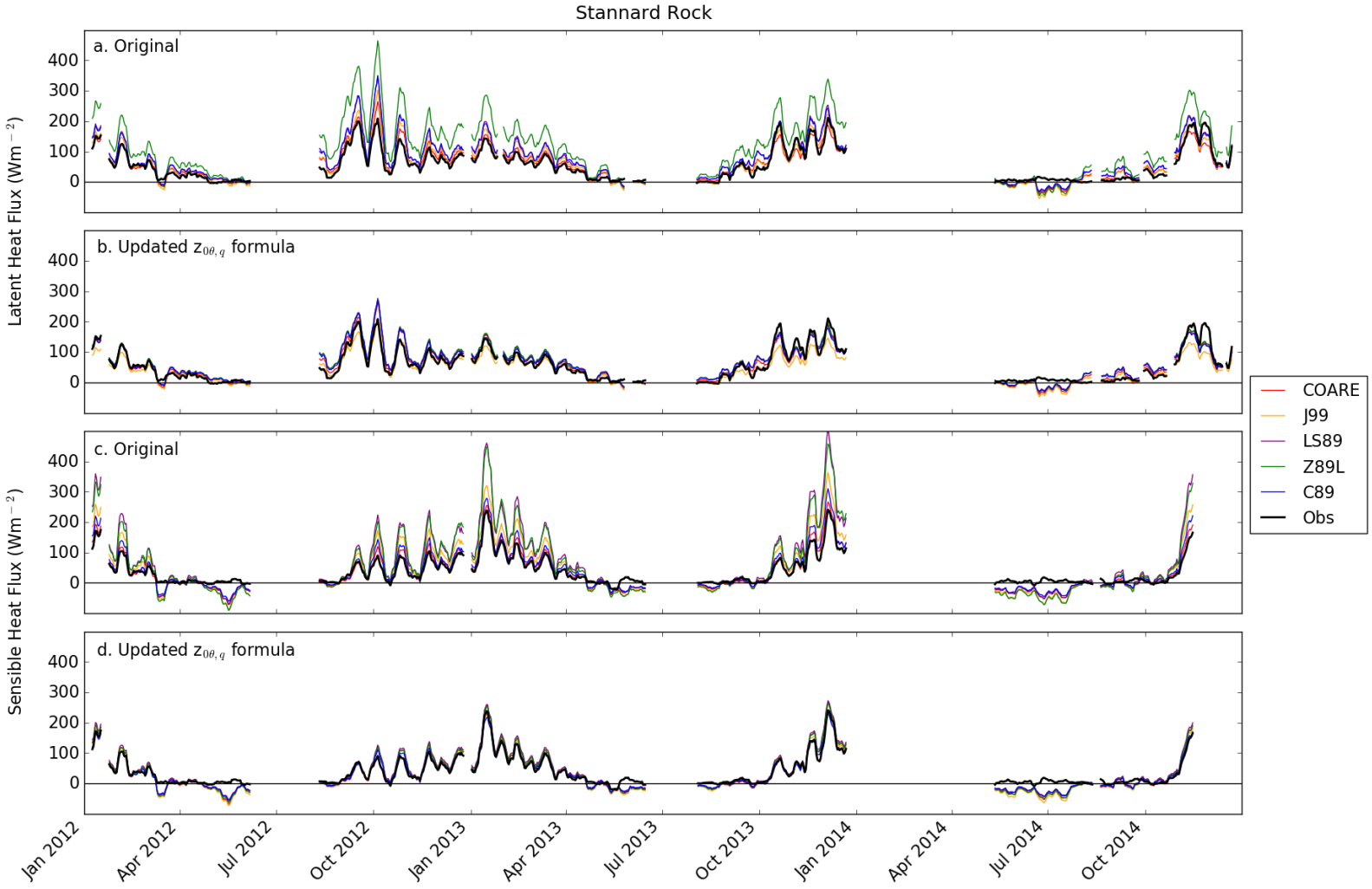


**Figure 4. 10-day running mean time series of latent ($\lambda E$) and sensible ($H$) heat fluxes at Stannard Rock. Black lines denote observed $\lambda E$ and $H$ and the same for (a), (b) and (c), (d), respectively. The $\lambda E$ and $H$ simulations employ the original $z_{0\theta,q}$ formula in (a), (c) and with the updated $z_{0\theta,q}$ formula in (b) and (d). The COARE simulation results are unchanged from (a) to (b) or from (c) to (d).**

