Journal of Hydrology 590 (2020) 125221

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Benchmarking large-scale evapotranspiration estimates: A perspective from a calibration-free complementary relationship approach and FLUXCOM

Ning Ma[a,b,c,*], Jozsef Szilagyi[d,e], Janos Jozsa[d,f]

[a] Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
[b] State Key Laboratory of Cryospheric Science, Chinese Academy of Sciences, Lanzhou, China
[c] Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China
[d] Department of Hydraulic and Water Resources Engineering, Budapest University of Technology and Economics, Budapest, Hungary
[e] Conservation and Survey Division, School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, USA
[f] MTA-BME Water Management Research Group, Budapest University of Technology and Economics, Budapest, Hungary

ARTICLE INFO

This manuscript was handled by G. Syme, Editor-in-Chief

Keywords:
Complementary relationship
FLUXCOM
Large-scale evapotranspiration
Benchmarking
Water balance

ABSTRACT

Accurately quantifying large-scale terrestrial evapotranspiration (ET) remains hampered by poor parameterization of the physical processes that relate to ET. Previous studies suggested that the calibration-free complementary relationship (CR) method that requires only routine meteorological data performed better than main-stream atmospheric reanalyses, land surface or remote sensing models in estimating large-scale ET. Here we simultaneously evaluate the latest machine learning-based upscaling of eddy-covariance measurements (FLUXCOM) and the CR estimates against the water-balance derived ET rates of 18 large Hydrologic Unit Code-2 (HUC2) and 327 medium HUC6 basins across the conterminous United States. Overall, CR and FLUXCOM perform comparably in representing the multiyear mean and temporal variations in annual ET at both, HUC2 and HUC6, scales for the 1979–2013 period. Such equally good skills also hold true for the 2003–2015 period, during which FLUXCOM was driven solely by remote sensing data. However, the CR generally captures the long-term linear tendencies in annual ET rates somewhat better than FLUXCOM. Because of its minimal data requirement, the calibration-free version of the CR may continue to serve as a benchmarking tool for large-scale ET simulations.

1. Introduction

A spatially and temporally explicit representation of land surface evapotranspiration (ET) at regional/global scales is essential to understanding the Earth's energy, water and carbon cycles (Jung et al., 2010; Lemordant et al., 2018; Miralles et al., 2019). However, estimation of large-scale ET remains saddled by difficulties in parameterizing the physical processes that control ET in land surface models (LSMs) (Ma et al., 2017; Stoy et al., 2019) and remote sensing (RS) algorithms (Liu, 2018; Zhang et al., 2016) because ET is affected by various environmental and biophysical factors (Sun et al., 2020; Teuling et al., 2019). Uncertainties can also result from poor accuracy in gridded vegetation and soil data due to the complexity/heterogeneity of terrestrial ecosystems, thus arguably presenting themselves as the key drivers of

Abbreviations: α, Priestley-Taylor coefficient, dimensionless; γ, psychrometric constant, hPa $^{\circ}C^{-1}$; Δ, slope of the saturation vapor pressure curve at T, hPa $^{\circ}C^{-1}$; Δ_{Tw}, slope of the saturation vapor pressure curve at T_{ws}, hPa $^{\circ}C^{-1}$; β, Bowen ratio, dimensionless; e_a, saturation vapor pressure at T_d, hPa; e_o, saturation vapor pressure at T, hPa; $e_{o,\ Tws}$, saturation vapor pressure at T_{ws}, hPa; $e_{o,\ Tdry}$, saturation vapor pressure at T_{dry}, hPa; $e_{o,\ Twb}$, saturation vapor pressure at T_{wb}, hPa; ET_p, potential evapotranspiration rate, mm d^{-1}; ET_{pmax}, dry-environment potential evapotranspiration rate, mm d^{-1}; ET_w, wet-environment evapotranspiration rate, mm d^{-1}; ET, actual evapotranspiration rate, mm d^{-1} in the CR model but mm yr^{-1} for evaluations of annual values; ET_{wb}, water-balance derived actual evapotranspiration rate, mm yr^{-1}; f_u, wind function mm d^{-1} hPa^{-1}; G, ground heat flux in water equivalent of mm d^{-1}; NSE, Nash-Sutcliffe efficiency, dimensionless; P, annual precipitation rate, mm yr^{-1}; Q, annual runoff, mm yr^{-1}; R_n, surface net radiation in water equivalent of mm d^{-1}; R, Pearson correlation coefficient, dimensionless; R_{IAV}, ratio of the interannual variability in modeled ET to that in ET_{wb}, dimensionless; RMSE, root mean square error, mm yr^{-1}; RB, relative bias, dimensionless; δS, change in terrestrial water storage, mm yr^{-1}; T, air temperature, °C; T_d, dew-point temperature, °C; T_{dry}, dry-environment air temperature, °C; T_w, wet-environment air temperature, °C; T_{wb}, wet-bulb temperature, °C; T_{ws}, wet-surface temperature, °C; u_2, wind speed at 2 m above the ground, m s^{-1}; u_{10}, wind speed at 10 m above the ground, m s^{-1}

* Corresponding author.
E-mail addresses: ma.n2007@aliyun.com, ningma@itpcas.ac.cn (N. Ma).

https://doi.org/10.1016/j.jhydrol.2020.125221
Received 1 March 2020; Received in revised form 19 June 2020; Accepted 21 June 2020
Available online 27 June 2020

errors in the current parameter-rich ET models (Boisier et al., 2014; Polhamus et al., 2013). Cai et al. (2019), for example, recently reported that the ELMv1 (i.e., the land component of the Energy Exascale Earth System Model of the U.S. Department of Energy) still failed to simulate a reasonable response of ET to deforestation due mainly to inappropriate soil water- and plant-related model parameters. The latter ones may be more intractable in large-scale ET modeling because current widely-used leaf area index (LAI) products [e.g., the Global Land Surface Satellite (Xiao et al., 2016), the Global Mapping LAI (Liu et al., 2012), the LAI3g (Zhu et al., 2013), and the Terrestrial Climate Data Record (Claverie et al., 2016)] were neither intraconsistent over time nor interconsistent with each other due possibly to National Oceanic and Atmospheric Administration satellite orbit changes or Moderate Resolution Imaging Spectroradiometer (MODIS) sensor degradation, as was pointed out by Jiang et al. (2017).

As an alternative of the data-demanding LSMs and RS models that require not only meteorological forcing but also vegetation and soil information for ET estimation, the complementary relationship (CR) (Bouchet, 1963) of evapotranspiration inherently accounts for the integrated effects of the soil-vegetation interface, thus making itself rely solely on standard atmospheric variables (i.e., air and dew point temperature, wind speed, and net radiation) to estimate ET. After more than three decades of working with linear CR models (e.g., the Advection-Aridity model of Brutsaert and Stricker (1979)), Brutsaert (2015), inspired by the work of Han et al. (2012), proposed a state-of-the-art nonlinear CR framework by introducing additional boundary conditions with physical constraints. Lately Han and Tian (2018) re-derived their own boundary conditions, proposing a sigmoid CR model that considers Han et al. (2012)'s model and the Advection-Aridity model as two special cases. These remarkably improved the CR's skill in estimating ET, and simultaneously avoid the need of employing an additional asymmetry coefficient [typically dependent on how wet-environment (ET_w) and potential evapotranspiration (ET_p) rates are represented] found in some linear CR models (Kahler and Brutsaert, 2006; Szilagyi, 2007; Ma et al., 2015). While Brutsaert (2015) nonlinear CR model has been widely used to estimate ET over a diverse range of ecosystems, the stumbling block of applying it across large spatial scales has been the optimization of its key model parameter, the Priestley-Taylor (PT) coefficient (α) (Priestley and Taylor, 1972) via ground-truth ET data (typically by eddy covariance (EC), Bowen-ratio measurements or water-balance-derived values). Even simple water-balances may be difficult to derive on a continental scale because of the presence of basins that are either ungauged or poorly gauged (Hrachowitz et al., 2013; Sivapalan, 2003) for precipitation and/or discharge. Lately Brutsaert et al. (2020) presented another version of the nonlinear CR with the help of seven globally calibrated fitting parameters including the PT α coefficient. Szilagyi et al. (2017) however had already proposed an alternative and efficient scheme for deriving the α value (without employing any additional parameters or measured/water-balance-derived ET data) in their calibration-free and upgraded version of the nonlinear CR of Brutsaert (2015), via the application of estimated temperature and humidity gradients between the wet surface and the air by inverting the Priestley-Taylor equation over wet areas that are automatically identified within a given (and possibly large) spatial domain. This novel version of the nonlinear CR thus avoids any "prior" information of ground-truth ET for the usual calibration of α, a definite advantage for large-scale ET estimation applications. A recent assessment by Ma and Szilagyi (2019) demonstrated that this calibration-free CR model performs better than the mainstream (at least those selected) LSMs, RS models, and atmospheric reanalyses in estimating large-scale ET, and is on a par with the FLUXNET Model Tree Ensemble (FLUXNET-MTE) product, which represents a spatial upscaling of EC measurements using a machine learning (ML) approach.

Indeed, the FLUXNET-MTE ET product (Jung et al., 2010) has been extensively used in not only the evaluations of ET rates produced by LSMs (Ma et al., 2017), RS models (Velpuri et al., 2013) and atmospheric reanalyses (Draper et al., 2018), but also in the investigations of long-term dynamics in global terrestrial water and ecosystems (Feng et al., 2017; Sun et al., 2016). Nevertheless, because FLUXNET-MTE only used one ML algorithm for upscaling, uncertainty still exists since different ML methods may have different responses during the training processes (Tramontana et al., 2016). To reduce multiple sources of uncertainties in empirical upscaling of EC measurements, Jung et al. (2019) recently produced global energy flux products (net radiation, sensible and latent heat) called FLUXCOM, employing an ensemble of multiple ML approaches. However, the FLUXCOM ET (i.e., latent heat) products were only compared with the Global Land Evaporation Amsterdam Model (Martens et al., 2017), LandFLUX-EVAL (Mueller et al., 2013) and FLUXNET-MTE (Jung et al., 2010) model data, but have not yet been evaluated against ground-truth ET results such as derived by a water-balance approach, which is often considered the most accurate way to validate any large-scale ET products (Liu, 2018; Rodell, 2004). In this context, the accuracy of the FLUXCOM ET products remains unknown. At the same time, while the above mentioned and water-balance verified calibration-free CR model may serve as a benchmarking tool for large-scale ET simulation (Ma and Szilagyi, 2019; Ma et al., 2019), and has been further employed for not only detecting long-term regional ET trends (Szilagyi, 2018a; Szilagyi and Jozsa, 2018) but also providing the baseline ET data for nationwide drought monitoring (Kim et al., 2019; Kyatengerwa et al., 2020), it is yet unclear how the FLUXCOM ET estimates compare with those of the CR.

The specific objective of the present study this way is a simultaneous assessment of the annual ET rates of the calibration-free CR and the latest ML-based ET product, FLUXCOM, against the water-balance results of 18 first-level two-digit Hydrological-Unit-Code (i.e., HUC2) and 327 third-level six-digit HUC (i.e., HUC6) basins across the conterminous United States (CONUS). By aiming at this goal, we also expand the previous work of Ma and Szilagyi (2019). The paper is structured as follows: Section 2 introduces the CR and FLUXCOM ET products as well as the water balance method within the planned evaluations. Section 3 considers the evaluation results with regard to the multi-year mean, long-term trends and temporal variations in annual ET. Section 4 presents discussions and the conclusion.

