An effective formulation for estimating wetland surface energy fluxes from weather data

Yi Wang¹ and Richard Michael Petrone¹

¹University of Waterloo

February 16, 2024

Abstract

In modelling evapotranspiration, the need for land surface variables including ground heat fluxes (G), surface temperature (Ts), surface relative humidity (RHs) and surface resistance often present a challenge due to land heterogeneity and limited measurements. This study introduces a simple formulation rooted in the shared physical basis of the maximum entropy model (MaxEnt), the Relative Humidity at Equilibrium (ETRHEQ) method, and the Surface Flux Equilibrium (SFE) method, and it estimates sensible (H) and latent fluxes (LE) in wetlands without requiring land surface variables or site-specific calibration, except for an assumed vegetation height. Further, it effectively estimates LE from half-hourly to monthly scales in FLUXNET and AmeriFlux wetland sites. While its performance in estimating H is less satisfactory due to loosely constrained boundary conditions, it shows promising potential for simultaneously and precisely estimating LE, H, G, Ts, and RHs from weather data in various ecosystems.

Hosted file

Manuscript 5_Abby_01.2024-RP.docx available at https://authorea.com/users/710669/articles/717372-an-effective-formulation-for-estimating-wetland-surface-energy-fluxes-from-weather-data

Hosted file

Manuscript 5_SI_Tables.docx available at https://authorea.com/users/710669/articles/717372-an-effective-formulation-for-estimating-wetland-surface-energy-fluxes-from-weather-data
An effective formulation for estimating wetland surface energy fluxes from weather data

Yi Wang¹, Richard. M. Petrone¹

¹Hydrometeorology Research Group, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1

Abstract: In modelling evapotranspiration, the need for land surface variables including ground heat fluxes (G), surface temperature ($T_s$), surface relative humidity ($RH_s$) and surface resistance often present a challenge due to land heterogeneity and limited measurements. This study introduces a simple formulation rooted in the shared physical basis of the maximum entropy model (MaxEnt), the Relative Humidity at Equilibrium (ETRHEQ) method, and the Surface Flux Equilibrium (SFE) method, and it estimates sensible (H) and latent fluxes (LE) in wetlands without requiring land surface variables or site-specific calibration, except for an assumed vegetation height. Further, it effectively estimates LE from half-hourly to monthly scales in FLUXNET and AmeriFlux wetland sites. While its performance in estimating H is less satisfactory due to loosely constrained boundary conditions, it shows promising potential for simultaneously and precisely estimating LE, H, G, $T_s$, and $RH_s$ from weather data in various ecosystems.

Key points:

1. The formulation is based on the principle of maximum Shannon information entropy production for turbulence fluxes.
2. The formulation does not require land surface variables or site-specific calibration; only an assumed vegetation height is needed.

3. The formulation effectively estimates LE from half-hourly to monthly scales.

**Plain language summary**: This study introduces a new method to predict how much water and heat wetlands transport to the atmosphere, a process that is usually complicated because it involves a lot of detailed information about land properties that are hard to measure. This new method does not need all those details, and instead just needs an estimate of how tall the plant canopy is. This method works extremely well for predicting water release into the air over periods ranging from half-hourly to monthly in FLUXNET and AmeriFlux wetland sites. Although this method is not perfect at predicting heat release due to some assumptions that have to be made about ground heat and surface temperature, it shows a lot of promise. With a bit of fine-tuning, it could be used to accurately measure both water and heat exchanges in various types of ecosystems, not just wetlands.

1. Introduction

The partitioning of energy on the land surface of terrestrial ecosystems into ground heat (G), sensible heat (H) and latent heat (LE) has long been recognized as a result of complex interactions between atmospheric and land surface properties (Duveiller et al., 2018; Forzieri et al., 2020; Williams and Torn, 2015; Wilson et al., 2002). At short temporal scales, it is impacted by plant physiological activities and boundary layer properties, and over the long term, the biogeochemical cycling, disturbance, and climate all have significant roles to play (Arneth et al., 2012; Green et al., 2017; Wilson et al., 2002). While the importance of land surface properties
cannot be overlooked, land surface variables are a challenge to parameterize due to land heterogeneity and varied physiological responses of vegetation to changing environmental conditions (Dickinson et al., 1991; Mueller and Seneviratne, 2014; Wang and Dickinson, 2012).

