

Water Resources Research



RESEARCH ARTICLE

10.1029/2019WR024867

Key Points:

- Nine gridded ET products were evaluated against water-balance estimates of the mean—and linear trend in—annual ET rates over HUC2 and HUC6 basins of the conterminous United States
- The calibration-free CR with only minimal meteorological data requirement performed the best overall
- Due to its performance, data efficiency, and simple structure the calibration-free CR could serve as a benchmarking and diagnostic tool for more complex models

Supporting Information:

- Supporting Information S1

Correspondence to:

J. Szilagyi,
jszilagy1@unl.edu

Citation:

Ma, N., & Szilagyi, J. (2019). The CR of evaporation: A calibration-free diagnostic and benchmarking tool for large-scale terrestrial evapotranspiration modeling. *Water Resources Research*, 55. <https://doi.org/10.1029/2019WR024867>

Received 28 JAN 2019

Accepted 27 JUL 2019

Accepted article online 1 AUG 2019

The CR of Evaporation: A Calibration-Free Diagnostic and Benchmarking Tool for Large-Scale Terrestrial Evapotranspiration Modeling

Ning Ma^{1,2} and Jozsef Szilagyi^{3,4}

¹Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China, ²RCE-TEA, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, ³Department of Hydraulic and Water Resources Engineering, Budapest University of Technology and Economics, Budapest, Hungary, ⁴Conservation and Survey Division, School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, USA

Abstract Monthly evapotranspiration (ET) rates for 1979–2015 were estimated by the latest, calibration-free version of the complementary relationship (CR) of evaporation over the conterminous United States. The results were compared to similar estimates of three land surface models (Noah, VIC, Mosaic), two reanalysis products (National Centers of Environmental Protection Reanalysis II, ERA-Interim), two remote-sensing-based (Global Land Evaporation Amsterdam Model, Penman-Monteith-Leuning) algorithms, and the spatially upscaled eddy-covariance ET measurements of FLUXNET-MTE. Model validations were performed via simplified water-balance derived ET rates employing Parameter-Elevation Regressions on Independent Slopes Model precipitation, United States Geological Survey two- and six-digit Hydrologic Unit Code (HUC2 and HUC6) discharge, and terrestrial water storage anomalies from Gravity Recovery and Climate Experiment, the latter for 2003–2015. The CR outperforms all other multiyear mean annual HUC2-averaged ET estimates with root-mean-square error = 51 mm/year, $R = 0.98$, relative bias of -1% , and Nash-Sutcliffe efficiency = 0.94, respectively. Inclusion of the Gravity Recovery and Climate Experiment data into the annual water balances for the shorter 2003–2015 period does not have much effect on model performance. Similarly, the CR outperforms all other models for the linear trend of the annual ET rates over the HUC2 basins. Over the significantly smaller HUC6 basins where the water-balance validation is more uncertain, the CR still outperforms all other models except FLUXNET-MTE, which has the advantage of possible local ET measurements, a benefit that clearly diminishes at the HUC2 scale. As the employed CR is calibration-free and requires only very few meteorological inputs, yet it yields superior ET performance at the regional scale, it may serve as a diagnostic and benchmarking tool for more complex and data intensive models of terrestrial evapotranspiration rates.

1. Introduction

Globally, about 70% of the precipitation over land returns to the atmosphere via terrestrial evapotranspiration (ET; Oki & Kanae, 2006), which at the same time consumes approximately 60% of the energy available at the land surface (Trenberth et al., 2009). The variations in ET are also regarded as an important indicator of hydrological responses to global warming since ET plays a key role in controlling terrestrial water availability and climatic conditions (Seager et al., 2007; Zeng et al., 2017).

While ET can be accurately monitored by a wide variety of ground measurements (e.g., lysimeters, energy balance Bowen ratio, eddy covariance [EC], and scintillometry; Allen et al., 2011), estimation of large-scale, long-term ET remains a difficult task since these in situ techniques typically cover short periods (mainly less than a decade) with limited spatial extent. As a result, a number of approaches employing land surface models (LSM; Cai et al., 2014; Ma et al., 2017; Xia et al., 2016), remote sensing (RS) algorithms (Fisher et al., 2008; Miralles et al., 2011; Mu et al., 2011), data assimilation systems (Lu et al., 2017; Xu et al., 2019), and/or geostatistical upscaling of ground ET measurements (Jung et al., 2011), all with inputs from meteorological as well as surface property data, have been devised to quantify ET over larger areas and longer periods. However, most methods mentioned above require a significant number of soil- and vegetation-related parameters as inputs, which are typically interpolated from limited point-scale

©2019. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

measurements/surveys and/or are retrieved from satellite/airborne observations (Masson et al., 2003), thus leading to additional uncertainties. These uncertainties emerge because (i) ground survey of soil profiles is yet relatively scarce in developing countries (Shangguan et al., 2013; Shi et al., 2004) and the simple average of those small-scale soil features will not result in a value representative at the scale of the modeling grid due to typically considerable spatial variability in them, although high-resolution spatially resolved and continuous soil properties maps become available recently in certain developed countries (Chaney et al., 2019) and (ii) translation of remote sensing signals into vegetation traits has great uncertainties because of the existing challenges in the treatment of confounding factors in spectrum-trait relations (Zurita-Milla et al., 2015). For example, a wide range of evaluations (Dai et al., 2019; Mölders, 2005; Teuling et al., 2009) have identified that different LSMs obviously diverge in accuracy for predicting surface fluxes due to empirical parameters (e.g., physiological, phenological, thermal, hydraulic, radiative) employed to represent multiple vegetation and soil types. Moreover, the poor scalability of the algorithms and the limited consideration of subgrid heterogeneity are also great challenges in LSMs for effectively representing the terrestrial hydrological processes, as was highlighted by Samaniego et al. (2017).

While state-of-the-art LSMs and RS models with advanced parameterizations of the surface fluxes are able to simulate large-scale ET in a satisfactory manner, they typically need detailed soil and vegetation data as inputs (e.g., Martens et al., 2017; Mu et al., 2011; Xia et al., 2012). For circumventing the difficulties in representing substantial heterogeneity in soil type, thickness, layering, vegetation cover, and rooting depth on the regional, continental, and/or global scales, the complementary relationship (CR) of evaporation (Bouchet, 1963) may be an appropriate choice for the estimation of ET because it does not require any surface, soil, or vegetation information as input.

The CR predicts the latent heat flux of the (vegetated or bare) land surface over a suitably long period of time, typically five days or longer (to filter out the effect of passing weather fronts (Morton, 1983) that may temporarily disturb the dynamic balance between land ET rate and the resulting moisture content of the air) by relating the ET rate of a small wet patch (ET_p), typically affected by horizontal energy advection (E_h), to that (ET_w) of a wet land surface of regional extent (thus with vanishing E_h) under the same assumed net surface radiation and wind conditions. For almost half a century following the pioneering work of Bouchet (1963), the CR relationship between three evaporation terms: actual ET, wet patch ET (ET_p), and wet-environment ET (ET_w), was thought to be linear in nature. Inspired by the work of Han et al. (2012) on the boundary conditions of the CR, Brutsaert (2015) formulated a more general, nonlinear version of the complementary relationship of evaporation.

Recently, Szilagyi et al. (2017), building on the latest developments of the nonlinear formulation of the CR by Brutsaert (2015) and Crago et al. (2016), proposed a calibration-free nonlinear CR model with more appropriate physical constraints. While the CR model of Brutsaert (2015) was originally intended to be generally applicable, it still contains a tunable parameter on top of its reliance of accurate plot-scale EC measurements (e.g., Brutsaert et al., 2017; Crago & Qualls, 2018; Han & Tian, 2018; Hu et al., 2018) for the calibration of a second model parameter. In the context of large-sample hydrological studies, however, Gupta et al. (2014) argued that the flexibility and transferability of a model cannot be clarified without evaluations over different regions encompassing various land surface and climatic conditions. This highlights the need of assessing any ET model's performance over multiple regions with hydroclimatic regimes that vary considerably. To this end, an appropriate way of setting the value of the key parameter, the Priestley-Taylor (PT) α (Priestley & Taylor, 1972) of the CR, becomes vital for large-scale model applications because it would be particularly difficult to calibrate this parameter at a grid-cell basis in continental- and/or global-scale applications due primarily to the unavailability of measured ET data on a cell-by-cell basis (Ma et al., 2019). Even if measured ET data from a limited number of locations were available for calibrating this model parameter, it may not be directly transferable to other regions or temporal scales (Clark et al., 2017; Samaniego et al., 2010; Zink et al., 2018). To avoid this problem, Szilagyi et al. (2017) proposed a method of identifying wet cells in regional-to-continental-scale grids of meteorological data for obtaining an appropriate constant value of the PT coefficient without resorting to any ET measurements (or their proxies). By inverting the Priestley-Taylor equation over those wet cells, and having an estimate of the wet surface temperature (Szilagyi, 2014), the PT α could be expressed with the help of temperature and humidity gradients between the wet surface and the air, thus making such a CR model calibration-free for

estimating large-scale ET. See the Appendix for a brief explanation of the CR theory and some of its historically widely used model versions.

While preliminary evaluations indicated that this calibration-free CR model could, though with a minimum number of input variables, yield ET estimates on a par with current LSM outputs (Szilagyi, 2018a, 2018b; Szilagyi et al., 2017), such a conclusion was based on an assessment of only one specific LSM product within the North American Regional Reanalysis data (Mesinger et al., 2006). Therefore, it is not clear whether this calibration-free CR model would perform better (or worse) than other ET products generated by (i) remote sensing (RS) models; (ii) atmospheric reanalysis; (iii) spatial upscaling of EC flux tower measurements; and (iv) other LSMs, for example, those widely employed in the second phase of the North American Land Data Assimilation System (NLDAS-2; Xia et al., 2012). Additionally, the comparisons made by Szilagyi et al. (2017) focused mainly on the mean annual ET rates of the six-digit Hydrologic Unit Code (HUC6) basins, while the CR model's (in long-term studies very important) ability to capture tendencies (e.g., linear trends) in annual ET rates in a wide variety of climates of the conterminous United States (CONUS) still remains largely unknown. In this context, a key question that needs also to be addressed is how well current main-stream ET products and also the CR are able to capture trends in annual ET.

In this study, to fully test whether the latest calibration-free CR model improves upon already existing ET estimates across CONUS, long-term (>30 years) water balance data from 18 two-digit HUC (HUC2) and 334 HUC6 basins were employed to evaluate the CR model when run with monthly aggregated inputs and to compare its performance to that of eight other ET products (i.e., three from LSMs, two from reanalysis, two from RS models, and an upscaling of EC measurements). The comparisons are focused on their skills in describing (i) the multiyear mean annual HUC2- and HUC6-averaged ET rates, (ii) the linear trends in their annual ET rates over the individual model coverage within the 37-year period of 1979–2015, and (iii) interannual variability of basin-averaged ET rates during corresponding periods. A series of statistical metrics were also applied to assess how the CR and the main-stream ET models perform in simulating large-scale, long-term ET rates over the CONUS.

2. Materials and Methods

2.1. Complementary Relationship-Based ET Estimates for 1979–2015

The Szilagyi et al. (2017) formulation of Brutsaert's nonlinear CR model (2015) was employed for large-scale ET estimation across the CONUS, relating two dimensionless evapotranspiration terms in the form

$$y = (2-X)X^2 \quad (1)$$

where y and X are defined as

$$X = \frac{E_p^{\max} - E_p}{E_p^{\max} - E_w} \frac{E_w}{E_p} \quad (2)$$

$$y = \frac{E}{E_p} \quad (3)$$

Here E is the actual, while E_p is the potential ET rate, that is, the evapotranspiration rate of a small wet patch in a drying (i.e., not fully wet) environment, typically expressed by the Penman (1948) equation as

$$E_p = \frac{\Delta(R_n - G)}{(\Delta + \gamma)} + \frac{\gamma f_u (e^* - e)}{(\Delta + \gamma)} \quad (4)$$

where Δ (hPa/°C) is the slope of the saturation vapor pressure curve at air temperature, T (°C), and γ is the psychrometric constant (hPa/°C). R_n and G are the net radiation and soil heat flux into the ground (the latter typically negligible at a monthly scale) in water equivalent of mm/day, respectively. e^* and e [= $e^*(T_d)$] are the saturation and actual vapor pressure of the air (hPa), T_d is the dew-point temperature, and f_u is the wind function containing the 2-m wind speed (u_2 ; m/s), that is,

$$f_u = 0.26(1 + 0.54u_2) \quad (5)$$

E_w is the wet-environment ET rate, observed over a regionally extensive well-watered surface, specified by the Priestley-Taylor equation (Priestley & Taylor, 1972), that is,

$$E_w = \alpha \frac{\Delta(T_w)}{\Delta(T_w) + \gamma} (R_n - G) \quad (6)$$

Note that equation (6) was derived for completely wet environments by Priestley and Taylor (1972), and therefore, Δ should be evaluated at the air temperature, T_w (°C), observed in a wet environment, instead of the typical, drying environment T (Szilagyi, 2014; Szilagyi & Jozsa, 2008). This is important since previous studies have found that the difference between these two may reach 5–10 °C (Huntington et al., 2011; Ma et al., 2015; Ortman, 2009; Szilagyi, 2014). By making use of a mild vertical air temperature gradient (de Vries, 1959; Szilagyi, 2014; Szilagyi & Jozsa, 2009) observable in wet environments (as R_n is consumed predominantly by the latent and not by sensible heat flux), T_w can be approximated by the wet surface temperature, T_{ws} (°C). Note that T_{ws} may still be larger than T when the air is close to saturation, but not T_w , due to the cooling effect of evaporation, and in such cases T_w should be capped by T (Szilagyi, 2014; Szilagyi & Jozsa, 2018). Szilagyi and Schepers (2014) demonstrated that the wet surface temperature is independent of areal extent; thus, T_{ws} can be obtained by iteration from the Bowen ratio (β ; Bowen, 1926) of a small wet patch (assuming that available energy for the wet patch is close to that of the drying surface) for which the Penman equation is valid, that is,

$$\beta = \frac{R_n - G - E_p}{E_p} \approx \gamma \frac{T_{ws} - T}{e^*(T_{ws}) - e^*(T_d)} \quad (7)$$

Here $e^*(T_{ws})$ is the saturation vapor pressure at T_{ws} . Parameter α in equation (6) is the dimensionless Priestley-Taylor coefficient, with typical values from the range of [1.1–1.32] (Morton, 1983). For large-scale model applications when measured ET is lacking for the calibration of α , Szilagyi et al. (2017) proposed a novel method (see Appendix for details) for assigning an appropriate value of α by automatically identifying wet grid cells and utilizing observed gridded T and humidity data over them. The α value of 1.15 derived by Szilagyi (2018b) for the conterminous United States was retained for the present CONUS-wide ET simulation.

E_p^{\max} is the maximum value that E_p can, in theory, reach during a complete dry-out (i.e., when e_a is negligible) of the land surface, that is,

$$E_p^{\max} = \frac{\Delta(T_{\text{dry}})}{\Delta(T_{\text{dry}}) + \gamma} (R_n - G) + \frac{\gamma}{\Delta(T_{\text{dry}}) + \gamma} f_u e^*(T_{\text{dry}}) \quad (8)$$

in which T_{dry} (°C) is the dry-environment air temperature. The latter can be estimated from the adiabat of an air parcel in contact with the drying surface under constant $R_n - G$ (Szilagyi, 2018a), that is,

$$T_{\text{dry}} = T_{\text{wb}} + \frac{e^*(T_{\text{wb}})}{\gamma} \quad (9)$$

where T_{wb} (°C) is the wet-bulb temperature. T_{wb} is derived by another iteration of writing out the Bowen ratio for adiabatic changes (Szilagyi, 2014) as

$$\gamma \frac{T_{\text{wb}} - T}{e^*(T_{\text{wb}}) - e^*(T_d)} = -1 \quad (10)$$

Please refer to the Appendix for (i) a graphical illustration of the saturation vapor pressure curve, and the different temperatures and related ET rates defined above; (ii) a brief historical description of how the CR evolved into equation (1) over the past 40 years; (iii) plots of selected CR functions over sample data; (iv) how assigning a value of α is performed without resorting to any calibration; and (v) a pseudo-algorithm.

Equation (1) was applied in a continuous monthly simulation over the 37-year period of 1979–2015 across CONUS, employing the 4-km spatial resolution Parameter-Elevation Regressions on Independent Slopes Model (PRISM; Daly et al., 1994) air and dew-point temperature data. The 32-km North American Regional Reanalysis surface net radiation and 10-m wind data (Mesinger et al., 2006) were linearly interpolated onto the PRISM grid employing a power transformation (Brutsaert, 1982) of the 10-m wind (u_{10}) values into $u_2 [=u_{10}(2/10)^{1/7}]$, required by equation (5). The CR-derived monthly ET rates were then aggregated into annual sums for further evaluations.

2.2. Long-Term Large-Scale ET Products

Eight long-term, publicly available ET products were selected to represent mainstream approaches in large-scale ET estimations, that is, LSMs, RS models, reanalysis, and spatial upscaling of EC measurements. Table 1 provides an overview of the eight products, which includes the following.