2. Data and Methods

2.1. Complementary-relationship-based ET estimates for CONUS

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

$$y = (2 - X)X^2 \tag{1}$$

where y and X are defined as

$$X = \frac{ET_{pmax} - ET_p}{ET_{pmax} - ET_w}\frac{ET_w}{ET_p} \tag{2}$$

$$y = \frac{ET}{ET_p} \tag{3}$$

here ET is the actual evapotranspiration rate, while ET_p the potential evapotranspiration rate, i.e., the evapotranspiration rate of a small wet patch in a drying (i.e., not fully wet) environment, estimated by the Penman (1948) equation

$$ET_p = \frac{\Delta(R_n - G)}{(\Delta + \gamma)} + \frac{\gamma f_u(e_o - e_a)}{(\Delta + \gamma)} \tag{4}$$

where Δ is the slope of the saturation vapor pressure curve at air temperature, T, and γ the psychrometric constant. R_n and G are net radiation and ground heat flux in water equivalent of mm d^{-1}, while e_o and e_a are the saturation vapor pressure at T and the dew-point

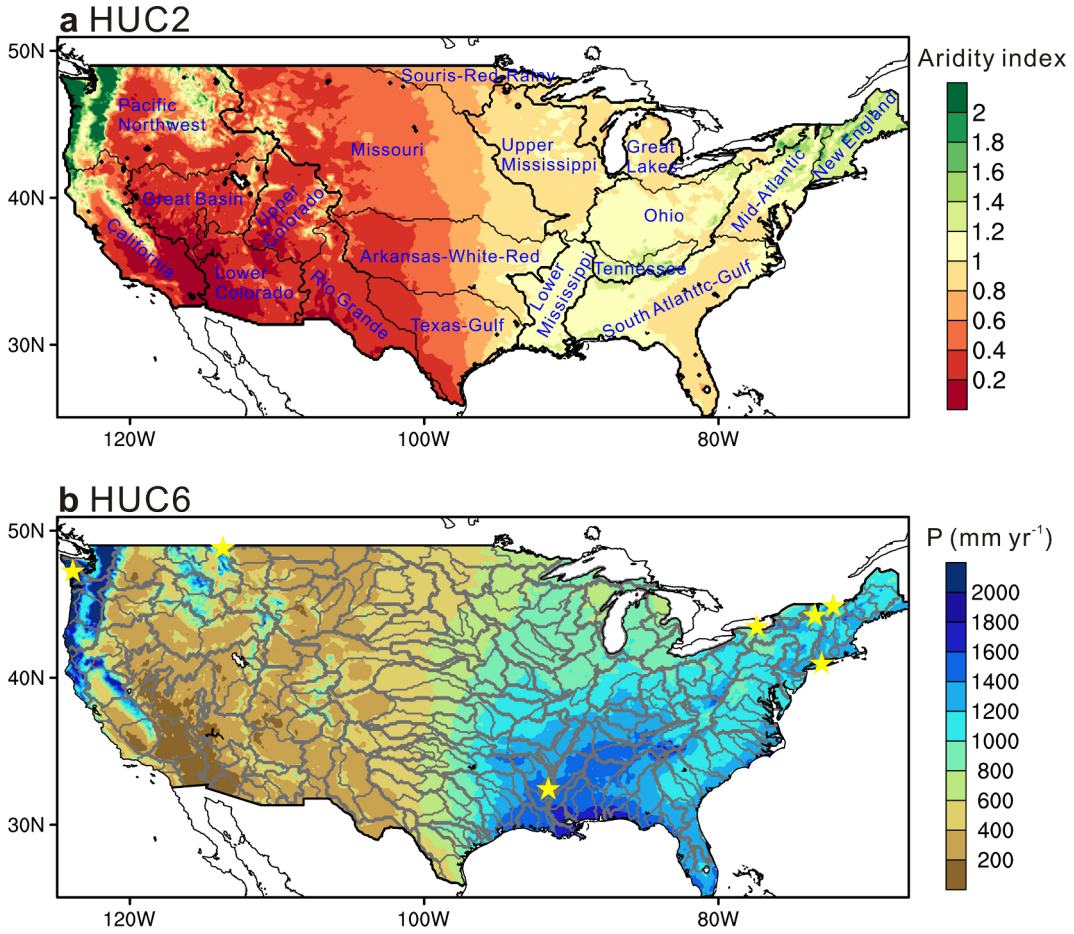

Fig. 1. Spatial distribution of the (a) 18 HUC2, and; (b) 334 HUC6 basins across the conterminous United States. Overlain of HUC2 are the multiyear mean annual aridity index, i.e., the ratio of multiyear mean annual PRISM precipitation to Penman potential evapotranspiration calculated by Eq. (4); while overlain of HUC6 are the multiyear mean annual precipitation from PRISM. The name of each HUC2 basin is shown. Seven HUC6 basins, marked with the yellow star, were left out from the analysis because of outlying ET_{wb} and/or Q values.

temperature (T_d), respectively. f_u is the wind function (Brutsaert, 1982) containing the 2-m wind speed (u_2), i.e.,

$$f_u = 0.26(1 + 0.54u_2) \tag{5}$$

ET_w is the wet-environment evapotranspiration rate, observed over a regionally extensive well-watered surface, specified by the Priestley and Taylor (1972) equation, i.e.,

$$ET_w = \alpha \frac{\Delta_{Tw}}{\Delta_{Tw} + \gamma}(R_n - G) \tag{6}$$

Note that Eq. (6) was derived for completely wet environments by Priestley and Taylor (1972), and therefore, Δ_{Tw} should be evaluated at the air temperature, T_w, observed in a wet environment, instead of the typical, drying environment T (Szilagyi, 2014). By making use of a mild vertical air temperature gradient (Szilagyi, 2014) observable in wet environments (as R_n is consumed predominantly by the latent heat flux), T_w can be approximated by the wet surface temperature, T_{ws}. Note also that T_{ws} may still be larger than T when the drying-environment air is near saturation, but not T_w, due to the cooling effect of evaporation, and in such cases T_w should be capped by T (Szilagyi and 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) 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, i.e.,

$$\beta = \frac{R_n - G - ET_p}{ET_p} \approx \gamma \frac{T_{ws} - T_a}{e_{o,\,Tws} - e_a} \tag{7}$$

here $e_{o,Tws}$ is the saturation vapor pressure at T_{ws}. For continental-scale model applications, the method of Szilagyi et al. (2017) for assigning an appropriate value of α by utilizing observed gridded T and humidity data over automatically-identified wet grid-cells was used [see the Appendix B of Ma and Szilagyi (2019) for details]. The α value of 1.15 derived by Szilagyi (2018b) for the CONUS was retained for the present monthly ET simulation.

ET_{pmax} in Eq. (2) is the maximum value that ET_p can, in theory, reach during a complete dry-out (i.e., when e_a becomes negligible) of the land surface, i.e.,

$$ET_{pmax} = \frac{\Delta_{Tdry}}{\Delta_{Tdry} + \gamma}(R_n - G) + \frac{\gamma}{\Delta_{Tdry} + \gamma} f_u\, e_{o,\,Tdry} \tag{8}$$

in which Δ_{Tdry} and $e_{o,Tdry}$ are the slope of the saturation vapor pressure curve and the saturated vapor pressure, respectively, at the dry-environment air temperature, T_{dry}. 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), i.e.,

$$T_{dry} = T_{wb} + \frac{e_{o,\,Twb}}{\gamma} \tag{9}$$

where $e_{o,Twb}$ is the saturated vapor pressure at the wet-bulb temperature, T_{wb}. T_{wb} can be derived by another iteration of writing out the

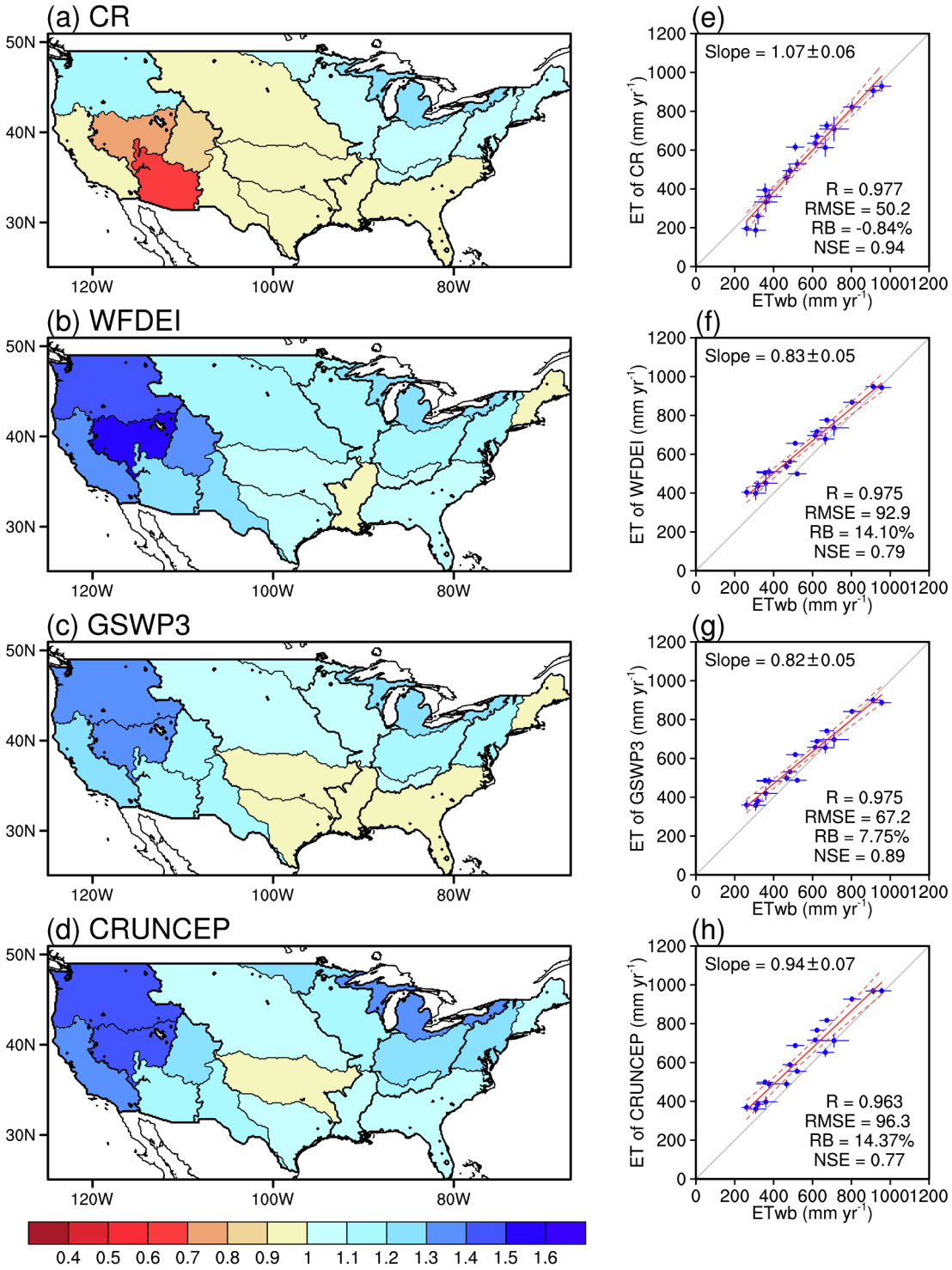

Fig. 2. Spatial pattern of the ratio of basin-wide multiyear (1979–2013) mean annual model ET of (a) CR and (b–d) three RS_METEO products of FLUXCOM to ET_{wb} of the 18 HUC2 basins and (e–h) the corresponding regression plots. The length of the whiskers represents the standard deviation of the 35 annual values of each basin. The strips around the red fitted line denote the 95% confidence intervals. RMSE is in millimeters per year.

Bowen ratio for adiabatic changes (Szilagyi, 2014) as

$$\gamma \frac{T_{wb} - T_a}{e_{o,\,Twb} - e_a} = -1 \tag{10}$$

Readers are kindly directed to Ma and Szilagyi (2019) for a detailed pseudocode that describes the necessary steps to run the model with routine meteorological data. Please, also refer to the Abbreviations for a complete list of the variables (including measurement units).