Recent studies proposed two methodologies, namely the Relative Humidity at Equilibrium (ETRHEQ) method and the Surface Flux Equilibrium (SFE) method, to estimate surface energy fluxes from near-surface atmospheric conditions (McColl et al., 2019; Salvucci and Gentine, 2013). ETRHEQ determines the optimal daily surface conductance that yields the most accurate ET predictions based on minimum vertical variance of relative humidity (RH) (Salvucci and Gentine, 2013), and SFE provides the solution of ETRHEQ at the steady state (McColl et al., 2019). The two methods are justified by strong land-atmospheric coupling wherein land surface properties are embedded in the near-surface atmospheric conditions (McColl and Rigden, 2020; McColl et al., 2019). Conversely, the conditions of the near-surface atmosphere are also reflected in land surface variables, which partly justifies another methodology called the maximum entropy model (MaxEnt) that estimates surface energy fluxes using only the surface temperature and surface relative humidity in addition to net radiation (Wang and Bras, 2011; Wang and Bras, 2009). Although the three models have shown success over a variety of ecosystems worldwide (Chen et al., 2021; McColl and Rigden, 2020; Rigden and Salvucci, 2015; Yang et al., 2022), each have their own limitations. ETRHEQ requires vegetation height and ground heat fluxes in addition to 24-hour subdaily weather measurements, to estimate latent and sensible fluxes at the daily scale (Rigden and Salvucci, 2015; Salvucci and Gentine, 2013). SFE, though it requires less parameters (i.e., only net radiation, ground heat flux, air temperature and air specific humidity), works for sites near or at the steady state and estimates energy fluxes at the daily or larger temporal scales (Chen et al., 2021; Kim et al., 2023; McColl and Rigden, 2020). The
MaxEnt model is formulated based on minimizing the dissipation function of turbulent fluxes (which is equivalent to maximizing Shannon information entropy production of the turbulent fluxes (Dewar, 2005)) and the Monin-Obukhov similarity theory (MOST)'s extremum solution (Wang and Bras, 2009), but the justification of extremum solution still requires further examination (Wang and Bras, 2010; Wang et al., 2023).

Wang et al. (2023) investigated the linkage of the three models and found that minimizing the dissipation function of energy fluxes in MaxEnt is equivalent to minimizing the vertical variance of RH in ETRHEQ. The empirical success of the three models is explained by the fact that far-from-equilibrium ecosystems progress toward a steady state (i.e., the SFE state) by minimizing dissipation, and this tendency is manifested through the vertical variance of RH (Wang et al., 2023). In addition, Wang et al. (2023) demonstrated that the connection among the three models is independent of Monin-Obukhov similarity theory (MOST)'s extremum solution (Wang et al., 2023), and proposed a more general formulation describing the dissipation function (D) of energy fluxes for both non-steady and steady states, as:

$$D = \frac{2G^2}{I_s} + \frac{2H^2}{I_a} + \frac{LE^2}{I_e}$$

(1)

where $I_s$, $I_a$ and $I_e$ are the thermal inertia parameters for G, H and LE, respectively; The parameterization of $I_s$ is provided in Huang et al. (2017) and Yang et al. (2022) in which $I_d$ is the thermal inertia of dry soil; $\theta$ is the volumetric soil moisture; $I_w$ is the thermal inertia of still liquid water; $\rho$ is the density of air; $c_p$ is the specific heat capacity of air; $g_a$ is the aerodynamic conductance; $\delta$ is the slope of the relation between saturated specific humidity and temperature,
\[ \gamma = \frac{c_p}{\lambda} \] with \( \lambda \) being the latent heat of vaporization of water; and \( \text{RH}_s \) is the surface relative humidity. The detailed formulation will be introduced in the next section.

The new formulation is denoted as MaxEnt-ETRHEQ, indicating the shared physical basis underlying MaxEnt and ETRHEQ. It appears to require both atmospheric and land surface variables at first glance. However, closer scrutiny revealed that land surface variables such as surface temperature, surface relative humidity and soil moisture are interlinked in the calculation of \( G, H \) and \( \text{LE} \) under energy closure. This interconnection renders the formulation self-constrained. Consequently, the energy fluxes and the land surface variables can be analytically determined by identifying the minimum value of \( D \) given suitable ranges of surface temperature and relative humidity. Therefore, MaxEnt-ETRHEQ has potential to estimate surface energy fluxes for various ecosystems, with minimal or no land surface information. But its effectiveness is yet to be examined. Leveraging our proficiency and background in wetland ecosystems, we demonstrate in this paper that MaxEnt-ETRHEQ is an effective formulation for estimating energy fluxes for wetland ecosystems, especially for estimating \( \text{LE} \) from subdaily to monthly scales, and it does not necessitates any land surface parameters; only an assumption regarding vegetation height is required.

2. Methods

2.1 The formulation of MaxEnt-ETRHEQ

The main formula of MaxEnt-ETRHEQ is given as Eq. 1. The required input parameters are atmospheric pressure (p), air temperature (\( T_a \)), wind speed (u), friction velocity (u*), air relative humidity (RH), net radiation (\( R_n \)), the height of the measurements of weather data (z) and
vegetation height ($z_{\text{veg}}$). Meanwhile, MaxEnt-ETRHEQ will automatically create two variables, surface temperature ($T_s$) and surface relative humidity ($R_{\text{H} s}$) within a pre-defined range for studied ecosystems (will be explained later).

The surface pressure ($p_s$) is calculated from the atmospheric pressure by rearranging the formulas used in ETRHEQ, as (Salvucci and Gentine, 2013):

$$p_s = \frac{p \exp\left(-\frac{gz}{R_d T_a}\right)}{\exp\left(-\frac{gz}{R_d T_a}\right)}$$

where $p_s$ is the surface pressure (Pa), $p$ is the atmospheric pressure (Pa), $g$ is the gravitational constant (9.8 m·s$^{-2}$), $z$ is the height of the measurements of weather data (m), $R_d$ is the gas constant for dry air (287 J·kg$^{-1}$·K$^{-1}$), and $T_a$ is the air temperature (K).