2.2.1. LSM-Based Products: Noah, VIC, and Mosaic

Three LSMs, Noah (Chen & Dudhia, 2001), VIC (Liang et al., 1994), and Mosaic (Koster & Suarez, 1996), from NLDAS-2 were selected to represent ET estimates from the LSMs. While these LSMs were developed by different groups for use in global climate models, the underlying principle they calculate ET is essentially the same. They all use a resistance-type stress factor to adjust the Penman-Monteith-based (when surface resistance is set to zero; Monteith, 1965) potential ET rate to represent actual ET. However, there are substantial differences among the models in parameterizations of soil water flow, stomatal conductance, root water uptake, etc., which all impact the estimated stress factor (Xia et al., 2016) in these LSMs (see Figure 2 for the differences in LSM-modeled ET rates).

All three LSMs are driven by NLDAS-2 atmospheric forcing (air temperature, specific humidity, wind speed, surface pressure, incoming solar and longwave radiation, and precipitation) at a spatial resolution of 0.125° (Xia et al., 2012), which are derived primarily from identical forcing fields of North American Regional Reanalysis data by the National Centers of Environmental Prediction (NCEP), except for precipitation. NLDAS-2 precipitation comes from the NOAA Climate Prediction Center's 0.125° gauge-based precipitation analysis, with monthly PRISM adjustments for orographic impacts. The soil parameters for these LSMs are determined by the global 1-km hybrid State Soil Geographic Database (Miller & White, 1998). The ground land cover classification for the NLDAS-2 LSMs was based on the global, 1-km vegetation database of the University of Maryland (Hansen et al., 2010), which was retrieved from the advanced very high resolution radiometer. For each 0.125° model cell, the vegetation field considers the relative frequency (RF) value of each vegetation class established on the 1-km resolution. Noah uses the predominant vegetation class, while VIC and Mosaic use subgrid vegetation tiles weighted by the RF of the classes. Additional vegetation greenness fractions were derived from the NOAA normalized difference vegetation index data (Gutman & Ignatov, 1998) on a multiyear-mean monthly basis without any interannual variation. For a thorough introduction on NLDAS-2 LSM forcing, parameters, configuration, and outputs, readers are suggested to refer to <https://ldas.gsfc.nasa.gov/nldas/NLDAS2model.php>. In the present study monthly ET rates from three NLDAS-2 LSMs (Noah, VIC, and Mosaic) were employed for 1979–2015 in a spatial resolution of 0.125° .

2.2.2. Reanalysis Products: NCEP-II and ERA-Interim

Reanalysis data are produced by a specific data assimilation scheme which combines various observations (e.g., ground stations, aircrafts, and/or satellites) and forecast outputs from weather prediction models. Two atmospheric reanalysis data, NCEP-DOE (Department of Energy) Reanalysis 2 (NCEP-II) and ERA-Interim, were selected in the present study. NCEP-II (Kanamitsu et al., 2002) is an improved version of the first global reanalysis data (i.e., NCEP-I) released by NCEP-NCAR (National Center for Atmospheric Research), which fixes numerous problems occurring in the latter, although both are based on the three-dimensional variational data assimilation scheme. ERA-Interim is a global atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts (Dee et al., 2011), which used the most advanced, that is, a four-dimensional variational data assimilation, technique. Three-dimensional variational scheme treats observations that occur within a certain time interval around the target-analysis time as occurring simultaneously with the target time. Four-dimensional variational one, however, uses temporal weights of those near-simultaneous measurements to estimate the target time value. Note that while the spatial resolution of reanalysis data is usually coarse, that is, 2.5° for NCEP-II and ~ 79 km for

Table 1
Basic Information of Eight ET Products Used for Comparisons in This Study

ET products	Category	Spatial resolution	Temporal resolution	Temporal coverage	References
Noah	Land surface model	0.125°	3 hr, monthly	1979–2015	Xia et al. (2012)
VIC	Land surface model	0.125°	3 hr, monthly	1979–2015	Xia et al. (2012)
Mosaic	Land surface model	0.125°	3 hr, monthly	1979–2015	Xia et al. (2012)
NCEP-II	Reanalysis	2.5°	24 hr, monthly	1979–2015	Kanamitsu et al. (2002)
ERA-Interim	Reanalysis	T255 (~79 km)	6 hr, monthly	1979–2015	Dee et al. (2011)
GLEAM	Remote sensing model	0.25°	Daily	1980–2015	Martens et al. (2017)
PML	Remote sensing model	0.5°	Monthly	1981–2012	Zhang et al. (2017)
FLUXNET-MTE	Upscaling of EC measurements	0.5°	Monthly	1982–2011	Jung et al. (2011)

ERA-Interim, they can, however, be produced in a near-real-time manner, supporting long-term hydroclimatic research. In the present study, monthly ET values from NCEP-II and ERA-Interim for 1979–2015 over the CONUS domain were employed.

2.2.3. RS-Based Products: Global Land Evaporation Amsterdam Model and Penman-Monteith-Leuning

The Global Land Evaporation Amsterdam Model (GLEAM; Miralles et al., 2011) is a two-step method for the estimation of ET, meaning that first the Priestley and Taylor (1972) potential ET rate is obtained which is subsequently adjusted by the help of a stress factor. The latter is a function of root-zone soil moisture and vegetation optical depth and could be derived using microwave satellite data. This model is actually designed using remote sensing observations only, but can also be driven by reanalysis- or station-based forcing to extend its temporal coverage. In the present study GLEAM v3.2a ET product (Martens et al., 2017), covering 1980–2015 on a daily time step with a spatial resolution of 0.25°, was selected, driven by net radiation and air temperature from ERA-Interim plus the Multi-Source Weighted-Ensemble Precipitation data, and snow water equivalent on top of the already mentioned inputs. The static input variables of GLEAM include the ground cover fraction from the Moderate Resolution Imaging Spectroradiometer and the soil properties from the Global Gridded Surfaces of Selected Soil Characteristics generated by the International Geosphere-Biosphere Program.

The Penman-Monteith-Leuning (PML) model combines a Penman-Monteith-type equation for plant transpiration, the Leuning model for canopy conductance (Leuning et al., 2008), and the Priestley-Taylor equation for soil evaporation (Zhang et al., 2017). The global scale 0.5°, monthly PML ET product for 1981–2012 was then generated by Zhang et al. (2017) using the Princeton Meteorological Forcing precipitation, air temperature, vapor pressure, shortwave and longwave downward radiation, wind speed (Sheffield et al., 2006), and the leaf area index data derived from advanced very high resolution radiometer normalized difference vegetation index (Zhu et al., 2013). The static land cover input for the PML model was also from International Geosphere-Biosphere Program. The PML-obtained ET rates were constrained by the Budyko framework using global streamflow observations at the mean annual scale to ensure an internal water balance (Zhang et al., 2016). It should be mentioned that GLEAM and PML also partition ET into its components (e.g., soil evaporation, plant transpiration), thus representing state-of-the-art RS-based ET products.

2.2.4. Spatially Upscaled Eddy-Covariance Measurement Product: FLUXNET-MTE

FLUXNET-MTE is a spatially upscaled ET product of global eddy-covariance measurements using a machine learning technique called model tree ensembles (MTE) (Jung et al., 2011). The MTE was trained using 29 explanatory variables with measured flux data of 198 FLUXNET towers across a wide range of biomes worldwide, which are mostly located in North America and Europe. FLUXNET-MTE generated a monthly, 0.5° latent heat flux data set for 1982–2011 over the continents. Inputs of the model comprise of (i) the fraction of absorbed photosynthetic active radiation derived from merged remote sensing products of Global Inventory Modeling and Mapping Studies, Sea-viewing Wide Field-of-view Sensor, and Moderate Resolution Imaging Spectroradiometer; (ii) meteorological data from the Climatic Research Unit and the Global Precipitation Climatology Centre; and (iii) a land cover data fused from multiproducts called SYNMAP (Jung et al., 2006). It should be noted that although nearly 30 predictor variables were used to train the MTE, most of them did not vary interannually. Therefore, the FLUXNET-MTE product tends to

underestimate the interannual variability of ET (Jung et al., 2011) that translates into subdued linear trend values (seen later).

2.3. Water-Balance-Derived ET of the HUC2 and HUC6 Basins Across CONUS

For a standard geographical framework for hydrological research and water resource management, the United States Geological Survey divides the country into successively smaller hydrological units with a unique “hydrological unit code” (HUC; Seaber et al., 1987) identifier. The CONUS has 18 first-level two-digit (i.e., HUC2) hydrological units (Figure 1), which comprise either the drainage area of a major river (e.g., Missouri region) or the combined drainage areas of a series of rivers (e.g., Texas-Gulf region). These HUC2 basins are also further divided into 204 second-level four-digit (i.e., HUC4) and 334 third-level six-digit basins (i.e., HUC6) based on surface topography (Seaber et al., 1987; Figure 1). In the present study, all (except the CONUS averaged mean annual ET rates of Figure 2) evaluations are based on basin-wide ET rates averaged over either HUC2 or HUC6 basins.

Basin-scale water-balance-based evapotranspiration rates (ET_{wb}) can be derived by

$$ET_{wb} = P - Q - \delta S \quad (11)$$

where P , Q , and δS are the basin-wide precipitation, stream discharge/runoff, and the changes in terrestrial water storage, respectively. The last term of equation (11) represents the combined changes of water storage in the soils, groundwater, open water bodies, and the snow/ice system, and usually considered small enough to be ignored at an annual (or longer) scale (Senay et al., 2011), that is,

$$ET_{wb} \approx P - Q \quad (12)$$

Equation (12), however, may not be accurate for smaller basins affected by significant (i) interbasin water exchanges (Szilagyi & Jozsa, 2018) and/or (ii) reservoir regulations of large (relative to mean flow) storage volumes (Han et al., 2015). While the Gravity Recovery and Climate Experiment (GRACE) data (Tapley et al., 2019) enable the community to estimate basin-scale δS in equation (11), its available time-coverage (i.e., 2002–2017) is too short to allow for a reliable estimation of the HUC2/HUC6 mean annual ET_{wb} values and especially for the estimation of long-term tendencies; therefore, equation (12) was employed for the 1979–2015 period in addition to equation (11) for the 2003–2015 time period.

Annual δS in the present study was calculated as the difference in terrestrial water storage anomaly of successive Decembers, the latter taken as the arithmetic average of three GRACE products processed by Geoforschungs Zentrum Potsdam; by the Center for Space Research at the University of Texas, Austin; and by the Jet Propulsion Laboratory in Pasadena, CA, after applying the relevant gain factors for each type of data (Landerer & Swenson, 2012). The monthly P and Q data came from PRISM precipitation (version AN81m; Daly et al., 1994) and United States Geological Survey runoff (at both HUC2 and HUC6 scales) records (Brakebill et al., 2011) for 1979–2015. Altogether seven HUC6 basins (Figure 1) were removed (about 2% of all HUC6 basins) from the analysis because their ET_{wb} values were outliers from similar values in their neighborhood.

Because of a considerable difference that exists in the spatial resolutions of the selected ET products, including the CR-based estimates, all ET data (also the PRISM P and GRACE-derived δS values) were first resampled into a common 0.125° grid using the nearest-neighbor method, followed by a spatial averaging to calculate basin-wide ET (as well as P and δS) rates for each HUC2 and HUC6 units (see below).

The performance of the nine ET products was measured by the (i) Pearson correlation coefficient (R), (ii) root-mean-square error ($RMSE$), (iii) relative bias (RB), (iv) ratio of standard deviations (SR) in modeled and water-balance derived values, and (v) Nash-Sutcliffe efficiency (NSE) against water-balance derived ET_{wb} rates for 18 HUC2 and 327 HUC6 basins. The comparisons involved the multiyear mean annual ET rates and the sum-of-squares-fitted linear trends in annual ET for the available data periods (see Table 1) of the ET products, and separately for the overlapping period of the GRACE data and the given ET product. In addition, the same performance measures were calculated for the annual ET time series of the 18 HUC2 basins to assess how the models capture interannual variability.

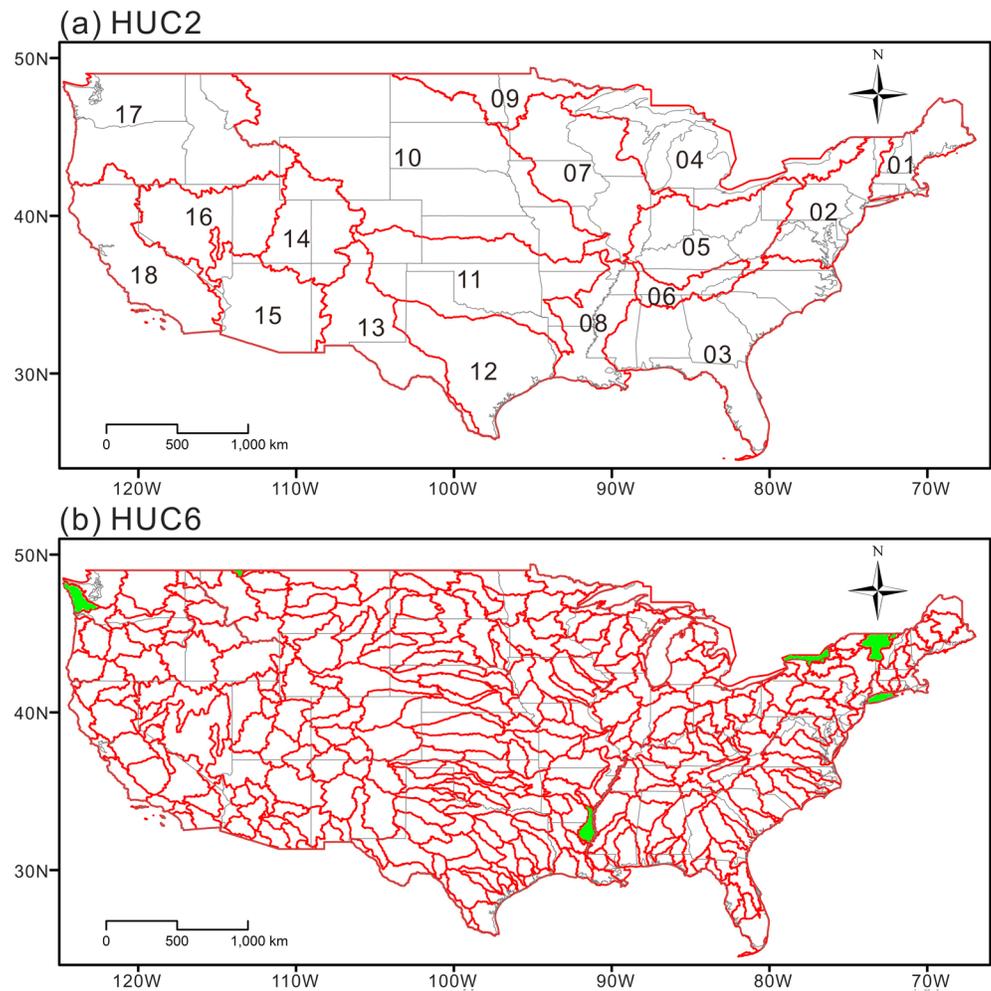


Figure 1. Distribution of the (a) 18 HUC2 and (b) 334 HUC6 basins across the conterminous United States. The green-colored HUC6 basins yielded outlying ET_{wb} values and were left out of the analysis. The names of the 18 HUC2 basins are 01 = New England, 02 = Mid-Atlantic, 03 = South Atlantic-Gulf, 04 = Great Lakes, 05 = Ohio, 06 = Tennessee, 07 = Upper Mississippi, 08 = Lower Mississippi, 09 = Souris-Red-Rainy, 10 = Missouri, 11 = Arkansas-White-Red, 12 = Texas Gulf, 13 = Rio Grande, 14 = Upper Colorado, 15 = Lower Colorado, 16 = Great Basin, 17 = Pacific Northwest, 18 = California.

3. Results

3.1. Spatial Pattern of the Multiyear Mean Annual ET From CR and Eight Other Products

There is no considerable difference in the spatial pattern of multiyear mean annual ET values among seven products (without VIC and NCEP-II), all generally displaying lower ET rates over the arid regions in the western CONUS and higher values in the humid regions of the southeastern CONUS (Figure 2). However, it seems that VIC simulates particularly low values over most parts of the eastern CONUS, while NCEP-II does the opposite. A country-averaged multiyear mean annual ET comparison indeed yields the highest value for NCEP-II with 838 ± 60 mm/year (mean \pm standard deviation), which is followed by ERA-Interim (653 ± 28 mm/year) and Mosaic (625 ± 23 mm/year), while VIC yields the smallest ET rate of only 435 ± 14 mm/year (Figure 2). CR yields 534 ± 19 mm/year, which is close to GLEAM's 530 ± 17 mm/year, both are “in the middle” ET rates when comparing all model values. As seen, FLUXNET-MTE displays the lowest interannual variations (10 mm/year) among all the models, mainly because the great majority of the explanatory variables used to train the MTE was assumed static over the years, thereby leading to the observed decreased interannual variability (Jung et al., 2011), which then probably also translates to reduced spatial variance of the resulting ET rates, seen later.