In the present study, the calibration-free CR model was applied in a continuous monthly simulation over the 37-year period of 1979–2015 across the CONUS, employing the 4-km spatial resolution Parameter-Elevation Regressions on Independent Slopes Model (PRISM) T and T_d data (Daly et al., 1994). The 32-km North American regional reanalysis (NARR) surface R_n (assuming G is negligible at a monthly scale) and 10-m wind speed (u_{10}) data (Mesinger et al., 2006) were linearly interpolated onto the PRISM grid employing a power transformation

Table 1
Statistical metrics for the evaluations of multiyear mean annual ET rates from the CR and FLUXCOM products for the HUC2 and HUC6 basins across CONUS. The best values for each period are in bold.

Periods	Products	HUC2				HUC6			
		R	RMSE*	RB	NSE	R	RMSE*	RB	NSE
1979–2013	CR	**0.977**	**50.2**	**−0.84%**	**0.94**	0.933	**89.0**	**0.45%**	**0.86**
	WFDEI	0.975	92.9	14.10%	0.79	**0.954**	107.3	12.60%	0.79
	GSWP3	0.975	67.2	7.75%	0.89	0.951	89.6	6.34%	**0.86**
	CRUNCEP	0.963	96.3	14.37%	0.77	0.924	116.4	12.41%	0.76
2003–2015	CR	**0.980**	**52.5**	**−0.45%**	**0.94**	0.937	91.2	1.22%	0.85
	RS	0.965	54.3	0.51%	0.93	**0.939**	**82.3**	**0.42%**	**0.88**

* RMSE is in mm yr^{-1}.

(Brutsaert, 1982) of the u_{10} values into u_2 ($=u_{10}(2/10)^{1/7}$), required by Eq. (5). The CR-derived monthly ET rates were aggregated into annual sums for further evaluations.

2.2. FLUXCOM ET products

FLUXCOM (Jung et al., 2019) represents a state-of-the-art ML-based upscaling of EC-measured surface fluxes where data from 224 flux towers around the world were used to train multiple (i.e., three to nine) ML methods. For details of the ML algorithms and their training, plus cross-validations of the results, refer to Tramontana et al. (2016). Briefly, FLUXCOM comprises of two groups of ET products, i.e., "RS" (at 0.0833° spatial and 8-day temporal resolutions) and "RS_METEO" (at 0.5° spatial and 1-day temporal resolutions). The RS product employed only remote sensing data (i.e., MODIS) to estimate ET for the period of 2001–2015, while the RS_METEO products used both gridded meteorological forcing and MODIS data for extended temporal coverages. The meteorological forcing applied to create the RS_METEO products include i) WATCH Forcing Data ERA-Interim (WFDEI for 1979–2013) (Weedon et al., 2015); ii) Global Soil Wetness Project 3 forcing (GSWP3 for 1950–2014) (Kim, 2017); and iii) a fused forcing of the Climate Research Unit and National Centers for Environmental Prediction (CRUNCEP for 1950–2016) (Wei et al., 2014). There is also a short RS_METEO product driven by radiation from the Clouds and the Earth's Radiant Energy System and precipitation from the Global Precipitation Climatology Project (Jung et al., 2019), but was excluded in the present study since it only covers 2001–2014. Readers are suggested to refer to Jung et al. (2019) for a thorough introduction on the FLUXCOM ET products. In the present study, the 0.0833° RS ensemble ET product during 2003–2015, generated from 27 members involving nine ML methods and three energy balance closure correction methods, was used. For the longer period, three 0.5° RS_METEO forcing-specific ensemble ET products, driven by (i) WFDEI; (ii) GSWP3, and; (iii) CRUNCEP, during the overlap period of 1979–2013 were applied. For each meteorological forcing, the ensemble product is based on nine members involving three ML methods and three energy balance closure correction methods. All FLUXCOM products were aggregated into annual sums for further evaluations.

2.3. Water-balance-derived HUC2 and HUC6 ET_{wb} rates

All ET products were evaluated against the water-balance-based evapotranspiration rates (ET_{wb}) at 18 HUC2 and 327 HUC6 basins across CONUS (Fig. 1), which were obtained as

$$ET_{wb} = P - Q - \delta S \tag{11}$$

in which P, Q, and δS are basin-wide annual precipitation, runoff and the change in terrestrial water storage, respectively. These basins are characterized by various hydroclimatic regimes including sub-tropical humid in the south-eastern CONUS, continental in the middle, oceanic in the north-western, and semi-arid in the south-western part (Fig. 1), which are appropriate for addressing the models' skill in estimating ET under diverse environmental conditions. The mean area of the HUC2 and HUC6 basins are ~430,000 km^2 and 23,000 km^2, respectively, representing large to medium watersheds.

In the present study, annual precipitation from PRISM (Daly et al., 1994) and HUC2 as well as HUC6 runoff data from United States Geological Survey during 1979–2015 were used for Equation (11). Considering the common temporal coverage of the FLUXCOM and CR products, the present evaluations concentrate on two periods, i.e., the longer period of 1979–2013 and the shorter one of 2003–2015. Over the longer one the CR and three RS_METEO products of FLUXCOM, while over the shorter one the RS product of FLUXCOM and the CR were evaluated.

For the 2003–2015 period, δS values derived from the Gravity Recovery and Climate Experiment (GRACE) (Tapley et al., 2004), available only after 2002, were applied when calculating HUC2- and HUC6-averaged ET_{wb}. Specifically, the annual δS was calculated as the difference in terrestrial water storage anomaly (TWSA) of successive Decembers in two continuous years, in which TWSA is the arithmetic mean value of three GRACE products processed by GeoforschungsZentrum Potsdam, the Center for Space Research at the University of Texas, Austin, and by the Jet Propulsion Laboratory with further consideration to the gain factors proposed by Landerer and Swenson (2012).

For the 1979–2013 period that GRACE cannot fully cover, we used the 0.5°, monthly TWSA data from the GRACE-REC (Humphrey and Gudmundsson, 2019), in which TWSA was reconstructed using a statistical model with inputs of precipitation and temperature at each global land grid for the past century. A series of evaluations against the sea level budget and streamflow datasets by Humphrey and Gudmundsson (2019) suggested that the GRACE-REC is a state-of-the-art long-term TWSA product, which is especially useful for the pre-GRACE era. The development of GRACE-REC involved two kinds of GRACE products (for training purpose) and three kinds of climate forcing, thus leading to six different versions of GRACE-REC (see Table 3 in Humphrey and Gudmundsson (2019) for detailed information about every version of GRACE-REC). In the present study, to calculate annual δS for 1979–2013, we used the version based on the GSWP3 precipitation and temperature data that were calibrated with TWSA from the Jet Propulsion Laboratory mascons.

Due to differences in the spatial resolutions between PRISM P, GRACE and GRACE-REC δS, CR, and the FLUXCOM ET products, we first used the nearest neighbor method to resample all products into the common 0.125° grid. Then we calculated the basin-wide ET rates by spatial averaging over each HUC2 or HUC6 basin using the 0.125° river-mask for the corresponding basins.

The multiyear mean ET_{wb} rates of the HUC2 and HUC6 basins are shown in Fig. S1. Note that seven out of the original 334 HUC6 basins were excluded (see Fig. 1b) in all evaluations because of outlier ET_{wb} and/or Q data in comparison with their neighbors.

To evaluate a given ET product against ET_{wb}, we focused on three

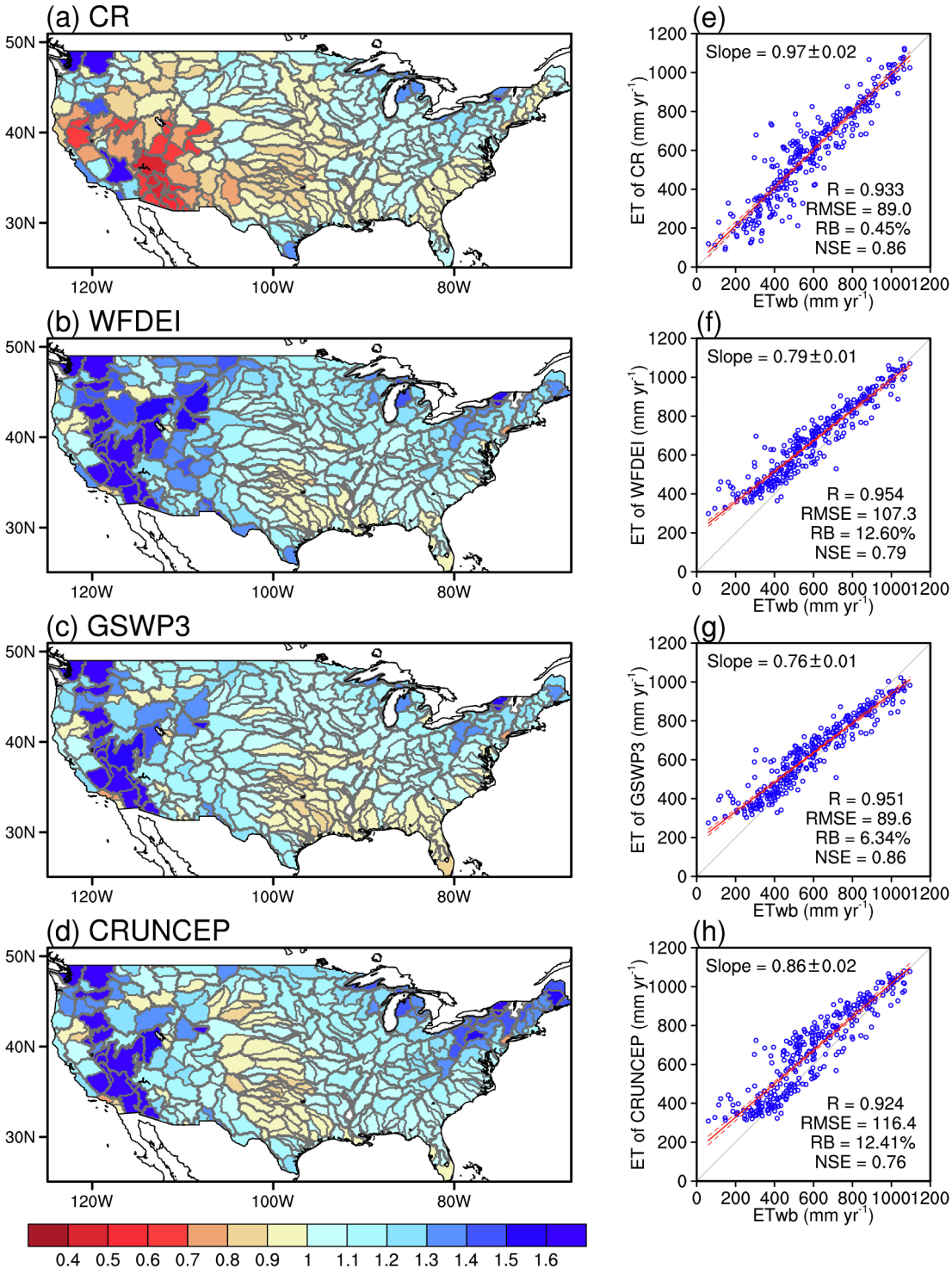

Fig. 3. Spatial pattern of the ratio of basin-wide multiyear (1979–2013) mean annual model ET of (a) CR and (b–d) three RS_METEO products of FLUXCOM to ET_{wb} of the 327 HUC6 basins and (e–h) the corresponding regression plots. The strips around the red fitted line denote the 95% confidence intervals. RMSE is in millimeters per year.

aspects: (i) the multiyear mean, (ii) the temporal variation in the annual values, and; (iii) the linear trends in annual ET rates over the given period. The employed statistical metrics (either spatial for the mean and trend, or temporal for the annual time-series) included the Pearson correlation coefficient (R), root mean square error (RMSE), relative bias (RB), and Nash-Sutcliffe efficiency (NSE) between ET_{wb} and basin-averaged ET rates of every product. Note that in the assessment of the trends, RB is not an effective measure (thus was excluded) because the denominator may be very small if ET_{wb} had little change over a given period. For the time-series of basin-wide ET, the ratio of the interannual variability (IAV) in modeled ET to that in ET_{wb} (i.e., R_{IAV}) was also calculated to assess how well models capture the IAV in annual ET_{wb} for the given period.