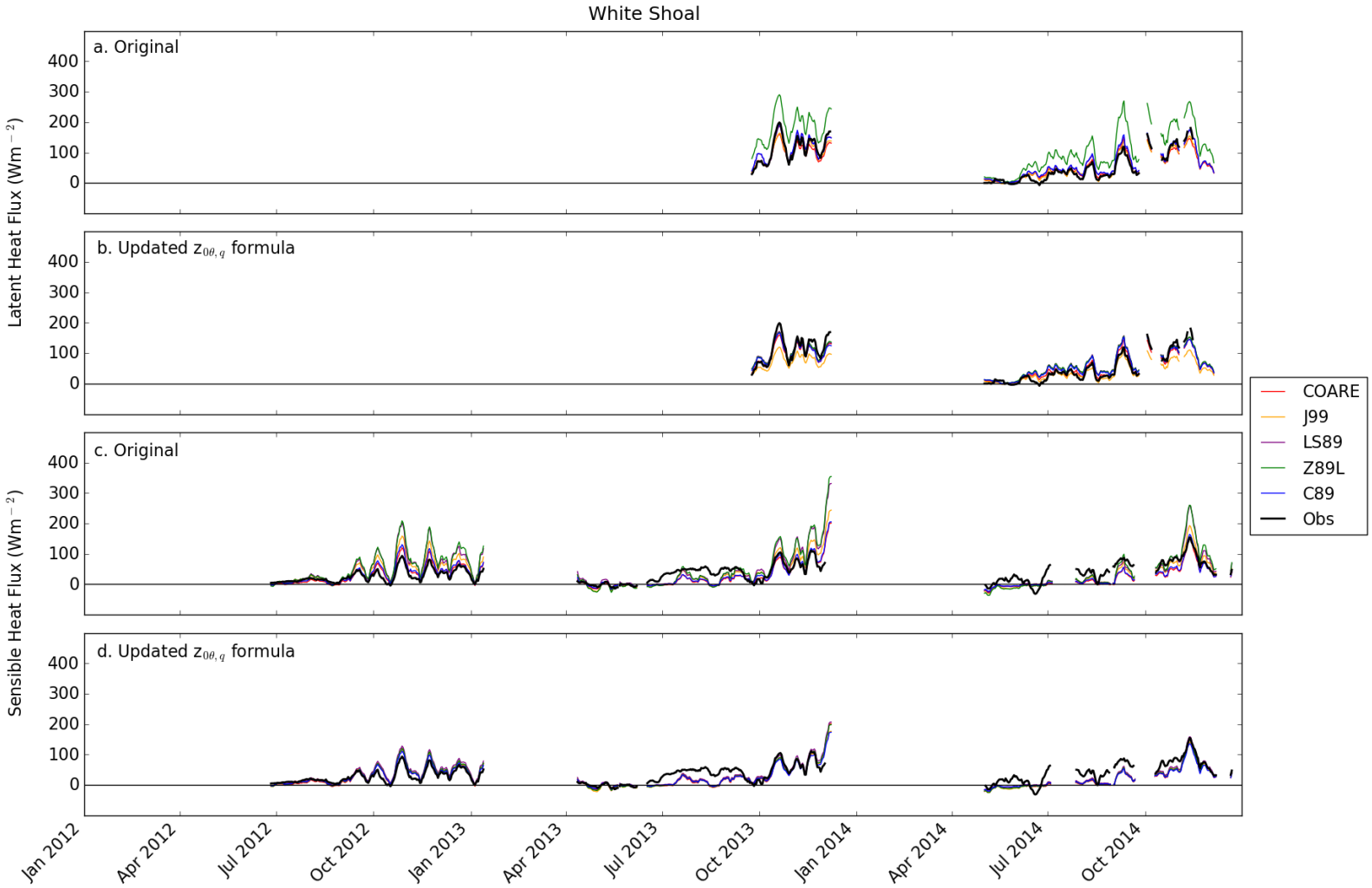


**Figure 5. The same as Figure 4, but at White Shoal.**

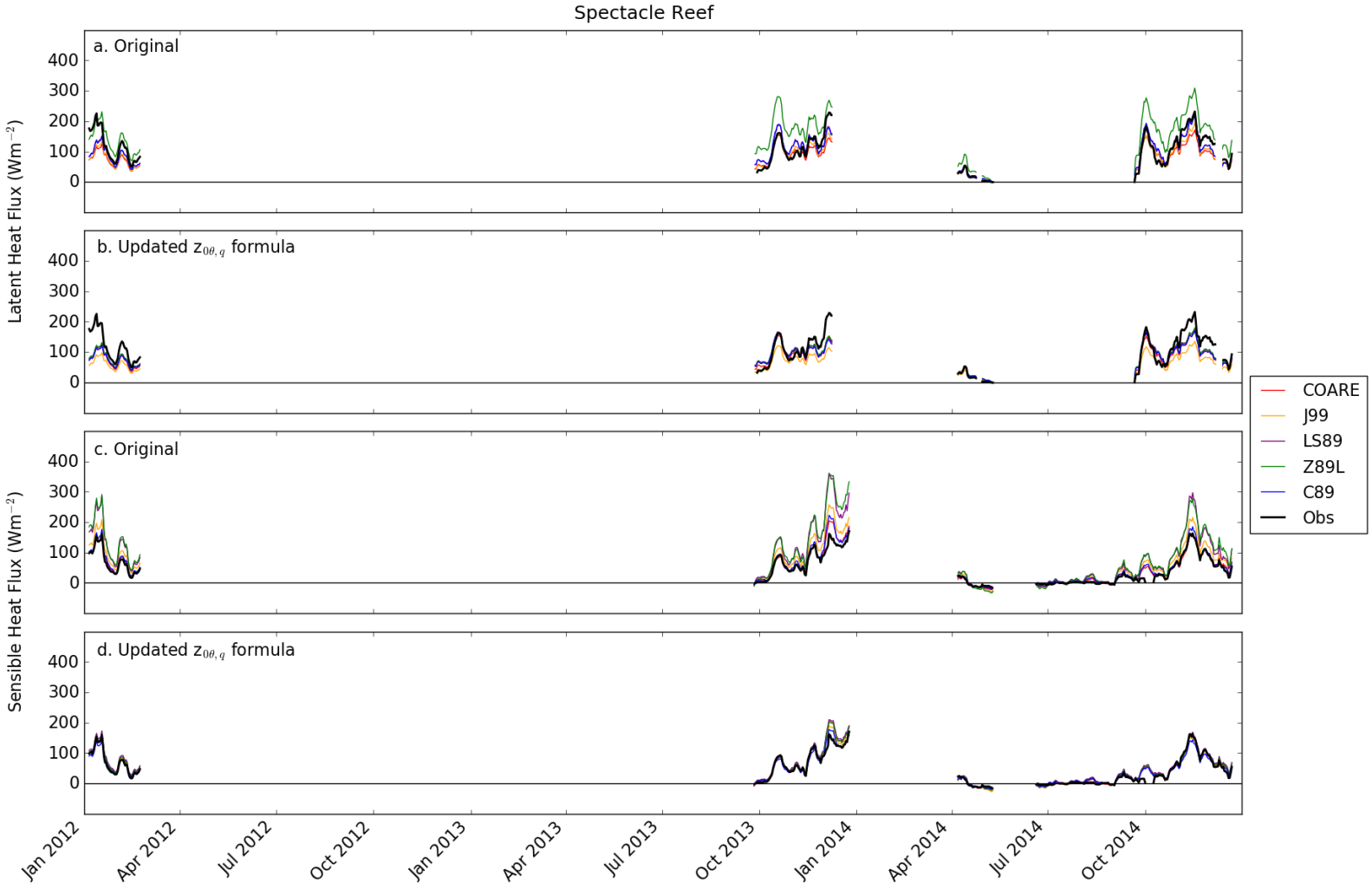


**Figure 6. The same as Figure 4, but at Spectacle Reef.