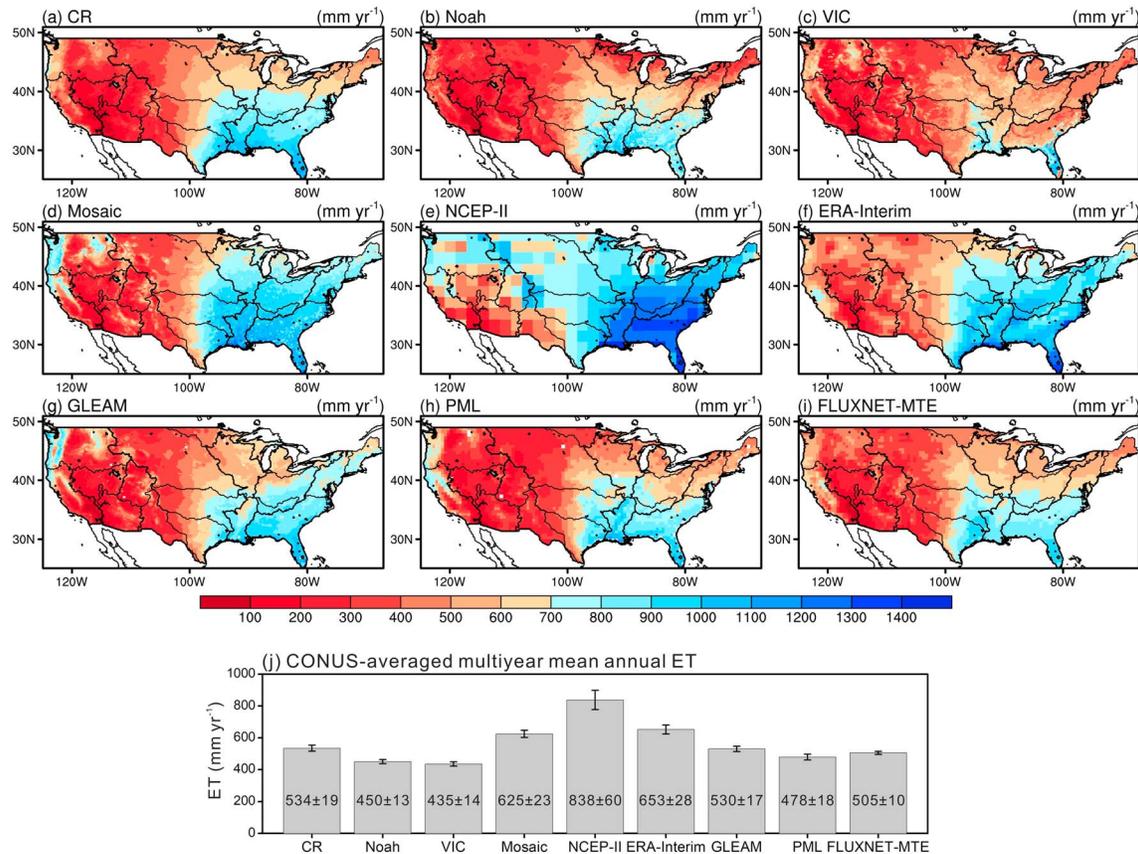


Figure 2. (a–i) Spatial distribution of the modeled multiyear mean annual ET rates (mm/year) and (j) their CONUS-averaged values (plus/minus standard deviation).

3.2. Basin-Scale Evaluation of the Multiyear Mean Annual ET Rates

The water-balance-based (equation (12)) multiyear mean annual ET rates (ET_{wb}) for the HUC2 basins are displayed in Figure 3. Not surprisingly, the Great Basin (#16; see Figure 1) and the Lower Colorado (#15) are the driest ($ET_{wb} = 260\text{--}305$ mm/year, respectively) while the South Atlantic-Gulf (#03) and Lower Mississippi (#08) the wettest ($ET_{wb} = 920\text{--}960$ mm/year, respectively) HUC2 basins. Figure 4 displays the evaluations of the multiyear mean annual ET estimates against the water-balance-based values over the 18 HUC2 basins for each product, with error bars denoting the interannual variability. Note that the averaging periods throughout the study follow the temporal coverage of the corresponding products (Table 1). While the multiyear mean annual values are not sensitive to the changes in the averaging periods (see Table S1), the trend values, calculated later, are.

In comparison with ET_{wb} , the CR performs the best in estimating the multiyear mean annual ET rate of the HUC2 basins with a NSE value of 0.94 and $RMSE$ value of 51.2 mm/year (Figure 4 and Table 2). FLUXNET-MTE and other two RS models (i.e., GLEAM and PML) also have NSE values larger than 0.86 and $RMSE$ values smaller than 77 mm/year. Among the LSMs, Noah ($NSE = 0.75$) appears to produce more accurate ET rates (with the highest $R = 0.99$ value, surpassing the CR model's $R = 0.98$ value) than VIC or Mosaic. The worst skill occurs in NCEP-II, which significantly overestimates ET_{wb} rates in all basins, thus being the only model with a negative NSE value (-1.47). Another reanalysis product, ERA-Interim ($NSE = 0.55$), has a very similar performance to that of Mosaic ($NSE = 0.65$).

Figure 5 illustrates the spatial pattern of the ratio of multiyear mean HUC2-averaged ET from nine products to water-balance derived ET_{wb} across CONUS. It can be seen that the errors of CR are within $\pm 10\%$ for most HUC2 basins with the largest exception for the Lower Colorado (#15). Noah and VIC generally underestimate the multiyear mean annual ET rates over almost all basins, with the most significant bias observed

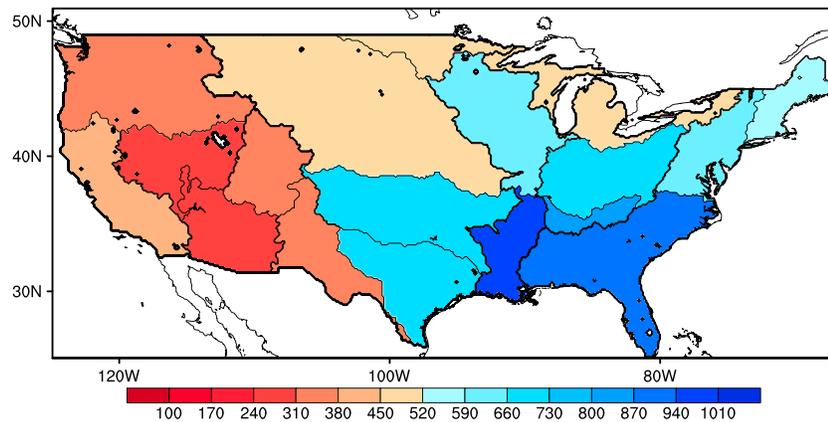


Figure 3. Spatial distribution of the water-balance derived multiyear (1979–2015) mean annual HUC2 ET_{wb} rates (mm/year).

in California and in the southeastern CONUS, respectively. On the contrary, Mosaic overestimates ET in most basins except for the Lower Colorado and Rio Grande, with the largest positive bias found in the Pacific Northwest (#17) and Great Lakes (#04). A positive bias of varying degree exists in all basins in the ERA-Interim data. The two RS models, PML and GLEAM, succeed the best [*SR* values of 0.98 and 1.01 (Figure 4 and Table 2), respectively], while VIC and FLUXNET-MTE the worst, in replicating the spatial

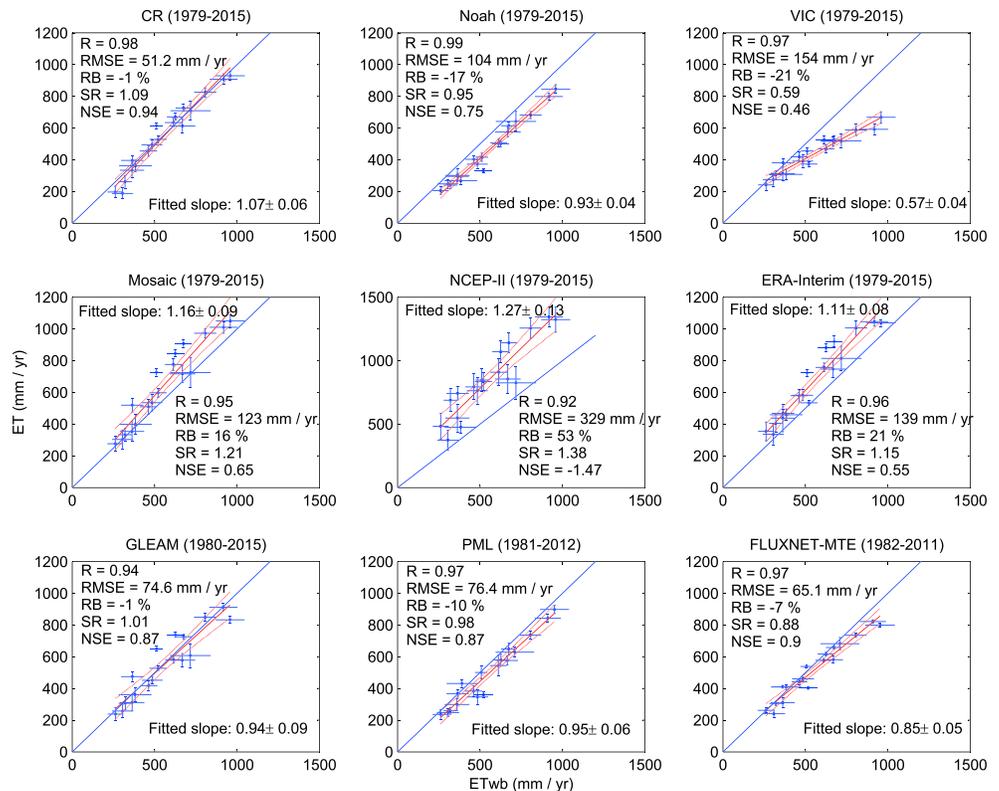


Figure 4. Regression plots of the HUC2-averaged multiyear mean annual ET rates of the nine ET products against water-balance ET_{wb} data. The period for calculating multiyear mean values of ET_{wb} follows the one in parenthesis for each product. The length of the whiskers denotes the standard deviation of the basin-averaged annual ET values during corresponding periods. The strips around the fitted lines (red) denote the 95% confidence intervals and the long blue line represents a 1:1 relationship.

Table 2
Performance Statistics of the Modeled Multiyear Mean Annual ET Rates for the Full and GRACE Periods of the 18 HUC2 and 327 HUC6 Basins Across CONUS

Category	Periods	Products	HUC2					HUC6							
			R	RMSE	RB	SR	NSE	R	RMSE	RB	SR	NSE			
Full period	1979–2015	CR	0.98	51.2	−1	109	0.94	0.93	89	2	104	0.86			
		Noah	0.99	104	−17	95	0.75	0.96	116	−16	93	0.76			
		VIC	0.97	154	−21	59	0.46	0.90	177	−21	63	0.45			
		Mosaic	0.95	123	16	121	0.65	0.92	140	16	111	0.66			
		NCEP-II	0.92	329	53	138	−1.47	0.81	353	50	134	−1.19			
		ERA-Interim	0.96	139	21	115	0.55	0.91	171	22	111	0.48			
GRACE period	1980–2015	GLEAM	0.94	74.6	−1	101	0.87	0.88	116	0	97	0.76			
	1981–2012	PML	0.97	76.4	−10	98	0.87	0.95	90.1	−8	98	0.86			
	2003–2015	FLUXNET-MTE	CR	0.98	54.2	−1	114	0.94	0.94	91	1	109	0.85		
			Noah	0.99	97.1	−16	96	0.79	0.96	110	−15	95	0.79		
			VIC	0.96	146	−20	62	0.53	0.89	170	−20	65	0.49		
			Mosaic	0.95	131	17	123	0.62	0.92	147	17	113	0.62		
			NCEP-II	0.91	387	64	143	−2.31	0.80	407	60	137	−1.92		
			ERA-Interim	0.97	120	17	116	0.68	0.91	156	18	112	0.57		
	2003–2012	PML	GLEAM	0.94	72	0	101	0.89	0.88	116	1	99	0.76		
			ERA-Interim	0.97	75.7	−10	95	0.87	0.94	91.7	−7	98	0.85		
			2003–2011	FLUXNET-MTE	GLEAM	0.97	75.7	−10	95	0.87	0.94	91.7	−7	98	0.85
					ERA-Interim	0.97	120	17	116	0.68	0.91	156	18	112	0.57
GLEAM					0.94	72	0	101	0.89	0.88	116	1	99	0.76	
PML					0.97	75.7	−10	95	0.87	0.94	91.7	−7	98	0.85	
ERA-Interim	0.97	120			17	116	0.68	0.91	156	18	112	0.57			
GLEAM	0.94	72			0	101	0.89	0.88	116	1	99	0.76			
2003–2011	FLUXNET-MTE	0.97	60.1	−5	88	0.92	0.95	80	−4	87	0.88				

Note. The units of RMSE, RB, and SR are mm/year, %, and %, respectively; R and NSE are dimensionless. Outstanding values in each category are emphasized.

variance found in the HUC2-averaged ET_{wb} multiyear means. GLEAM has an obvious positive bias in the Pacific Northwest (#17), similar to Mosaic.

When switching to the GRACE period of 2003–2015 for inclusion of the δS values, CR still produces the highest NSE (0.94) and lowest RMSE (54.2 mm/year) values (Table 2) for the HUC2 basins, followed by FLUXNET-MTE (0.92 and 60.1 mm/year) and GLEAM (0.89 and 72 mm/year). NCEP-II (RB = 64%), ERA-Interim, and Mosaic (for both, RB = 17%) still overestimate the water-balance derived ET_{wb} values,

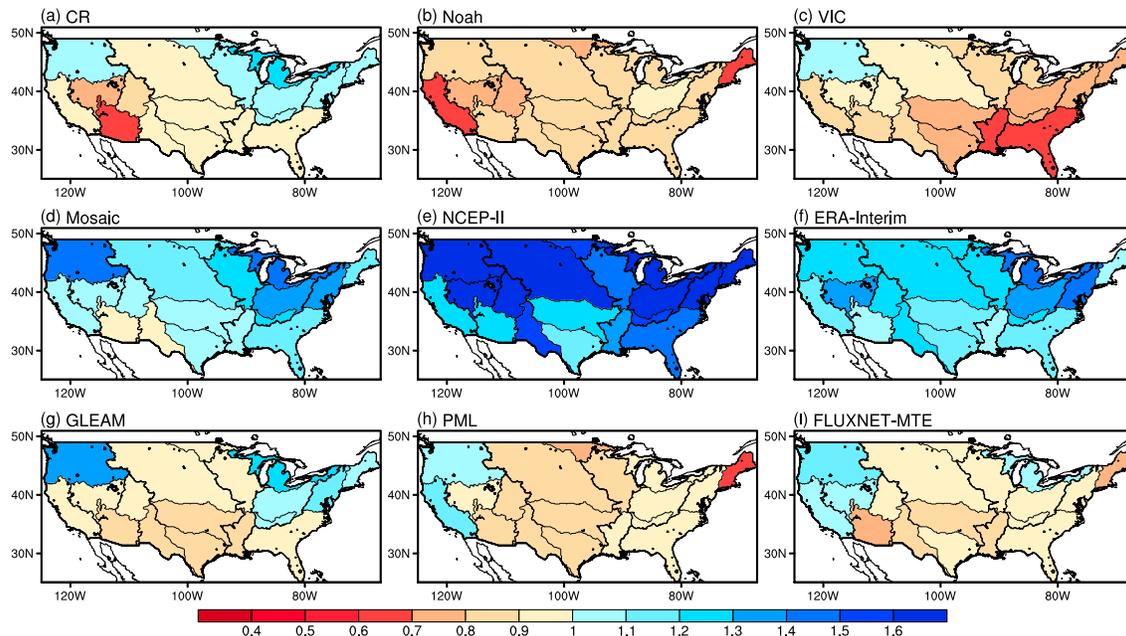


Figure 5. Spatial distribution of the ratio of multiyear mean annual HUC2-averaged modeled ET to ET_{wb} rates. In each panel, the period for calculating the multiyear mean values of ET_{wb} follows that of the corresponding ET product.

while VIC ($RB = -20\%$) and Noah ($RB = -16\%$) underestimate them. The two models most on target with their ET estimates are GLEAM ($RB = -1\%$ for 1980–2015 and $RB = 0\%$ for 2003–2015) and CR ($RB = -1\%$ in both cases), while the spatial variance is also reproduced the best by GLEAM ($SR = 1.01$) and the worst by VIC and FLUXNET-MTE (Table 2).

The water-balance derived (equation (12)) multiyear mean annual ET rates (ET_{wb}) for the HUC6 basins are displayed in Figure 6. While the wettest HUC6 basins ($ET_{wb} > 1,000$ mm/year) are found within the two wettest HUC2 basins, that is, the South Atlantic-Gulf (#03) and Lower Mississippi (#08), the driest ($ET_{wb} < 100$ mm/year) basins are found in the California (#18) HUC2 basin.