Saturation vapor pressure ($e^*$) is calculated from integrated Clausius–Clapeyron relation, as (Salvucci and Gentine, 2013):

$$e^*(T_a) = 611.2 \times \exp\left(\frac{17.67 \times (T_a - 273.15)}{T_a - 29.65}\right)$$

$$e^*(T_s) = 611.2 \times \exp\left(\frac{17.67 \times (T_s - 273.15)}{T_s - 29.65}\right)$$

where $e^*(T_a)$ and $e^*(T_s)$ are saturation vapor pressure (Pa) at air temperature ($T_a$, K) and surface temperature ($T_s$, K), respectively.

Saturated specific humidity ($q^*$) is related to the saturation vapor pressure ($e^*$) through the following equations (Salvucci and Gentine, 2013):

$$q^*(T_a) = \frac{\epsilon e^*(T_a)}{p-(1-\epsilon)e^*(T_a)}$$

$$q^*(T_s) = \frac{\epsilon e^*(T_s)}{p_s-(1-\epsilon)e^*(T_s)}$$

where $\epsilon$ is the dimensionless ratio of the gas constant for dry air to water vapor (0.622).
Using Eq. 5 and 6, the slope of the relation between saturated specific humidity \( q^* \) and temperature \( T \) can be linearly extrapolated following (Kim et al., 2021; McColl et al., 2019):

\[
\delta = \frac{q^*(T_s)-q^*(T_a)}{T_s-T_a}
\]  
(7)

where \( q^*(T_s) \) and \( q^*(T_a) \) are the surface and atmospheric saturated specific humidity (kg·kg\(^{-1}\)), respectively, and \( T_s \) and \( T_a \) are the surface and air temperature (K), respectively.

The sensible and latent heat fluxes are calculated using the flux gradient equations, as (Kim et al., 2021):

\[
H = \rho c_p g_a (T_s - T_a)
\]  
(8)

\[
LE = \lambda \rho g_a (q_s - q_a)
\]  
(9)

where \( H \) and \( LE \) are the sensible and latent heats (W·m\(^{-2}\)), \( \rho \) is the density of air \( (\rho = \frac{p}{R_d T_a}) \), kg·m\(^{-3}\), \( c_p \) is the specific heat of air at constant pressure \( (1004.7 \text{ J·kg}^{-1}·\text{°C}^{-1}) \), \( g_a \) is the aerodynamic conductance accounting for atmospheric stability \( (\text{m·s}^{-1}) \), \( \lambda \) is the latent heat of vaporization \( (2.502 \times 10^6 \text{ J·kg}^{-1}) \), \( q_s \) is the surface specific humidity \( (q_s = \text{RH}_s \cdot q^*(T_s), \text{kg·kg}^{-1}) \), and \( q_a \) is the air specific humidity \( (q_a = \text{RH} \cdot q^*(T_a), \text{kg·kg}^{-1}) \).

The aerodynamic conductance under the neutral atmospheric condition \( (g_{a,n}) \) is given by Allen et al. (1998), as:

\[
g_{a,n} = \frac{\kappa^2 u}{\ln\left(\frac{z-d}{z_{om}}\right) \ln\left(\frac{z-d}{z_{oh}}\right)}
\]  
(10)

with \( \kappa \) being the von Karman constant (0.41), \( u \) being the wind speed (m·s\(^{-2}\)), \( z \) being the height of height of the measurements of wind speed (m), \( d \) being the zero-plane displacement height.
(m), \( z_{om} \) is the roughness length governing momentum transfer (m), and \( z_{oh} \) is the roughness length governing transfer of heat and vapour (m).

When no vegetation is present in the study sites \((z_{veg} = 0 \text{ m})\), \( d \) is set as 0 m, with both \( z_{om} \) and \( z_{oh} \) being set as 0.001 m; whereas in the presence of vegetation, \( d \) is set as 0.7 of \( z_{veg} \), with \( z_{om} \) being 0.1 of \( z_{veg} \), and \( z_{oh} \) being estimated using \( \kappa B^{-1} \) approach, following Salvucci and Gentine (2013):

\[
\kappa B^{-1} = \ln\left(\frac{z_{om}}{z_{oh}}\right) \cong \kappa (6Re^{\frac{1}{4}} - 5) \tag{11}
\]

where \( Re \) is the roughness Reynolds number \((Re = \frac{u^*z_{om}}{\nu})\), with \( u^* \) is the friction velocity \((\text{m} \cdot \text{s}^{-2})\) and \( \nu \) being the kinematic viscosity, as \( 1.45 \times 10^{-5} \text{ m}^2 \cdot \text{s}^{-1} \).