**

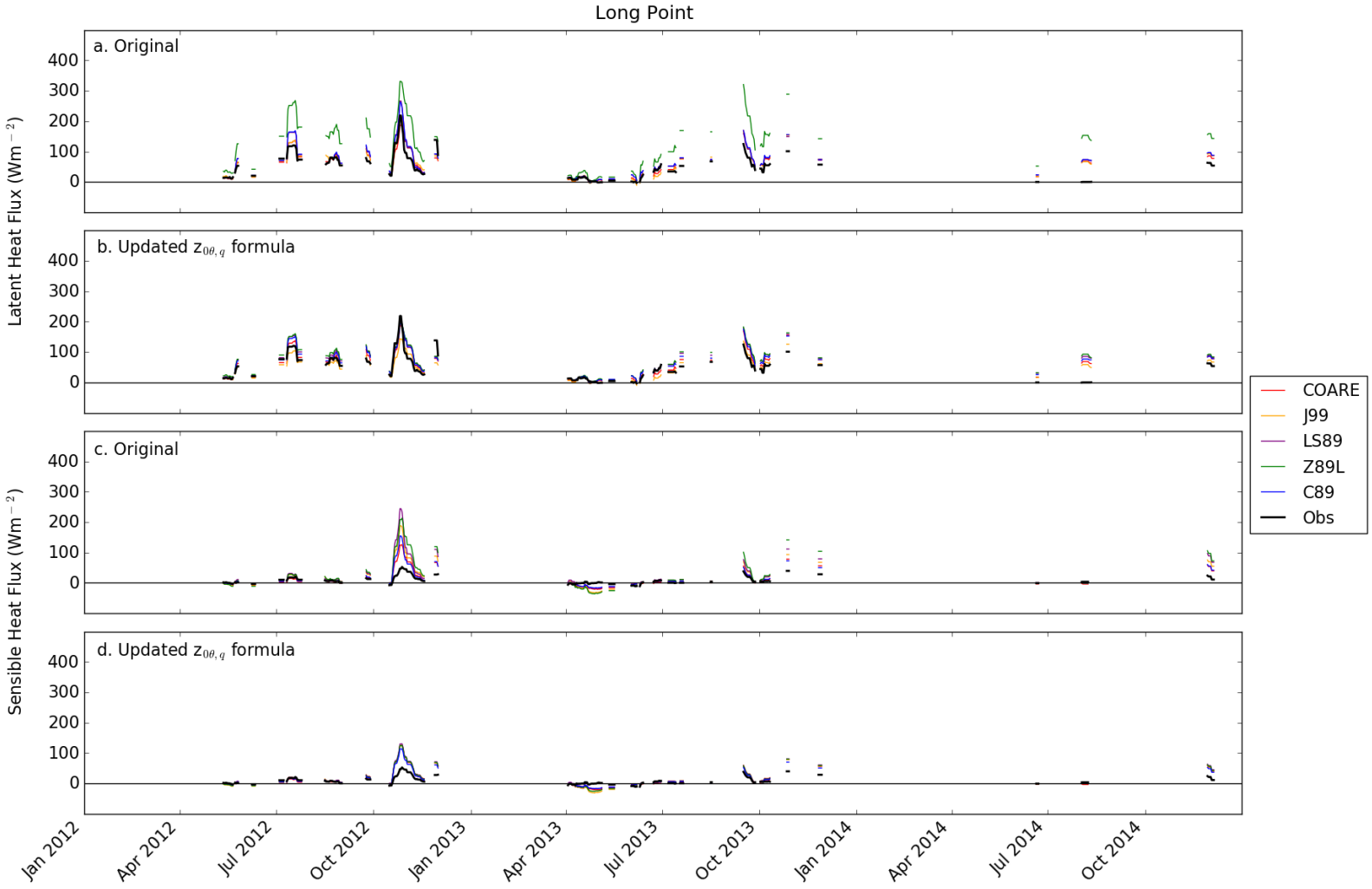


**Figure 7. The same as Figure 4, but at Long Point.**

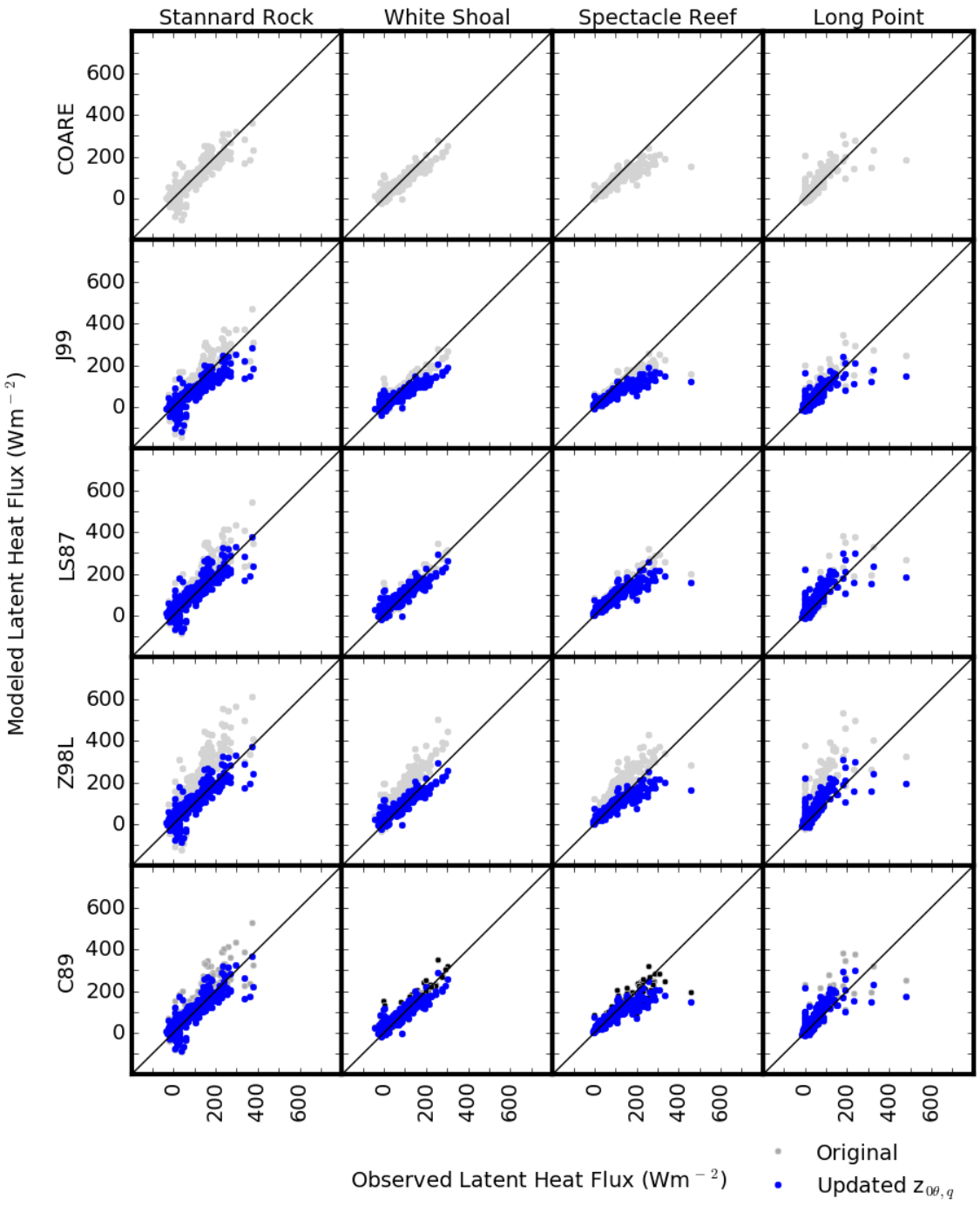


**Figure 8. Scatter plots of latent heat flux ($\lambda E$) comparing the observed ($x$-axis) and the simulated ($y$-axis) daily mean fluxes. Each row shows comparisons with a specific algorithm at the four stations, while each column shows comparisons with the five algorithms at a specific station. Grey and blue dots indicate the results with the original and updated $z_{0\theta,q}$ formulae, respectively.**

