

Surface Flux Equilibrium Estimates of Evapotranspiration at Large Spatial Scales

SHILIU CHEN,^{a,b} KAIGHIN A. MCCOLL,^{a,c} ALEXIS BERG,^a AND YUEFEI HUANG^{b,d}

^a *Department of Earth and Planetary Sciences, Harvard University, Cambridge, Massachusetts*

^b *State Key Laboratory of Hydrosience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing, China*

^c *School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts*

^d *State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University, Qinghai, China*

(Manuscript received 17 August 2020, in final form 4 January 2021)

ABSTRACT: A recent theory proposes that inland continental regions are in a state of surface flux equilibrium (SFE), in which tight coupling between the land and atmosphere allow estimation of the Bowen ratio at daily to monthly time scales solely from atmospheric measurements, without calibration, even when the land surface strongly constrains the surface energy budget. However, since the theory has only been evaluated at quasi-point spatial scales using eddy covariance measurements with limited global coverage, it is unclear if it is applicable to the larger spatial scales relevant to studies of global climate. In this study, SFE estimates of the Bowen ratio are combined with satellite observations of surface net radiation to obtain large-scale estimates of latent heat flux λE . When evaluated against multiyear mean annual λE obtained from catchment water balance estimates from 221 catchments across the United States, the resulting error statistics are comparable to those in the catchment water balance estimates themselves. The theory is then used to diagnostically estimate λE using historical simulations from 26 CMIP6 models. The resulting SFE estimates are typically at least as accurate as the CMIP6 model's simulated λE , when compared with catchment water balance estimates. Globally, there is broad spatial and temporal agreement between CMIP6 model SFE estimates and the CMIP6 model's simulated λE , although SFE likely overestimates λE in some arid regions. We conclude that SFE applies reasonably at large spatial scales relevant to climate studies, and is broadly reproduced in climate models.

KEYWORDS: Evaporation; Evapotranspiration; Hydrologic cycle; Latent heating/cooling

1. Introduction

Evapotranspiration E is a major flux in the terrestrial water, energy, and carbon cycles. On average, approximately 60% of precipitation falling on land returns to the atmosphere directly through terrestrial E (Oki and Kanae 2006). The associated latent heat flux λE (where λ is the latent heat of vaporization of water) cools the land surface and lower atmosphere. Plants link carbon assimilation with transpiration through stomatal regulation (Berry et al. 2010). In addition, E is of practical importance in water resources management and agriculture (e.g., DeLucia et al. 2019).

Evapotranspiration is constrained by both the land (soil moisture, plant physiology, surface temperature) and atmosphere (vertical gradients in air temperature and humidity, stratified turbulent transport). The land surface is highly heterogeneous, which complicates modeling at global scales. As a result of these modeling challenges, climate models systematically overestimate E at annual time scales (Mueller and Seneviratne 2014). They also underestimate it at shorter time scales in many inland continental regions during the Northern Hemisphere summer (Mueller and Seneviratne 2014; Ma et al. 2018). Compared with climate models, which are prognostic, diagnostic E products have also been developed (e.g., Mueller et al. 2013). Such products often take advantage of new satellite observations (e.g., Fisher et al. 2008; Martens et al. 2017).

While satellite E products have proved useful in a variety of applications, a recent intercomparison study found substantial disagreement between diagnostic E products (Miralles et al. 2016). Furthermore, when evaluated against global long-term catchment water balance estimates, none of the diagnostic E products demonstrated unequivocal improvements over E obtained from a reanalysis (Fig. 10 of Miralles et al. 2016).

Since there is often a trade-off between model simplicity and accuracy, much attention has been paid to improving model accuracy by developing more complex E models. However, different models will be necessary for different purposes, and simple models have their own advantages. For example, simple models can be more useful for developing understanding of the governing physics (Held 2005; Jeevanjee et al. 2017; Maher et al. 2019), an approach adopted fruitfully in many previous studies of E (De Bruin 1983; McNaughton and Spriggs 1986; Culf 1994; Brutsaert and Parlange 1998; Betts 2000; Raupach 2000, 2001; McColl et al. 2019). In addition, relatively simple models may be more useful in settings where input data required by more complex models are not available. For example, land surface observations—including soil moisture, soil texture, and various vegetation properties (height, water content and functional type)—are often required as inputs to models of E , at spatial resolutions that are sufficiently fine to resolve the (typically considerable) spatial variability of the heterogeneous land surface. Such observations are seldom routinely available, particularly at large spatial scales, and so parameterizations must be introduced that add considerable error (e.g., Rigden et al. 2018; Trugman et al. 2018). Algorithms based on the “complementary

Corresponding author: Kaighin A. McColl, kmccoll@seas.harvard.edu

DOI: 10.1175/JHM-D-20-0204.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy \(www.ametsoc.org/PUBSReuseLicenses\)](https://www.ametsoc.org/PUBSReuseLicenses).

Brought to you by Harvard Library Information and Technical Services | Unauthenticated | Downloaded 03/23/21 02:56 PM UTC

relationship” (Bouchet 1963; Morton 1969; Brutsaert and Stricker 1979; Brutsaert 2015; Ma and Szilagyi 2019) and the Evapotranspiration from Relative Humidity at Equilibrium (ETRHEQ) method (Salvucci and Gentine 2013; Rigden and Salvucci 2015) have allowed estimation of E with substantially fewer data inputs compared with more complex methods. Even in applications where model accuracy is paramount, if there is substantial uncertainty in reference measurements used to evaluate model performance, it may be difficult to justify substantial model complexity due to the risk of overfitting. For this reason, if a simple model exhibits errors comparable to those in the reference measurements themselves—an upper bound on the performance of any model—then the simple model is preferable to more complex models.

In this study, we focus on a maximally simple diagnostic model of continental E at daily to monthly time scales. The model, called surface flux equilibrium (SFE; McColl et al. 2019; McColl and Rigden 2020), assumes strong coupling between the land and atmosphere, such that higher near-surface air temperatures and specific humidities are mainly caused by higher sensible heat fluxes and latent heat fluxes at the land surface, respectively (rather than atmospheric mechanisms, such as convergence of heat and moisture). This model is attractive because it bypasses the need to explicitly account for the complexity of the land surface; instead, land surface heterogeneity becomes embedded in the near-surface atmospheric state, allowing the estimation of land surface quantities from readily available atmospheric measurements, without calibration. The specific prediction made by SFE is that, within this tightly coupled system, an approximate balance exists between the surface moistening and heating terms in the near-surface relative humidity budget; this assumed balance leads to a simple equation for the Bowen ratio (defined as $\beta = H/\lambda E$) or, equivalently, the evaporative fraction [$EF = (1 + \beta)^{-1}$]. When combined with observations of net radiation, it can be used to estimate latent heat flux without any calibration parameters or land surface inputs, even when the land surface state substantially constrains E .

While SFE is, to our knowledge, the simplest model of actual E at daily to monthly time scales over inland continental regions, it is far from obvious that its assumptions are reasonable, and that its predictions will be accurate. While SFE is not expected to hold in coastal regions or at subdaily time scales, a recent comparison with state-of-the-art eddy covariance measurements from around the world demonstrated that errors in SFE estimates were indistinguishable from those in the eddy covariance measurements themselves at a majority of inland continental sites (McColl and Rigden 2020). However, the eddy covariance measurements used in that study apply to quasi-point spatial scales and have limited global coverage. The dominant physical processes controlling E vary substantially with spatial scale, particularly since surface heterogeneity is unavoidable at larger scales (Jarvis and McNaughton 1986; Baldocchi et al. 1991; Raupach and Finnigan 1995; Brutsaert 1998; Mahrt 2000; Taylor et al. 2013; Li and Wang 2019; Bou-Zeid et al. 2020). For example, consider an idealized case in which a small pan of water is placed in an otherwise dry desert. At the scale of the pan, E will be very high and limited by atmospheric water demand, since there is an abundant supply of

water at the land surface provided by the pan. At larger scales, however, E will be very low and limited by land surface water supply, since the land surface water supply is negligible when averaged over both the pan and the much larger desert.

To what extent does SFE apply at larger spatial scales relevant to climate studies? If it applies reasonably at these scales, to what extent is it reproduced in climate models? In response to these questions, in this study, we evaluate the performance of SFE predictions at large spatial scales [$O(10^2\text{--}10^4)$ km²] relevant to studies of climate. To do this, SFE estimates are compared with the most accurate large-scale estimate of E : catchment water balance estimates of multiyear mean annual E . Since the catchment water balance estimates are not entirely free from error, a simple error propagation analysis is incorporated into the comparison. The performance of SFE estimates are benchmarked against equivalent estimates from a reanalysis, and from two other simple models of E : the Budyko and Priestley–Taylor equations. While simple, it is shown that the Budyko equation is indistinguishable from a perfect model in terms of estimated error statistics, due to unavoidable uncertainties in the catchment water balance reference estimates; comparisons with a broader suite of more complex satellite or reanalysis products would, therefore, be no more statistically meaningful than comparison with the Budyko equation. The comparison with the Priestley–Taylor equation is used as a test of the statistical power of our analysis: since the Priestley–Taylor equation is not expected to provide accurate estimates of E over most land surfaces, our analysis should clearly demonstrate that SFE estimates perform better than Priestley–Taylor estimates. If it does not, this would imply that errors in the catchment water balance estimates preclude a statistically meaningful assessment of the performance of SFE. After demonstrating that SFE performs well at large scales, we also evaluate the extent to which SFE is an emergent feature within a suite of climate models. The aim of this comparison is not to evaluate SFE, but to test the degree to which climate models reproduce SFE. It is shown that climate models do, on the whole, reproduce SFE reasonably well; that is, given climate model outputs of near-surface air temperature and specific humidity, and surface net radiation, the climate models are able to estimate λE using SFE reasonably accurately. These results provide further empirical support for the robustness of SFE, and for its application to studies of continental climate.

This manuscript is organized as follows. In section 2, the data and climate models used to estimate and evaluate SFE at the catchment scale are presented, along with the E models used to benchmark its performance. In section 3, the performance comparisons are presented and discussed, along with known limitations of our analysis and its relation to previous studies. We conclude with a summary in section 4, and a brief discussion of future research opportunities using SFE.