To account for atmospheric stability, the actual aerodynamic conductance \((g_a, \text{ m} \cdot \text{s}^{-1})\) is calculated following Merlin et al. (2016), as:

\[
g_a = (1 + R_i)^\eta \cdot g_{a,n} \tag{12}
\]

\[
R_i = \frac{\beta_{\text{thermal}} \cdot g \cdot (T_s - T_a)}{T_a u^2} \tag{13}
\]

where \( \beta_{\text{thermal}} \) is the thermal expansion coefficient, and \( \beta_{\text{thermal}} = 5 \) was used following Choudhury et al. (1986) and Merlin et al. (2011); \( g \) is the gravitational constant \((9.8 \text{ m} \cdot \text{s}^{-2})\), \( T_s \) is the surface soil temperature (K), \( T_a \) is the air temperature (K). In Eq. (11), the coefficient \( \eta \) is set to 0.75 in unstable conditions \((T_s > T_a)\) and to 2 in stable conditions \((T_s < T_a)\); \( u \) is the wind speed \((\text{m} \cdot \text{s}^{-1})\) and \( z \) is the height (m) at which wind speed was measured.

The ground heat flux \((G, \text{ W} \cdot \text{m}^{-2})\) is calculated using energy balance equation as:

\[
G = R_n - H - LE \tag{14}
\]
where $R_n$ is the net radiation (W·m$^{-2}$), and $H$ and $LE$ are calculated based on Eq. 8 and 9.

The parameterization of thermal inertias ($I_s$, $I_a$, and $I_e$) is provided in Eq. 1. To minimize the land surface parameters needed in the MaxEnt-ETRHEQ formulation, $I_s$ is set as a constant (1300 J·m$^{-2}$·K$^{-1}$·s$^{-1/2}$, i.e., tiu) following Rigden and Salvucci (2017). It is postulated that such a constant is acceptable, because: (1) Rigden and Salvucci (2015) stated that the optimal range of $I_s$ was between 300 and 1000 tiu for AmeriFlux sites, and as $I_s$ increases with wetter soils, it should be slightly higher than the optimal range; (2) Rigden and Salvucci (2017) used the calibrated $I_s$ of 1300 tiu for their study sites across united states; (3) the modelling results agree well with the eddy covariance measurements (presented in the results section); and (4) using measured soil moisture did not significantly improve the modelling performance (presented in Table S4).

The last step is to specify appropriate ranges of $G$, $T_s$, and $R_{H_s}$. Without this specification, $G$ could become unrealistically large, which does not occur in the real world. After specifying the ranges, the dissipation function $D$ is computed for every set of input weather data and every possible paring of $G$, $T_s$, and $R_{H_s}$. The selection of the optimal set of $G$, $T_s$, and $R_{H_s}$ will be done through finding the minimum $D$. Once these optimal values are found, $H$ and $LE$ are concurrently determined through the calculations from Eq.1 to 14.

2.2 The boundary conditions for wetland ecosystems

The upper limit of $G$ for wetland ecosystems is set as 0.20 of $R_n$ based on the empirical relationship between $G$ and $R_n$ used in GLEAM model: $G/R_n = 0.20$ for short vegetation ($0.05 \text{ m} < z_{veg} < 1 \text{ m}$) and $G/R_n = 0.15$ for tall vegetation ($Z_{veg} > 1 \text{ m}$) (Miralles et al., 2011). $T_s$ and $T_a$ at the 2 m height above land surface may differ by several °C, but the difference between maximum $T_s$ and maximum $T_a$ may vary up to 30 °C (Good et al., 2017; Mildrexler et al., 2011),
so $T_s$ is set to $T_a \pm 30$ °C. Typically, RHs must be higher than RH for evapotranspiration to occur.

As evapotranspiration progresses, RH tends to increase while RHs decreases until the ecosystem reaches the surface flux equilibrium state ($\text{RH}_{eq} = \text{RH} = \text{RH}_s$). This suggests that there exists a boundary for RHs, which falls within the range of RH and RH$_{eq}$. To estimate RH$_{eq}$, the Priestley-Taylor equation for water body (i.e., the left of the equals sign of Eq. 15) (Priestley and Taylor, 1972) is combined with the PM$_{RH}$ equation under RH=RH$_s$ (the right of the equals sign of Eq. 15) (Kim et al., 2021) to determine the maximum RH$_{eq}$, as:

$$1.26\lambda(R_n - G)\frac{\Delta}{\Delta + \gamma'} = \frac{\text{RH}_{eq}\Delta}{\text{RH}_{eq}\Delta + \gamma'}(R_n - G)$$

where 1.26 is the Priestley-Taylor coefficient for open water; $\lambda$ is the latent heat of vaporization ($2.502 \times 10^6$ J·kg$^{-1}$); $R_n$ is the net radiation (W·m$^{-2}$); $G$ is the ground heat flux (W·m$^{-2}$); $\Delta$ is the slope of the relation between saturation vapor pressure and temperature (Pa·°C$^{-1}$); $\gamma'$ is the psychrometric constant ($\gamma' = \frac{p_c\epsilon}{\lambda}$, with $p$ being is the air pressure (Pa), $c_p$ being the specific heat of air at constant pressure (1004.7 J·kg$^{-1}$·°C$^{-1}$), $\epsilon$ being the dimensionless ratio of the gas constant for dry air to water vapor (0.622), and $\lambda$ being the latent heat of vaporization ($2.502 \times 10^6$ J·kg$^{-1}$)); and RH$_{eq}$ is the equilibrium RH of a saturated wetland ecosystem. Rearranging Eq. 15 leads to the expression of RH$_{eq}$, as:

$$\text{RH}_{eq} = \frac{1.26\lambda\gamma'}{\Delta(1-1.26\lambda) + \gamma'}$$

There are multiple ways to estimate $\Delta$. In this study, the method provided in the FAO Penman-Monteith equation is chosen to estimate $\Delta$ from $T_a$, as (Allen et al., 1998):
where 1000 is a unit conversion coefficient, $T_a$ is the air temperature (K).