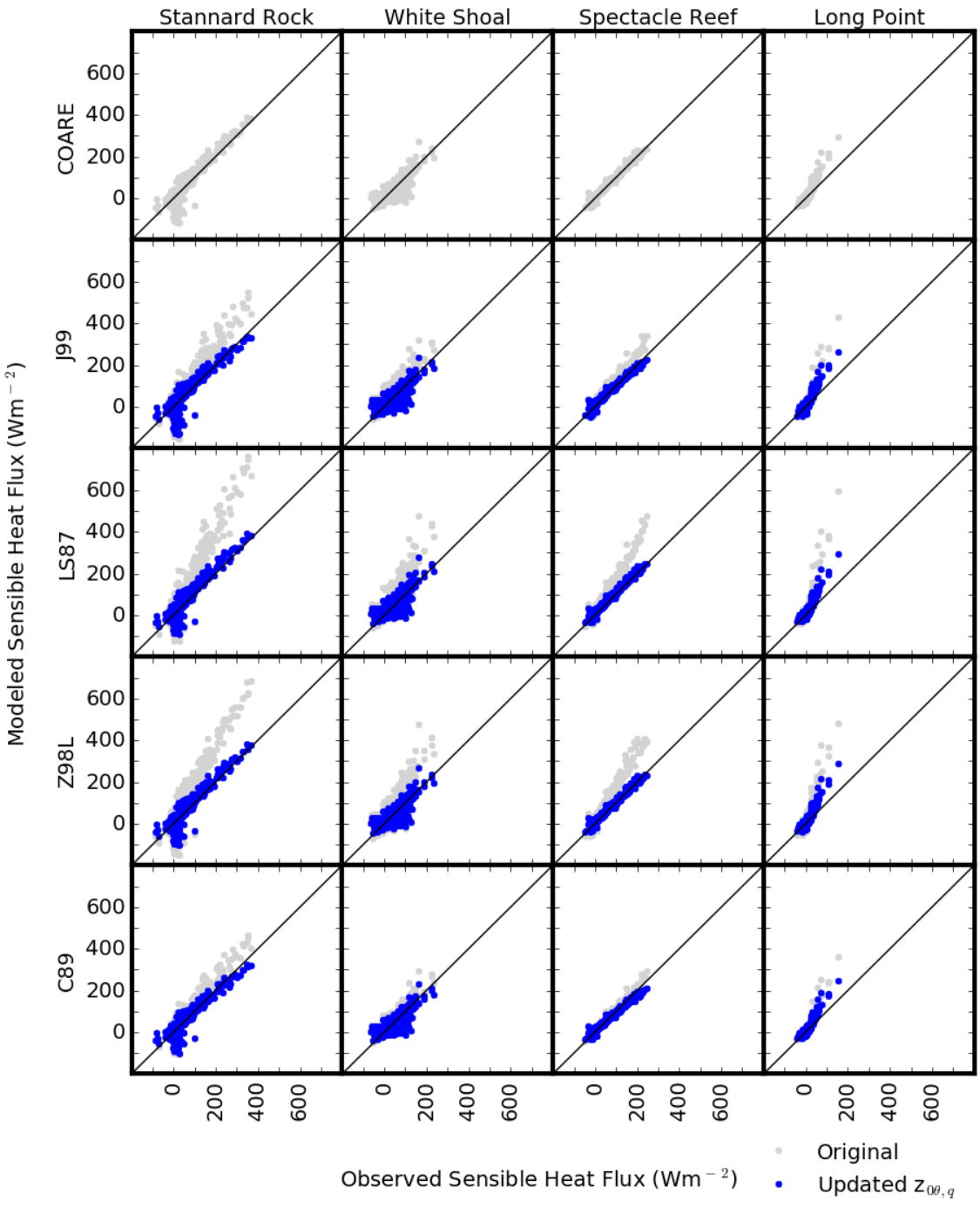


**Figure 9. The same as Figure 8, but for sensible heat flux (*H*).**

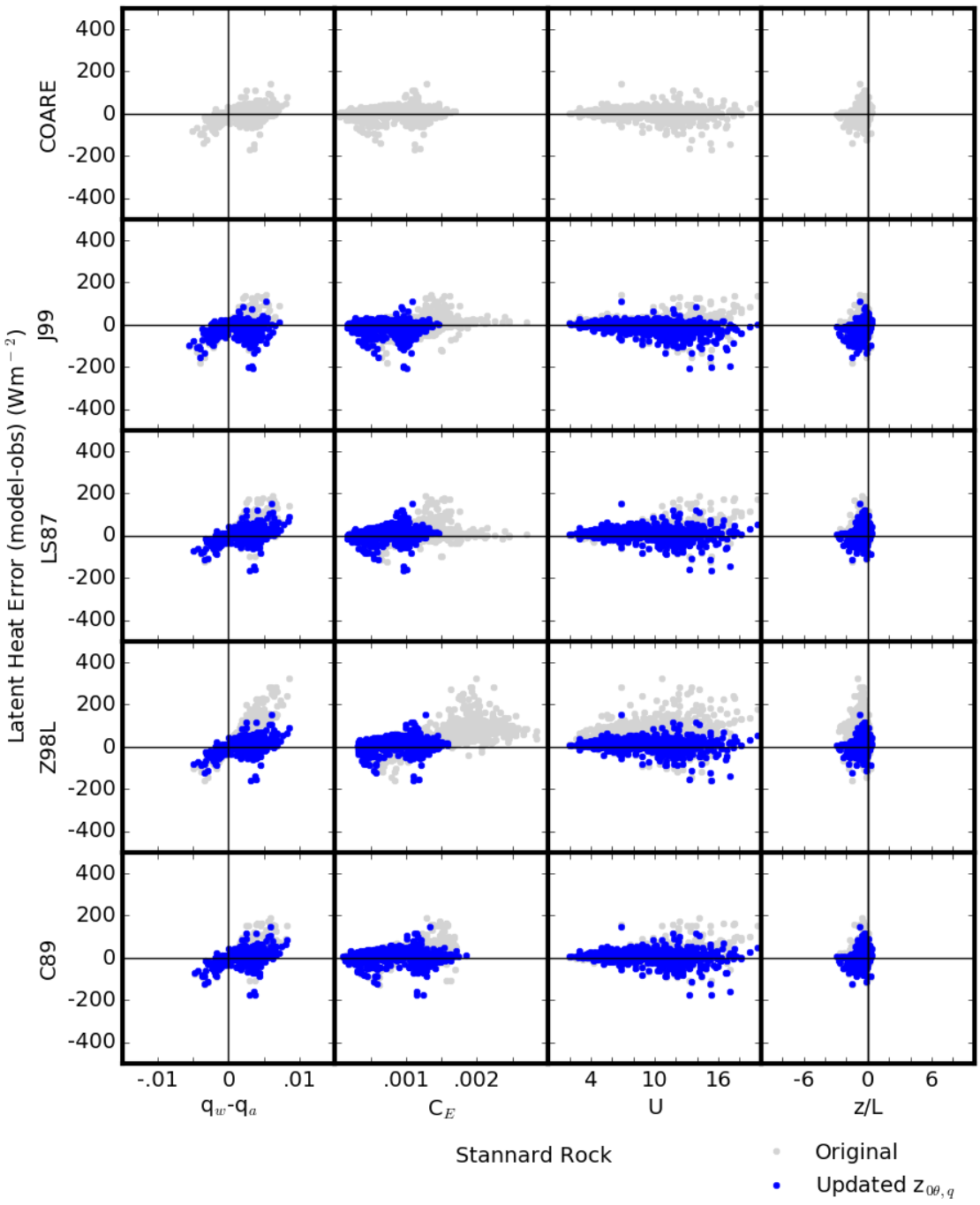


**Figure 10. Errors in daily mean latent heat flux (*y*-axis) versus specific humidity difference between the water surface and air at the sensor height $q_w$-$q_a$ [kg kg$^{-1}$], transfer coefficient $C_E$ [-], wind speed $U$ [m s$^{-1}$], and stability factor $z/L$ (*x*-axis) for the five algorithms at Stannard Rock. Grey and blue dots indicate the results using the original and updated $z_{0\theta,q}$ formulae, respectively.**