2. Methods and data

a. Estimation of multiyear mean annual E from catchment water balance

Averaging over multiyear time scales, annual changes in water stored within a catchment are typically small in comparison

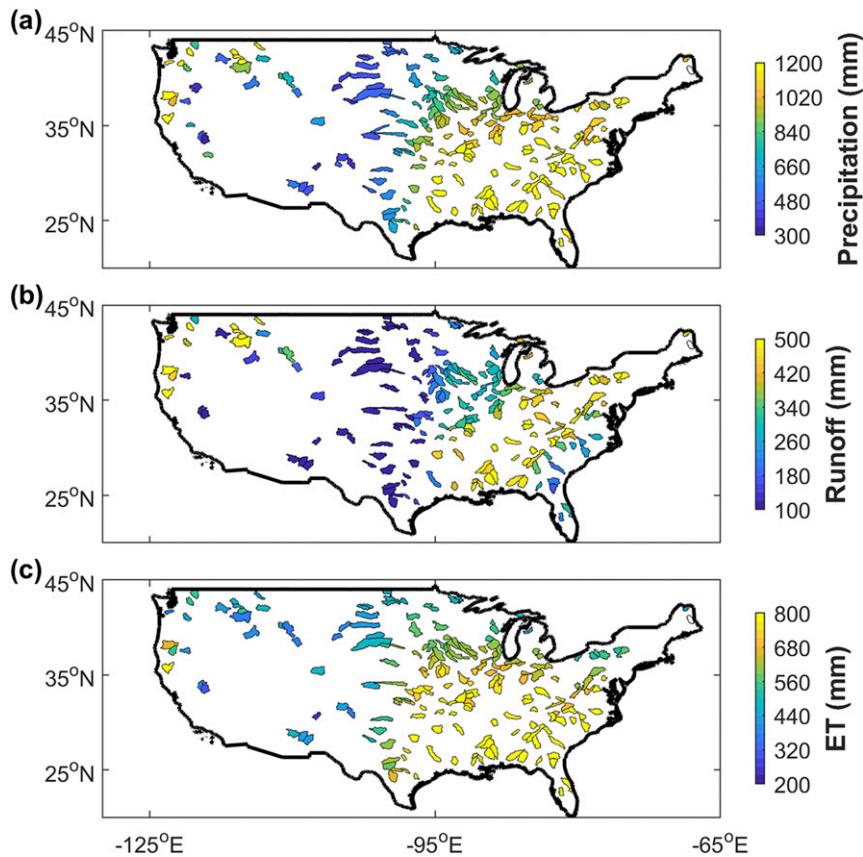


FIG. 1. Spatial distribution of the multiyear (2001–14) annual mean (a) precipitation, (b) runoff, and (c) evapotranspiration E for each catchment used in this study. Evapotranspiration E is estimated as the difference between observed precipitation and runoff.

to annual fluxes of water in and out of the catchment. If one further assumes that precipitation P , E , and runoff Q are the dominant fluxes, and that all runoff exits the catchment at its gauged outlet, then the catchment's multiyear mean annual water balance reduces to $E = P - Q$, where Q is measured as streamflow at the catchment's outlet. Since measurements of P and Q are more prevalent than measurements of E , this approach is often used to estimate multiyear-mean annual E at catchment scales (e.g., Jung et al. 2010; Miralles et al. 2016). The assumptions made in this approach are substantially violated at shorter time scales. For example, the assumption of no change in storage can be significantly inaccurate even when applied to estimating annual E for a single year (Han et al. 2015); neglecting positive changes in storage will positively bias estimates of E obtained using this method. However, for multiyear mean annual E , this approach arguably remains the most accurate available at large spatial scales (Vinukollu et al. 2011). For this reason, our analyses focus on multiyear mean annual time scales when using catchment water balance E estimates. We consider shorter time scales in more detail in later sections using climate model outputs.

The catchments used in this study are shown in Fig. 1, and were obtained from the Model Parameter Estimation Experiment (MOPEX) dataset (Duan et al. 2006). Since SFE is not expected

to hold in coastal regions (McCull et al. 2019; McCull and Rigden 2020), catchments within 250 km of coasts and large waterbodies were removed. Moreover, since smaller catchments are more likely to violate the assumptions of the catchment water balance mentioned above, catchments smaller than 2048 km² were excluded, consistent with previous studies (Yin et al. 2019). This resulted in 221 catchments retained for further analysis (Fig. 1), spanning a broad range of climates, land cover types, and basin sizes (2053–25 791 km²). Figure 1 maps the three major fluxes in each catchment's multiyear mean annual water balance, where E is estimated as the difference between precipitation and runoff (further details on the data used in this figure are provided in the next section).

b. Data

The surface net radiation data used are obtained from the NASA Clouds and Earth's Radiant Energy System (CERES) EBAF Level 3b Edition 4.1 product (Kato et al. 2018). CERES observations have been globally validated (Jia et al. 2018), and used extensively in other studies of global E (Vinukollu et al. 2011; Miralles et al. 2016). The CERES data are provided at a monthly temporal resolution, and a 1° spatial resolution. To match the finer spatial resolution of other forcing data used in this study (1/8°), the CERES data are interpolated to a 1/8° grid

using nearest-neighbor resampling. We use data spanning the period 2001–14 in this study.

Daily streamflow gauge observations from 2001 to 2014 were obtained for each catchment from the U.S. Geological Survey (USGS) National Water Information System (<http://waterdata.usgs.gov/usa/nwis/sw>). The daily streamflow data were converted to annual averages. Only years with no missing daily data were retained in the analysis.

Precipitation data were obtained over the period 2001–14 for each catchment from the Parameter-Elevation Regressions on Independent Slopes Model (PRISM, <http://prism.oregonstate.edu>, Daly et al. 1994). The PRISM product interpolates daily observations from 13 000 ground stations across the United States onto a 4-km resolution grid. The PRISM precipitation data were also aggregated onto 1/8° grids for consistency with other forcing datasets.

All other meteorological data, including near-surface air temperature, specific humidity and surface pressure, were obtained from the North American Regional Reanalysis (NARR), available at a monthly temporal resolution and 1/8° spatial resolution (NARR, Mesinger et al. 2006). The NARR latent heat flux was also obtained for comparison with estimates obtained from SFE. The data are outputs from the NARR model cycling, rather than direct observations, although the system itself is constrained by observations. Given the absence of observations at different heights, we used reanalysis estimates of near-surface atmospheric quantities at a height of 2 m at all locations.

c. Evaluation of E models

Surface flux equilibrium (McColl et al. 2019; McColl and Rigden 2020) hypothesizes that, at daily to monthly time scales, the latent heat flux λE can be estimated using the relation

$$\lambda E = \frac{\lambda^2 q}{\lambda^2 q + c_p R_v T^2} R_n, \quad (1)$$

where λ is the latent heat of vaporization of water (J kg^{-1}); R_n is net radiation (W m^{-2}); q and T are the specific humidity (—) and temperature (K) of the near-surface atmosphere, respectively; c_p is the specific heat capacity of the near-surface atmosphere at constant pressure ($\text{J kg}^{-1} \text{K}^{-1}$); and $R_v = 461.5 \text{ J kg}^{-1} \text{K}^{-1}$ is the ideal gas constant for water vapor. While λ and c_p are both technically functions of temperature, their dependence on temperature is weak for typical atmospheric conditions on Earth and they are commonly treated as constants, with $\lambda \approx 2.5008 \times 10^6 \text{ J kg}^{-1}$ and $c_p \approx 1005 \text{ J kg}^{-1} \text{K}^{-1}$. Furthermore, the ground heat flux has been neglected here, as it is typically much smaller than R_n at time scales longer than a day. This implies that λE can be estimated, without calibration of any free parameters, solely from observations of q , T , and R_n using Eq. (1). We used Eq. (1) to estimate λE for each month, and for each pixel within a given catchment, and then spatially and temporally aggregated to catchment and multiyear scales, respectively.

To benchmark the performance of the SFE estimates at catchment scales, we compared it to three other E estimates: the Priestley–Taylor and Budyko equations, and the reanalysis (NARR). The Priestley–Taylor and Budyko equations were

chosen for comparison because, like SFE, they do not require any land surface variables as inputs, and are often implemented without calibration of free parameters. While it might seem strange to benchmark against these relatively simple models rather than more detailed models, it will be shown that the SFE and Budyko models both perform better than the reanalysis, which is substantially more complicated. More broadly, it will be shown that the SFE and Budyko models are indistinguishable from a perfect model in terms of estimated error statistics, due to unavoidable uncertainties in the reference dataset. The comparison with the Priestley–Taylor equation provides a test of the statistical power of our analysis: if our analysis is not able to demonstrate that the Priestley–Taylor equation performs relatively poorly (since it is not expected to accurately estimate E over most water-limited land surfaces), then the analysis is overwhelmed by uncertainties in the catchment water balance reference dataset.

The Priestley–Taylor equation (Priestley and Taylor 1972) provides an estimate of E from a saturated land surface, and is given by

$$\lambda E = \alpha \frac{\varepsilon}{\varepsilon + 1} R_n = \alpha \frac{\lambda^2 q^*(T)}{\lambda^2 q^*(T) + c_p R_v T^2} R_n, \quad (2)$$

where $q^*(T)$ is the saturation specific humidity (—), $\varepsilon = (\lambda/c_p)[\lambda q^*(T)/R_v T^2]$ is the normalized slope of the Clausius–Clapeyron relation (—), and α is the Priestley–Taylor parameter (—). A common choice for this parameter, which we adopt here, is $\alpha = 1.26$ (Priestley and Taylor 1972). Since the Priestley–Taylor equation only requires T and R_n as inputs, it is simpler than Eq. (1), provided α is treated as a known constant. It should be noted that estimates of α based on field observations vary considerably (Shuttleworth and Calder 1979).

The Budyko relation provides an estimate of multiyear mean annual λE at large scales [Eq. (107) of Budyko 1958]:

$$\lambda E = \lambda P \sqrt{\frac{R_n}{\lambda P} \tanh\left(\frac{\lambda P}{R_n}\right) \left[1 - \exp\left(-\frac{R_n}{\lambda P}\right)\right]}, \quad (3)$$

where P is multiyear mean annual precipitation ($\text{kg m}^{-2} \text{s}^{-1}$). This equation is also arguably simpler than Eq. (1), as it only requires P and R_n as inputs. Other functional forms have also been proposed (e.g., Yang et al. 2008):

$$\lambda E = \lambda P \left[\left(\frac{R_n}{\lambda P} \right)^{-n} + 1 \right]^{-1/n}, \quad (4)$$

where n is a (positive) calibration parameter, which we obtain by fitting to the catchment water balance data described in the previous section. In terms of inputs, both Eqs. (3) and (4) require P and R_n as inputs, meaning they are both somewhat simpler than Eq. (1), which requires T , q , and R_n . However, Eq. (4) includes a calibration parameter, whereas Eq. (1) has no parameters. Furthermore, unlike SFE, neither form of the Budyko relation generalizes to time scales shorter than the multiyear mean annual.