It is important to recognize that the range for $G$, $T_s$, and $RH_s$ can be refined in various ways. The ranges defined above are just simple examples to determine the plausible ranges of these parameters in wetland ecosystems, achieving more realistic results while reducing computation time. The true ranges for $G$, $T_s$, and $RH_s$ might be more constrained than these estimated values.

And many models, especially the models of $G$ (e.g., the models listed in Purdy et al. (2016), can be coupled with MaxEnt-ETRHEQ formulation. Exploring the potential enhancement of MaxEnt-ETRHEQ's performance by integrating these models presents an intriguing subject for future research.

3. Data and model evaluation

3.1 Data

All wetland sites classified as WET under the Vegetation IGBP category from the FLUXNET 2015 (Pastorello et al., 2020) and AmeriFlux (ameriflux.lbl.gov) FULLSET data products, shared under the CC-BY-4.0 license, were chosen for this study. The characteristics of the sites include latitude, longitude, elevation, mean measurement height, mean vegetation height, mean annual temperature, mean annual precipitation, and the distance to the coast (Table S1 and Table S2). Sites within 25 miles (~40 km) of the coast were removed, as ETRHEQ does not perform well in coastal regions (Rigden and Salvucci, 2015). In addition, the sites without the measurements of $Rn$ and $G$ were removed. The filter process results in 11 sites, including CZ-wet (Dušek et al., 2016), DE-SfN (Klatt et al., 2016), DE-Zrk (Sachs et al., 2016), FI-Lom (Aurela et al., 2016),

For every site, its fullset product encompasses five separate datasheets, containing measurements of atmospheric variables and energy fluxes at half-hourly, daily, weekly, monthly, and annual scales. At each temporal scale, \( u^* \) ("USTAR"), RH ("RH"), and \( R_n \) ("NETRAD") as well as gap-filled atmospheric measurements (denoted with the ".F" qualifier), including \( p \) ("PA.F"), \( T_a \) ("TA.F"), \( u \) ("WS.F"), and VPD ("VPD.F"), and the energy fluxes with marginal distribution sampling gap-filling method, which are \( G \) ("G_F_MDS"), \( H \) ("H_F_MDS"), and \( LE \) ("LE_F_MDS") were obtained. The names enclosed in double quotes within brackets in the above sentence represent the variable names in the data products. RH at daily or larger scales was not directly available, so it was estimated from VPD and \( T_a \) using the Clausius–Clapeyron relation. Besides \( z \) and \( z_{veg} \) were provided in Table S1 and S2. For sites where \( z_{veg} \) is not available, a value of 0.5 to \( z_{veg} \) was assigned. The focus here was limited to temporal scales between half-hourly and monthly, due to a lack of adequate sites and measurements for conducting a robust analysis at the yearly level. At the half-hourly scale, data with poor quality (i.e., the quality flag (QC) >1) were removed. At coarser temporal resolutions, i.e., from daily to monthly, only the measured data (QC=0) or the gap-filled data where over 80% measured or good quality gap-filled (QC=1) records aggregated from finer temporal resolutions were included, consistent with Kim et al. (2023). As a result, FI-Lom were removed from daily to monthly scales due to the lack of the quality flag for \( G \). At the monthly scale, DE-stfN was also removed because only one measurement was available. In addition, measurements were also removed if the surface energy imbalance was greater than 50 W·m\(^{-2}\)
or $R_n - G$ was negative (Kim et al., 2023). The amount of data after all filters from half-hourly to monthly scales are presented in Table S3. The model was also run for sites where soil moisture measurements were available (i.e., US-BZB, US-BZF, US-BZo and US-ICs), to assess whether incorporating soil moisture would enhance the model's performance, and the error statistics are provided in Table S4.

### 3.2 Model evaluation

The root-mean-square error (RMSE), slope and intercept (i.e., bias) of the fitted linear relationship between modelled and measured energy fluxes, and the coefficient of determination ($R^2$) were used as metrics to evaluate model performance. The evaluation was made of the measurements without energy closure correction (specifically, “H_F_MDS” and “LE_F_MDS” in the data product) and with correction using the energy balance closure correction factor on the assumption that the Bowen ratio is correct (Pastorello et al., 2020) (the corrected energy fluxes are “H_CORR” and “LE_CORR” in the data product), respectively. In addition, H_F_MDS and LE_F_MDS were compared with H and LE calculated as the residual of the energy balance (i.e., $H_{re} = R_n - G_{F\_MDS} - LE_{F\_MDS}$, and $LE_{re} = R_n - G_{F\_MDS} - H_{F\_MDS}$) to assess the how energy imbalance and the inherent uncertainty in the eddy covariance measurements affect the evaluation of the performance of the model. If there is no energy closure problems in the eddy covariance measurements, there would be a perfect fit between the measurements and the residuals of the energy balance for each energy flux. This represents the highest level of performance that can be expected from any model in comparison to eddy covariance measurements, as explained in McColl and Rigden (2020). However, comparisons with other models such as Penman-Monteith, Priestley-Taylor, MaxEnt, ETRHEQ, SFE, or MEP-SFE were not conducted because MaxEnt-ETRHEQ is still in its early stage, and this paper is intended to
provide a possible way to utilize it for wetland ecosystems. Further, MaxEnt-ETRHEQ is unique as it does not require G or T_s as inputs, unlike other models. However, inter-model comparisons will be considered in future research.