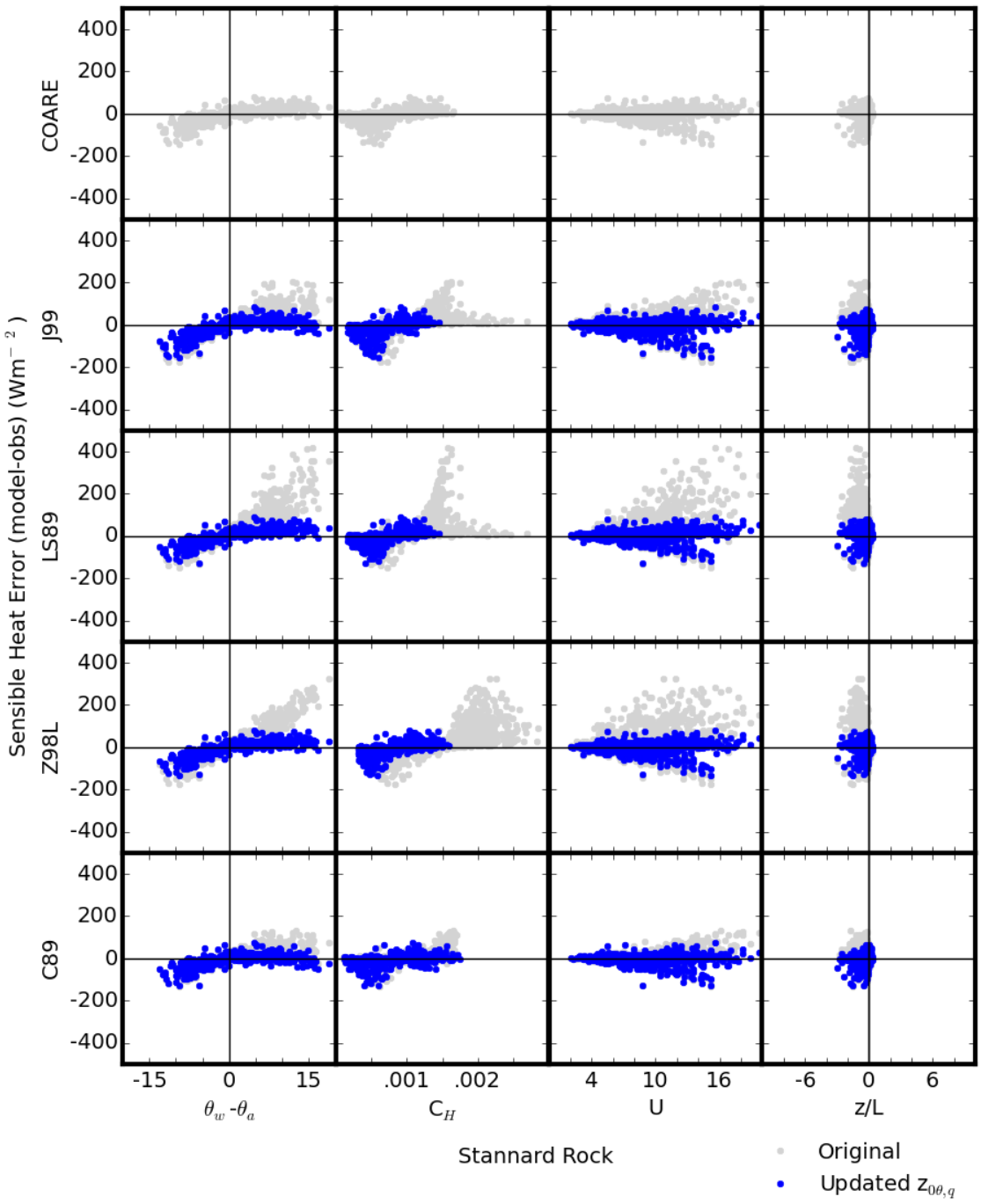


**Figure 11. Errors in daily mean sensible heat flux (*y*-axis) versus potential temperature difference between the water surface and air at the sensor height $\theta_w$-$\theta_a$ [$^o$C], transfer coefficient $C_H$ [-], wind speed $U$ [m s$^{-1}$], and stability factor $z/L$ (*x*-axis) for the five algorithms at Stannard Rock. Grey and blue dots indicate the results with the original and updated $z_{0\theta,q}$ formulae, respectively.**