The root-mean-square error (RMSE) and mean bias (B) were estimated for each model, with respect to the catchment

water balance estimate of E (which we also refer to as the “reference” E). Some of the deviation between modeled and reference E is due to errors in the catchment water balance E estimate, particularly due to sampling errors in the precipitation data used (Milly and Dunne 2002; Miralles et al. 2016). Therefore, even a perfect E model would exhibit imperfect error statistics when compared to catchment water balance E estimates. To quantify this effect and to benchmark the performance of each model, a simple error analysis was conducted. The error model $\hat{E} = P(1 + \delta) - Q$ was used, where δ is a fixed bias, and \hat{E} is the catchment water balance estimate. This error model emphasizes fixed biases [such as systematic negative biases from gauge undercatch (Legates and Willmott 1990), and undersampling of mountainous regions (Milly and Dunne 2002)] and ignores random errors, which are presumably significantly reduced by averaging over multiyear time scales. Furthermore, errors in Q are neglected, since they are likely much smaller than those in estimates of precipitation. Milly and Dunne (2002) conducted a detailed analysis of errors in catchment-scale precipitation estimates. They found that “a 10% to 20% bias in precipitation is typical” among the basins in their study, with the bias considerably larger in many cases. If $\delta = -15\%$ is used as a representative precipitation bias, this implies that even a perfect model of E will exhibit the following error statistics:

$$\begin{aligned} \text{RMSE}_0 &= \sqrt{\overline{(E - \hat{E})^2}} = \sqrt{\overline{\{(P - Q) - [P(1 + \delta) - Q]\}^2}} \\ &= \sqrt{\overline{P^2 \delta^2}} = 0.15 \sqrt{\frac{1}{N} \sum_{i=1}^N P_i^2}, \end{aligned} \quad (5)$$

$$B_0 = \overline{E - \hat{E}} = -\overline{P\delta} = 0.15 \frac{1}{N} \sum_{i=1}^N P_i, \quad (6)$$

where averaging is performed over N catchments used in the study. While the PRISM precipitation product used in this study makes corrections for orography, in general, precipitation products typically still underestimate precipitation even after making such corrections (Beck et al. 2020). RMSE_0 and B_0 are our best estimates of the error statistics of a perfect E model, when compared against the catchment water balance estimates of E .

d. Comparison with CMIP6 model simulations

We also evaluate the performance of SFE within climate models, using historical Coupled Model Intercomparison Project phase 6 (CMIP6; Eyring et al. 2016) outputs over the period 2001–14. Specifically, monthly outputs of variables tas (near-surface air temperature), huss (near-surface specific humidity), hfls (latent heat flux), and hfss (sensible heat flux) were used, with available energy estimated as the sum of latent and sensible heat fluxes. At time of writing, 26 CMIP6 models provided these outputs over the required time period, and so our analysis is restricted to those models. The focus of this study is on comparison with observations where possible, so we limit our analysis to historical simulations, which can be compared with observed climatology, rather than future projections.

When comparing each CMIP6 model to the catchment water balance estimate of E , catchments smaller than the model’s spatial resolution were discarded. Since spatial resolution varied between climate models, the number of catchments used in the comparison was different for each model.

The comparison was repeated using only catchments that were sufficiently large to be resolved by all CMIP6 models, resulting in $N = 8$ catchments. Since it is difficult to estimate validation statistics with such a small sample size, and our analysis is not focused on differences between CMIP6 models, we elected to focus on the analysis which uses the maximum number of catchments allowed by each model’s spatial resolution, even though this results in different catchments being used in the validation of each model. However, results are also presented for the case in which a consistent set of catchments are used for each CMIP6 model.

3. Results and discussion

a. Evaluation of SFE at the catchment scale

Figure 2 presents the comparison of the five E models—SFE [Eq. (1)], Priestley–Taylor [Eq. (2)], the reanalysis (NARR), the Budyko equation [Eq. (3)], and the calibrated Budyko equation [Eq. (4)]—with the reference catchment water balance λE estimate. SFE (Fig. 2a) and both Budyko equations (Figs. 2d,e) perform well. In particular, the estimated RMSE and B for these three models are comparable to (in fact, lower than) the values expected for a model with no error, after accounting for errors in the reference λE [$\text{RMSE}_0 = 12 \text{ W m}^{-2}$ and $B_0 = 11 \text{ W m}^{-2}$, based on Eqs. (5) and (6), respectively]. This does not imply that the SFE and Budyko equations are perfect models of multiyear mean annual E . Rather, it implies that they all perform at least as well as the most accurate model our analysis is capable of assessing. Differences in error statistics between models that occur below the RMSE_0 and B_0 thresholds cannot be distinguished from errors in the reference dataset, and should not be attributed to differences in model performance. This is particularly important for models with calibration parameters, such as the calibrated Budyko model (Fig. 2d), which are susceptible to overfitting. For this reason, we do not compare SFE to a broader set of more complicated E models.

In contrast to SFE and the Budyko models, the Priestley–Taylor equation performs much less well (Fig. 2b). Both RMSE and B are substantially higher than equivalent estimates for the SFE and Budyko models. They are also larger than RMSE_0 and B_0 , respectively. This is expected, since the Priestley–Taylor equation is a model of E from a saturated surface, and most catchments are not saturated. However, it demonstrates that the reference E , while subject to errors, is capable of distinguishing between E models that are known to be broadly accurate at multiyear mean time scales (the Budyko equations) and those that are expected to be inaccurate (the Priestley–Taylor equation). Therefore, the good performance of SFE is not an artifact caused by overwhelmingly large observation errors in the catchment water balance E estimates.

Surprisingly, SFE also performs substantially better than the reanalysis estimate of λE (Fig. 2c). Like the Priestley–Taylor

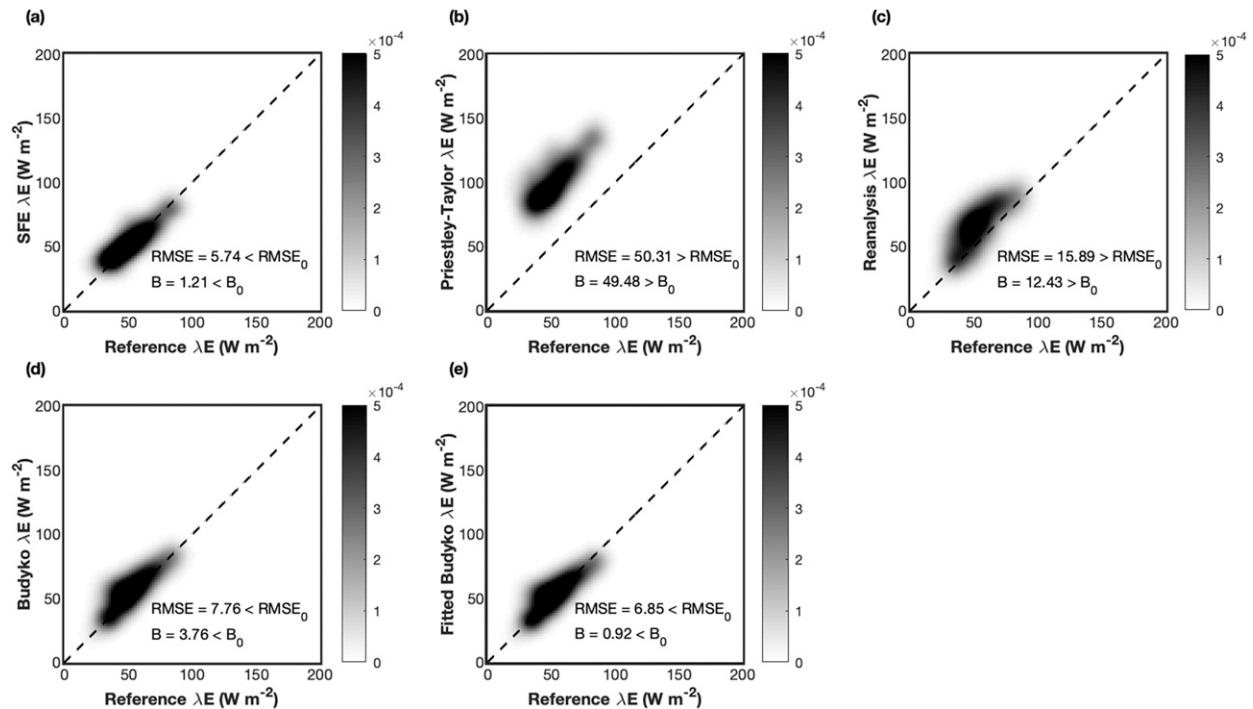


FIG. 2. Comparison with reference multiyear mean annual λE of equivalent estimates from (a) SFE, (b) the Priestley–Taylor equation, (c) the North American Regional Reanalysis (NARR), (d) the uncalibrated Budyko equation, and (e) the calibrated Budyko equation. Reference multiyear mean annual E is obtained from a catchment water balance. RMSE is root-mean-square error, and $RMSE_0$ is the RMSE expected solely from errors in the reference λE (12 W m^{-2}); B is mean bias, and B_0 is the bias expected solely from errors in the reference λE (11 W m^{-2}). Shaded areas and color bars show the estimated joint empirical distribution functions of variables listed on the horizontal and vertical axes. Dashed diagonal lines are 1:1 lines. For reference, $1 \text{ W m}^{-2} = 12.872 \text{ mm yr}^{-1}$.

equation, the reanalysis exhibits errors that are unlikely to be artifacts caused by errors in the catchment water balance estimate ($RMSE > RMSE_0$, $B > B_0$). These results are consistent with recent work that also identified a positive bias in NARR E at multiyear mean annual time scales (Yin et al. 2019), and demonstrate that more complex products are not necessarily more accurate. For completeness, a comparison between the reanalysis and SFE is provided across North America in Figs. 3a–c. Time series at three representative sites are also provided (Figs. 3d–f). In some drier sites, the SFE estimate is greater than the reanalysis estimate (Fig. 3d); at other sites, the reanalysis estimate is greater than the SFE estimate (Fig. 3f). In general, it is difficult to attribute differences between the products (Fig. 3c) to errors in either the reanalysis or SFE estimate. The limited validation in Fig. 2 suggests that, if anything, differences between SFE and reanalysis estimates are more likely to be due to errors in the reanalysis than in the SFE estimate. For this reason, more detailed comparisons between the reanalysis and SFE are not pursued further in this study.