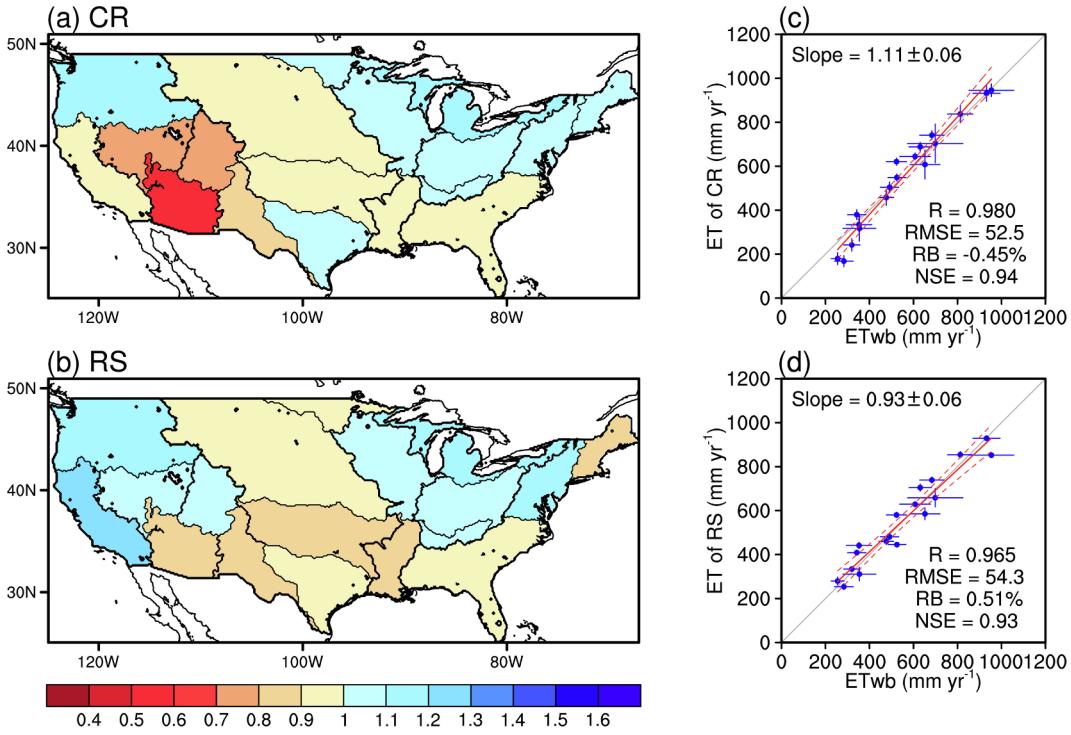

Fig. 4. Spatial pattern of the ratio of basin-wide multiyear (2003–2015) mean annual model ET from (a) CR and (b) the RS product of FLUXCOM to ET_{wb} of the 18 HUC2 basins and (c–d) the corresponding regression plots. The length of the whiskers represents the standard deviation of the 13 annual values of each basin. The strips around the red fitted line denote the 95% confidence intervals. RMSE is in millimeters per year.

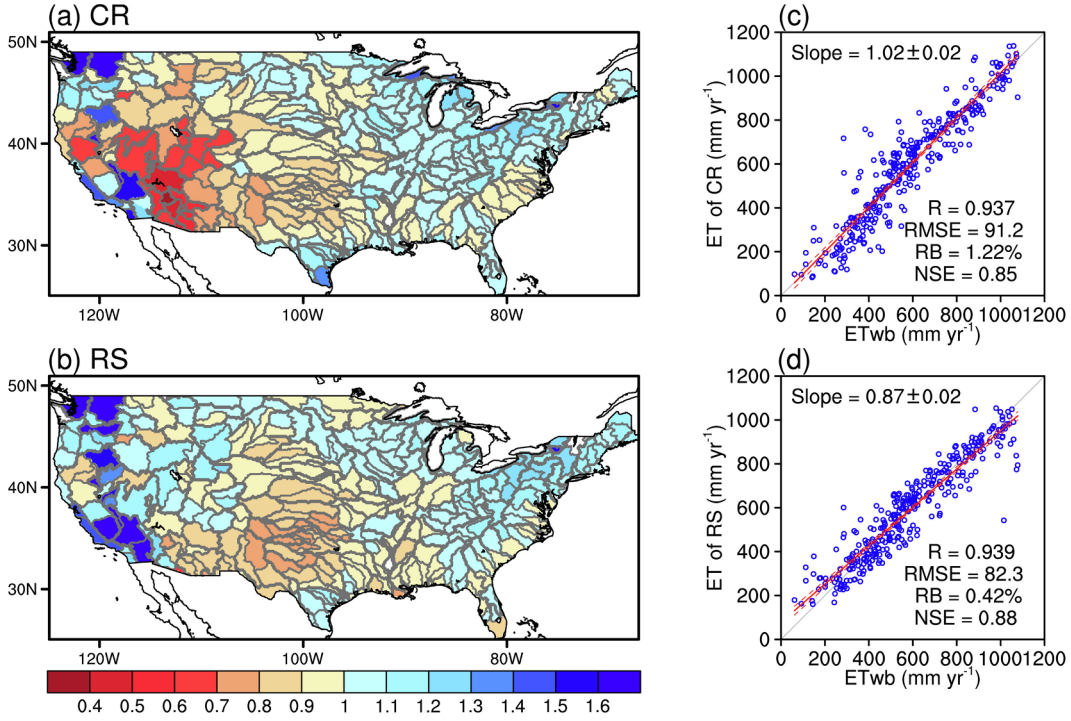

Fig. 5. Spatial pattern of the ratio of basin-wide multiyear (2003–2015) mean annual model ET from (a) CR and (b) the RS product of FLUXCOM to ET_{wb} of the 327 HUC6 basins and (c–d) the corresponding regression plots. The strips around the red fitted line denote the 95% confidence intervals. RMSE is in millimeters per year.

3. Results

3.1. Spatial assessment of the multiyear mean annual ET rates

The CR displays the best performance with the highest NSE (0.94) and R (0.977), and the lowest RMSE (50.2 mm yr^{-1}) and RB (−0.84%) values in comparison with the multiyear (1979–2013) mean ET_{wb} rates over the 18 HUC2 basins (Fig. 2 & Table 1). This is closely followed by the RS_METEO product driven by GSWP3 with the NSE value dropping to 0.89 and the RMSE and RB values jumping to 67.2 mm yr^{-1} and 7.75%, respectively. The other two RS_METEO products forced by WFDEI and CRUNCEP exhibit weaker performances with similar NSE

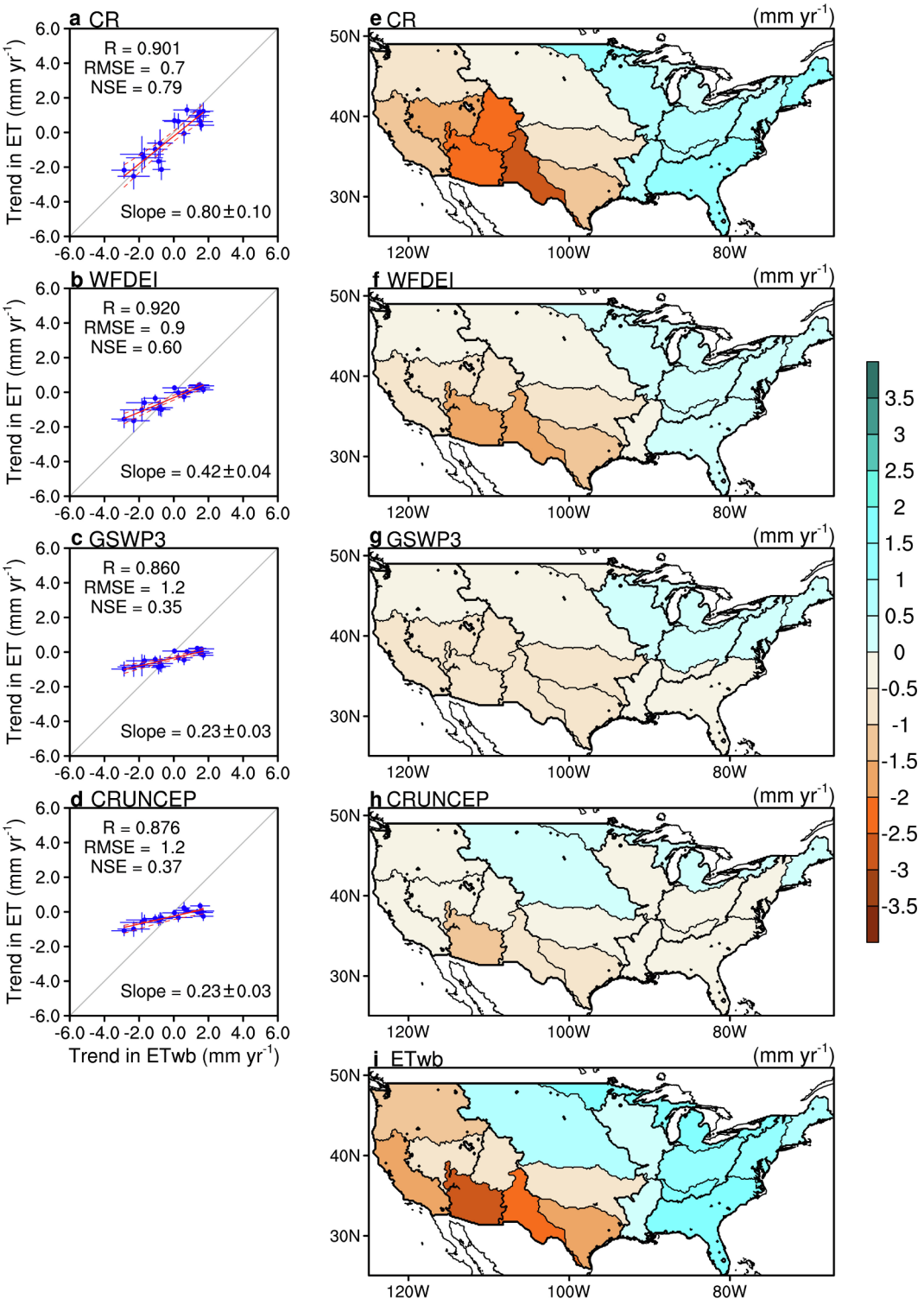

Fig. 6. Regression plots of the trends in HUC2-averaged annual model ET from (a) CR and (b–d) three RS_METEO products of FLUXCOM against ET_{wb} over the period of 1979–2013. The length of the whiskers represents the standard error in the estimated slope value. The strips around the red fitted lines denote the 95% confidence intervals. RMSE is in millimeters per year. The maps display the spatial distribution of the trends from (e) CR, (f–h) three RS_METEO products of FLUXCOM and (i) ET_{wb}.

(0.79 and 0.77, respectively), RMSE (both more than 90 mm yr^{-1}) and RB (both roughly 14%) values among them. Based on the spatial pattern of the ratio of ET to ET_{wb}, both CR and GSWP3 have errors within ± 10% over most HUC2 basins (Fig. 2a & c) except for (i) an underestimation by 39% over the Lower Colorado basin in the CR, and; (ii) an overestimation by roughly 37% over the Great Basin and Northwest basins in the GSWP3 (see Fig. 1a for HUC2 basin identification). WFDEI and CRUNCEP (Fig. 2b & d) tend to overestimate ET in the majority of HUC2 basins, especially noticeable for the western CONUS. The most significant positive bias occurred in the Great Basin for, both, WFDEI and CRUNCEP by 55% and 41%, respectively.