All the analysis was conducted on R 4.3.0 (R Core Team 2023). The R scripts, which contain the codes for calculating the distance from study sites to the coast, modelling the energy fluxes using MaxEnt-ETRHEQ for each site, and creating the figures presented in this paper, are all available at Wang (2024).

4. Results

MaxEnt-ETREHQ provides highly accurate predictions for LE from half-hourly to monthly scales (Figure 1), with slopes ranging from 0.86 to 1.08 and biases ranging from 4.00 to 6.34 W·m⁻². When the energy balance residuals (e.g., H_re and LE_re) were used to compare with the measurements (H_F_MDS and LE_F_MDS), their values of R² and the proximity of the slopes to 1 show similar levels with MaxEnt-ETRHEQ, but their bias, which is around 14 to 16 W·m⁻² and RMSE, which around 23 to 28 W·m⁻² (Table 1), were slightly larger than those of MaxEnt-ETRHEQ (bias: 4 to 7 W·m⁻² and RMSE: 11 to 27 W·m⁻²). In this sense, the error statistics of MaxEnt-ETRHEQ for estimating LE are slightly better than the errors statistics from the eddy covariance measurements (Table 1).
Figure 1. Modelled H and LE versus measurements (H_{obs} and LE_{obs}) without energy balance closure correction from half-hourly to monthly scales. The blue lines represent the fitted linear regressions. The black lines are 1:1 lines. The color of the points represents the density of the data ranging from low (purple) to medium (red) to high (yellow).

On the other hand, the model does not predict H with the same performance as it predicts LE, especially at the half-hourly scale (Figure 1). But when the time scale becomes larger, the
performance on estimating H is improved (Figure 1 and Table 1). Overall, MaxEnt-ETRHEQ
tends to overestimate H when H is low and underestimate H when it is high (Figure 1). Given the
estimation of LE was quite accurate, the less satisfactory performance of MaxEnt-ETRHEQ for
H can be attributed to less accurately defined boundary conditions for G and T_s. In the current
model setting, G was limited to up to 20% percent of Rn based on the GLEAM model that was
designed for daily applications (Purdy et al., 2016). This explains why MaxEnt-ETRHEQ
performs better in estimating H at daily and larger time scales. But at most study sites, G often
exceeds 20% of Rn when Rn is exceptionally low (e.g., less than 50 W·m⁻²), and frequently falls
below 10% of Rn when Rn is high (greater than 400 W·m⁻²). Therefore, when Rn is low, G is
underestimated, leading to an overestimation of H, and vice versa.

At the half-hourly scale, there is a spike of estimated H when measured H is near zero (Figure 1).
The spike is only from the US-BZo site (Figure S1) that happened during the night when Rn,
G_F_MDS, H_F_MDS and LE_F_MDS were all negative, and the absolute value of G was at
least 10 times larger than that of Rn. These energy fluxes suggest that US-BZo likely
encountered intense convective weather at these periods, characterized by air that was warmer
and more humid than the surface, accompanied by significant condensation. Under these weather
circumstances, the current MaxEnt-ETRHEQ formulations were unsuitable for H, LE, and g_a.

Determining the applicability of Eq. 1 in such conditions and devising revisions for the
calculations of H, LE, and g_a need to be addressed in future research.

When H and LE observations are adjusted to force energy balance closure (i.e., H_CORR and
LE_CORR), the performance of MaxEnt-ETRHEQ did not seem improve overall. For example,
the bias in H or LE estimations decrease when the energy fluxes are corrected at all temporal
scales, but the R², RMSE and slopes deteriorate (Table 1). This is because the energy balance
closure correction results in higher $H$ and $LE$ for most of the study sites. While this adjustment could result in more accurate energy fluxes, it also has the potential for overcorrection as diagnosed in Mauder et al. (2018). Consequently, the actual performance of MaxEnt-ETRHEQ in estimating $H$ and $LE$ should be in between its performance when compared to uncorrected fluxes and its performance when evaluated against corrected fluxes.

### Table 1. Summary of modelled fluxes against the energy balance corrected fluxes, and measured, uncorrected fluxes against the residuals of energy balance from half-hourly to monthly scales.