Why not compare SFE with a larger set of E products? In our view, a finer-grained comparison is not justified for at least two reasons. First, there is little reason to believe that other diagnostic products will perform better than the reanalysis used here. A recent comprehensive intercomparison study found that none of the products evaluated demonstrated clear

superiority over estimates obtained from a reanalysis (Miralles et al. 2016). Second, the significant errors present in even the most accurate large-scale estimates of multiyear mean annual E (the catchment water balance estimate) preclude a more nuanced comparison between products. Differences in validation statistics between E estimates are not statistically meaningful if they are within the observation error of the catchment water balance estimate. In our notation, this corresponds to $RMSE \leq RMSE_0$ and $B \leq B_0$. Since errors in SFE lie below these thresholds when compared with the catchment water balance estimates (Fig. 2a), such a comparison would not be capable of distinguishing differences in performance with another E estimate, unless that estimate displays errors above the thresholds $RMSE_0$ and B_0 [such as the Priestley–Taylor equation or the reanalysis (Figs. 2b,c)]. While the choices of $RMSE_0$ and B_0 are themselves uncertain, accounting for this additional uncertainty in the analysis would only make it less likely that differences in performance between models are statistically meaningful.

The large uncertainty regarding model performance is an unsatisfying reality of estimating large-scale E . In view of this uncertainty, it is even more prudent to use simple models. Unlike model performance, model complexity can be evaluated with high precision. In this respect, SFE stands out as simpler than most other methods, while maintaining relatively low errors. The Budyko methods are simpler, but only apply to

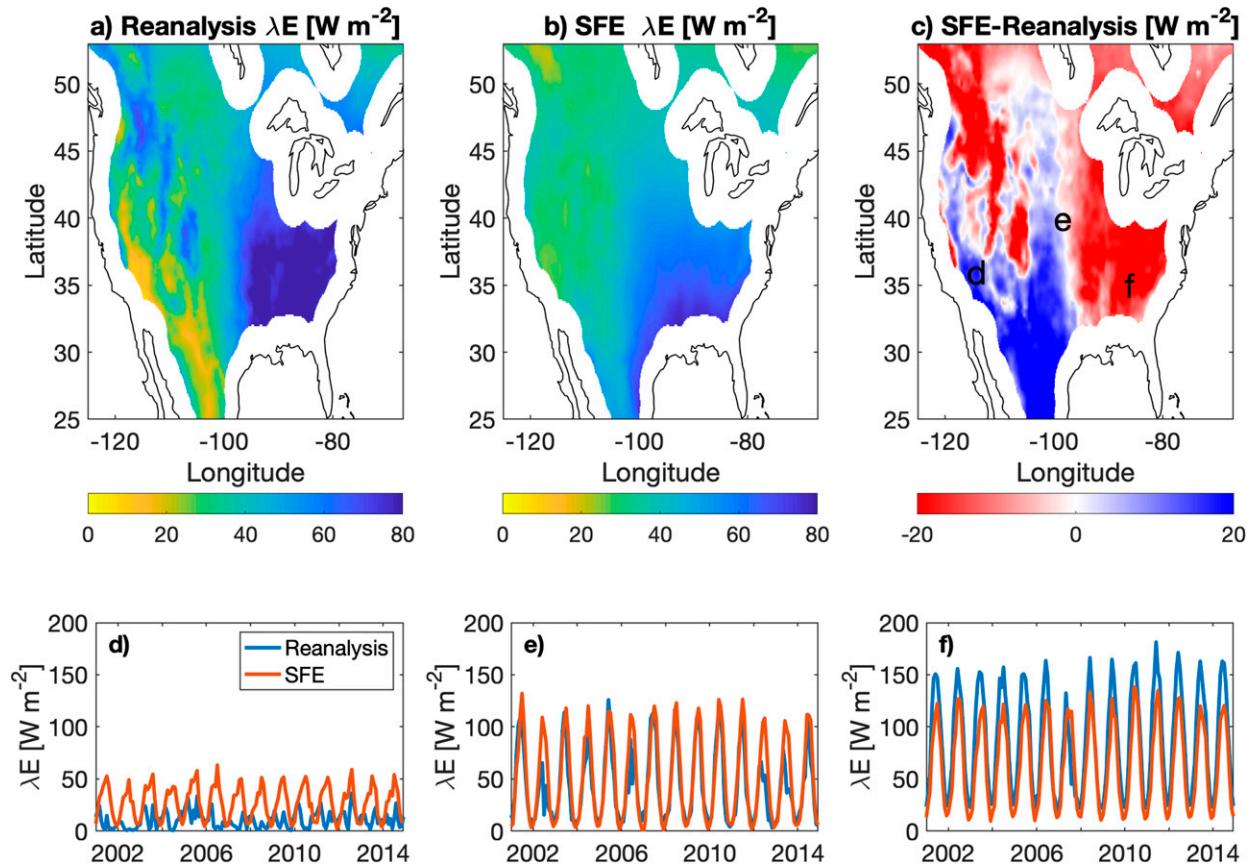


FIG. 3. Comparison over North America between NARR reanalysis λE and λE estimated using SFE with NARR air temperature, NARR specific humidity, and CERES net radiation as inputs. (a)–(c) Maps show multiyear annual mean λE , with regions within 250 km of a coast or large waterbody removed, since SFE is not expected to hold in coastal regions. (d)–(f) Time series are presented for three locations (grid cells in the NARR reanalysis), identified on the map in (c). For reference, $1 \text{ W m}^{-2} = 12.872 \text{ mm yr}^{-1}$.

multiyear mean annual time scales, whereas SFE is applicable at daily to monthly time scales (McCull and Rigden 2020). Priestley–Taylor E is also simpler than SFE, and also applies at daily to monthly time scales, but its performance is much worse than that of both SFE, and that of a perfect model ($\text{RMSE} > \text{RMSE}_0$, $B > B_0$).

In summary, these results demonstrate that SFE is able to estimate multiyear mean annual E quite well at spatial scales relevant to climate studies. Errors in SFE estimates are comparable to errors in the catchment water balance estimates themselves. This work builds on previous work that established SFE is also relatively accurate at estimating daily to monthly E at the quasi-point scale provided by eddy covariance data (McCull and Rigden 2020).

b. Evaluation of CMIP6 model output at the catchment scale

Overall, SFE estimates of λE are typically at least as accurate (when forced with climate model T , q , and R_n) compared with the climate model's own simulated λE , where the catchment water balance is used as a reference dataset (Figs. 4–6). The climate model λE RMSE is distinguishable from that of a perfect model (greater than RMSE_0) for 73% of CMIP6

models, whereas the equivalent SFE estimates are typically below the threshold ($\text{RMSE} > \text{RMSE}_0$ for 8% of CMIP6 models). The thresholds RMSE_0 and B_0 vary between CMIP6 models because different catchments are used to estimate RMSE and B for each model. For mean bias B , the CMIP6 model λE mean bias is mostly indistinguishable from that of a perfect model, with $B > B_0$ for only 19% of models. None of the SFE estimates exhibit a mean bias distinguishable from that of a perfect model.

Repeating the analysis, but restricting it to a common set of catchments results in $N = 8$ catchments. This allows a fair comparison between CMIP6 models, but at the cost of reducing the precision of the estimates of RMSE and B due to the low sample size. The results are broadly similar for both analyses, although with some differences (Fig. 7): $\text{RMSE} > \text{RMSE}_0$ for 92% of models, compared to 53% when using SFE estimates; and $B > B_0$ for 27% of models, compared to 4% when using SFE estimates. Since the focus of this study is on SFE, rather than differences between CMIP6 models, we focus on results in Fig. 6, rather than Fig. 7 for the remainder of the manuscript.

These results demonstrate that SFE is recreated quite well within a broad range of climate models. Surprisingly, in many cases, the SFE λE estimate, forced with climate model T , q , and

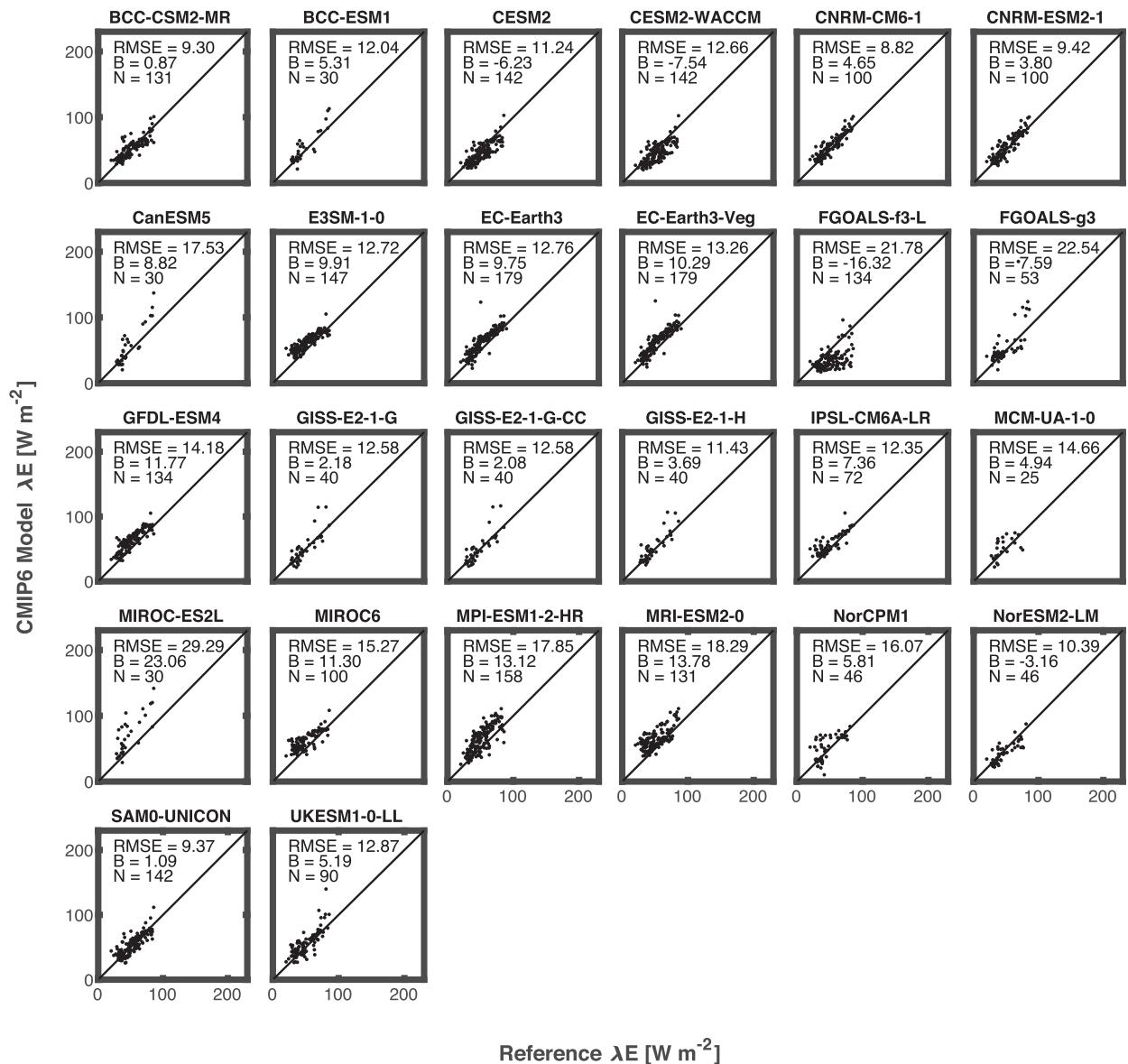


FIG. 4. Comparison between multiyear mean annual CMIP6 model λE and reference λE for 16 different CMIP6 models. Each plot corresponds to the multiyear (2001–14) mean annual E at one catchment. RMSE is root-mean-square error; B is mean bias, and N is the number of catchments used in each comparison. Solid diagonal lines are 1:1 lines. CMIP6 model names are listed above each plot. For reference, $1 \text{ W m}^{-2} = 12.872 \text{ mm yr}^{-1}$.