The CR continues to yield the smallest RMSE (89.0 mm yr^{-1}) and

Table 2
Statistical metrics for the evaluations of trends in annual ET rates over given periods from the CR and FLUXCOM products for the HUC2 and HUC6 basins across CONUS. The best values for each period are in bold.

Periods	Products	HUC2			HUC6		
		R	RMSE*	NSE	R	RMSE*	NSE
1979–2013	CR	0.901	**0.7**	**0.79**	0.553	**1.7**	0.29
	WFDEI	**0.920**	0.9	0.60	**0.632**	**1.7**	**0.31**
	GSWP3	0.860	1.2	0.35	0.599	1.8	0.21
	CRUNCEP	0.876	1.2	0.37	0.573	1.8	0.20
2003–2015	CR	**0.608**	**2.4**	**0.08**	**0.473**	**5.4**	**0.14**
	RS	0.448	3.4	−0.79	0.391	6.3	−0.19

* RMSE is in mm yr^{-1}.

RB (0.45%) values (Fig. 3) for the 35-year average of the ET rates among the much smaller HUC6 catchments. For the NSE value, CR and GSWP3 are comparable with an identical value of 0.86 (Table 1), which is higher than other products'. Both, WFDEI and GSWP3, display larger RMSE and RB values, while producing better R values (0.954 and 0.951, respectively) than CR (0.933). CRUNCEP still performs the weakest with the largest RMSE and the lowest NSE and R values. Spatially, widespread positive biases occur in all three RS_METEO products of FLUXCOM (Fig. 3b, c & d), with a reduced range, as can also be seen from their fitted slopes significantly smaller than unity (Fig. 3f, g & h). In contrast, CR appears to have a balanced performance across CONUS with a slope (0.97) close to identity (Fig. 3e).

The RS product of FLUXCOM over the shorter temporal coverage of 2003–2015 performs similarly to the CR at the HUC2 scale in describing the multiyear mean ET (Fig. 4), as was evidenced by comparable statistical metrics in Table 1. However, at the HUC6 scale during the same 13-year period (Fig. 5), the RS product of FLUXCOM improves a bit upon the CR due probably to benefitting from direct EC measurements at this smaller scale, as can be seen from the higher NSE (0.88 vs. 0.85) and smaller RMSE (82.3 vs. 91.2 mm yr^{-1}) values (Table 1). Note also that a significant underestimation of ET in partial regions of Arizona and Nevada exists in the CR model, which to a large extent is mitigated in the RS product of FLUXCOM, partly due to a flatter slope in those values (i.e., 0.87 vs. 1.02 for CR) (Fig. 5). The reason for underestimation of the CR in these arid catchments is not yet clear, but if ET rates are patchy in time (and in space as well), meaning very high rates during and shortly after relatively few precipitation events (and more persistently from phreatophyte vegetation), and low rates otherwise due to the water-conserving behavior of semi-arid vegetation, then the monthly mean input values (e.g., air humidity) of the CR may miss or significantly smooth out such sudden spikes. This effect is not so severe over the larger HUC2 basins (c.f., Fig. 4), as they would contain mountainous regions where precipitation is more abundant. Of course, the daily EC measurements employed by FLUXCOM would not be so affected.

3.2. Assessing the trends in annual ET

The CR is particularly robust in estimating trends in HUC2-averaged annual ET rates during 1979–2013, as evidenced by the highest NSE (0.79) and the lowest RMSE (0.7 mm yr^{-1}) values in Fig. 6 and Table 2. However, all three RS_METEO products of FLUXCOM tend to predict very narrow ranges (seen in reduced slope values) for the (increasing/decreasing) trends when compared to ET_{wb} values of the 18 HUC2 basins (Fig. 6b, c & d), thus underestimating long-term tendencies. The spatial distributions of the trends in annual ET rates corroborates this: GSWP3 and CRUNCEP produce much milder ones than ET_{wb} at most HUC2 basins. Such a discrepancy is especially true for (i) the southwestern CONUS where ET_{wb} decreased obviously (Fig. 6i) but only slight corresponding changes are found in these two products (Fig. 6g & h), and; (ii) the southeastern CONUS, where trends are opposite between ET_{wb} and these two products (see Fig. 6g, h & i). While the biases in trends diminish to some extent in the WFDEI product (Fig. 6b & Table 2), it still underestimates the decreasing trend of ET_{wb} in the southwestern US (Fig. 6f). On the whole, CR accurately describes the trends in annual ET rates over most HUC2 basins exceptthe Missouri basin with a trend opposite to the one in ET_{wb} (see Fig. 6e & i).

At the HUC6 scale where ET_{wb} is more uncertain [e.g., due to possible (i) differences in catchment boundaries between surface- and groundwater; (ii) enhanced role of interannual reservoir storage changes in the water balance], the skills of all models estimating the trends in ET rates during this 35-year period appear inferior to similar measures found at HUC2, as can be seen from the decreased NSE and increased RMSE values (Table 2). The highest NSE (0.31) and R (0.632) values are provided by WFDEI, yielding an RMSE value (1.7 mm yr^{-1}) identical to that of the CR, the latter also producing the largest slope value (0.37) (Fig. 7 & Table 2). In a spatial manner, however, none of the RS_METEO products of FLUXCOM could well capture the significant decreasing trend in ET_{wb} over western Texas, southern Arizona and New Mexico (Fig. 7i) during 1979–2013, while the CR was the only one showing magnitudes of change similar to those in ET_{wb} (Fig. 7e). Note again that the trends from all three RS_METEO products of FLUXCOM seem much milder than those from ET_{wb} due possibly to their underestimated IAV of annual ET rates (see Section 3.3), as can be seen from the milder fitted slopes ranging from 0.13 to 0.22 (Fig. 7b, c & d). In general, GSWP3 and CRUNCEP exhibit poorer performance for the trends at the HUC6 scale with NSE values of about 0.2 (Fig. 7c & d), consistent with their results at the HUC2 scale.

During the short period of 2003–2015, the CR's ability (NSE = 0.08 and RMSE = 2.4 mm yr^{-1}) in estimating the ET trends at the HUC2 scale is again better than that of the RS product of FLUXCOM (Table 2), the latter producing a negative (−0.79) NSE value (Fig. 8b). Moreover, the spatial pattern of trends by the CR (Fig. 8c) generally follows that in ET_{wb} (Fig. 8e), the latter exhibiting overall increasing trends for most HUC2 basins in the eastern CONUS. However, the RS product of FLUXCOM yields decreasing trends for most parts of the eastern CONUS (Fig. 8d).

The performance of the CR in estimating trends in annual ET rates during 2003–2015 does not change much at the HUC6 scale, with NSE and RMSE values of 0.14 and 5.4 mm yr^{-1}, respectively (Fig. 9a). The RS product displays again weaker performance with a negative (−0.19) NSE value (Fig. 9b & Table 2) due mainly to its large errors in the eastern CONUS (see Fig. 9d & e), similar to above-mentioned contrasting trends for the HUC2 basins.

3.3. Assessing the temporal variations in annual ET

The arithmetic averages of the statistical metrics of the CR and FLUXCOM performance in simulating the annual ET time series of the 18 HUC2 or 327 HUC6 basins were used for assessing the temporal variations in annual ET. The only exception is the NSE value where the median was computed due to the occasional large negative values for basins where the difference in multiyear modeled and water-balance ET averages is large.

For capturing temporal variations in annual ET rates during 1979–2013, the CR and FLUXCOM exhibit overall similar performances in terms of RMSE, RB and NSE values for both HUC2 (Fig. 10) and HUC6 (Fig. S2) scales. The IAV in ET_{wb} was unanimously underestimated, the strongest in the three RS_METEO products of FLUXCOM with R_{IAV} values less than 50%, while CR achieves about 80% (Fig. 10d). Over the period of 2003–2015 (Figs. 10f–j & S2f–j), the CR also yields better performance in R_{IAV} (~60% and ~30% for CR and RS products, respectively), while in the NSE, RB and RMSE values the CR again performs similar to the RS product of FLUXCOM in both spatial scales.

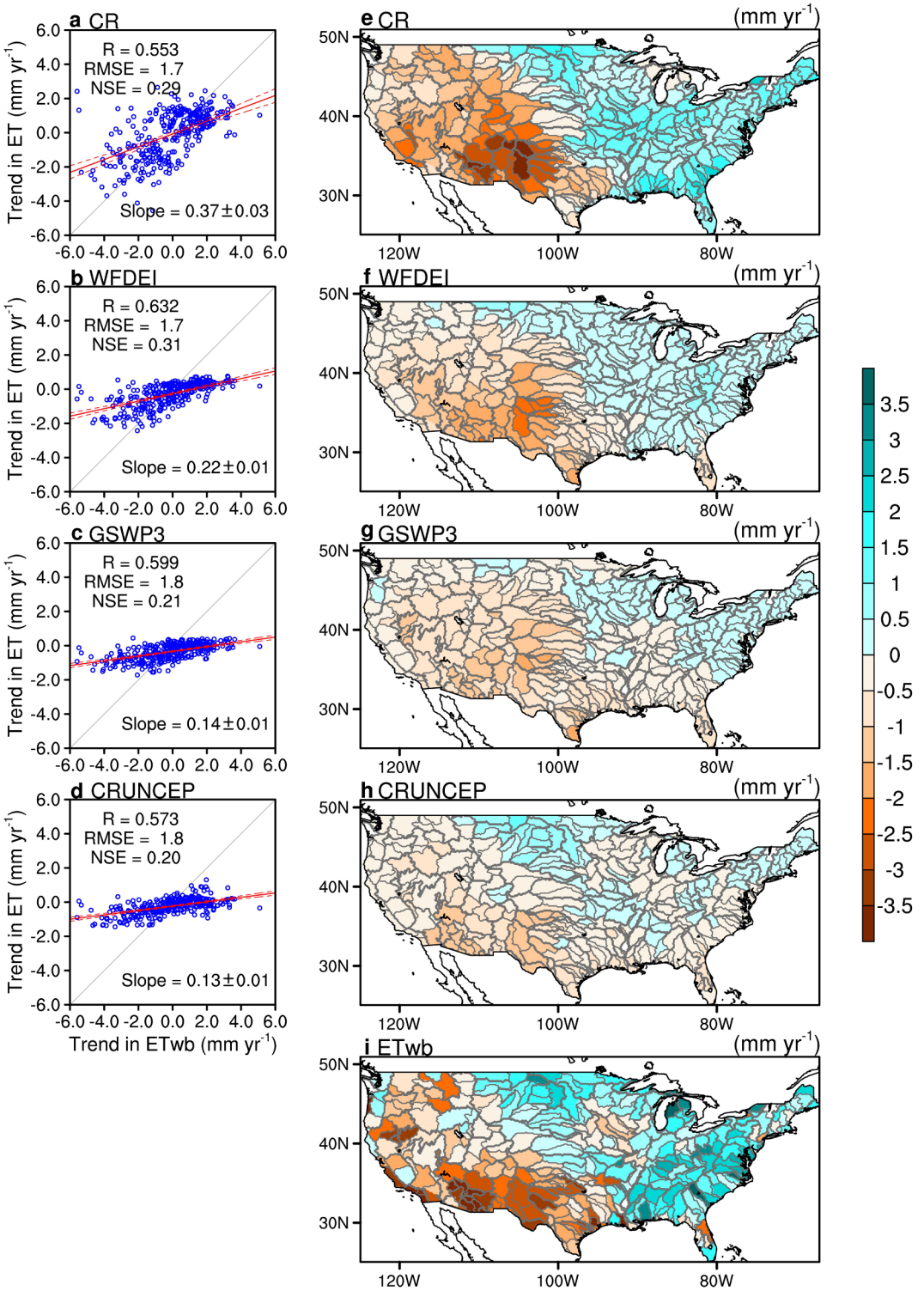

Fig. 7. Regression plots of the trends (1979–2013) in basin-wide annual model ET from (a) CR and (b–d) three RS_METEO products of FLUXCOM against ET_{wb} of the 327 HUC6 basins. The strips around the red fitted lines denote the 95% confidence intervals. RMSE is in millimeters per year. The maps display the spatial distribution of the trends from (e) CR, (f–h) three RS_METEO products of FLUXCOM and (i) ET_{wb}.