<table>
<thead>
<tr>
<th>Temporal scales</th>
<th>Variables</th>
<th>x</th>
<th>y</th>
<th>Slope</th>
<th>Intercept (bias)</th>
<th>$R^2$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Half-hourly</td>
<td>H_CORR</td>
<td>Modelled H</td>
<td>0.27</td>
<td>7.62</td>
<td>0.47</td>
<td>43.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE_CORR</td>
<td>Modelled LE</td>
<td>0.76</td>
<td>4.64</td>
<td>0.78</td>
<td>33.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H_F_MDS</td>
<td>$H_{re}$</td>
<td>0.90</td>
<td>12.80</td>
<td>0.69</td>
<td>27.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE_F_MDS</td>
<td>$LE_{re}$</td>
<td>0.88</td>
<td>14.60</td>
<td>0.76</td>
<td>27.87</td>
<td></td>
</tr>
<tr>
<td>Daily</td>
<td>H_CORR</td>
<td>Modelled H</td>
<td>0.56</td>
<td>9.50</td>
<td>0.58</td>
<td>19.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE_CORR</td>
<td>Modelled LE</td>
<td>0.72</td>
<td>5.59</td>
<td>0.74</td>
<td>20.66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H_F_MDS</td>
<td>$H_{re}$</td>
<td>1.11</td>
<td>15.60</td>
<td>0.74</td>
<td>23.65</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE_F_MDS</td>
<td>$LE_{re}$</td>
<td>1.08</td>
<td>15.10</td>
<td>0.82</td>
<td>23.65</td>
<td></td>
</tr>
<tr>
<td>Weekly</td>
<td>H_CORR</td>
<td>Modelled H</td>
<td>0.58</td>
<td>11.70</td>
<td>0.51</td>
<td>17.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE_CORR</td>
<td>Modelled LE</td>
<td>0.70</td>
<td>4.95</td>
<td>0.78</td>
<td>18.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H_F_MDS</td>
<td>$H_{re}$</td>
<td>1.15</td>
<td>16.70</td>
<td>0.74</td>
<td>24.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE_F_MDS</td>
<td>$LE_{re}$</td>
<td>1.13</td>
<td>15.40</td>
<td>0.84</td>
<td>24.08</td>
<td></td>
</tr>
<tr>
<td>Monthly</td>
<td>H_CORR</td>
<td>Modelled H</td>
<td>0.62</td>
<td>11.70</td>
<td>0.51</td>
<td>16.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE_CORR</td>
<td>Modelled LE</td>
<td>0.68</td>
<td>4.23</td>
<td>0.81</td>
<td>18.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>H_F_MDS</td>
<td>$H_{re}$</td>
<td>1.23</td>
<td>16.60</td>
<td>0.75</td>
<td>24.70</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LE_F_MDS</td>
<td>$LE_{re}$</td>
<td>1.18</td>
<td>15.10</td>
<td>0.86</td>
<td>24.70</td>
<td></td>
</tr>
</tbody>
</table>

Finally, the performance of MaxEnt-ETRHEQ in estimating $LE$ at individual sites throughout the temporal scales are also quite accurate. Figure 2 presents the half-hourly predictions of $LE$ at each site, and shows that despite varying accuracy across different sites, MaxEnt-ETRHEQ demonstrates high precision in predicting half-hourly $LE$, with $R^2$ values between 0.74 to 0.89 and RMSE ranging from 20.03 to 53.23 $W \cdot m^{-2}$. However, the predictions of $H$ at various
temporal scales were not satisfactory (Figure S1, Figure S2 and Figure S3). Nevertheless, when
the time scale is at the daily, weekly or monthly, both H and LE estimations are improved (Figure
3, Figure S2 and Figure S3). Considering that no site-specific calibration was made, and no $T_s$ or
$G$ were used as inputs, the performance of MaxEnt-ETRHEQ at individual sites were excellent.

Figure 2. Modelled LE versus measured LE without energy balance closure correction at
the half-hourly scale at the study sites. The blue lines represent the fitted linear regressions.
Black lines represent 1:1 lines.
5. Model advantages and limitations

The main advantage of MaxEnt-ETRHEQ is that it does not require land surface measurements like G and T<sub>s</sub>, which outcompetes most evapotranspiration (ET) models. While it could be argued that ET from wetland ecosystems closely approximates potential ET, which can be easily calculated using the Priestley-Taylor or Penman-Monteith equations for saturated water surfaces, the computation of potential ET (PET) still necessitates at least G as input. Additionally,
wetlands may not consistently be in a state of saturation (Streich, 2019), and using these equations could lead to substantial bias.

Moreover, MaxEnt-ETRHEQ is capable of providing estimates of LE at half-hourly intervals, distinguishing it from most equilibrium-based models that require equilibration times that typically extend beyond a daily timeframe, including the SFE model (McColl and Rigden, 2020) and the SFE-MEP model (Kim et al., 2023). The highly accurate half-hourly LE estimates provided by MaxEnt-ETRHEQ mean that the model is capable of precisely capturing the sub-daily fluctuations of ET. Many land surface models have shown considerable inaccuracies in sub-daily LE estimates, typically underestimating LE in the morning and overestimating it in the afternoon, owing to insufficient parameterizations of stomatal conductance and plant hydraulics (Matheny et al., 2014). MaxEnt-ETRHEQ and its underlying mechanism (i.e., the maximum entropy production) may provide new perspectives to enhance the performance of these models.