R_n , exhibits better error statistics than the climate model's own simulated λE . More specifically, the SFE λE estimate has lower RMSE compared with the climate model's simulated λE for 100% of models, using the results presented in Fig. 6, and for 85% of models, using the results presented in Fig. 7. This is likely due to several factors. First, the difference in error statistics is not necessarily statistically significant, given uncertainties in the catchment water balance ET estimate mentioned previously. Second, it is possible that near-surface atmospheric variables, used as inputs to the SFE estimate of λE , may have received more attention in model development, compared with surface fluxes, such as λE , since measurements of near-surface atmospheric variables are

more common than those for surface fluxes. If near-surface atmospheric quantities are relatively more accurate in models, compared with surface fluxes, this may partially explain the superior performance of the SFE estimate in some cases. Third, averaging over the period 2001–14 is an imperfect estimate of the climatology that may neglect, for example, some decadal variability, and it is possible that this may bias the results.

c. Global spatial comparison between SFE and CMIP6 model output

To further evaluate the extent to which SFE is reproduced in CMIP6 models, the model simulated E is compared directly

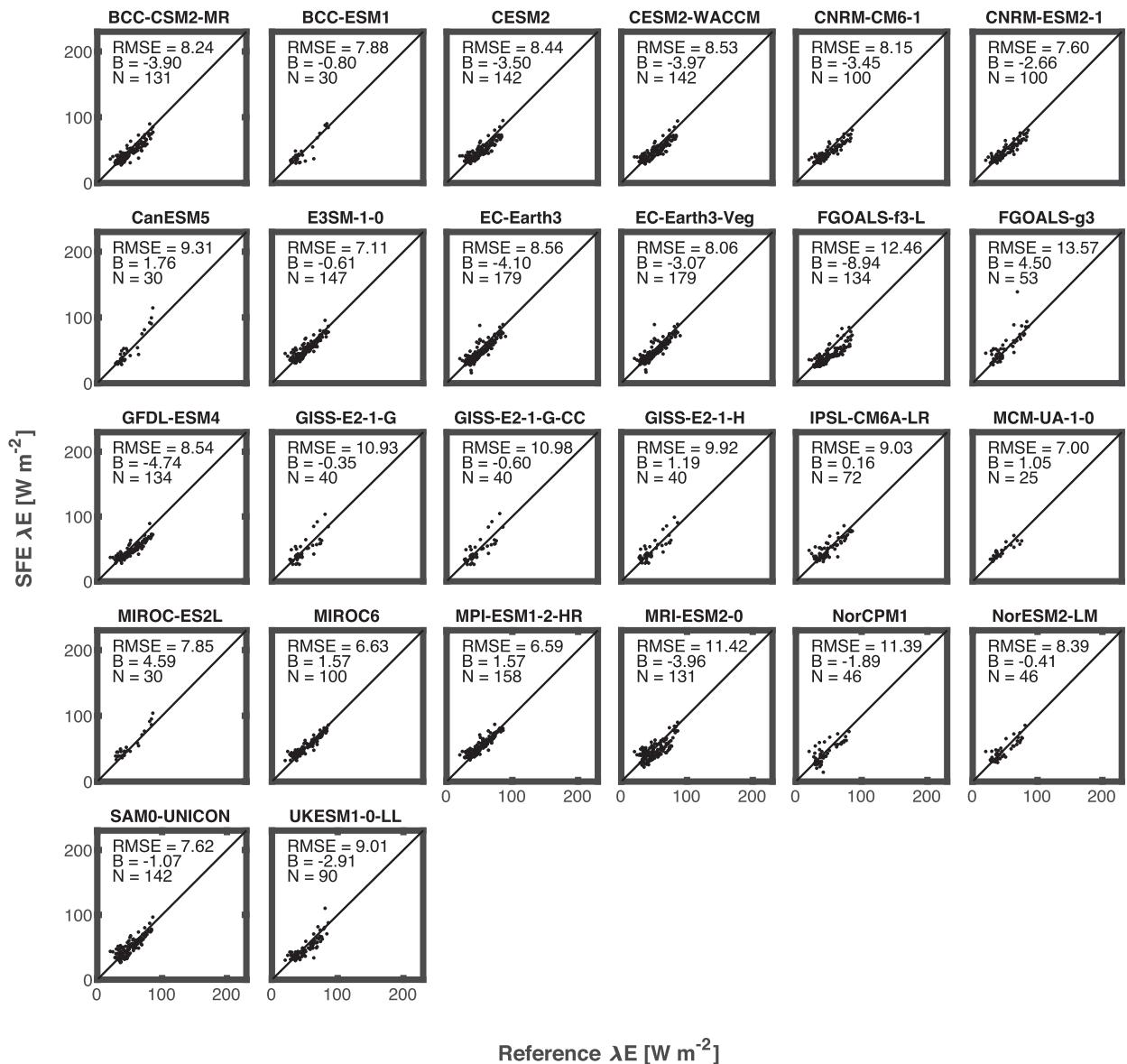


FIG. 5. As in Fig. 4, but comparing multiyear mean annual SFE estimates to the reference estimate, where SFE estimates are obtained using CMIP6 model outputs of near-surface air temperature, specific humidity, and net radiation as inputs to Eq. (1). For reference, $1 \text{ W m}^{-2} = 12.872 \text{ mm yr}^{-1}$.

with E estimated within the model using SFE. Specifically, maps of multimodel mean annual λE over the period 2001–14 are estimated across all land areas, using all 26 available CMIP6 models (Fig. 8a). Then, for each CMIP6 model, Eq. (1) is also used to estimate λE , using CMIP6 model outputs of q , T , and R_n . The resulting multimodel mean annual estimate obtained from SFE is presented in Fig. 8b. Differences in the SFE estimates compared with the CMIP6 model estimates are quantified in terms of mean bias (Fig. 8c) and RMSE (Fig. 8d).

Differences between the two estimates may be due to deficiencies in the SFE theory in particular regions, or to deficiencies in the CMIP6 models. Like other E models (e.g., Salvucci and Gentile 2013; Miralles et al. 2016), SFE tends to

overestimate in regions where λE is particularly low, and underestimate in regions where λE is particularly high (McColl and Rigden 2020). This is consistent with positive biases in SFE over the Sahara, the Middle East, and most of Australia (Fig. 8c). There are limited ground observations of E in these regions, and none at the scales relevant to climate models. However, it has been shown that SFE does systematically overestimate λE when compared with eddy covariance observations during the dry season in northern Australia (McColl and Rigden 2020). On the other hand, the land-atmosphere feedback mechanisms that are essential to SFE have known deficiencies in climate models (e.g., Green et al. 2017). Overall, we are not able to definitively adjudicate between the two

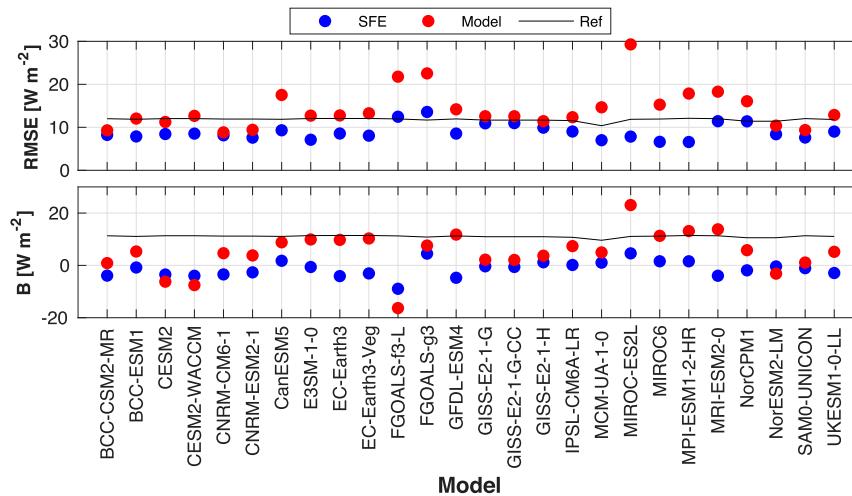


FIG. 6. Summary of statistical metrics presented in Figs. 4 and 5. RMSE is root-mean-square error; B is mean bias. Ref is the expected error due to errors in the catchment water balance estimates, and differs between models because a different set of catchments was used in each comparison, depending on each model's spatial resolution. For reference, $1 \text{ W m}^{-2} = 12.872 \text{ mm yr}^{-1}$.

explanations. It is likely that they both contribute to the observed differences. Differences between the SFE and CMIP6 model estimates are relatively small, as a fraction of the total λE (Fig. 8d compared with Fig. 8a), at least in regions where most of the world's E occurs.

d. Temporal comparison between SFE and CMIP6 model output at focus sites

To assess the degree to which SFE accurately represents temporal variability in E , monthly time series of CMIP6 model-simulated λE are compared with equivalent SFE estimates at eight “focus sites” (Fig. 9). The focus sites

are chosen to span a broad range of continents, ecosystem types and annual mean E . In nonarid regions (Figs. 9a–f), the SFE time series estimates agree well with the CMIP6 simulations. At the two arid sites in the Sahara and central Australia (Figs. 9g,h), SFE consistently overestimates λE relative to the CMIP6 simulations. Overall, these results demonstrate that, outside the most arid regions, the SFE estimates reproduce the seasonal dynamics of the CMIP6 models' simulated λE quite well. This significantly differentiates SFE from the Budyko model described in section 2, which is not capable of estimating λE at monthly time scales.