4. Discussions and Conclusion

4.1. Uncertainties in model validation using the water balance method

It must be recognized that the reported model performances may be confounded by possible errors embedded in ET_{wb} (Carter et al., 2018; Liu, 2018). While δS has been considered for both short (2003–2015) and long (1979–2013) periods, the coarse spatial resolutions of GRACE and GRACE-REC would probably cause biases in ET_{wb} of small watersheds, which may be particularly true for the HUC6 basins. Besides, while PRISM P is one of the most thoroughly validated land-measured precipitation dataset for CONUS because it considers the effects of elevation, rain shadows, and coastal proximity in the interpolation process (Daly et al., 2008), the ubiquitous underestimation in

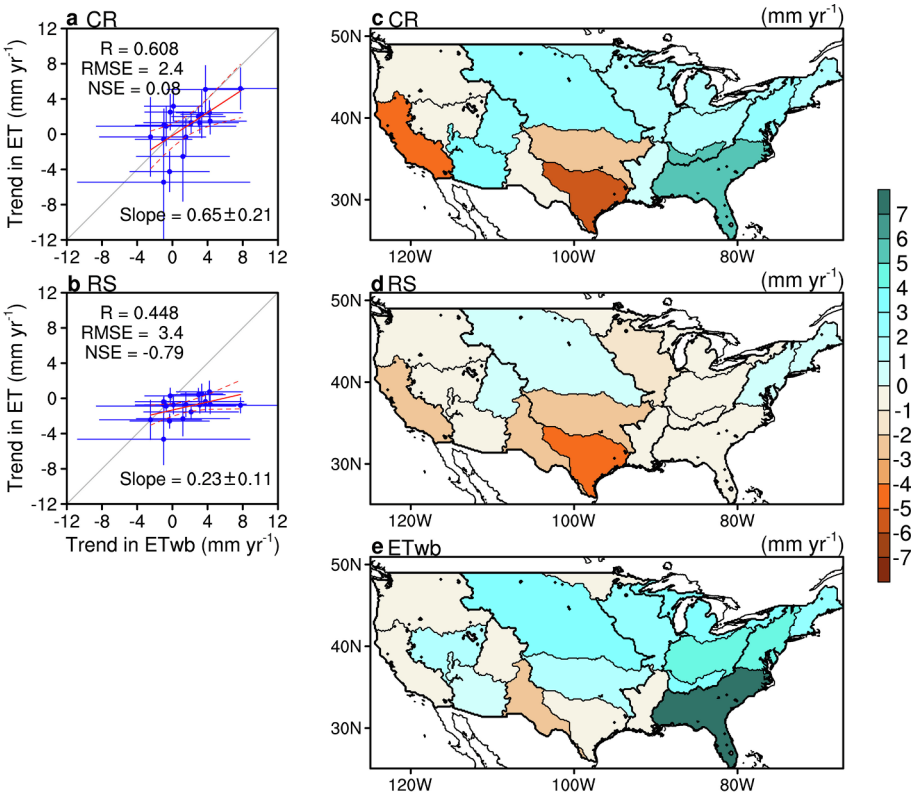

Fig. 8. Regression plots of the trends in HUC2-averaged annual model ET from (a) CR and (b) the RS product of FLUXCOM against those in ET_{wb} over the period of 2003–2015. The length of the whiskers represents the standard error in the estimated slope value. The strips around the red fitted lines denote the 95% confidence intervals. RMSE is in millimeters per year. The maps display the spatial distribution of the trends from (c) CR, (d) the RS product of FLUXCOM and (e) ET_{wb}.

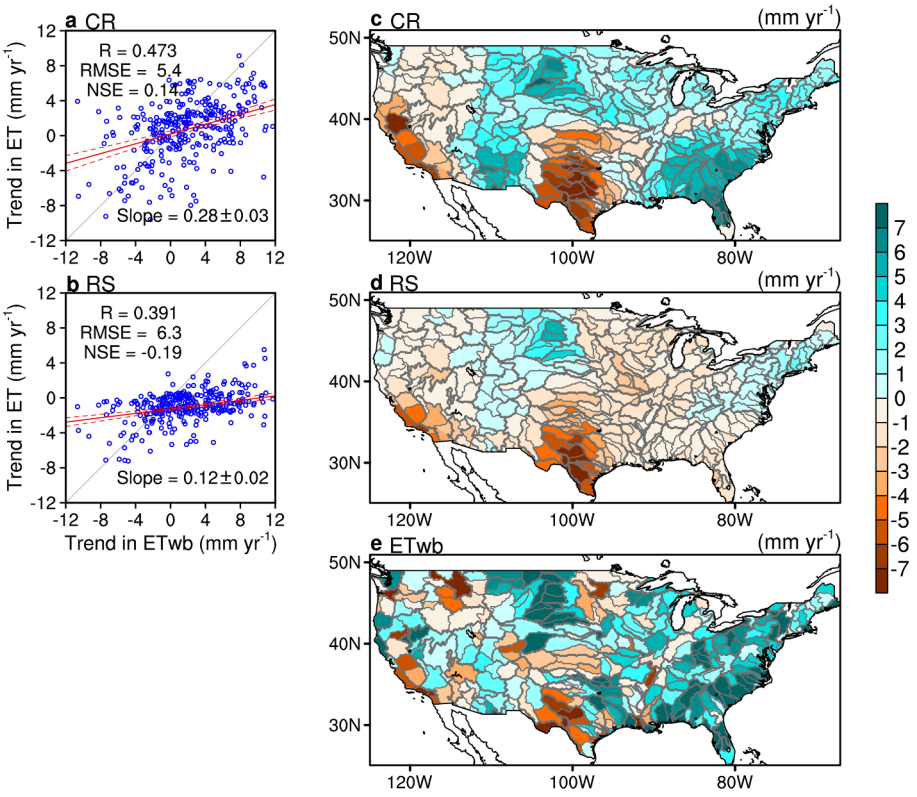

Fig. 9. Regression plots of the trends (2003–2015) in basin-wide annual model ET from (a) CR and (b) the RS product of FLUXCOM against those in ET_{wb} of the 327 HUC6 basins. The strips around the red fitted lines denote the 95% confidence intervals. RMSE is in millimeters per year. The maps display the spatial distribution of the trends from (c) CR, (d) the RS product of FLUXCOM and (e) ET_{wb}.

precipitation measurements due to, e.g., wind-induced undercatch as well as wetting and evaporation losses (Yang et al., 2005; Sieck et al., 2007), could lead to uncertainties in any gridded precipitation data (Lundquist et al., 2019). However, the errors found in the PRISM P estimates are mostly within 5 to 15% of the recorded ground station values (Jeton et al., 2005; Daly et al., 2008, 2017). While quantifying the errors in the ET_{wb} values emerging from precipitation and/or δS is beyond the scope of the present study, a detailed analysis of such uncertainties in the water balance approach would certainly benefit the overall validation of large-scale ET products.

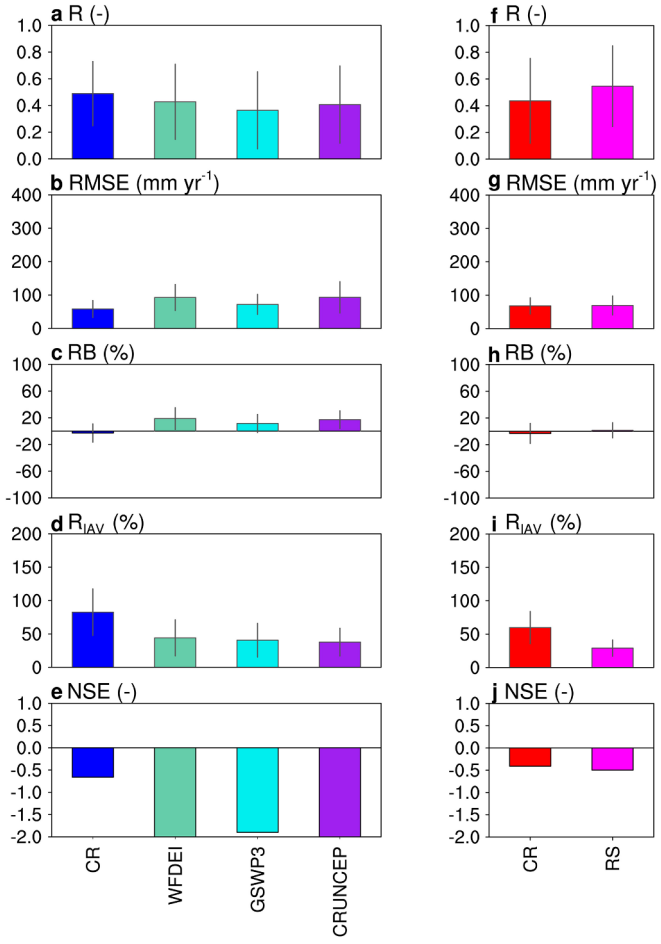

Fig. 10. Arithmetic averages (with error bars denoting the standard deviation) of the statistical metrics (R, RMSE, RB, R_{IAV}) of the modeled (a–e for 1979–2013; f–j for 2003–2015) annual ET time-series from the 18 HUC2 basins. For NSE the medians are displayed due to possible large negative values in individual HUC2 basins where the difference between mean modeled ET and ET_{wb} becomes large.

4.2. Relative merits of FLUXCOM and CR

The recently available powerful ML techniques have stimulated substantial advances in hydrological research such as the estimation of precipitation (Tang et al., 2018), soil moisture (Kolassa et al., 2018), snow water equivalent (Bair et al., 2018), runoff (Yaseen et al., 2019) and drought forecasting (Deo et al., 2017). In particular, the ML-based upscaling of EC-flux measurements to regional (e.g., Fang et al., 2020) or global scales (e.g., Jung et al., 2019) has greatly boosted our understanding of ET at extended spatial and temporal dimensions, although the low density of EC towers in certain areas (e.g., Africa, South America, West and South Asia) may not be sufficient to exploit the capability of ML to extract relevant characteristics of space structure. Note also that while the RS product of FLXUCOM takes full advantage of remotely sensed data and exhibits excellent skills in representing the spatial pattern of the multiyear mean ET rates, it performs poorly for the trends in annual ET (see Figs. 8 & 9) due possibly to the calibration degradation of MODIS Collection 5 data (Wang et al., 2012), in which the Enhanced Vegetation Index (a key variant used for training the RS product of FLUXCOM) has not been well calibrated (Lyapustin et al., 2014). This indicates that incorporating further advanced satellite observations may benefit RS-based machine learning of ET in the future. In addition, the uncertainty in any ET models can also result from the uncertainty in the meteorological forcing. For example, the evaluation by Tang et al. (2017) suggested that CRUNCEP still has obvious biases in its temperature and precipitation data when compared with the station-measured results of 90 cities in United States, though CRUNCEP has improved a little bit upon the NCEP I and NCEP II reanalyses, which might be the reason why CRUNCEP-based RS_METEO product of FLUXCOM exhibits larger biases in ET estimates across CONUS (see Tables 1 & 2).