However, MaxEnt-ETRHEQ is still in its early stages, as further efforts are required to accurately refine the ranges of G, T_s and RH_s. However, that does not mean that these land surface variables ought to be inputs for MaxEnt-ETRHEQ. Rather, identifying appropriate boundary conditions for these variables should suffice. With growing evidence showing the interactions between land surface variables like G, T_s, soil moisture, soil thermal inertia and vegetation properties and near-surface atmospheric conditions (Bennett et al., 2008; Chu et al., 2018; Purdy et al., 2016; Wang and Bras, 1999; Wang and Bou-Zeid, 2012), developing physical models to describe these linkages and determining the limiting cases for G, T_s and RH_s are not far off. Once these boundary conditions are defined properly, MaxEnt-ETRHEQ will be capable of simultaneously estimating not only H, LE, and G, but also T_s and RH_s. Thus, it opens up a promising avenue for future research.
In addition, it may be argued that MaxEnt-ETRHEQ relies on empirical parameters like the parameterization of $I_s$ and $g_a$. Indeed, most models for estimating surface energy fluxes are largely based on empirical approaches, particularly in calculating parameters such as displacement height, roughness length for momentum and heat transfer, and aerodynamic conductance. Furthermore, when these models are scaled up for application over extensive areas, the reliance on parameters that have been either assumed or previously calibrated becomes inevitable. Therefore, the use of empirical parameterizations in MaxEnt-ETRHEQ should not be viewed as shortcomings. Instead, it underscores the critical need for further research aimed at refining these parameters to enhance the model's accuracy.

6. Conclusion

The goal of this paper is to demonstrate the effectiveness of a newly developed formulation grounded in the principle of maximum Shannon information entropy production theory for estimating surface energy fluxes in wetland ecosystems. The formulation requires neither land surface variables nor site-specific calibration, except for a presumed vegetation height, and it effectively estimates LE from half-hourly to monthly scales in the FLUXNET and AmeriFlux wetland sites. While its estimation on $H$ is less satisfactory due to roughly constrained boundary conditions for $G$ and $T_s$, the formulation holds promise for concurrently and accurately estimating $LE$, $H$, $G$, $T_s$ and $RH_s$ for various ecosystems if limiting cases of $G$, $T_s$ and $RH_s$ are properly established. Overall, the formulation contributes new insights into developing earth systems models.
Open research

All datasets in this study, as well as the R scripts used for modeling and data visualization, are publicly available. For access to the specific datasets used in this study, please refer to the FLUXNET database (http://www.fluxnet.org) and the AmeriFlux network (http://ameriflux.lbl.gov). For the data analysis, the R programming language version 4.3.0 (R Core Team 2023) was employed. The codes can be accessed on Wang, Y. (2024). R scripts for the submission by Wang and Petrone, "An effective formulation for estimating wetland surface energy fluxes from weather data". Zenodo. https://doi.org/10.5281/zenodo.10602494.

Acknowledgement

The authors wish to express their gratitude to the FLUXNET community and AmeriFlux network for providing access to the invaluable datasets. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The FLUXNET eddy covariance data processing and harmonization was carried out by the ICOS Ecosystem Thematic Center, AmeriFlux Management Project and Fluxdata project of FLUXNET, with the support of CDIAC, and the OzFlux, ChinaFlux and AsiaFlux offices. We also acknowledge the following AmeriFlux sites for their data records: US-BZB, US-BZF, US-BZo and US-ICs. In addition, funding for AmeriFlux data resources was provided by the U.S. Department of Energy’s Office of Science.
We would like to thank the Mountain Water Futures project of the Global Water Futures programme (Canada First Research Excellence Fund), Alberta Innovates – Energy and Environment Solutions (Grant #s: AIES-2335, 2335 20007947; Rooney, Petrone), and the Canadian Natural Science and Engineering Research Council (NSERC) Discovery (RGPN-04182-2017; Petrone) and CREATE (463960-2015; Petrone) grants programs for the financial support.

References:


Desai, A., 2016. FLUXNET2015 US-Los Lost Creek, FluxNet; Univ. of Wisconsin, Madison, WI (United States).


Dušek, J. et al., 2016. FLUXNET2015 CZ-wet Trebon (CZECHWET), FluxNet; Global Change Research Institute CAS.


Euskirchen, E., 2022. AmeriFlux FLUXNET-1F US-BZo Bonanza Creek Old Thermokarst Bog, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States ….

Euskirchen, E., Shaver, G. and Bret-Harte, S., 2016. AmeriFlux AmeriFlux US-ICs Imnavait Creek watershed wet sedge tundra, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States ….


Streich, S.C., 2019. The hydrological function of a mountain valley-bottom peatland under drought conditions, University of Saskatchewan.


Figure S1. Half-hourly estimates of H versus measured H. without energy balance closure correction at the half-hourly scale at the study sites. Blue lines represent the fitted linear regressions. Black lines represent 1:1 lines.
Figure S2. Modelled H and LE and measured H and LE without energy balance closure correction at the weekly scale. Blue lines represent the fitted linear regressions. Black lines represent 1:1 lines.
Figure S3. Modelled H and LE versus measured H and LE without energy balance closure correction at the monthly scale. Blue lines represent the fitted linear regressions. Black lines represent 1:1 lines.