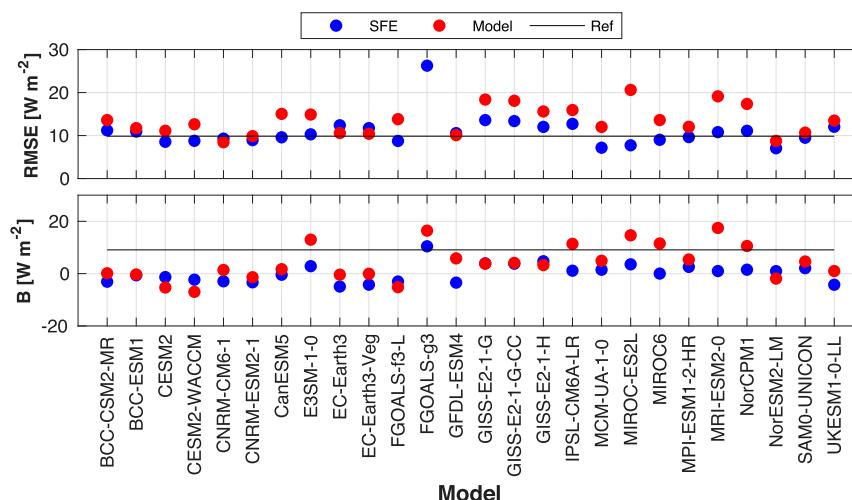


FIG. 7. As in Fig. 6, but using a consistent set of eight catchments to estimate validation statistics for each model. The RMSE for FGOALS-f3-L is 41.9 W m^{-2} ; this value is not visible on these axes, which are kept the same as Fig. 6 for easier comparison. For reference, $1 \text{ W m}^{-2} = 12.872 \text{ mm yr}^{-1}$.

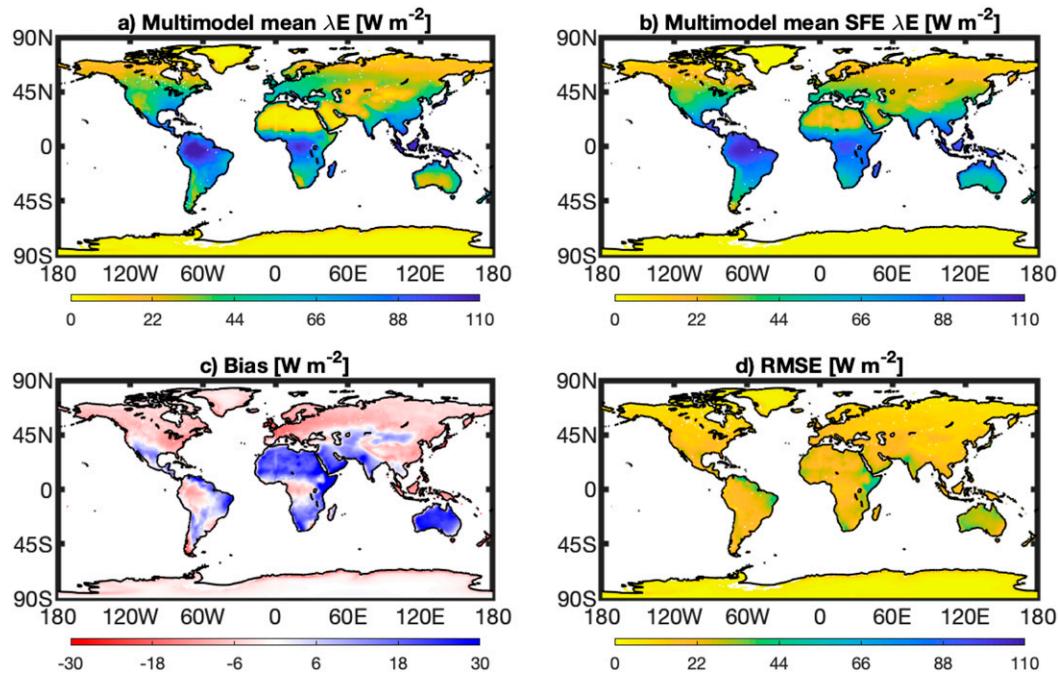


FIG. 8. Maps of CMIP6 multimodel, multiyear mean annual (a) λE ; (b) SFE λE , estimated using CMIP6 model near-surface air temperature, specific humidity, and net radiation as inputs; (c) mean bias (SFE λE minus CMIP6 model λE); and (d) RMSE between SFE and CMIP6 model λE . The same color bar scale is used for (a), (b), and (d) to allow visual comparison of RMSE and mean λE . For reference, $1 \text{ W m}^{-2} = 12.872 \text{ mm yr}^{-1}$.

e. Limitations

This section discusses limitations of our analysis, beyond those already discussed in previous sections. SFE is not expected to hold at subdaily time scales, since the atmosphere does not respond instantaneously to changes in surface fluxes; at subdaily time scales, E must still be estimated implicitly by numerically solving the surface energy budget coupled with diffusive expressions for surface fluxes, or by using explicit, approximate solutions of the same set of equations (Monteith 1965; Penman 1948; McColl 2020). SFE is also not expected to hold outside inland continental regions. In coastal land regions, moisture and heat convergence contribute significantly to the near-surface relative humidity budget, in violation of the assumptions of SFE. In addition, it is not expected that SFE will hold over oceans, where the dominant terms in the boundary layer moisture and temperature budgets are fundamentally different to those over land. Furthermore, we do not necessarily expect that SFE will continue to hold in substantially different past or future climates. Even in inland continental regions, SFE systematically overestimates E somewhat when E is low, and underestimates it when E is high in both climate models (Fig. 8c) and observations (McColl and Rigden 2020), although such biases are observed in more complex models, too (Salvucci and Gentine 2013; Miralles et al. 2016). Nevertheless, SFE clearly explains much of the observed spatial and temporal variability in λE within both existing observations (Fig. 2; McColl and Rigden 2020) and CMIP6 models (Figs. 8 and 9).

f. Relation to previous work

In an important early study, Barton (1979) proposed that the latent heat flux from an unsaturated surface could be estimated using the equation

$$\lambda E = \alpha \frac{\sigma \varepsilon}{\sigma \varepsilon + 1} R_n, \quad (7)$$

where σ is the ratio of the vapor pressure to the saturated vapor pressure evaluated at the evaporating surface [this corresponds to Eq. (7) of Barton (1979), but with ground heat flux neglected, for consistency with assumptions used in this study]. This relation is remarkably similar to Eq. (1), which can be written as

$$\lambda E = \frac{\text{RH} \varepsilon}{\text{RH} \varepsilon + 1} R_n, \quad (8)$$

where RH is the near-surface relative humidity, typically estimated several meters above the evaporating surface. However, Barton's equation differs from Eq. (1) in several respects. First, it retains the empirical Priestley–Taylor parameter, whereas SFE discards it. Second, it uses the quantity σ rather than RH, and these two quantities can differ substantially. Barton specifically acknowledged this point in his study: “With nonsaturated surfaces the value of σ can be vastly different to the relative humidity just above the surface; thus to avoid confusion, σ is not referred to as the surface relative humidity” (italics from the original). Third, σ is a property of the evaporating surface, meaning that land surface information is required to estimate it (Barton relates it to

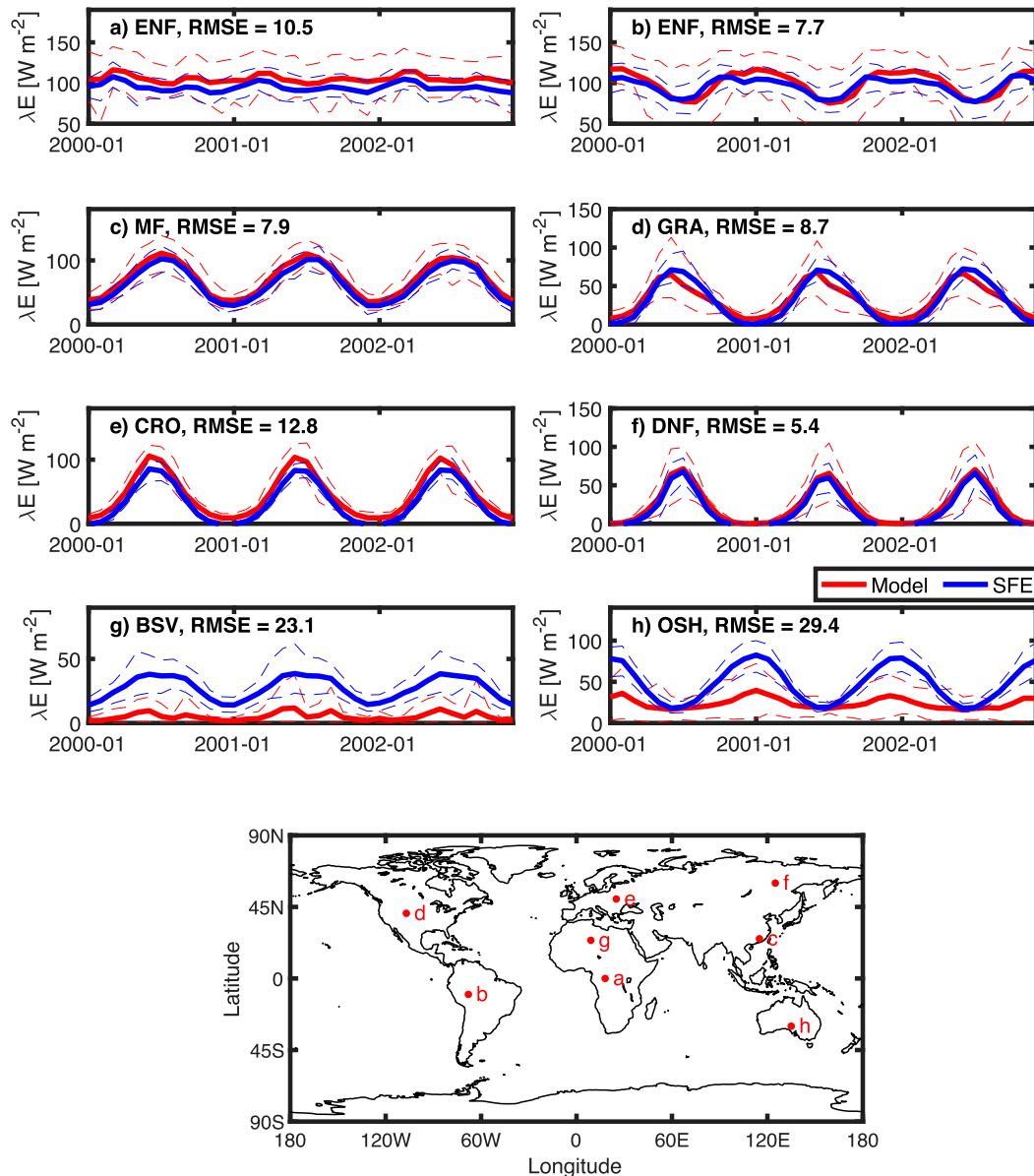


FIG. 9. Time series (2000–02) at eight sites of multimodel mean λE obtained directly from CMIP6 models (red) and from SFE (blue), using CMIP6 model near-surface air temperature, specific humidity, and net radiation as inputs. Locations of each site are shown on the map. Dashed lines are 90% confidence intervals. RMSE is root-mean-square error between CMIP6 models and SFE. Land cover types for each site are displayed in the top-left corner of each subplot: ENF is evergreen needleleaf forest, MF is mixed forest, GRA is grasslands, DNF is deciduous needleleaf forest, CRO is cropland, BSV is barren/sparse vegetation, and OSH is open shrubland. For reference, $1 \text{ W m}^{-2} = 12.872 \text{ mm yr}^{-1}$.

soil moisture), whereas no land surface information is required by SFE. The key insight of SFE is that land–atmosphere feedbacks embed information on water limitation occurring at the evaporating surface in the near-surface atmospheric state. Thus, while the two equations appear similar, the differences between them are rather significant, both in terms of their practical application, and in terms of their governing physics.