While the CR performs comparably to (for the trends better than) FLUXCOM in the estimation of ET across the CONUS, it is worthwhile to highlight that FLUXCOM produced not only ET but also net radiation and sensible heat fluxes (Jung et al., 2019) as well as carbon fluxes including gross primary production and net ecosystem exchange (Jung et al., 2020) with a worldwide coverage over the last few decades. This way FLUXCOM complements a holistic view of land-atmosphere energy, water, and carbon exchanges at a global scale that may not be otherwise possible from any other single dataset, which is particularly vital for the developing countries where EC flux towers are spatially sparser and temporally shorter than in developed ones (Chu et al., 2017). The present validations suggest that the GSWP3-drived RS_METEO product and the RS product of FLUXCOM perform well in representing the multiyear mean ET (c.f., Figs. 2 & 4). We therefore believe that FLUXCOM is certainly valuable for not only LSMs and RS model evaluations but also provides fundamental constraints of global energy and carbon cycling.

The most important merit of CR is that it relies solely on standard atmospheric forcing (Han and Tian, 2020) to estimate ET without any precipitation data, the latter being the most uncertain meteorological variable to predict e.g., in climate models. Besides, CR avoids the possible uncertainties in any gridded vegetation and soil data, which are usually key inputs for most LSM- and RS-based ET models. However, a weakness of the present calibration-free CR as a benchmark for calibration/verification of other large-scale ET models is that it is not recommended to be employed on a daily or sub-daily bases. In fact, Morton (1983) suggested not using the CR for periods shorter than five days because large-scale weather fronts may bring air masses over the land with a moisture signature decoupled from the underlying surface, which thus may temporarily disrupt the dynamic equilibrium of air humidity and surface fluxes in the land–atmosphere system. Hence, an additional temporal aggregation of model simulations is necessary for models that are driven by daily or sub-daily forcing before validating them against the CR. Note also that the CR is not able to predict ET in a prognostic mode (i.e., ET rates at a future time of $t + dt$) from the soil moisture and vegetation status obtained at time t since it estimates the cause (ET) from the effect (moisture content and temperature of the air) (Ma and Szilagyi, 2019). However, the CR can be employed for reality checks of the climate models' future scenarios to see whether the projected state of the atmosphere is indeed in balance (via the CR) with the assumed ET rates produced by the climate models.

In conclusion it can be stated that the CR is on a par with the FLUXCOM ET products no matter which meteorological forcing and/or satellite data are used in the latter. The CR however performs somewhat better in representing the long-term ET trends than FLUXCOM. While uncertainties exist in the CR-based ET results over certain regions, the CR could still serve as a benchmarking tool for verifications of the LSMs, RS, and ML-based large-scale ET estimates due to its robust performance, minimal data requirement and calibration-free nature. However, it should be emphasized again that FLUXCOM contains not only ET but also other energy and carbon fluxes with a global coverage and thus, continues to provide the community with a valuable reference dataset for studies associated with large-scale terrestrial energy, water and carbon cycles.

CRediT authorship contribution statement

Ning Ma: Conceptualization, Methodology, Formal analysis, Software, Investigation, Resources, Data curation, Writing - original draft, Visualization, Funding acquisition. **Jozsef Szilagyi:**

Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing - review & editing. **Janos Jozsa:** Funding acquisition, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The manuscript benefitted greatly from the comments by Martin Jung. This research was jointly supported by the National Key Research and Development Program of China (2017YFA0603101), National Natural Science Foundation of China (41801047), Open Research Fund Program of State Key Laboratory of Cryospheric Science, Northwest Institute of Eco-Environment and Resources, CAS (SKLCS-OP-2020-11), China Postdoctoral Science Foundation (2019T120139), and BME-Water Sciences & Disaster Prevention FIKP grant of EMMI (BME FIKP-VIZ). We thank the two anonymous reviewers for their comments that improved an earlier version of the manuscript. All data used in this study can be freely accessed from the websites as follows: USGS HUC2 and HUC6 runoff (https://waterwatch.usgs.gov/?id=wwds_runoff); PRISM data (http://prism.oregonstate.edu/); NARR data (www.esrl.noaa.gov/psd/data/gridded/data.narr.html); GRACE (http://grace.jpl.nasa.gov), GRACE_REC (https://doi.org/10.6084/m9.figshare.7670849). The FLUXCOM ET products are available from the MPI-BGC data portal (https://www.bgc-jena.mpg.de/geodb/projects/Home.php).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2020.125221.

References

Bair, E.H., Abreu Calfa, A., Rittger, K., Dozier, J., 2018. Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan. Cryosphere 12, 1579–1594. https://doi.org/10.5194/tc-12-1579-2018.

Boisier, J.P., de Noblet-Ducoudré, N., Ciais, P., 2014. Historical land-use-induced evapotranspiration changes estimated from present-day observations and reconstructed land-cover maps. Hydrol. Earth Syst. Sci. 18, 3571–3590. https://doi.org/10.5194/hess-18-3571-2014.

Bouchet, R.J., 1963. Evapotranspiration réelle et potentielle, signification climatique. Int. Assoc. Sci. Hydrol. Publ. 62, 134–142.

Brutsaert, W., 1982. Evaporation into the Atmosphere: Theory, History and Applications. Springer, New York, p. 299.

Brutsaert, W., 2015. A generalized complementary principle with physical constraints for land-surface evaporation. Water Resour. Res. 51, 8087–8093. https://doi.org/10.1002/2015WR017720.

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

Brutsaert, W., Cheng, L., Zhang, L., 2020. Spatial distribution of global landscape evaporation in the early twenty first century by means of a generalized complementary approach. J. Hydrometeorol. 21, 287–298. https://doi.org/10.1175/jhm-d-19-0208.1.

Cai, X., Riley, W.J., Zhu, Q., Tang, J., Zeng, Z., Bisht, G., Randerson, J.T., 2019. Improving representation of deforestation effects on evapotranspiration in the E3SM Land Model. J Adv. Model. Earth Syst. 11, 2412–2427. https://doi.org/10.1029/2018ms001551.

Carter, E., Hain, C., Anderson, M., Steinschnerider, S., 2018. A water balance based, spatiotemporal evaluation of terrestrial evapotranspiration products across the contiguous United States. J. Hydrometeorol. 19, 891–905. https://doi.org/10.1175/JHM-D-17-0186.1.

Chu, H., Baldocchi, D.D., John, R., Wolf, S., Reichstein, M., 2017. Fluxes all of the time? a primer on the temporal representativeness of FLUXNET. J. Geophys. Res. Biogeosci. 122, 289–307. https://doi.org/10.1002/2016jg003576.

Claverie, M., Matthews, J., Vermote, E., Justice, C., 2016. A 30+ year AVHRR LAI and FAPAR climate data record: algorithm description and validation. Remote Sens. 8, 1–12.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. 28, 2031–2064. https://doi.org/10.1002/joc.1688.

Daly, C., Neilson, R.P., Phillips, D.L., 1994. A statistical topographic model for mapping climatological precipitation over mountainous terrain. J. Appl. Meteorol. 33, 140–158.

Daly, C., Slater, M.E., Roberti, J.A., Laseter, S.H., Swift Jr., L.W., 2017. High-resolution precipitation mapping in a montainous watershed: ground truth for evaluating uncertainty in a national precipitation dataset. Int. J. Climatol. 37, 124–137. https://doi.org/10.1002/joc.4986.

Deo, R.C., Kisi, O., Singh, V.P., 2017. Drought forecasting in eastern Australia using multivariate adaptive regression spline, least square support vector machine and M5Tree model. Atmos. Res. 184, 149–175. https://doi.org/10.1016/j.atmosres.2016.10.004.

Draper, C.S., Reichle, R.H., Koster, R.D., 2018. Assessment of MERRA-2 land surface energy flux estimates. J. Clim. 31, 671–691. https://doi.org/10.1175/jcli-d-17-0121.1.

Fang, B., Lei, H., Zhang, Y., Quan, Q., Yang, D., 2020. Spatio-temporal patterns of evapotranspiration based on upscaling eddy covariance measurements in the dryland of the North China Plain. Agric. For. Meteorol. 281, 107844. https://doi.org/10.1016/j.agrformet.2019.107844.

Feng, H., Zou, B., Luo, J., 2017. Coverage-dependent amplifiers of vegetation change on global water cycle dynamics. J. Hydrol. 550, 220–229. https://doi.org/10.1016/j.jhydrol.2017.04.056.

Han, S., Hu, H., Tian, F., 2012. A nonlinear function approach for the normalized complementary relationship evaporation model. Hydrol. Process. 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 Resour. Res. 54 (7), 5050–5068.

Han, S., Tian, F., 2020. A review of the complementary principle of evaporation: from the original linear relationship to generalized nonlinear functions. Hydrol. Earth Syst. Sci. 24, 2269–2285.

Hrachowitz, M., Savenije, H.H.G., Blöschl, G., McDonnell, J.J., Sivapalan, M., Pomeroy, J.W., Arheimer, B., Blume, T., Clark, M.P., Ehret, U., Fenicia, F., Freer, J.E., Gelfan, A., Gupta, H.V., Hughes, D.A., Hut, R.W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P.A., Uhlenbrook, S., Wagener, T., Winsemius, H.C., Woods, R.A., Zehe, E., Cudennec, C., 2013. A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrol. Sci. J. 58, 1198–1255. https://doi.org/10.1080/02626667.2013.803183.

Humphrey, V., Gudmundsson, L., 2019. GRACE-REC: a reconstruction of climate-driven water storage changes over the last century. Earth Syst. Sci. Data 11, 1153–1170. https://doi.org/10.5194/essd-11-1153-2019.

Jeton, A.E., Watkins, S.A., Lopes, T.J., Huntington, J., 2005. Evaluation of Precipitation Estimates from PRISM for the 1961-90 and 1971-2000 Data Sets, Nevada. United States Geological Survey Scientific Investigations Rep. 2005-5291, 26 pp., https://pubs.usgs.gov/sir/2005/5291/.

Jiang, C., Ryu, Y., Fang, H., Myneni, R., Claverie, M., Zhu, Z., 2017. Inconsistencies of interannual variability and trends in long-term satellite leaf area index products. Glob. Change Biol. 23, 4133–4146. https://doi.org/10.1111/gcb.13787.

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., Reichstein, M., 2019. The FLUXCOM ensemble of global land-atmosphere energy fluxes. Sci. Data 6, 1–14. https://doi.org/10.1038/s41597-019-0076-8.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S.I., Sheffield, J., Goulden, M.L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A.J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B.E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A.D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., Zhang, K., 2010. Recent decline in the global land evapotranspiration trend due to limited moisture supply. Nature 467, 951–954. https://doi.org/10.1038/nature09396.

Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N., Chevallier, F., Gans, F., Groll, D.S., Haverd, V., Koehler, P., Ichii, K., Jain, A.K., Liu, J., Lombardozzi, D., Nabel, J.E.M.S., Nelson, J.A., Pallandt, M., Papale, D., Peters, W., Pongratz, J., Rödenbeck, C., Sitch, S., Tramontana, G., Walker, A., Weber, U., Reichstein, M., 2020. Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach. Biogeosciences 17, 1343–1365. https://doi.org/10.5194/bg-17-1343-2020.

Kahler, D.M., Brutsaert, W., 2006. Complementary relationship between daily evaporation in the environment and pan evaporation. Water Resour. Res. 42, W05413. https://doi.org/10.1029/2005WR004541.

Kim, D., Lee, W.-S., Kim, S.T., Chun, J.A., 2019. Historical drought assessment over the contiguous United States using the generalized complementary principle of evapotranspiration. Water Resour. Res. 55, 6244–6267. https://doi.org/10.1029/2019WR024991.

Kim, H. (2017). Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1). Data Integration and Analysis System (DIAS). doi: 10.20783/DIAS.501.