For the HUC6 basins the FLUXNET-MTE ET rates exhibit the highest NSE value of 0.87 (Figure 7 and Table 2), closely followed by those of CR and PML ($NSE = 0.86$ for both). The $RMSE$ is the smallest also for FLUXNET-MTE (85.2 mm/year) followed by CR (89 mm/year) and PML (90.1 mm/year) again. For relative bias, GLEAM and CR have the smallest values (0% and 2%, respectively), followed by FLUXNET-MTE with -6% . The R value is the highest though for Noah (0.96), followed by FLUXNET-MTE and PML (both of 0.95). The worst performer is NCEP-II with the only negative NSE value and a significant overestimation, yielding an RB value of 50% (Figure 7). VIC and ERA-Interim are also among the worst performers with a significant underestimation ($RB = -21\%$) and overestimation ($RB = 22\%$), leading to the lowest positive NSE values of 0.45 and 0.48, respectively (Figure 7). The spatial variance is best captured by the two RS models (PML and GLEAM) plus CR with SR values of 0.98, 0.97, and 1.04, respectively (Table 2).

Again, by switching to the GRACE period for using equation (11), there is no change in the best NSE and $RMSE$ performance order among FLUXNET-MTE, CR, and PML (Table 2). For RB , GLEAM and CR yield the same 1% value. The order for the highest R values remains the same as before; however, CR comes up to the level of PML with an R value of 0.94. NCEP-II performs even worse than before, while VIC and ERA-Interim both improve a bit. For reproducing spatial variability both RS models continue to dominate with the highest SR .

Figure 8 presents how the different models predict the multiyear mean annual ET rates of the individual HUC6 basins in a spatially referenced manner. As seen, three models, CR, GLEAM, and FLUXNET-MTE, yield a similar spatial pattern in the modeled ET rates, relative to ET_{wb} values. All of them tend to overestimate ET_{wb} in the northwestern and underestimate it in the southwestern CONUS. Mosaic and ERA-Interim overestimate ET_{wb} in the majority of the HUC6 basins, while NCEP-II seriously overestimates it in almost every basin. The most serious underestimation of ET_{wb} at differing locations can be seen in the CR, Noah, and VIC models.

3.3. Spatial Pattern of the Trends in Annual ET Rates From CR and Eight Other Products

While not all products (mainly PML and FLUXNET-MTE) cover the full period of 1979–2015, almost all products demonstrate that annual ET increased over the majority of the upper north-eastern third of the CONUS, while decreasing trends dominated the western parts (Figure 9). Note again that the trend of a given product is for the period it covers, displayed in Figure 9. It appears that the trends in the reanalysis products, NCEP-II and ERA-Interim, may be unreasonably strong when compared with other products (i.e., the increasing trends in the former and the decreasing ones in the latter). The spatial pattern of the trends from the CR are generally consistent with those of LSMs (Noah, VIC, Mosaic) over most regions, except for the South Atlantic-Gulf HUC2 (#03) basin (e.g., Florida, Georgia, and Alabama) where the CR displays significant increasing trends but the LSMs do not, although GLEAM, PML, and FLUXNET-MTE (even with a shortened modeling period) do so as well. Similarly, in most parts of California the CR-derived ET rates display significant decreasing trends, less apparent in other model outputs, except ERA-Interim.

3.4. Basin-Scale Evaluation of the Trends in Annual ET Rates

For every HUC2/HUC6 basin, a linear trend in the basin-averaged annual ET_{wb} values is calculated for the data coverages of the different ET products and also for their overlap with the GRACE period (2003–2015). Figure 10 displays the spatial distribution of the resulting linear tendencies in annual ET_{wb} over 1979–2015 for the 18 HUC2 basins. West of the Lower Mississippi (#08) and Missouri (#10) HUC2 basins the tendencies are negative, with the largest ones (between -2 and -2.5 mm/year) found

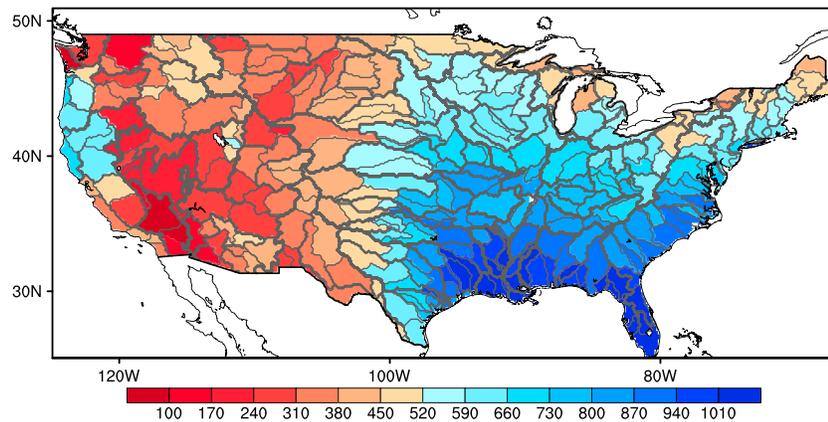


Figure 6. Spatial distribution of the water-balance derived multiyear (1979–2015) mean annual HUC6 ET_{wb} rates (mm/year).

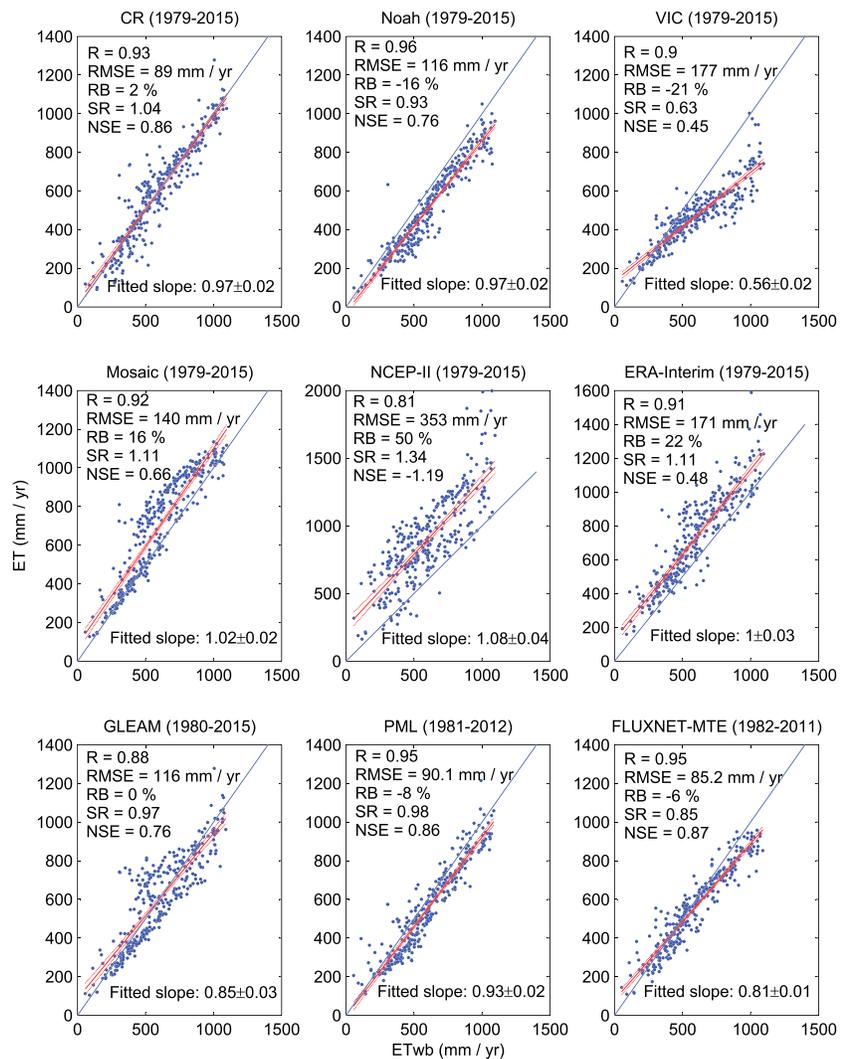


Figure 7. Regression plots of the HUC6-averaged multiyear mean annual ET rates of the nine ET products against water-balance ET_{wb} data. The period for calculating multiyear mean values of ET_{wb} follows the one in parenthesis for each product. The strips around the fitted lines (red) denote the 95% confidence intervals and the long blue line represents a 1:1 relationship.

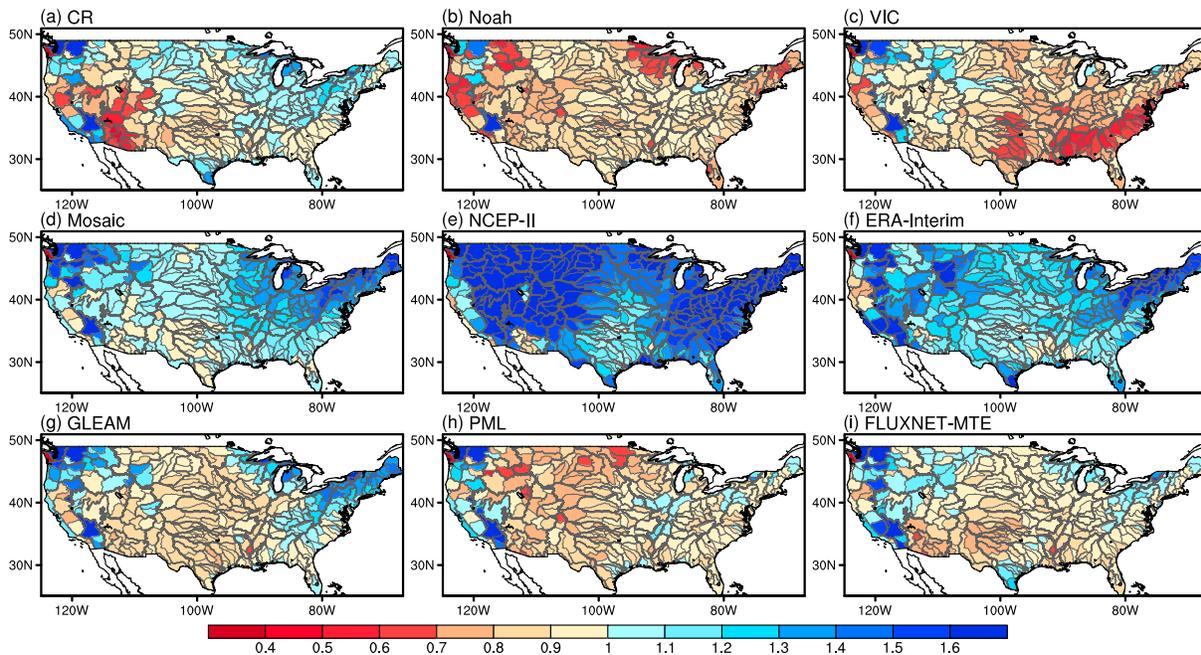


Figure 8. Spatial distribution of the ratio of multiyear mean annual HUC6-averaged modeled ET to ET_{wb} rates. In each panel, the period for calculating the multi-year mean values of ET_{wb} follows that of the corresponding ET product.

in California (#18) and the Lower Colorado (#15), while the largest positive trend (2.3 mm/year) is found in the South-Atlantic Gulf (#03) basin.

The evaluations of the modeled linear trends at the HUC2-basin scale are presented in Figures 11 and 12 as well as in Table 3. It can be seen that the CR captures the trends most effectively among all the models and model periods considered; that is, CR produces the highest *NSE* (0.80), *R* (0.91), and the lowest *RMSE*

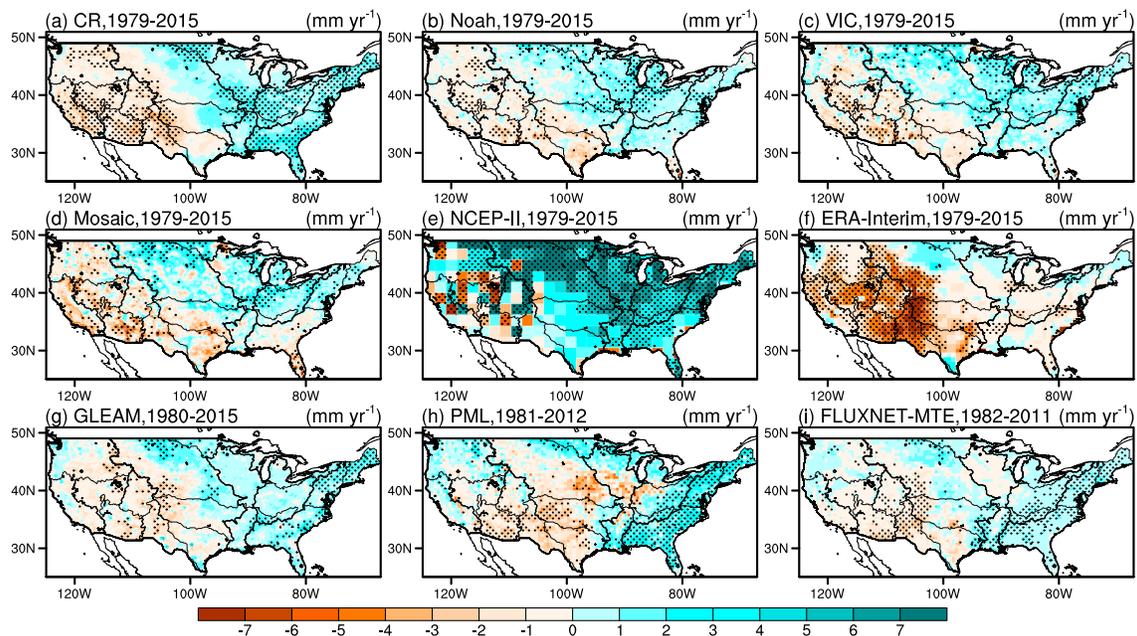


Figure 9. Spatial distribution of the linear tendencies (mm/year) in annual ET sums of the nine different ET products considered. The stippling denote trends that are statistically significant ($p < 5\%$) in the Student's *t* test.

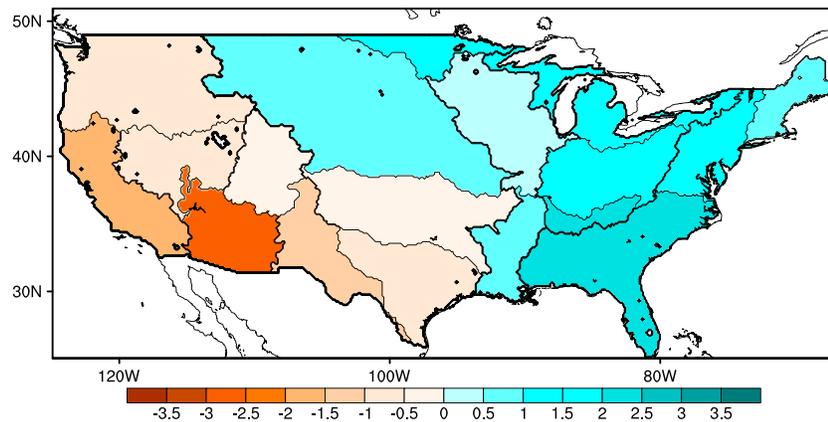


Figure 10. Spatial distribution of the linear tendencies (mm/year) in annual ET_{wb} sums during 1979–2015 for the HUC2 basins.

(0.63 mm/year) values for the period of 1979–2015. Note the large uncertainty in the water-balance-derived trend values (long horizontal whiskers in Figures 11 and 12). Unlike for mean annual ET rates, VIC and Noah also perform relatively well in the estimation of the trends with NSE values of 0.74 and 0.63, respectively. The largest errors are found in the two reanalysis, NCEP-II and ERA-Interim, data (Figure 11 and Table 3), leading to negative NSE values (Figure 11). Note that ERA-Interim is the only product that yields decreasing trends in most basins of the eastern CONUS, contrary to what is observed in ET_{wb} data (Figures 10 and 13). Among the RS models, GLEAM and PML provide reasonable estimates for the trends in annual ET; however, the former predicts a very narrow range for the trends in the HUC2 basins (see Figure 12 and $SR = 0.49$ in Table 3), similar to that of FLUXNET-MTE, this latter one due, most probably, to its underestimation of the interannual variability of the ET values (see section 3.5). When

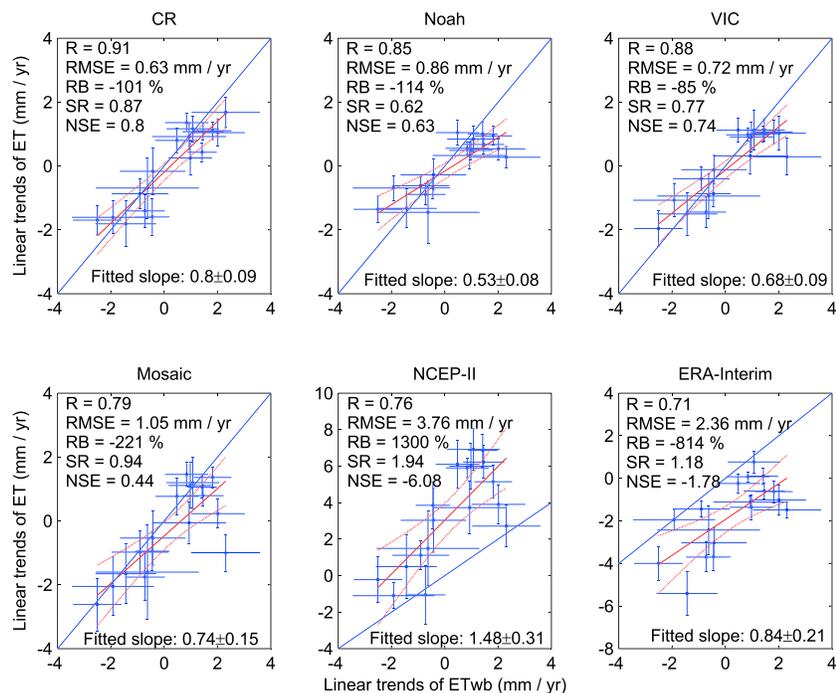


Figure 11. Regression plots of the linear trend values (mm/year) in modeled HUC2-averaged annual ET sums against those in ET_{wb} over 1979–2015. The length of the whiskers denotes the standard error in the estimated slope value. The strips around the fitted lines (red) denote the 95% confidence intervals and the long blue line represents a 1:1 relationship.