A similar class of approaches based on the “complementary relationship” is also capable of estimating E using minimal

land surface information. While there are many variants of the complementary relationship found in the literature, to our knowledge, none are as simple as SFE: in addition to inputs required by SFE, they also either require a parameter (e.g., Kahler and Brutsaert 2006), or land surface information such as soil moisture (e.g., Aminzadeh et al. 2016), or wind speed (e.g., Ma and Szilagyi 2019).

A further advantage of SFE is that it can be used to estimate the evaporative fraction ($EF = \lambda E/R_n$) without requiring R_n as

an input, simply by dividing Eq. (1) by R_n . The evaporative fraction is a useful diagnostic of surface energy balance partitioning (e.g., Gentile et al. 2011). In contrast, the Budyko relations [Eqs. (3) and (4)] cannot be written in this way: dividing these equations by R_n does not eliminate dependence on R_n on the right-hand side of the equation. This property of SFE is useful because R_n is the variable for which fewest observations are typically available. This limitation has motivated the recent development of a variant of the Budyko relation that is capable of estimating EF without requiring R_n (Yin et al. 2019). Like other versions of the Budyko relation, this relation applies to multiyear mean annual values. In contrast, SFE also holds at daily and monthly time scales.

4. Summary and conclusions

This study has tested the empirical validity of SFE—a maximally simple model of the Bowen ratio in continental regions at daily to monthly time scales—at large spatial scales relevant to studies of climate. The only inputs required by the model are near-surface air temperature and specific humidity; information on land surface constraints is assumed to become embedded in the near-surface atmospheric state by strong land–atmosphere coupling. By combining this prediction with observations of net radiation, estimates of λE are obtained.

The accuracy of the SFE λE estimates was evaluated by comparing with estimates of multiyear mean annual λE obtained from catchment water balances for 221 catchments across the United States. While catchment water balance estimates of multiyear mean annual λE are widely regarded as one of the more accurate estimates of large-scale λE , they contain their own errors that can be significant; hence, even a perfect model would display errors when evaluated against the catchment water balance λE estimate. By using a reasonable estimate of errors inherent to the catchment water balance λE , it was shown that the SFE λE estimate displayed error statistics that were indistinguishable from those of a perfect model. Two different functional forms of the Budyko relation were similarly indistinguishable from a perfect model. On the other hand, estimates from the Priestley–Taylor equation and a reanalysis were not. While the poorer performance of the Priestley–Taylor equation was expected, since it applies to saturated surfaces and most catchments are not saturated, it demonstrated that our comparison was capable of broadly distinguishing between accurate and inaccurate models of λE .

After establishing the reasonable performance of SFE at large scales, we evaluated its performance in 26 climate models from CMIP6 historical simulations. Model SFE estimates [obtained using Eq. (1) with CMIP6 model outputs of q , T , and R_n] were typically at least as accurate as the model's own λE output, when evaluated against catchment water balance λE estimates. The model's SFE estimate and its simulated λE output displayed broad agreement both in space and time. An exception to this was found in very dry regions, such as the Sahara, central Australia, and parts of the Middle East, where the SFE estimate was systematically higher. This difference may be due to errors in the SFE estimate, or in the climate models, or a combination of both.

Future studies should investigate the degree to which SFE holds in past and future climates. If SFE holds reasonably, it may be possible to reconstruct time series of the Bowen ratio using readily available weather station data, which have more substantial spatial and temporal coverage compared with eddy covariance data. It may also be possible to combine proxy records of temperature and humidity with SFE to reconstruct the Bowen ratio further back in time, to better constrain the terrestrial water cycle on paleoclimatic time scales. Comparisons between SFE E estimates and a broader set of reanalysis and satellite E products may also be useful.

Acknowledgments. The China Scholarship Council funded S.C.'s visit to K.A.M.'s group at Harvard. K.A.M. acknowledges funding from a Winokur Seed Grant in the Environmental Sciences from the Harvard University Center for the Environment, and from the Dean's Competitive Fund for Promising Scholarship from Harvard University. We thank Dan Chavas for making his `dist_from_coast` code publicly available, and three anonymous reviewers for providing feedback on the manuscript. All data used in this study are publicly available. Catchment boundary data from the MOPEX experiment (Duan et al. 2006) are available from https://www.nws.noaa.gov/ohd/mopex/mo_datasets.htm. Surface net radiation observations from the CERES product (Kato et al. 2018) are available from <https://ceres.larc.nasa.gov/data/>. Streamflow data are available from the USGS at <http://waterdata.usgs.gov/usa/nwis/sw>. The PRISM precipitation data (Daly et al. 1994) are available at <http://prism.oregonstate.edu>. Data from the NARR (Mesinger et al. 2006) are available at <https://psl.noaa.gov/data/gridded/data.narr.monolevel.html>. The CMIP6 (Eyring et al. 2016) model outputs are available at <https://esgf-node.llnl.gov/search/cmip6/>.

REFERENCES

- Aminzadeh, M., M. L. Roderick, and D. Or, 2016: A generalized complementary relationship between actual and potential evaporation defined by a reference surface temperature. *Water Resour. Res.*, **52**, 385–406, <https://doi.org/10.1002/2015WR017969>.
- Baldocchi, D. D., R. J. Luxmoore, and J. L. Hatfield, 1991: Discerning the forest from the trees: An essay on scaling canopy stomatal conductance. *Agric. For. Meteorol.*, **54**, 197–226, [https://doi.org/10.1016/0168-1923\(91\)90006-C](https://doi.org/10.1016/0168-1923(91)90006-C).
- Barton, I. J., 1979: A parameterization of the evaporation from non-saturated surfaces. *J. Appl. Meteorol.*, **18**, 43–47, [https://doi.org/10.1175/1520-0450\(1979\)018<0043:APOTEF>2.0.CO;2](https://doi.org/10.1175/1520-0450(1979)018<0043:APOTEF>2.0.CO;2).
- Beck, H. E., E. F. Wood, T. R. McVicar, M. Zambrano-Bigiarini, C. Alvarez-Garreton, O. M. Baez-Villanueva, J. Sheffield, and D. N. Karger, 2020: Bias correction of global high-resolution precipitation climatologies using streamflow observations from 9372 catchments. *J. Climate*, **33**, 1299–1315, <https://doi.org/10.1175/JCLI-D-19-0332.1>.
- Berry, J. A., D. J. Beerling, and P. J. Franks, 2010: Stomata: Key players in the Earth system, past and present. *Curr. Opin. Plant Biol.*, **13**, 232–239, <https://doi.org/10.1016/j.pbi.2010.04.013>.
- Betts, A. K., 2000: Idealized model for equilibrium boundary layer over land. *J. Hydrometeorol.*, **1**, 507–523, [https://doi.org/10.1175/1525-7541\(2000\)001<0507:IMFEBL>2.0.CO;2](https://doi.org/10.1175/1525-7541(2000)001<0507:IMFEBL>2.0.CO;2).