Kolassa, J., Reichle, R.H., Liu, Q., Alemohammad, S.H., Gentine, P., Aida, K., Asanuma, J., Bircher, S., Caldwell, T., Colliander, A., Cosh, M., Collins, C.H., Jackson, T.J., Martinez-Fernandez, J., McNairn, H., Pacheco, A., Thibeault, M., Walker, J.P., 2018. Estimating surface soil moisture from SMAP observations using a Neural Network technique. Remote Sens. Environ. 204, 43–59. https://doi.org/10.1016/j.rse.2017.10.045.

Kyatengerwa, C., Kim, D., Choi, M., 2020. A national-scale drought assessment in Uganda based on evapotranspiration deficits from the Bouchet hypothesis. J. Hydrol. 580. https://doi.org/10.1016/j.jhydrol.2019.124348.

Landerer, F.W., Swenson, S.C., 2012. Accuracy of scaled GRACE terrestrial water storage estimates. Water Resour. Res. 48. https://doi.org/10.1029/2011wr011453.
Lemordant, L., Gentine, P., Swann, A.S., Cook, B.I., Scheff, J., 2018. Critical impact of vegetation physiology on the continental hydrologic cycle in response to increasing CO2. Proc. Natl. Acad. Sci. U. S. A. 115, 4093–4098. https://doi.org/10.1073/pnas.1720712115.
Liu, W., 2018. Evaluating remotely sensed monthly evapotranspiration against water balance estimates at basin scale in the Tibetan Plateau. Hydrol. Res. 49 (6), 1977–1990.
Liu, Y., Liu, R., Chen, J.M., 2012. Retrospective retrieval of long-term consistent global leaf area index (1981–2011) from combined AVHRR and MODIS data. J. Geophys. Res. Biogeosci. 117. https://doi.org/10.1029/2012JG002084.
Lundquist, J., Hughes, M., Gutmann, E., Kapnick, S., 2019. Our skill in modeling mountain rain and snow is bypassing the skill of our observational networks. Bull. Am. Meteor. Soc. 100, 2473–2490. https://doi.org/10.1175/bams-d-19-0001.1.
Lyapustin, A., Wang, Y., Xiong, X., Meister, G., Platnick, S., Levy, R., Franz, B., Korkin, S., Hilker, T., Tucker, J., Hall, F., Sellers, P., Wu, A., Angal, A., 2014. Scientific impact of MODIS C5 calibration degradation and C6+ improvements. Atmos. Meas. Tech. 7, 4353–4365. https://doi.org/10.5194/amt-7-4353-2014.
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. J. Geophys. Res. Atmos. 122, 12245–12268. https://doi.org/10.1002/2017JD027597.
Ma, N., Szilagyi, J., 2019. The CR of evaporation: A calibration-free diagnostic and benchmarking tool for large-scale terrestrial evapotranspiration modeling. Water Resour. Res. 55, 7246–7274. https://doi.org/10.1029/2019wr024867.
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. J. Geophys. Res. Atmos. 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 Resour. Res. 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., Beck, H.E., Dorigo, W.A., Verhoest, N.E.C., 2017. GLEAM v3: satellite-based land evaporation and root-zone soil moisture. Geosci. Model Dev. 10, 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017.
Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jović, D., Woollen, J., Rogers, E., Berbery, E.H., Ek, M.B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., Shi, W., 2006. North American regional reanalysis. Bull. Amer. Meteor. Soc. 87, 343–360. https://doi.org/10.1175/bams-87-3-343.
Miralles, D.G., Gentine, P., Seneviratne, S.I., Teuling, A.J., 2019. Land-atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges. Ann. New York Acad. Sci. 1436, 19–35. https://doi.org/10.1111/nyas.13912.
Morton, F.I., 1983. Operational estimates of areal evapotranspiration and their significance to the science and practice of hydrology. J. Hydrol. 66 (1–4), 1–76. https://doi.org/10.1016/0022-1694(83)90177-4.
Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P.A., Dolman, A.J., Fisher, J.B., Jung, M., Ludwig, F., Maignan, F., Miralles, D.G., McCabe, M.F., Reichstein, M., Sheffield, J., Wang, K., Wood, E.F., Zhang, Y., Seneviratne, S.I., 2013. Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. Hydrol. Earth Syst. Sci. 17, 3707–3720. https://doi.org/10.5194/hess-17-3707-2013.
Penman, H.L., 1948. Natural evaporation from open water, bare soil and grass. Proc. R. Soc. Lond. Ser. A 193, 120–145.
Polhamus, A., Fisher, J.B., Tu, K.P., 2013. What controls the error structure in evapotranspiration models? Agric. For. Meteorol. 169, 12–24. https://doi.org/10.1016/j.agrformet.2012.10.002.
Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Wea. Rev. 100, 81–92.
Rodell, M., 2004. Basin scale estimates of evapotranspiration using GRACE and other observations. Geophys. Res. Lett. 31. https://doi.org/10.1029/2004gl020873.
Sieck, L.C., Burges, S.J., Steiner, M., 2007. Challenges in obtaining reliable measurements of point rainfall. Water Resour. Res. 43. https://doi.org/10.1029/2005wr004519.
Sivapalan, M., 2003. Prediction in ungauged basins: a grand challenge for theoretical hydrology. Hydrol. Process. 17, 3163–3170. https://doi.org/10.1002/hyp.5155.
Sun, S., Song, Z., Chen, X., Wang, T., Zhang, Y., Zhang, D., Zhang, H., Hao, Q., Chen, B., 2020. Multimodel-based analyses of evapotranspiration and its controls in China over the last three decades. Ecohydrol. 13. https://doi.org/10.1002/eco.2195.
Sun, Y., Piao, S., Huang, M., Ciais, P., Zeng, Z., Cheng, L., Li, X., Zhang, X., Mao, J., Peng, S., Poulter, B., Shi, X., Wang, X., Wang, Y.-P., Zeng, H., 2016. Global patterns and climate drivers of water-use efficiency in terrestrial ecosystems deduced from satellite-based datasets and carbon cycle models. Glob. Ecol. Biogeogr. 25, 311–323. https://doi.org/10.1111/geb.12411.
Stoy, P.C., El-Madany, T.S., Fisher, J.B., Gentine, P., Gerken, T., Good, S.P., Klosterhalfen, A., Liu, S., Miralles, D.G., Perez-Priego, O., Rigden, A.J., Skaggs, T.H., Wohlfahrt, G., Anderson, R.G., Coenders-Gerrits, A.M.J., Jung, M., Maes, W.H., Mammarella, I., Mauder, M., Migliavacca, M., Nelson, J.A., Poyatos, R., Reichstein, M., Scott, R.L., Wolf, S., 2019. Reviews and syntheses: turning the challenges of partitioning ecosystem evaporation and transpiration into opportunities. Biogeosciences 16 (19), 3747–3775.
Szilagyi, J., 2007. On the inherent asymmetric nature of the complementary relationship of evaporation. Geophys. Res. Lett. 34. https://doi.org/10.1029/2006GL028708.
Szilagyi, J., 2014. Temperature corrections in the Priestley-Taylor equation of evaporation. J. Hydrol. 519, 455–464. https://doi.org/10.1016/j.jhydrol.2014.07.040.
Szilagyi, J., 2018a. Anthropogenic hydrological cycle disturbance at a regional scale: State-wide evapotranspiration trends (1979–2015) across Nebraska, USA. J. Hydrol. 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. Hydrol. Res. 49, 648–657. https://doi.org/10.2166/nh.2017.078.
Szilagyi, J., Crago, R., Qualls, R., 2017. A calibration-free formulation of the complementary relationship of evaporation for continental-scale hydrology. J. Geophys. Res. Atmos. 122, 264–278. https://doi.org/10.1002/2016jd025611.
Szilagyi, J., Jozsa, J., 2018. Evapotranspiration trends (1979–2015) in the Central Valley of California, USA: Contrasting tendencies during 1981–2007. Water Resour. Res. 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. Geophys. Res. Lett. 41, 435–441. https://doi.org/10.1002/2013gl058979.
Tang, D., Ma, C., Wang, Y., Xu, X., 2017. Multiscale evaluation of NCEP and CRUNCEP data sets at 90 large U.S. cities. J. Geophys. Res. Atmos. 122, 7433–7444. https://doi.org/10.1002/2016jd026165.
Tang, G., Long, D., Behrangi, A., Wang, C., Hong, Y., 2018. Exploring deep neural networks to retrieve rain and snow in high latitudes using multisensor and reanalysis data. Water Resour. Res. 54, 8253–8278. https://doi.org/10.1029/2018wr023830.
Tapley, B.D., Bettadpur, S., Ries, J.C., Thompson, P.F., Watkins, M.M., 2004. GRACE measurements of mass variability in the Earth system. Science 305, 503–505.
Teuling, A.J., de Badts, E.A.G., Jansen, F.A., Fuchs, R., Buitink, J., Hoek van Dijke, A.J., Sterling, S.M., 2019. Climate change, reforestation/afforestation, and urbanization impacts on evapotranspiration and streamflow in Europe. Hydrol. Earth Syst. Sci. 23, 3631–3652. https://doi.org/10.5194/hess-23-3631-2019.
Tramontana, G., Jung, M., Schwalm, C.R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M.A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., Papale, D., 2016. Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. Biogeosciences 13, 4291–4313. https://doi.org/10.5194/bg-13-4291-2016.
Velpuri, N.M., Senay, G.B., Singh, R.K., Bohms, S., Verdin, J.P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: using point and gridded FLUXNET and water balance ET. Remote Sens. Environ. 139, 35–49. https://doi.org/10.1016/j.rse.2013.07.013.
Wang, D., Morton, D., Masek, J., Wu, A., Nagol, J., Xiong, X., Levy, R., Vermote, E., Wolfe, R., 2012. Impact of sensor degradation on the MODIS NDVI time series. Remote Sens. Environ. 119, 55–61. https://doi.org/10.1016/j.rse.2011.12.001.
Weedon, G.P., Balsamo, G., Bellouin, N., Gomes, S., Best, M.J., Viterbo, P., 2015. The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. Water Resour. Res. 50, 7505–7514. https://doi.org/10.1002/2014WR015638.
Wei, Y., Liu, S., Huntzinger, D.N., Michalak, A.M., Viovy, N., Post, W.M., Schwalm, C.R., Schaefer, K., Jacobson, A.R., Lu, C., Tian, H., Ricciuto, D.M., Cook, R.B., Mao, J., Shi, X., 2014. The North American Carbon Program Multi-scale Synthesis and Terrestrial Model Intercomparison Project-Part 2: Environmental driver data. Geosci. Model Dev. 7, 2875–2893. https://doi.org/10.5194/gmd-7-2875-2014.
Xiao, Z., Liang, S., Wang, J., Xiang, Y., Zhao, X., Song, J., 2016. Long-time-series global land surface satellite leaf area index product derived from MODIS and AVHRR surface reflectance. IEEE Trans. Geosci. Remote Sens. 54, 5301–5318.
Yang, D., Kane, D., Zhang, Z., Legates, D., Goodison, B., 2005. Bias corrections of long-term (1973–2004) daily precipitation data over the northern regions. Geophys. Res. Lett. 32, L19501. https://doi.org/10.1029/2005gl024057.
Yaseen, Z.M., Sulaiman, S.O., Deo, R.C., Chau, K.-W., 2019. An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. J. Hydrol. 569, 387–408. https://doi.org/10.1016/j.jhydrol.2018.11.069.
Zhang, K., Kimball, J.S., Running, S.W., 2016. A review of remote sensing based actual evapotranspiration estimation. Wiley Interdiscip. Rev. Water 3, 834–853. https://doi.org/10.1002/wat2.1168.
Zhu, Z., Bi, J., Pan, Y., Ganguly, S., Anav, A., Xu, L., Samanta, A., Piao, S., Nemani, R., Myneni, R., 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 Sens. 5, 927–948. https://doi.org/10.3390/rs5020927.