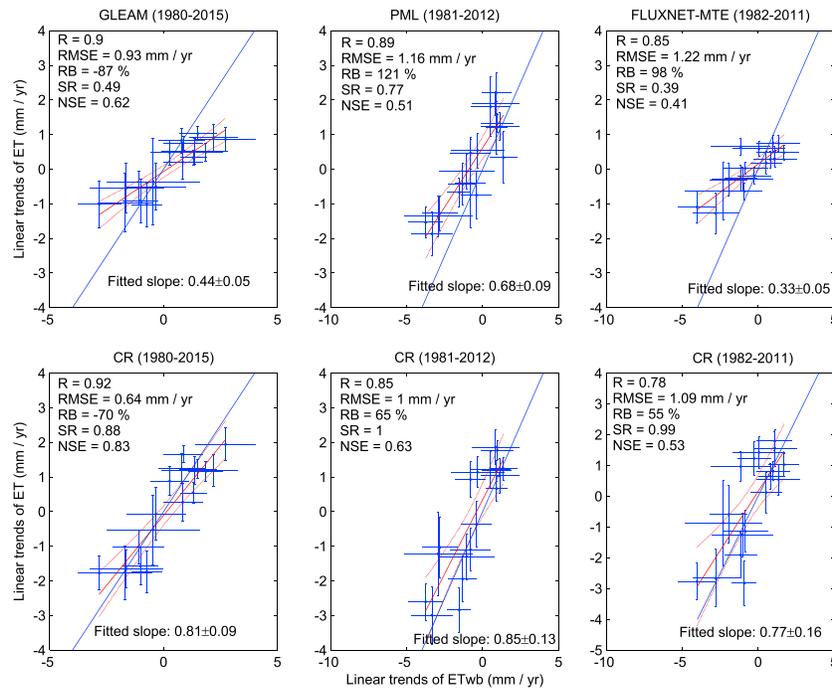


Figure 12. Regression plots of the linear trend values (mm/year) in GLEAM-, PML-, and FLUXNET-MTE-modeled HUC2-averaged annual ET sums against those in ET_{wb} over the different model periods (shown in parenthesis). For comparison, regressions for the CR ET values over the same periods are also displayed. The length of the whiskers denotes the standard error in the estimated slope value. The strips around the fitted lines (red) denote the 95% confidence intervals and the long blue line represents a 1:1 relationship.

Table 3

Performance Statistics of the Linear Trend Values in Modeled Annual ET Rates for the Full and GRACE Periods of the 18 HUC2 and 327 HUC6 Basins Across CONUS

Category	Periods	Products	HUC2					HUC6				
			R	RMSE	RB	SR	NSE	R	RMSE	RB	SR	NSE
Full period	1979–2015	CR	0.91	0.63	-101	87	0.80	0.58	1.63	-43	66	0.32
		Noah	0.85	0.86	-114	62	0.63	0.6	1.62	-124	56	0.32
		VIC	0.88	0.72	-85	77	0.74	0.64	1.53	-83	61	0.40
		Mosaic	0.79	1.05	-221	94	0.44	0.58	1.84	-242	85	0.13
		NCEP-II	0.76	3.76	1300	194	-6.08	0.49	4.35	-1134	164	-3.84
	1980–2015	ERA-Interim	0.71	2.36	-814	118	-1.78	0.4	2.88	-655	99	-1.12
		CR	0.92	0.64	-70	88	0.83	0.61	1.66	-20	67	0.36
		GLEAM	0.9	0.93	-87	49	0.62	0.61	1.7	-71	45	0.34
		CR	0.85	1	65	100	0.63	0.51	2.19	92	73	0.11
		PML	0.89	1.16	121	77	0.51	0.59	2.11	113	70	0.17
1982–2012	CR	0.78	1.09	55	99	0.53	0.41	2.35	98	69	0.02	
	FLUXNET-MTE	0.85	1.22	98	39	0.41	0.51	2.21	105	31	0.13	
	CR	0.78	1.09	98	39	0.41	0.51	2.21	105	31	0.13	
GRACE period	2003–2015	CR	0.61	2.49	-45	108	0.07	0.47	5.36	-64	59	0.14
		Noah	0.51	3.29	-124	105	-0.62	0.29	6.85	-130	74	-0.41
		VIC	0.53	3.16	-110	109	-0.49	0.3	6.63	-113	73	-0.32
		Mosaic	0.45	4.71	-186	155	-0.23	0.25	8.67	-174	114	-1.26
		NCEP-II	0.3	6.21	-137	233	-4.78	0.17	9.36	-129	134	-1.63
		ERA-Interim	0.6	4.3	-223	108	-1.77	0.25	7.39	-172	69	-0.64
	2003–2012	GLEAM	0.64	2.46	-80	91	0.09	0.5	5.41	-85	54	0.12
		CR	0.77	7.58	-223	67	0.22	0.66	10.55	-205	59	0.13
		PML	0.55	7.29	-60	55	0.28	0.46	10.33	-75	53	0.17
		CR	0.82	4.62	-253	91	0.58	0.64	9.53	-1650	69	0.34
		FLUXNET-MTE	0.88	4.66	3	44	0.58	0.64	9.47	-313	39	0.35
		CR	0.82	4.62	-253	91	0.58	0.64	9.53	-1650	69	0.34
		FLUXNET-MTE	0.88	4.66	3	44	0.58	0.64	9.47	-313	39	0.35
2003–2011	CR	0.82	4.62	-253	91	0.58	0.64	9.53	-1650	69	0.34	
	FLUXNET-MTE	0.88	4.66	3	44	0.58	0.64	9.47	-313	39	0.35	
	CR	0.82	4.62	-253	91	0.58	0.64	9.53	-1650	69	0.34	

Note. The units of RMSE, RB, and SR are mm/year, %, and %, respectively; R and NSE are dimensionless. Outstanding values in each category are emphasized.

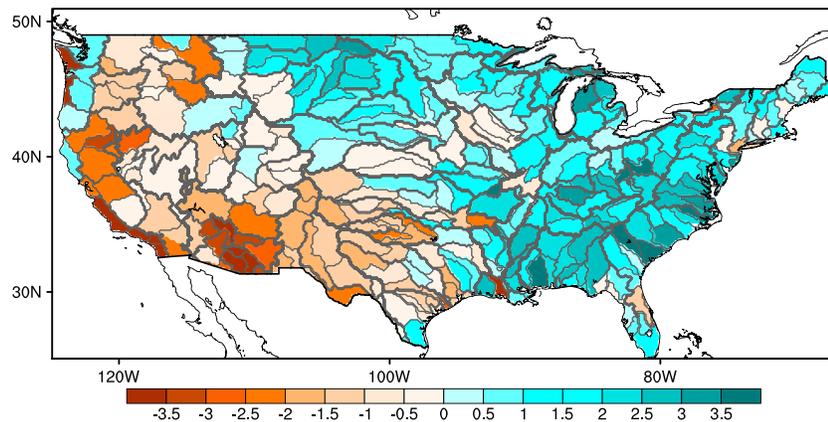


Figure 13. Spatial distribution of the linear tendencies (mm/year) in annual ET_{wb} sums during 1979–2015 for the HUC6 basins.

considering the three products that do not cover the full range of 1979–2015 (Figure 12), the CR yields better performance in every (except R) statistical measure (Table 3) employed. In the model evaluation of trends, RB should not be weighed heavily since the denominator may be very small if ET_{wb} demonstrates little change over a given period.

The linear trend estimates really deteriorate for the GRACE period (Figures S1 and S2 and Table 3) due primarily to the shortness (13 years altogether at best, but only 9 years for FLUXNET-MTE and 10 years for PML) of the periods (notice the enlarged horizontal whiskers in the ET_{wb} trend estimates of Figures S1 and S2 in comparison with those in Figures 11 and 12). Note that only the CR and GLEAM produce (similar) positive NSE values for the 13-year period, and also the smallest $RMSE$ and largest R values. For the even shorter periods in Figure S2, there is not much difference between the models.

Figure 13 displays the spatial distribution of the HUC6 linear trend values during 1979–2015 found in the water-balance derived annual ET_{wb} rates. It can be seen that most of the trends in the annual HUC6-averaged ET_{wb} are within the $[-4-4$ mm/year] interval. The largest positive trend (8.6 mm/year) again is found for a HUC6 watershed within the South Atlantic Gulf (#03) HUC2 basin, while the opposite (-7.6 mm/year) in the Pacific Northwest (#17) HUC2 basin. In addition, ET_{wb} rates of certain HUC6 basins in the coastal regions of California and in southern Arizona decreased the most significantly.

From Figures S3 and S4 and Table 3 it becomes clear that the skill of all products in predicting the water-balance derived trend deteriorates significantly due to the large uncertainty in the HUC6 water-balance-obtained trends caused by large interannual variability in the data. The resulting whiskers are particularly long and the number of data points is so numerous that displaying these uncertainties became impossible in these graphs. The largest NSE (0.4) and the smallest $RMSE$ (1.53 mm/year) values are provided by VIC, followed by an overall equal performance by Noah and CR (0.32 and 1.6 mm/year, respectively) for the 1979–2015 period (Figure S3 and Table 3). Similar to HUC2-basin performance, the reanalysis models also end up with negative NSE values for the HUC6 basins. For the shortened periods (Figure S4), the CR and the relevant models (GLEAM, PML, FLUXNET-MTE) behave about the same during the corresponding periods. With regard to the whole GRACE period of 2003–2015, only the CR and GLEAM could produce positive NSE values of 0.14 and 0.12 (Table 3), respectively; while for the shortened GRACE periods the models involved perform about the same again for the HUC6 basins (Table 3).

3.5. Basin-Scale Evaluation of the Temporal Variations of Annual ET Rates

In order to understand better how the different ET models capture interannual variability (IAV) of the ET values, annual time series from the nine ET products plus that of ET_{wb} are displayed in Figures S5 and S6 for each HUC2 basin and each period considered. As expected, ET_{wb} in general display reduced variability for the GRACE period due to changes in annual water storage. As seen, FLUXNET-MTE data displays by far the smallest IAV. While all ET products fail to reproduce the IAV of ET_{wb} in the eastern part of CONUS

(i.e., left column), it appears that most of them show a more comparable IAV to that of ET_{wb} in the generally more arid western part of the CONUS (i.e., right column). When averaged for all 18 HUC2 basins and over all periods, the mean SR values of most ET products are similar (50–70%)—with the best performing model being Mosaic ($SR = 76\%$)—except for NCEP-II which significantly overestimates the IAV of ET_{wb} (Figure S7), and FLUXNET-MTE doing the opposite.

Figure S7 and Table S2 also display additional model skills in simulating temporal variations in HUC2 basin annual ET rates. As seen, the CR produced the smallest period-combined $RMSE$ (71 mm/year) and largest NSE (−0.24%) values in comparison with the other eight products. Note that for NSE the median was computed instead of the arithmetic average due to frequently observed large negative values in individual HUC2 basins which happens when the mean of model ET differs significantly from water-balance one. Noah yielded the highest overall R value (0.58), while GLEAM and CR succeeded the most (RB of −0.3% and −3%, respectively) with being on target with the HUC2-basin model-ET means.

4. Discussions

While the water-balance method is often regarded the most accurate avenue for deriving basin-scale ET rates (Liu et al., 2016), the uncertainties in such “ground-truth” data are not negligible. First, although the subsurface leakage to adjacent basins should be minimal for HUC2 basins, this may not be true for all HUC6 basins (e.g., in the Sand Hills of Nebraska) whose mean area are only approximately 23,000 km² (Bleed & Flowerday, 1989). Besides, in watersheds with an obvious trend in the stored water (e.g., due to groundwater pumping) volumes, equation (12) may be inaccurate. Similarly, large-capacity water storage reservoirs may disturb the annual water balance on smaller watersheds. Second, the United States Geological Survey runoff value for each hydrological unit was generated by weighting the stream discharge data collected at HUC8 stream gauges with the corresponding drainage area and subsequently averaging the so-derived values over the (i.e., HUC6 or HUC2 in our case) basins (Brakebill et al., 2011). Although this data set is regarded as a close surrogate of the natural runoff in hydroclimate studies (Ashfaq et al., 2013; Szilagyi, 2018a), those basins with relatively few discharge gauging stations may be less reliable. Note also that water transfer between watersheds is not always negligible, see, for example, the Colorado River water transfer into California (Hanak et al., 2011).

Third, PRISM precipitation data (Daly et al., 1994) may not be perfectly appropriate for computing long-term tendencies for a given location (or grid cell) since certain stations experienced location and/or equipment changes, which may potentially lead to inhomogeneity in the time series of the data in that location/cell. This effect however vanishes for the HUC2 and most likely for the majority of the HUC6 basins as well, because such changes can affect only basins with very few precipitation stations. Fourth, the original spatial resolution of GRACE plus the nine ET products differ significantly (range of 0.125° to 2.5°). Accuracy of the basin-wide ET estimates therefore may deteriorate to some extent over small watersheds. For this reason, basin-wide ET_{wb} and modeled ET rates may be more reliable for HUC2 than for HUC6 basins. In general, future detailed studies on the accuracy of long-term, multi-level basin-wide P , Q , and δS could benefit a comprehensive evaluation of the large-scale ET products.

It should be noted that the accuracy of any ET product is highly dependent on the quality of inputs used to drive the model. With regard to the eight ET products considered in the present study, only Noah, VIC, and Mosaic from NLDAS-2 were intended exclusively for the CONUS where high-resolution, more reliable atmospheric forcing, soil and vegetation data could be obtained (Xia et al., 2012), while the other five ET products cover the whole globe with a limited number of available atmospheric forcing and soil/vegetation property data sets. For example, atmospheric forcing of the GLEAM v3.2a product (Martens et al., 2017) includes precipitation from the Multi-Source Weighted Ensemble Precipitation data set (Beck et al., 2017) together with air temperature and radiation data from the ERA-Interim reanalysis (Dee et al., 2011). It is therefore believed that the quality of such global forcing is not as high as those developed exclusively for the CONUS, that is, PRISM temperature and precipitation data. Note that while NLDAS-2 LSMs were driven by better atmospheric forcing and soil/vegetation data across the CONUS (Xia et al., 2012), they did not show much improved skills in comparison with GLEAM, PML, and FLUXNET-MTE, indicating great challenges due to not only parameter uncertainties (Clark et al., 2017; Mendoza et al., 2015; Mölders, 2005) but also the representation of various physical processes related to ET, such as vegetation dynamics (Ma et al., 2017),

surface exchange coefficients (LeMone et al., 2008), and root water uptake (Wang et al., 2018), which would potentially be upgraded in future NLDAS LSMs (Xia et al., 2016). Finally, most LSMs and reanalysis products do not consider anthropogenic activities such as agricultural irrigation, which may impact both regional ET and precipitation to a large extent (Szilagyi, 2018a; Szilagyi & Jozsa, 2018).

5. Summary and Conclusions

This study evaluated the performance of the latest calibration-free CR model together with eight other mainstream methods (LSMs, RS-based models, reanalysis products, and a spatial upscaling of eddy-covariance measurements) of estimating monthly ET rates and their multiyear linear tendencies with the help of water-balance-derived annual ET rates (ET_{wb}) over 18 HUC2 and 327 HUC6 basins within the CONUS.

At the HUC2 basin level where the ET_{wb} estimates are expected to be more accurate (due to a higher number of precipitation stations, more accurate discharge measurements in larger rivers, and a smaller relative effect of any possible interbasin water transfers and/or reservoir regulations on the water balance), the CR produced the highest *NSE* and best *RMSE* values among the nine products for the multiyear mean annual ET rates, independent of the different evaluation periods (due to somewhat different data coverage among models). The CR's *RB* value was matched/exceeded only by GLEAM, the latter also excelling in *SR*. Similarly, the CR's (spatial) *R* value was only surpassed by Noah. Still at the HUC2 level, the CR produced the highest *R* and *NSE* values together with the best *RMSE* for the periods considered in the estimation of the linear tendencies in annual ET rates. In the *RB* value FLUXNET-MTE excelled, followed by the CR, while in the *SR* value Noah and CR did so.

At the HUC6 level where the water-balance derived ET rates are less reliable for the above reasons, all models' performance worsened. For the mean annual values FLUXNET-MTE achieved the highest *NSE* and best *RMSE* values closely followed by CR and PML, in that order. The *RB* value of GLEAM reflected almost perfectly unbiased estimates, again closely followed by CR. In the *R* value Noah excelled, followed by FLUXNET-MTE, PML, and CR, in that order. However, *SR* was the best for PML and GLEAM in both full and GRACE periods, followed by CR. For the linear tendencies all model-performance measures further declined. Yet the CR managed to produce the lowest *RB* and the best *RMSE* values overall. CR, VIC, and FLUXNET-MTE produced about the same *R* values, while CR together with VIC and GLEAM showed comparable *NSE* values during the full period. The *SR* value was the best for ERA-Interim and Mosaic for the full and GRACE period, respectively.