- Bouchet, R., 1963: Evapotranspiration réelle, évapotranspiration potentielle, et production agricole. *Ann. Agron.*, **14**, 743–824.
- Bou-Zeid, E., W. Anderson, G. G. Katul, and L. Mahrt, 2020: The persistent challenge of surface heterogeneity in boundary-layer meteorology: A review. *Bound.-Layer Meteor.*, **177**, 227–245, <https://doi.org/10.1007/s10546-020-00551-8>.
- Brutsaert, W., 1998: Land-surface water vapor and sensible heat flux: Spatial variability, homogeneity, and measurement scales. *Water Resour. Res.*, **34**, 2433–2442, <https://doi.org/10.1029/98WR01340>.
- , 2015: A generalized complementary principle with physical constraints for land-surface evaporation. *Water Resour. Res.*, **51**, 8087–8093, <https://doi.org/10.1002/2015WR017720>.
- , and H. Stricker, 1979: An advection-aridity approach to estimate actual regional evapotranspiration. *Water Resour. Res.*, **15**, 443–450, <https://doi.org/10.1029/WR015i002p00443>.
- , and M. B. Parlange, 1998: Hydrologic cycle explains the evaporation paradox. *Nature*, **396**, 30, <https://doi.org/10.1038/23845>.
- Budyko, M. I., 1958: The heat balance of the Earth's surface. Weather Bureau Doc., 259 pp.
- Culf, A. D., 1994: Equilibrium evaporation beneath a growing convective boundary layer. *Bound.-Layer Meteor.*, **70**, 37–49, <https://doi.org/10.1007/BF00712522>.
- Daly, C., R. P. Neilson, and D. L. Phillips, 1994: A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *J. Appl. Meteor.*, **33**, 140–158, [https://doi.org/10.1175/1520-0450\(1994\)033<0140:ASTMFM>2.0.CO;2](https://doi.org/10.1175/1520-0450(1994)033<0140:ASTMFM>2.0.CO;2).
- De Bruin, H. R., 1983: A model for the Priestley-Taylor parameter. *J. Climate Appl. Meteor.*, **22**, 572–578, [https://doi.org/10.1175/1520-0450\(1983\)022<0572:AMFTPT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1983)022<0572:AMFTPT>2.0.CO;2).
- DeLucia, E. H., and Coauthors, 2019: Are we approaching a water ceiling to maize yields in the United States? *Ecosphere*, **10**, e02773, <https://doi.org/10.1002/ecs2.2773>.
- Duan, Q., and Coauthors, 2006: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *J. Hydrol.*, **320**, 3–17, <https://doi.org/10.1016/j.jhydrol.2005.07.031>.
- Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, <https://doi.org/10.5194/gmd-9-1937-2016>.
- Fisher, J. B., K. P. Tu, and D. D. Baldocchi, 2008: Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sens. Environ.*, **112**, 901–919, <https://doi.org/10.1016/j.rse.2007.06.025>.
- Gentine, P., D. Entekhabi, and J. Polcher, 2011: The diurnal behavior of evaporative fraction in the soil-vegetation-atmospheric boundary layer continuum. *J. Hydrometeorol.*, **12**, 1530–1546, <https://doi.org/10.1175/2011JHM1261.1>.
- Green, J. K., A. G. Konings, S. H. Alemohammad, J. Berry, D. Entekhabi, J. Kolassa, J.-E. Lee, and P. Gentine, 2017: Regionally strong feedbacks between the atmosphere and terrestrial biosphere. *Nat. Geosci.*, **10**, 410–414, <https://doi.org/10.1038/ngeo2957>.
- Han, E., W. T. Crow, C. R. Hain, and M. C. Anderson, 2015: On the use of a water balance to evaluate interannual terrestrial ET variability. *J. Hydrometeorol.*, **16**, 1102–1108, <https://doi.org/10.1175/JHM-D-14-0175.1>.
- Held, I. M., 2005: The gap between simulation and understanding in climate modeling. *Bull. Amer. Meteor. Soc.*, **86**, 1609–1614, <https://doi.org/10.1175/BAMS-86-11-1609>.
- Jarvis, P. G., and K. G. McNaughton, 1986: Stomatal control of transpiration: Scaling up from leaf to region. *Adv. Ecol. Res.*, **15**, 1–49, [https://doi.org/10.1016/S0065-2504\(08\)60119-1](https://doi.org/10.1016/S0065-2504(08)60119-1).
- Jeevanjee, N., P. Hassanzadeh, S. Hill, and A. Sheshadri, 2017: A perspective on climate model hierarchies. *J. Adv. Model. Earth Syst.*, **9**, 1760–1771, <https://doi.org/10.1002/2017MS001038>.
- Jia, A., S. Liang, B. Jiang X. Zhang, and G. Wang, 2018: Comprehensive assessment of global surface net radiation products and uncertainty analysis. *J. Geophys. Res. Atmos.*, **123**, 1970–1989, <https://doi.org/10.1002/2017JD027903>.
- Jung, M., and Coauthors, 2010: Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, **467**, 951–954, <https://doi.org/10.1038/nature09396>.
- Kahler, D. M., and W. Brutsaert, 2006: Complementary relationship between daily evaporation in the environment and pan evaporation. *Water Resour. Res.*, **42**, W05413, <https://doi.org/10.1029/2005WR004541>.
- Kato, S., and Coauthors, 2018: Surface irradiances of edition 4.0 Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) data product. *J. Climate*, **31**, 4501–4527, <https://doi.org/10.1175/JCLI-D-17-0523.1>.
- Legates, D. R., and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected, global precipitation. *Int. J. Climatol.*, **10**, 111–127, <https://doi.org/10.1002/joc.3370100202>.
- Li, D., and L. Wang, 2019: Sensitivity of surface temperature to land use and land cover change-induced biophysical changes: The scale issue. *Geophys. Res. Lett.*, **46**, 9678–9689, <https://doi.org/10.1029/2019GL084861>.
- Ma, H.-Y., and Coauthors, 2018: CAUSES: On the role of surface energy budget errors to the warm surface air temperature error over the central United States. *J. Geophys. Res. Atmos.*, **123**, 2888–2909, <https://doi.org/10.1002/2017JD027194>.
- Ma, N., and J. Szilagyi, 2019: The CR of evaporation: A calibration-free diagnostic and benchmarking tool for large-scale terrestrial evapotranspiration modeling. *Water Resour. Res.*, **55**, 7246–7274, <https://doi.org/10.1029/2019WR024867>.
- Maher, P., and Coauthors, 2019: Model hierarchies for understanding atmospheric circulation. *Rev. Geophys.*, **57**, 250–280, <https://doi.org/10.1029/2018RG000607>.
- Mahrt, L., 2000: Surface heterogeneity and vertical structure of the boundary layer. *Bound.-Layer Meteor.*, **96**, 33–62, <https://doi.org/10.1023/A:1002482332477>.
- Martens, B., and Coauthors, 2017: GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.*, **10**, 1903–1925, <https://doi.org/10.5194/gmd-10-1903-2017>.
- McCull, K. A., 2020: Practical and theoretical benefits of an alternative to the Penman-Monteith evapotranspiration equation. *Water Resour. Res.*, **56**, e2020WR027106, <https://doi.org/10.1029/2020WR027106>.
- , and A. J. Rigden, 2020: Emergent simplicity of continental evapotranspiration. *Geophys. Res. Lett.*, **47**, e2020GL087101, <https://doi.org/10.1029/2020GL087101>.
- , G. D. Salvucci, and P. Gentine, 2019: Surface flux equilibrium theory explains an empirical estimate of water-limited daily evapotranspiration. *J. Adv. Model. Earth Syst.*, **11**, 2036–2049, <https://doi.org/10.1029/2019MS001685>.
- McNaughton, K. G., and T. W. Spriggs, 1986: A mixed-layer model for regional evaporation. *Bound.-Layer Meteor.*, **34**, 243–262, <https://doi.org/10.1007/BF00122381>.
- Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, <https://doi.org/10.1175/BAMS-87-3-343>.

- Milly, P. C. D., and K. A. Dunne, 2002: Macroscale water fluxes 1. Quantifying errors in the estimation of basin mean precipitation. *Water Resour. Res.*, **38**, 1205, <https://doi.org/10.1029/2001WR000759>.
- Miralles, D. G., and Coauthors, 2016: The WACMOS-ET project – Part II: Evaluation of global terrestrial evaporation data sets. *Hydrol. Earth Syst. Sci.*, **20**, 823–842, <https://doi.org/10.5194/hess-20-823-2016>.
- Monteith, J. L., 1965: Evaporation and environment. *The State and Movement of Water in Living Organisms*, G. Fogg, Ed., Cambridge University Press, 205–234.
- Morton, F. I., 1969: Potential evaporation as a manifestation of regional evaporation. *Water Resour. Res.*, **5**, 1244–1255, <https://doi.org/10.1029/WR005i006p01244>.
- Mueller, B., and S. I. Seneviratne, 2014: Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophys. Res. Lett.*, **41**, 128–134, <https://doi.org/10.1002/2013GL058055>.
- , and Coauthors, 2013: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. *Hydrol. Earth Syst. Sci.*, **17**, 3707–3720, <https://doi.org/10.5194/hess-17-3707-2013>.
- Oki, T., and S. Kanae, 2006: Global hydrological cycles and world water resources. *Science*, **313**, 1068–1072, <https://doi.org/10.1126/science.1128845>.
- Penman, H., 1948: Natural evaporation from open water, bare soil and grass. *Proc. Roy. Soc. London*, **193A**, 120–145, <https://doi.org/10.1098/rspa.1948.0037>.
- Priestley, C. H. B., and R. J. Taylor, 1972: On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Wea. Rev.*, **100**, 81–92, [https://doi.org/10.1175/1520-0493\(1972\)100<0081:OTAOSH>2.3.CO;2](https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2).
- Raupach, M. R., 2000: Equilibrium evaporation and the convective boundary layer. *Bound.-Layer Meteorol.*, **96**, 107–142, <https://doi.org/10.1023/A:1002675729075>.
- , 2001: Combination theory and equilibrium evaporation. *Quart. J. Roy. Meteor. Soc.*, **127**, 1149–1181, <https://doi.org/10.1002/qj.49712757402>.
- , and J. J. Finnigan, 1995: Scale issues in boundary-layer meteorology: Surface energy balances in heterogeneous terrain. *Hydrol. Processes*, **9**, 589–612, <https://doi.org/10.1002/hyp.3360090509>.
- Rigden, A. J., and G. D. Salvucci, 2015: Evapotranspiration based on equilibrated relative humidity (ETRHEQ): Evaluation over the continental U.S. *Water Resour. Res.*, **51**, 2951–2973, <https://doi.org/10.1002/2014WR016072>.
- , D. Li, and G. D. Salvucci, 2018: Dependence of thermal roughness length on friction velocity across land cover types: A synthesis analysis using AmeriFlux data. *Agric. For. Meteorol.*, **249**, 512–519, <https://doi.org/10.1016/j.agrformet.2017.06.003>.
- Salvucci, G. D., and P. Gentile, 2013: Emergent relation between surface vapor conductance and relative humidity profiles yields evaporation rates from weather data. *Proc. Natl. Acad. Sci. USA*, **110**, 6287–6291, <https://doi.org/10.1073/pnas.1215844110>.
- Shuttleworth, W. J., and I. R. Calder, 1979: Has the Priestley–Taylor equation any relevance to forest evaporation? *J. Appl. Meteorol.*, **18**, 639–646, [https://doi.org/10.1175/1520-0450\(1979\)018<0639:HTPTEA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1979)018<0639:HTPTEA>2.0.CO;2).
- Taylor, C. M., C. E. Birch, D. J. Parker, N. Dixon, F. Guichard, G. Nikulin, and G. M. S. Lister, 2013: Modeling soil moisture–precipitation feedback in the Sahel: Importance of spatial scale versus convective parameterization. *Geophys. Res. Lett.*, **40**, 6213–6218, <https://doi.org/10.1002/2013GL058511>.
- Trugman, A. T., D. Medvigy, J. S. Mankin, and W. R. L. Anderegg, 2018: Soil moisture stress as a major driver of carbon cycle uncertainty. *Geophys. Res. Lett.*, **45**, 6495–6503, <https://doi.org/10.1029/2018GL078131>.
- Vinukollu, R. K., E. F. Wood, C. R. Ferguson, and J. B. Fisher, 2011: Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches. *Remote Sens. Environ.*, **115**, 801–823, <https://doi.org/10.1016/j.rse.2010.11.006>.
- Yang, H., D. Yang, Z. Lei, and F. Sun, 2008: New analytical derivation of the mean annual water-energy balance equation. *Water Resour. Res.*, **44**, W03410, <https://doi.org/10.1029/2007WR006135>.
- Yin, J., S. Calabrese, E. Daly, and A. Porporato, 2019: The energy side of Budyko: Surface-energy partitioning from hydrological observations. *Geophys. Res. Lett.*, **46**, 7456–7463, <https://doi.org/10.1029/2019GL083373>.