Trend-performance statistics are sensitive to the actual data period (especially if the number of data points is low) when the fitting is performed in the typical least sum-of-square-sense (employed here too), as such fitting is not robust (well known in statistics) since outliers exert a disproportionately large weight in the fitting process making it sensitive to such values and errors in the data. Note that here trends served as another tool of comparing model performance, and not the focus of analysis of this study.

Among the models and periods considered CR, GLEAM, PML, and FLUXNET-MTE never produced any negative value for the spatial *NSE*.

When the combined (i.e., for full and GRACE periods) performance statistics of the modeled annual ET time series of the 18 HUC2 basins are considered (Table S2) then CR excelled in *RMSE* and *NSE*, GLEAM in *RB*, Noah in *R*, and Mosaic in *SR*. Only VIC and PML never excelled in any category.

It can be stated that altogether the CR outperformed all other models (even the one which utilizes measured ET fluxes) considered in this study for the estimation of the multiyear mean annual ET value and linear tendency in the annual values at the HUC2 level across the CONUS and was only outdone by FLUXNET-MTE at the HUC6 level. The CR was especially strong with the estimation of the linear trend for the HUC2 basins. This seems quite a feat if one considers that the model was not calibrated against water-balance data as it is calibration-free (however, setting the PT α value is required; detailed in the Appendix) and employs only a minimum number of input variables, such as net surface radiation, wind speed, and air- and dew-point temperature (or any other humidity measure) which also make it easily applicable on a global scale (with the PT α possibly set by, e.g., continents). Future research may clarify if setting the PT α may require further regionalization techniques such as discussed by Samaniego et al. (2010). As of today it is not clear what surface

and/or boundary layer flow parameters the value of the PT α depends on despite of several exploratory studies (de Bruin, 1983; de Bruin & Keijman, 1979; Heerwaarden et al., 2009; Lhomme, 1997; Szilagyi, 2015).

For these reasons, it is believed that the CR model here described could serve as a benchmark to other ET estimation models requiring a significantly larger number of input variables often obtained by state-of-the-art data acquisition (often remotely sensed) and assimilation systems while running on strong computer platforms. These authors believe that this CR could especially benefit the LSMs of current climate models with their calibration (when water-balance data or surface flux measurements of latent and sensible heat are absent or of poor quality) and “back-evaluation” to verify that the atmospheric measurements are indeed in balance with the LSM-predicted latent heat fluxes, especially in future climate scenarios when water balance data are nonexistent for verification. Note that the CR is only a diagnostic tool for the ET rates, as it estimates the cause (ET) from the effect (moisture content and temperature of the air), and therefore, it cannot be employed for predictions in the same sense as those models that predict the ET rate at a future time (i.e., at $t + dt$) from, for example, the soil moisture and vegetation status obtained at time t .

The success of the present version of the CR lies in its tracking the state of an air parcel, in contact with the surface, via adiabatic processes. It directly relates the surface latent heat flux averaged over a suitable spatial and temporal scale (allowing for a full adjustment of the atmospheric boundary layer to surface properties and fluxes) to the average state of the air overlying the land surface and not through a proxy, like soil moisture. It also makes use of the possible limit values in regional evaporation rates under a given surface-available energy and wind conditions together with how fast evaporation rates may depart from their limit values with changing atmospheric conditions of air humidity and temperature, that is, the very same atmospheric response that changing ET rates triggered in the first place.

Appendix A

Under a constant $R_n - G$ term at the drying surface the sum of the latent (LE ; i.e., the evaporation rate expressed in energy flux units) and sensible (H) heat fluxes is also constant, from which it follows that a change in one will cause a similar but opposite change (δ) in the other, that is, $\delta LE = -\delta H$. The corresponding response in terms of vapor pressure (e) and temperature (T) of the air in contact with the surface will follow the adiabatic line of Figure A2 (Monteith, 1981). With H changing, the surface temperature (T_s) will also change proportionally (proportionality denoted from here on by the \propto symbol), which will affect the air temperature, so that one can write $-\delta LE = \delta H \propto \delta T_s \propto \delta T = -\delta e/\gamma$ (the latter from the adiabat), thus obtaining $\delta LE \propto \delta e/\gamma$. By normalizing the changes with their maximum range, and taking advantage of the linear nature of the adiabat, yields the following nondimensional form

$$\frac{E - E_{dry}}{E_w - E_{dry}} = \frac{E}{E_w} \propto \frac{e - e_{dry}}{e(T_w) - e_{dry}} = \frac{e}{e(T_w)} \quad (A1)$$

Under a constant $R_n - G$ and f_u terms T_w (similar to T_{wb}) stays constant [and thus $e(T_w)$ too, which then can be replaced by any other constant in the proportionality, such as $e^*(T_w)$ or $e^*(T_{wb})$] under adiabatic changes of the surface air (Monteith, 1981; Szilagyi & Schepers, 2014). Therefore, the proportionality (equation (A1)) can be written as

$$\frac{E}{E_w} \propto \frac{e^*(T_d)}{e^*(T_w)} \quad (A2)$$

Szilagyi et al. (2017) employed a polynomial approach, similar to Brutsaert (2015), for the proper functional form of the proportionality in equation (A2) by considering expected boundary conditions for the function, but could not improve upon some already existing versions of the CR.

In the classical versions of the CR (Brutsaert & Stricker, 1979; Morton, 1983), changes in the vapor pressure are replaced by corresponding changes in the E_p term normalized also by E_w as E is in equation (A2) and are inserted into a first order (i.e., linear) polynomial (as a first guess for the unknown functional relationship) by considering that $1 \leq E_p/E_w \leq 2$, while E/E_w takes up values of 1 and 0, respectively, at the limit/boundary

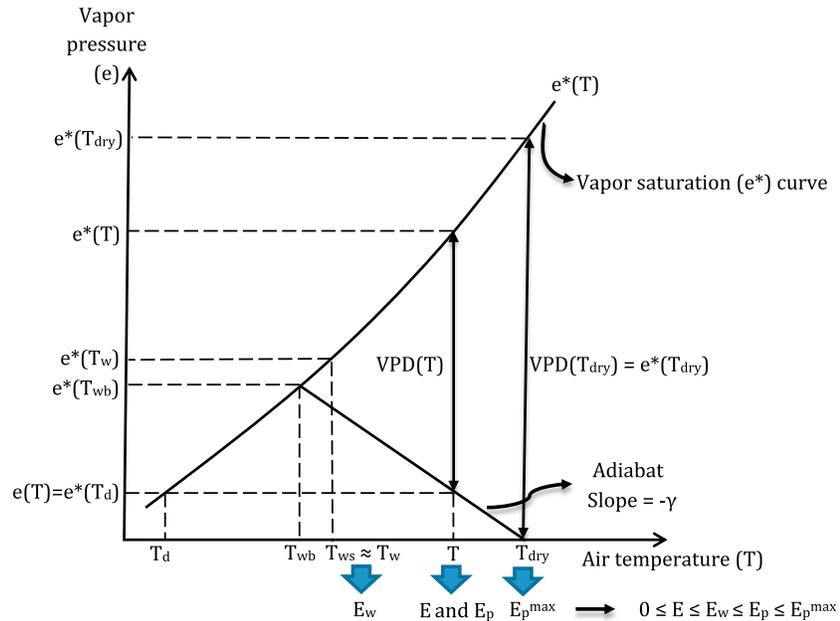


Figure A1. Air temperature (T)–vapor pressure (e) relationships. See the text for the definitions of the different terms displayed.

values of 1 and 2. Note that the upper limit value of E_p/E_w assumes, in addition to linearity, a symmetrical relationship between $\delta(E/E_w)$ and $\delta(E_p/E_w)$ around the shared value of unity (Figure A2). The solution for the resulting set of linear equations thus becomes

$$\frac{E}{E_w} = 2 - \frac{E_p}{E_w} \quad (\text{A3})$$

The replacement of $e(T) [=e^*(T_d)]$ by a linear function of E_p in equation (A2) works because the second term of the Penman equation for E_p contains the vapor pressure deficit [VPD = $e^*(T) - e(T)$] which changes in accordance (Figure A1) with $e(T)$ (but opposite in sign) along the adiabatic line [and slightly moderated by the change in the $\gamma/(\Delta + \gamma)$ term of equation (4) with temperature]. As seen from Figure A1, the VPD term is much more sensitive to temperature changes than the vapor pressure term and may explain its better success with predicting ET.

Equation (A3) yields two curves (one for E/E_w and another for E_p/E_w as functions of E_w/E_p , the latter a possible measure of water availability on the land surface) in Figure A2a, axisymmetric around the reference level of unity, when $E = E_w = E_p$. As the land dries out of a completely wet stage (the [1, 1] coordinate point in the graph), E is decreasing (therefore, T is increasing and with it e decreases along the adiabatic line of Figure A1) with water availability, and therefore, E_p is increasing due to corresponding increases in VPD.

With a different grouping of the evaporation terms, equation (A3) can equally be brought into a form such as

$$\frac{E}{E_p} = 2 \frac{E_w}{E_p} - 1 \quad (\text{A4})$$

displayed in Figure A2b. Note that the only way of ensuring that the theoretical lines overlap the measured data is via the choice of the PT α value of equation (6). As a result, the same water-balance derived E_{wb}/E_p values may appear in different locations (are shifted horizontally) in the graphs, depending on the α value that fits the given CR method.

As Kahler and Brutsaert (2006) and Szilagyi (2007) noted, there is no physical necessity for the CR to be symmetric since E_p can be interpreted in several ways, the Penman equation just one of them. E_p can equally be

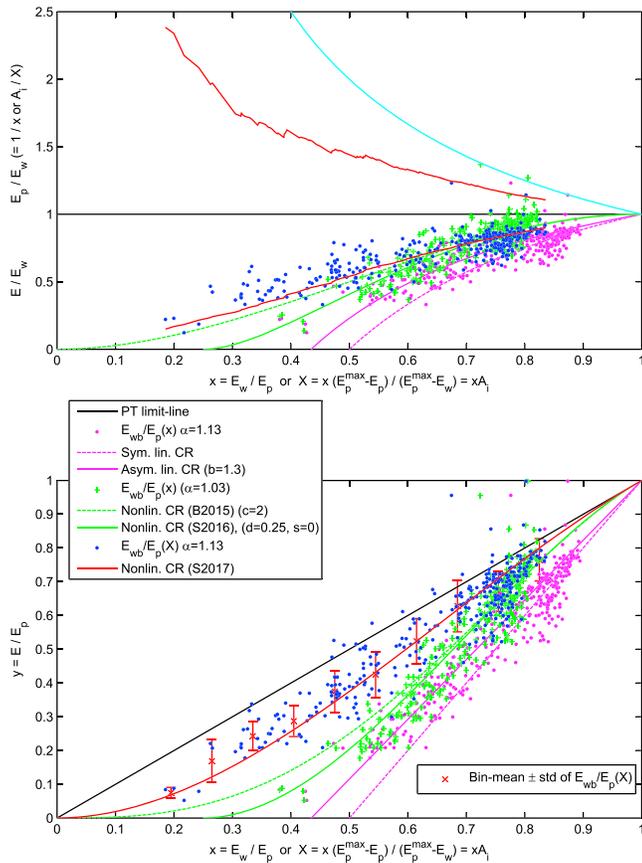


Figure A2. Theoretical relationships between the nondimensional terms of the different CR versions discussed in the Appendix, overlain a sample data set of mean annual HUC6 values across CONUS (data source: Szilagyi et al., 2017). No parameters (where appropriate) were calibrated for the data displayed which here serve only a demonstrational purpose between theoretical lines and data points. B2015 = Brutsaert (2015), S2016 = Szilagyi et al. (2016), S2017 = Szilagyi et al. (2017).

monotonic functional relationship between E/E_p and E_w/E_p as well as (ii) making sure that $y \leq x$ for $0 \leq x \leq 1$, which means that $E \leq E_w$ must always be met. With the $c = 2$ choice, displayed in Figure A2, one obtains

$$\frac{E}{E_p} = \left(\frac{E_w}{E_p}\right)^3 \left(3 - 2\frac{E_w}{E_p}\right) \quad (\text{A9})$$

and

$$\frac{E}{E_w} = \left(\frac{E_w}{E_p}\right)^2 \left(3 - 2\frac{E_w}{E_p}\right), \quad (\text{A10})$$

respectively. Szilagyi et al. (2016) pointed out that unless E_w is zero, the E_w/E_p term will always be larger than zero (thus, the lower BC is misconstrued in general) as the VPD term in the Penman equation is bounded by $e^*(T_{\text{dry}})$ (see Figure A1) and wind velocities at 2-m height above the ground are also limited in value. As a remedy, BC (ii) can be replaced by (ii) $y = 0$ at $x = d$ ($0 \leq d < 1$) since y may vanish already at $x = d$ (Szilagyi et al., 2016). To make the model even more flexible, Szilagyi et al. (2016) let the slope (s) of the nonlinear CR function be larger than zero at the origin, d , via the new BC (iv)

represented by different evaporation-pan data or by different wind functions (Ma et al., 2015) in the Penman equation. Therefore, Kahler and Brutsaert (2006) modified equation (A3) but still keeping its linear nature between its terms. The unity and zero values of E/E_p will now take place at the corresponding limit/boundary values of $1 \leq E_p/E_w \leq 1 + b$, where $b > 0$. The solution thus becomes

$$\frac{E}{E_w} = \frac{1+b}{b} - \frac{1}{b} \frac{E_p}{E_w} \quad (\text{A5})$$

which translates into

$$\frac{E}{E_p} = \frac{1+b}{b} \frac{E_w}{E_p} - \frac{1}{b}. \quad (\text{A6})$$

See Figure A2 for the corresponding curves with the $b = 1.3$ choice, as illustration. The CR curves for E/E_w and E_p/E_w in Figure A2a are no longer symmetric.

By allowing for a higher-order (i.e., third-order) polynomial between $y = E/E_p$ and $x = E_w/E_p$ and taking into consideration the following boundary conditions (BC): (i) $y = 1$ at $x = 1$, (ii) $y = 0$ at $x = 0$, (iii) $dy/dx = 1$ at $x = 1$, and (iv) $dy/dx = 0$ at $x = 0$, Brutsaert (2015) obtained the $y = 2x^2 - x^3$ nonlinear CR. When written out in the original variables it yields

$$\frac{E}{E_p} = \left(\frac{E_w}{E_p}\right)^2 \left(2 - \frac{E_w}{E_p}\right) \quad (\text{A7})$$

or equally

$$\frac{E}{E_w} = \frac{E_w}{E_p} \left(2 - \frac{E_w}{E_p}\right). \quad (\text{A8})$$

To make it more flexible Brutsaert (2015) introduced an additional parameter, c , which thus yields $y = (2 - c)x^2 + (2c - 1)x^3 - cx^4$. Szilagyi et al. (2016) warned that $-1 \leq c \leq 2$ must be met for (i) a

$dy/dx = s$ at $x = d$, with $s \geq 0$. Solution of the third-order polynomial of $y = a_3x^3 + a_2x^2 + a_1x + a_0$ thus becomes

$$\begin{aligned} a_3 &= (d-s+ds+1)(d-1)^{-3}, a_2 = (2s-d^2s-2d^2-2d-ds-2)(d-1)^{-3}, \\ a_1 &= (d^3+2d^2s+d^2+4d-s-ds)(d-1)^{-3}, a_0 = (ds-2d^2-d^2s)(d-1)^{-3} \end{aligned} \quad (A11)$$

with the admissible range for the slope as $0 \leq s \leq (2d+1)/(1-d)$, resulting from the (monotony and PT limit line) constraints above on the behavior of the expected CR function. Note that equation (A11) is valid for $x \geq d$; when $x < d$, $y = 0$ is expected (Szilagyi et al., 2016). For the E/E_w term the polynomial becomes (with the same coefficients above)

$$\frac{E}{E_w} = a_3x^2 + a_2x + a_1 + a_0x^{-1}. \quad (A12)$$

See Figure A2 for the two sample curves when $d = 0.25$ and $s = 0$.

The nonlinear CR solution of Szilagyi et al. (2016) improves upon the unrealistic boundary value choice of $E_w/E_p = 0$ by Brutsaert (2015) by exploiting that E_p is bounded. The unrealistic zero value of x by Brutsaert (2015) leads to similarly unrealistic low values of the PT α ($=1.03$ in Figure A2) in Brutsaert's (2015) solution (see Brutsaert et al., 2017), and also to the present choice of the largest possible value of c ($=2$), the latter making the CR curve as wide as possible in Figure A2, both necessary to bring observations match with theory in Figure A2. The problem though with the improved BC of Szilagyi et al. (2016) becomes that it is just a constant, while the maximum value (E_p^{\max}) of the E_p term may change by the measurement period. These E_p^{\max} values for each measurement period however can be estimated by the Penman equation through the $e^*(T_d) = 0$ choice, and by the help of the adiabatic line of Figure A1 for the estimation of the corresponding dry-environment air temperature, T_{dry} (equation (9)), reached when the environment becomes completely dry, or is very close to it. A proper scaling of the $x = E_w/E_p$ values thus becomes as $X = (E_p^{\max} - E_p)(E_p^{\max} - E_w)^{-1}x$, which ensures that X indeed is zero when E_p has reached its maximum value of E_p^{\max} (Crago et al., 2016; Szilagyi et al., 2017) and allows X become unity when the environment is wet, that is, $E = E_w = E_p$. The $A_i = (E_p^{\max} - E_p)(E_p^{\max} - E_w)^{-1}$ term itself can be considered as an index of aridity (A_i), having a value of zero under total lack of moisture conditions and a value of unity in a wet environment. See Szilagyi et al. (2017) for the mean annual value of A_i over CONUS. By employing the same BC's Brutsaert (2015) prescribed but now for X , one obtains (Szilagyi et al., 2017)

$$\frac{E}{E_p} = X^2(2-X) \quad (A13)$$

or equally

$$\frac{E}{E_w} = A_i X(2-X). \quad (A14)$$

Note the slight fluctuations in the equation (A14) curve of Figure A2a and in the corresponding E_p/E_w curve ($=A_i/X$) due to the emergence of A_i outside of X (but not in equation (A13)).

Let us finally discuss Brutsaert's (2015) BCs of (iii) $dy/dX = 1$ at $X = 1$ and (iv) $dy/dX = 0$ at $X = 0$, employed now for X . Let us use finite differences for the derivatives and denote the new values the variables acquire through the changes with an index of "2". At $X = 1$ one can write

$$\delta y = \delta \left(\frac{E}{E_p} \right) = \frac{E_2}{E_{p2}} - 1 = \frac{E_2 - E_{p2}}{E_{p2}} = \frac{[1 - \phi(0)]E_w - E_{p2}}{E_{p2}} \quad (A15)$$

where the ϕ symbol means a positive number very close to the one shown in its argumentum. The corresponding change in X is

$$\delta X = \frac{E_p^{max} - E_{p2}}{E_p^{max} - E_w} \frac{E_w}{E_{p2}} - 1 = \frac{[1 - \phi(0)]E_w - E_{p2}}{E_{p2}} \quad (A16)$$

Thus, dy/dX at $X = 1$ is unity. The same changes at $X = 0$ are

$$\delta y = \delta \left(\frac{E}{E_p} \right) = \frac{E_2}{E_{p2}} - \frac{0}{E_p^{max}} = \frac{E_2}{E_{p2}} = \phi(0) \quad (A17)$$

while

$$\delta X = \frac{E_p^{max} - E_{p2}}{E_p^{max} - E_w} \frac{E_w}{E_{p2}} - 0 = \frac{E_w}{E_p^{max} - E_w} \frac{E_p^{max} - E_{p2}}{E_{p2}} \quad (A18)$$

where the change, $E_p^{max} - E_{p2}$, in E_p is likely much larger than the corresponding change in E (and thus in T , therefore in e along the adiabat), because of the steep slope of the $e^*(T)$ curve at high temperatures causing a large change in VPD and thus in E_p due to a relatively small change in T (see Figure A1). As a result, $dy/dX = 0$ at $X = 0$ can be expected as the $E_w (E_p^{max} - E_w)^{-1}$ term can never be too small (will not vanish).

Finally, the nonlinear CR model (equations (A13) and (A14)) with BCs (i) $y = 1$ at $X = 1$, (ii) $y = 0$ at $X = 0$, (iii) $dy/dX = u (\geq 1)$ at $X = 1$, and (iv) $dy/dX = v (> 0)$ at $X = 0$ were derived and tested (the results not displayed) with calibrated values of u and v with the help of water-balance data of the HUC6 watersheds of CONUS to see if relaxing BCs (iii) and (iv) for X would result in better ET predictions. None of these calibrated models performed significantly better than the calibration-free CR version applied in this study, which shows at the very least that the current formulation is indeed highly robust and the assumptions/approximations as a whole, employed in its derivation, are not far-fetched from reality.

Appendix B

Estimation of a spatially and temporally constant value of α can be obtained with the help of the Priestley and Taylor (1972), equation (6), inserted into the Bowen ratio (Bowen, 1926) written for a wet environment as

$$\frac{R_n - G - E_w}{E_w} = \frac{1 - \alpha \frac{\Delta(T_w)}{\Delta(T_w) + \gamma}}{\alpha \frac{\Delta(T_w)}{\Delta(T_w) + \gamma}} = \gamma \frac{T_{ws} - T_w}{e^*(T_{ws}) - e_a} \quad (B1)$$

in which all variables are defined in section 2.1. After rearrangement of equation (B1) one obtains

$$\alpha = \frac{[\Delta(T_w) + \gamma][e^*(T_{ws}) - e_a]}{\Delta(T_w)\{\gamma[T_{ws} - T_w] + [e^*(T_{ws}) - e_a]\}} \quad (B2)$$

For large-scale gridded data, it is possible to identify wet cells as they must satisfy certain predefined requirements, thereby calculating their α values that must fall within the theoretical limits of $\{1, [\Delta(T_w) + \gamma]/\Delta(T_w)\}$ (Priestley & Taylor, 1972). Note that when a cell is wet then T_w in equations (B1) and (B2) equals the measured T , therefore these equations must be supplied by T during the identification process of the possible wet cells. For cells where the measured T is not equal to T_w (i.e., are not wet), the α value of equation (B2) will likely be outside the specified interval. Following Szilagyi et al. (2017), a grid cell is considered “wet” provided that (i) $T_{ws} > T + 2$ °C (note that if a cell is not wet then $T_{ws} < T$ or $T_{ws} \approx T$ is likely), (ii) relative humidity (RH) is larger than 90%, and (iii) the equation (B2) calculated α value falls into the above specified interval. $RH = [100 e^*(T_d)/e^*(T)]$ was derived using T and T_d data from PRISM. T_{ws} comes from equation (7) and here should not be limited by T . The average value of α resulting from such wet cells yielded a value of 1.15 for CONUS, which was retained for use in the current CR model.

Appendix C

Pseudocode of ET Rate (mm/day) Estimations

The calculations may be performed by temporally averaged input values over intervals of integer multiples of a day. Temporal averaging of the input variables five days or longer are expected to yield the most accurate ET estimates due to the assumptions inherent in the complementary relationship (CR) of evaporation. The optimal spatial scale for the CR is 1 km² and up.

Input variables:

Air temperature (°C) T

Dew-point temperature (°C) T_d

2-m wind velocity (m/s) u_2

Net surface radiation (MJ·m⁻²·d⁻¹) R_n

Soil heat flux (MJ·m⁻²·d⁻¹) G

Constants:

Specific heat of dry air $c_p = 1.005 \times 10^{-3} \text{ MJ} \cdot \text{kg}^{-1} \cdot \text{K}^{-1}$

Specific gas constant of dry air $R_a = 287 \text{ J} \cdot \text{kg}^{-1} \cdot \text{K}^{-1}$

Gravitational acceleration $g = 9.81 \text{ m/s}^2$

Standard temperature lapse rate $l_r = 0.0065 \text{ K/m}$

Latent heat of vaporization for water (at 10 °C) $L_v = 2.48 \text{ MJ/kg}$

Density of water $\rho_w \approx 1,000 \text{ kg/m}^3$

Formulas:

Tetens equation of the saturation vapor pressure (hPa) $e^* = 6.108e^{\frac{17.27T}{237.3+T}}$; T in °C

Slope of the vapor-pressure curve (hPa/K) $\Delta = \frac{4098e^*}{(237.3+T)^2}$; T in °C

Barometric pressure (hPa) at elevation z (m) $P = 1,013 \left(\frac{T+273.16}{T+273.16+l_r z} \right)^{\frac{g}{l_r R_a}}$

Psychrometric constant (hPa/K) $\gamma = \frac{c_p P}{0.622 L_v}$

From wind speed (u_h) at h m above ground to 2-m wind (u_2) $u_2 = u_h \left(\frac{2}{h} \right)^{1/7}$

Energy flux to water-depth equivalent R_n [mm/day] = $10^3 R_n$ [MJ·m⁻²·d⁻¹] $\rho_w^{-1} L_v^{-1}$; the same applies for G

Data input: T, T_d, u_2, G, R_n

Output: Estimated actual ET rate, E , in mm/day

Start

1. Convert R_n and G into water-depth equivalent of mm/day
2. Calculate E_p by equations (4) and (5)
3. Calculate T_{ws} by iteratively solving equation (7), cap the value of T_{ws} by T , if $T_{ws} > T$
4. Calculate T_{wb} by iteratively solving equation (10)
5. Calculate T_{dry} by equation (9)
6. With gridded large-scale data of the input variables follow Appendix B for obtaining α , otherwise choose a trial value from [1.1–1.32]
7. Assign the value of T_{ws} to T_w and calculate E_w by equation (6)
8. Calculate E_p^{\max} by equation (8)
9. Calculate the actual ET rate (E) by equations (1)–(3)

End

The value of α can be calibrated (provided that measured ET rates are available) by trial and error relatively fast as it is typically satisfactory to specify it to two decimals, and as such, there are only 23 distinct values within the [1.1–1.32] interval.

Acknowledgments

This research was supported by the BME-Water Sciences & Disaster Prevention FIKP grant of EMMI (BME FIKP-VIZ), the National Key Research and Development Program of China (2017YFA0603101), the National Natural Science Foundation of China (41801047, 41430748, and 41661144025), and the China Postdoctoral Science Foundation (2019T120139, 2017LH032, and 2017M620069). We thank Xitian Cai (Lawrence Berkeley National Laboratory) for his help in producing HUC6 river mask. Special appreciations are given to the three anonymous reviewers for their valuable comments resulting in an improved manuscript. All data used in this study can be accessed from the websites as follows: NARR data (www.esrl.noaa.gov/psd/data/gridded/data.narr.html); PRISM temperature, humidity, and precipitation (<http://prism.oregonstate.edu/>); USGS HUC2 and HUC6 runoff (https://waterwatch.usgs.gov/?id=wwds_runoff); GRACE (<http://grace.jpl.nasa.gov/>); Noah (https://disc.gsfc.nasa.gov/datasets/NLDAS_NOAH0125_M_V002/summary?keywords=NLDAS); VIC (https://disc.gsfc.nasa.gov/datasets/NLDAS_VIC0125_M_V002/summary?keywords=NLDAS); Mosaic (https://disc.gsfc.nasa.gov/datasets/NLDAS_MOS0125_M_V002/summary?keywords=NLDAS); NCEP-II (<https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html>); ERA-Interim (<https://www.ecmwf.int/en/forecasts/datasets/archive-datasets/reanalysis-datasets/era-interim>); GLEAM (<https://www.gleam.eu/#home>); PML (<https://data.csiro.au/dap/landingpage?pid=csiro:17375&v=2&d=true>); and FLUXNET-MTE (<https://www.bgc-jena.mpg.de/geodb/projects/Home.php>). The CR-modeled ET product and the HUC2- and HUC6-averaged ET_{wb} , P , Q , and δS data are available from <https://digitalcommons.unl.edu/natrespapers/986/>.

References

Allen, R. G., Pereira, L. S., Howell, T. A., & Jensen, M. E. (2011). Evapotranspiration information reporting: I. Factors governing measurement accuracy. *Agricultural Water Management*, 98(6), 899–920. <https://doi.org/10.1016/j.agwat.2010.12.015>

Ashfaq, M., Ghosh, S., Kao, S.-C., Bowling, L. C., Mote, P., Touma, D., et al. (2013). Near-term acceleration of hydroclimatic change in the western U.S. *Journal of Geophysical Research: Atmospheres*, 118, 10,676–10,693. <https://doi.org/10.1002/jgrd.50816>

Beck, H. E., van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & de Roo, A. (2017). MSWEP: 3-hourly 0.25 global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, 21(1), 589–615. <https://doi.org/10.5194/hess-21-589-2017>

Bleed, A., & Flowerday, C. (Eds) (1989). *An atlas of the Sand Hills*. University of Nebraska-Lincoln: Conservation and Survey Division. ISBN 1-56161-000-3

Bouchet, R. J. (1963). Evapotranspiration réelle et potentielle, signification climatique. *International Association of Scientific Hydrology, Publication*, 62, 134–142.

Bowen, I. S. (1926). The ratio of heat losses by conduction and by evaporation from any water surface. *Physics Review*, 27(6), 779–787. <https://doi.org/10.1103/PhysRev.27.779>

Brakebill, J., Wolock, D., & Terziotti, S. (2011). Digital hydrologic networks supporting applications related to spatially referenced regression modeling. *Journal of the American Water Resources Association*, 47(5), 916–932. <https://doi.org/10.1111/j.1752-1688.2011.00578.x>

Brutsaert, W. (1982). *Evaporation into the atmosphere: Theory, history and applications*. Dordrecht, Holland: D. Reider. <https://doi.org/10.1007/978-94-017-1497-6>

Brutsaert, W. (2015). A generalized complementary principle with physical constraints for land-surface evaporation. *Water Resources Research*, 51, 8087–8093. <https://doi.org/10.1002/2015WR017720>

Brutsaert, W., Li, W., Takahashi, A., Hiyama, T., Zhang, L., & Liu, W. (2017). Nonlinear advection-aridity method for landscape evaporation and its application during the growing season in the southern Loess Plateau of the Yellow River basin. *Water Resources Research*, 53, 270–282. <https://doi.org/10.1002/2016WR019472>

Brutsaert, W., & Stricker, H. (1979). An advection-aridity approach to estimate actual regional evapotranspiration. *Water Resources Research*, 15(2), 443–450. <https://doi.org/10.1029/WR015i002p00443>

Cai, X., Yang, Z.-L., Xia, Y., Huang, M., Wei, H., Leung, L. R., & Ek, M. B. (2014). Assessment of simulated water balance from Noah, Noah-MP, CLM, and VIC over CONUS using the NLDAS test bed. *Journal of Geophysical Research: Atmospheres*, 119, 13,751–13,770. <https://doi.org/10.1002/2014JD022113>

Chaney, N. W., Minasny, B., Herman, J. D., Nauman, T. W., Brungard, C. W., Morgan, C. L. S., et al. (2019). POLARIS soil properties: 30-m probabilistic maps of soil properties over the contiguous United States. *Water Resources Research*, 55, 2916–2938. <https://doi.org/10.1029/2018wr022797>

Chen, F., & Dudhia, J. (2001). Coupling an advanced land surface-hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*, 129(4), 569–585. [https://doi.org/10.1175/1520-0493\(2001\)129<0569:CAALSH>2.0.CO;2](https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2)

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., et al. (2017). The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism. *Hydrology and Earth System Sciences*, 21(7), 3427–3440. <https://doi.org/10.5194/hess-21-3427-2017>

Crago, R., & Qualls, R. (2018). Evaluation of the generalized and rescaled complementary evaporation relationships. *Water Resources Research*, 54, 8086–8102. <https://doi.org/10.1029/2018WR023401>

Crago, R., Szilagyi, J., Qualls, R., & Huntington, J. (2016). Rescaling the complementary relationship for land surface evaporation. *Water Resources Research*, 52, 8461–8471. <https://doi.org/10.1002/2016WR019753>

Dai, Y., Yuan, H., Xin, Q., Wang, D., Shangguan, W., Zhang, S., et al. (2019). Different representations of canopy structure—A large source of uncertainty in global land surface modeling. *Agricultural and Forest Meteorology*, 269–270, 119–135. <https://doi.org/10.1016/j.agrformet.2019.02.006>

Daly, C., Neilson, R. P., & Phillip, D. L. (1994). A statistical topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology*, 33(2), 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. A. R. (1983). A model for the Priestley-Taylor parameter α . *Journal of Climatology and Applied Meteorology*, 22(4), 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)

de Bruin, H. A. R., & Keijman, J. Q. (1979). The Priestley-Taylor evaporation model applied to a large, shallow lake in the Netherlands. *Journal of Applied Meteorology*, 18(7), 898–903. [https://doi.org/10.1175/1520-0450\(1979\)018<0898:TPTEMA>2.0.CO;2](https://doi.org/10.1175/1520-0450(1979)018<0898:TPTEMA>2.0.CO;2)

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), 553–597. <https://doi.org/10.1002/qj.828>

de Vries, D. A. (1959). The influence of irrigation on the energy balance and the climate near the ground. *Journal of Meteorology*, 16(3), 256–270. [https://doi.org/10.1175/1520-0469\(1959\)016<0256:TIOIOT>2.0.CO;2](https://doi.org/10.1175/1520-0469(1959)016<0256:TIOIOT>2.0.CO;2)

Fisher, J. B., Tu, K. P., & Baldocchi, D. D. (2008). Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. *Remote Sensing of Environment*, 112(3), 901–919. <https://doi.org/10.1016/j.rse.2007.06.025>

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., & Andréassian, V. (2014). Large-sample hydrology: A need to balance depth with breadth. *Hydrology and Earth System Sciences*, 18(2), 463–477. <https://doi.org/10.5194/hess-18-463-2014>

Gutman, G., & Ignatov, A. (1998). The derivation of green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of Remote Sensing*, 19(8), 1533–1543. <https://doi.org/10.1080/014311698215333>

- Han, E., Crow, W. T., Hain, C. R., & Anderson, M. C. (2015). On the use of a water balance to evaluate interannual terrestrial ET variability. *Journal of Hydrometeorology*, *16*(3), 1102–1108. <https://doi.org/10.1175/jhm-d-14-0175.1>
- Han, S., Hu, H., & Tian, F. (2012). A nonlinear function approach for the normalized relationship evaporation model. *Hydrological Processes*, *26*, 3973–3981. <https://doi.org/10.1002/hyp.8414>
- Han, S., & Tian, F. (2018). Derivation of a sigmoid generalized complementary function for evaporation with physical constraints. *Water Resources Research*, *54*, 5050–5068. <https://doi.org/10.1029/2017WR021755>
- Hanak, E., Lund, J., Dinar, A., Gray, B., Howitt, R., Mount, J., et al. (2011). *Managing California's water: From conflict to reconciliation*. Public Policy Institute of California.
- Hansen, M. C., Defries, R. S., Townshend, J. R. G., & Sohlberg, R. (2010). Global land cover classification at 1 km spatial resolution using a classification tree approach. *International Journal of Remote Sensing*, *21*(6-7), 1331–1364. <https://doi.org/10.1080/014311600210209>
- Heerwaarden, C. C., Arellano, J. V., Moene, A. F., & Holtlag, A. A. M. (2009). Interactions between dry-air entrainment, surface evaporation and convective boundary-layer development. *Quarterly Journal of the Royal Meteorological Society*, *135*(642), 1277–1291. <https://doi.org/10.1002/qj.431>
- Hu, Z., Wang, G., Sun, X., Zhu, W., Song, C., Huang, K., & Chen, X. (2018). Spatial-temporal patterns of evapotranspiration along an elevation gradient on Mount Gongga, southwest China. *Water Resources Research*, *54*, 4180–4192. <https://doi.org/10.1029/2018WR022645>
- Huntington, J. L., Szilagyi, J., Tyler, S. W., & Pohl, G. M. (2011). Evaluating the complementary relationship for estimating evaporation from arid shrublands. *Water Resources Research*, *47*, W05533. <https://doi.org/10.1029/2010WR009874>
- Jung, M., Henkel, K., Herold, M., & Churkina, G. (2006). Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sensing of Environment*, *101*(4), 534–553. <https://doi.org/10.1016/j.rse.2006.01.020>
- Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., et al. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research*, *116*, G00J07. <https://doi.org/10.1029/2010jg001566>
- Kahler, D. M., & Brutsaert, W. (2006). Complementary relationship between daily evaporation in the environment and pan evaporation. *Water Resources Research*, *42*, W05413. <https://doi.org/10.1029/2005WR004541>
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J. J., Fiorino, M., & Potter, G. L. (2002). NCEP-DOE AMIP-II Reanalysis (R2). *Bulletin of the American Meteorological Society*, *83*(11), 1631–1644. <https://doi.org/10.1175/BAMS-83-11>
- Koster, R., & Suarez, M. (1996). Energy and water balance calculations in the Mosaic LSM. Report, 58 pp. Goddard Space Flight Center, National Aeronautics and Space Administration, Greenbelt, Maryland.
- Landerer, F. W., & Swenson, S. C. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, *48*, W04531. <https://doi.org/10.1029/2011wr011453>
- LeMone, M. A., Tewari, M., Chen, F., Alfieri, J. G., & Niyogi, D. (2008). Evaluation of the Noah land surface model using data from a fair-weather IHOP_2002 day with heterogeneous surface fluxes. *Monthly Weather Review*, *136*(12), 4915–4941. <https://doi.org/10.1175/2008mwr2354.1>
- Leuning, R., Zhang, Y. Q., Rajaud, A., Cleugh, H., & Tu, K. (2008). A simple surface conductance model to estimate regional evaporation using MODIS leaf area index and the Penman-Monteith equation. *Water Resources Research*, *44*, W10419. <https://doi.org/10.1029/2007WR006562>
- Lhomme, J. P. (1997). A theoretical basis for the Priestley-Taylor coefficient. *Boundary-Layer Meteorology*, *82*(2), 179–191. <https://doi.org/10.1023/A:1000281114105>
- Liang, X., Lettenmaier, D. P., Wood, E. F., & Burges, S. J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, *99*(D7), 14,415–14,428. <https://doi.org/10.1029/94JD00483>
- Liu, W., Wang, L., Zhou, J., Li, Y., Sun, F., Fu, G., et al. (2016). A worldwide evaluation of basin-scale evapotranspiration estimates against the water balance method. *Journal of Hydrology*, *538*, 82–95. <https://doi.org/10.1016/j.jhydrol.2016.04.006>
- Lu, Y., Steele-Dunne, S. C., Farhadi, L., & van de Giesen, N. (2017). Mapping surface heat fluxes by assimilating SMAP soil moisture and GOES land surface temperature data. *Water Resources Research*, *53*, 10,858–10,877. <https://doi.org/10.1002/2017WR021415>
- Ma, N., Niu, G.-Y., Xia, Y., Cai, X., Zhang, Y., Ma, Y., & Fang, Y. (2017). A systematic evaluation of Noah-MP in simulating land-atmosphere energy, water, and carbon exchanges over the continental United States. *Journal of Geophysical Research: Atmospheres*, *122*, 12,245–12,268. <https://doi.org/10.1002/2017JD027597>
- Ma, N., Szilagyi, J., Zhang, Y., & Liu, W. (2019). Complementary-relationship-based modeling of terrestrial evapotranspiration across China during 1982–2012: Validations and spatiotemporal analyses. *Journal of Geophysical Research: Atmospheres*, *124*, 4326–4351. <https://doi.org/10.1029/2018JD029850>
- Ma, N., Zhang, Y., Szilagyi, J., Guo, Y., Zhai, J., & Gao, H. (2015). Evaluating the complementary relationship of evapotranspiration in the alpine steppe of the Tibetan Plateau. *Water Resources Research*, *51*, 1069–1083. <https://doi.org/10.1002/2014wr015493>
- Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., et al. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925. <https://doi.org/10.5194/gmd-10-1903-2017>
- Masson, V., Champeaux, J.-J., Chauvin, F., Meriguet, C., & Lacaze, R. (2003). A global database of land surface parameters at 1 km resolution in meteorological and climate models. *Journal of Climate*, *9*, 1261–1282.
- Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., & Gupta, H. (2015). Are we unnecessarily constraining the agility of complex process-based models? *Water Resources Research*, *51*, 716–728. <https://doi.org/10.1002/2014wr015820>
- Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P. C., Ebisuzaki, W., et al. (2006). North American Regional Reanalysis. *Bulletin of the American Meteorological Society*, *87*(3), 343–360. <https://doi.org/10.1175/BAMS-87-3-343>
- Miller, D. A., & White, R. A. (1998). A conterminous United States multilayer soil characteristics data set for regional climate and hydrology modeling. *Earth Interactions*, *2*(2), 1–26. [https://doi.org/10.1175/1087-3562\(1998\)002<0001:ACUSMS>2.3.CO;2](https://doi.org/10.1175/1087-3562(1998)002<0001:ACUSMS>2.3.CO;2)
- Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., & Dolman, A. J. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, *15*(2), 453–469. <https://doi.org/10.5194/hess-15-453-2011>
- Mölders, N. (2005). Plant- and soil-parameter-caused uncertainty of predicted surface fluxes. *Monthly Weather Review*, *133*(12), 3498–3516. <https://doi.org/10.1175/MWR3046.1>
- Monteith, J. L. (1965). *Evaporation and environment, paper presented at Symposia of the Society for Experimental Biology*. Cambridge University Press.

- Monteith, J. L. (1981). Evaporation and surface temperature. *Quarterly Journal of the Royal Meteorological Society*, *107*(451), 1–27. <https://doi.org/10.1002/qj.49710745102>
- Morton, F. I. (1983). Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology. *Journal of Hydrology*, *66*(1-4), 1–76. [https://doi.org/10.1016/0022-1694\(83\)90177-4](https://doi.org/10.1016/0022-1694(83)90177-4)
- Mu, Q., Zhao, M., & Running, S. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, *115*(8), 1781–1800. <https://doi.org/10.1016/j.rse.2011.02.019>
- Oki, T., & Kanae, S. (2006). Global hydrological cycles and world water resources. *Science*, *313*(5790), 1068–1072. <https://doi.org/10.1126/science.1128845>
- Ortman, C. L., (2009). Okavango river basin technical diagnostic analysis: Environmental flow modul specialist report. The Permanent Okavango River Basin Water Commission, Maun, Botswana.
- Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *193*, 120–145. <https://doi.org/10.1098/rspa.1948.0037>
- Priestley, C. H. B., & Taylor, R. J. (1972). On the assessment of surface heat flux and evaporation using large-scale parameters. *Monthly Weather Review*, *100*(2), 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)
- Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*, W05523. <https://doi.org/10.1029/2008wr007327>
- Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., et al. (2017). Toward seamless hydrologic predictions across spatial scales. *Hydrology and Earth System Sciences*, *21*(9), 4323–4346. <https://doi.org/10.5194/hess-21-4323-2017>
- Seaber, P. R., Kapinos, F. P., & Knapp, G. L. (1987). Hydrologic unit maps. *U. S. Geological Survey Water-Supply Paper*, *2294*, 1–66.
- Seager, R., Ting, M., Held, I., Kushnir, Y., Lu, J., Vecchi, G., et al. (2007). Model projections of an imminent transition to a more arid climate in southwestern North America. *Science*, *316*(5828), 1181–1184. <https://doi.org/10.1126/science.1139601>
- Senay, G. B., Leake, S., Nagler, P. L., Artan, G., Dickinson, J., Cordova, J. T., & Glenn, E. P. (2011). Estimating basin scale evapotranspiration (ET) by water balance and remote sensing methods. *Hydrological Processes*, *25*(26), 4037–4049. <https://doi.org/10.1002/hyp.8379>
- Shangguan, W., Dai, Y., Liu, B., Zhu, A., Duan, Q., Wu, L., et al. (2013). A China data set of soil properties for land surface modeling. *Journal of Advances in Modeling Earth Systems*, *5*, 212–224. <https://doi.org/10.1002/jame.20026>
- Sheffield, J., Goteti, G., & Wood, E. (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, *19*(13), 3088–3111. <https://doi.org/10.1175/JCLI3790.1>
- Shi, X. Z., Yu, D. S., Warner, E. D., Pan, X. Z., Petersen, G. W., Gong, Z. G., & Weindorf, D. C. (2004). Soil database of 1:1,000,000 digital soil survey and reference system of the Chinese genetic soil classification system. *Soil Horizons*, *45*(4), 129. <https://doi.org/10.2136/sh2004.4.0129>
- Szilagyi, J. (2007). On the inherent asymmetric nature of the complementary relationship of evaporation. *Geophysical Research Letters*, *34*, L02405. <https://doi.org/10.1029/2006GL02870>
- Szilagyi, J. (2014). Temperature corrections in the Priestley-Taylor equation of evaporation. *Journal of Hydrology*, *519*, 455–464. <https://doi.org/10.1016/j.jhydrol.2014.07.040>
- Szilagyi, J. (2015). Water balance validation of a temperature-dependent parameter value of the Priestley-Taylor equation of evapotranspiration. *Journal of Hydrology and Environmental Research*, *3*(1), 1–5.
- Szilagyi, J. (2018a). Anthropogenic hydrological cycle disturbance at a regional scale: State-wide evapotranspiration trends (1979–2015) across Nebraska, USA. *Journal of Hydrology*, *557*, 600–612. <https://doi.org/10.1016/j.jhydrol.2017.12.062>
- Szilagyi, J. (2018b). A calibration-free, robust estimation of monthly land surface evapotranspiration rates for continental-scale hydrology. *Hydrology Research*, *49*(3), 648–657. <https://doi.org/10.2166/nh.2017.078>
- Szilagyi, J., Crago, R., & Qualls, R. J. (2016). Testing the generalized complementary relationship of evaporation with continental-scale long-term water-balance data. *Journal of Hydrology*, *540*, 914–922. <https://doi.org/10.1016/j.jhydrol.2016.07.001>
- Szilagyi, J., Crago, R., & Qualls, R. J. (2017). A calibration-free formulation of the complementary relationship of evaporation for continental-scale hydrology. *Journal of Geophysical Research: Atmospheres*, *122*, 264–278. <https://doi.org/10.1002/2016jd025611>
- Szilagyi, J., & Jozsa, J. (2008). New findings about the complementary relationship based evaporation estimation methods. *Journal of Hydrology*, *354*(1–4), 171–186. <https://doi.org/10.1016/j.jhydrol.2008.03.008>
- Szilagyi, J., & Jozsa, J. (2009). Analytical solution of the coupled 2-D turbulent heat and vapor transport equations and the complementary relationship of evaporation. *Journal of Hydrology*, *372*(1-4), 61–67. <https://doi.org/10.1016/j.jhydrol.2009.03.035>
- Szilagyi, J., & Jozsa, J. (2018). Evapotranspiration trends (1979–2015) in the Central Valley of California, USA: Contrasting tendencies during 1981–2007. *Water Resources Research*, *54*, 5620–5635. <https://doi.org/10.1029/2018WR022704>
- Szilagyi, J., & Schepers, A. (2014). Coupled heat and vapor transport: The thermostat effect of a freely evaporating land surface. *Geophysical Research Letters*, *41*, 435–441. <https://doi.org/10.1002/2013GL058979>
- Tapley, B. D., Watkins, M. M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., et al. (2019). Contributions of GRACE to understanding climate change. *Nature Climate Change*, *9*(5), 358–369. <https://doi.org/10.1038/s41558-019-0456-2>
- Teuling, A. J., Uijlenhoet, R., van den Hurk, B., & Seneviratne, S. I. (2009). Parameter sensitivity in LSMs: An Analysis using stochastic soil moisture models and ELDAS soil parameters. *Journal of Hydrometeorology*, *10*(3), 751–765. <https://doi.org/10.1175/2008jhm1033.1>
- Trenberth, K. E., Fasullo, J. T., & Kiehl, J. (2009). Earth's global energy budget. *Bulletin of the American Meteorological Society*, *90*(3), 311–324. <https://doi.org/10.1175/2008bams2634.1>
- Wang, P., Niu, G.-Y., Fang, Y.-H., Wu, R.-J., Yu, J.-J., Yuan, G.-F., et al. (2018). Implementing dynamic root optimization in Noah-MP for simulating phreatophytic root water uptake. *Water Resources Research*, *54*, 1560–1575. <https://doi.org/10.1002/2017WR021061>
- Xia, Y., Cosgrove, B. A., Mitchell, K. E., Peters-Lidard, C. D., Ek, M. B., Brewer, M., et al. (2016). Basin-scale assessment of the land surface water budget in the National Centers for Environmental Prediction operational and research NLDAS-2 systems. *Journal of Geophysical Research: Atmospheres*, *121*, 2750–2779. <https://doi.org/10.1002/2015JD023733>
- Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., et al. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research*, *117*, D03109. <https://doi.org/10.1029/2011jd016048>
- Xu, T., He, X., Bateni, S. M., Auligne, T., Liu, S., Xu, Z., et al. (2019). Mapping regional turbulent heat fluxes via variational assimilation of land surface temperature data from polar orbiting satellites. *Remote Sensing of Environment*, *221*, 444–461. <https://doi.org/10.1016/j.rse.2018.11.023>
- Zeng, Z., Piao, S., Li, L. Z. X., Zhou, L., Ciais, P., Wang, T., et al. (2017). Climate mitigation from vegetation biophysical feedbacks during the past three decades. *Nature Climate Change*, *7*(6), 432–436. <https://doi.org/10.1038/nclimate3299>

- Zhang, Y., Chiew, F. H. S., Peña-Arancibia, J., Sun, F., Li, H., & Leuning, R. (2017). Global variation of transpiration and soil evaporation and the role of their major climate drivers. *Journal of Geophysical Research: Atmospheres*, *122*, 6868–6881. <https://doi.org/10.1002/2017JD027025>
- Zhang, Y., Peña-Arancibia, J. L., McVicar, T. R., Chiew, F. H. S., Vaze, J., Liu, C., et al. (2016). Multi-decadal trends in global terrestrial evapotranspiration and its components. *Scientific Reports*, *6*(1), 19124. <https://doi.org/10.1038/srep19124>
- Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., et al. (2013). Global data sets of vegetation leaf area index (LAI)3g and fraction of photosynthetically active radiation (FPAR)3g derived from Global Inventory Modeling and Mapping Studies (GIMMS) normalized difference vegetation index (NDVI3g) for the period 1981 to 2011. *Remote Sensing*, *5*, 927–948. <https://doi.org/10.3390/rs5020927>
- Zink, M., Mai, J., Cuntz, M., & Samaniego, L. (2018). Conditioning a hydrologic model using patterns of remotely sensed land surface temperature. *Water Resources Research*, *54*, 2976–2998. <https://doi.org/10.1002/2017WR021346>
- Zurita-Milla, R., Laurent, V. C. E., & van Gijzel, J. A. E. (2015). Visualizing the ill-posedness of the inversion of a canopy radiativetransfer model: A case study for Sentinel-2R. *International Journal of Applied Earth Observation and Geoinformation*, *43*, 7–18. <https://doi.org/10.1016/j.jag.2015.02.003>