A continental-scale evaluation of the calibration-free complementary relationship with physical, machine-learning, and land-surface models

Daeha Kim[1], Minha Choi[2], Jong Ahn Chun[3]

5 [1]Department of Civil Engineering, Jeonbuk National University, Jeonju, Jeollabuk-do, 54896, Republic of Korea
[2]Environment and Remote Sensing Laboratory, Department of Water Resources, Graduate School of Water Resources, Sungkyunkwan University, Suwon, Gyeonggi-do, 16419, Republic of Korea
[3]Prediction Research Department, APEC Climate Center, Busan, 48058, Republic of Korea

10 *Correspondence to*: Jong Ahn Chun (jachun@apcc21.org)

Abstract. The widespread negative correlation between the atmospheric vapor pressure deficit and soil moisture lends strong support to the complementary relationship (CR) of evapotranspiration. While it has showed outstanding performance in predicting actual evapotranspiration (ET_a) over land surfaces, the calibration-free CR formulation has not been tested in the Australian continent dominantly under (semi-)arid climates. In this work, we comparatively evaluated its predictive performance with seven physical, machine-learning, and land surface models for the continent at a 0.5°×0.5° grid resolution. Results showed that the calibration-free CR that forces a single parameter to everywhere produced considerable biases when comparing to water-balance ET_a (ET_{wb}). The CR method was unlikely to outperform the other physical, machine-learning, and land surface models, overrating ET_a in (semi-)humid coastal areas for 2002-2012 while underestimating in arid inland locations. By calibrating the parameter against water-balance ET_a independent of the simulation period, the CR method became able to outperform the other models in reproducing the spatial variation of the mean annual ET_{wb} and the interannual variation of the continental means of ET_{wb}. However, interannual the grid-scale variability and trends were captured unacceptably even after the calibration. The calibrated parameters for the CR method were significantly correlated with the mean net radiation, temperature, and wind speed, implying that (multi-)decadal climatic variability could diversify the optimal parameters for the CR method. The other physical, machine-learning, and land surface models provided a consistent indication with the prior global-scale assessments. We also argued that at least some surface information is necessary for the CR method to describe long-term hydrologic cycles at the grid scale.

1 Introduction

Terrestrial evapotranspiration (ET_a) links water, energy, and carbon exchanges between lands and the atmosphere. On the global scale, more than 60% of land precipitation (P) returns to the atmosphere through plants' vascular systems and 30 soil pores, consuming over 70% of surface net radiation for the phase change of water (Trenberth et al., 2009; 2007). Due to

the warming atmosphere, the upward latent heat flux has received growing attention, because it can control surface water availability, plants productivity, and ecosystem sustainability (Pareek et al., 2020; Kyatengerwa et al., 2020; Jasechko, 2018; Swann et al., 2016). Interacting with the atmosphere, changes in ET_a could substantially increase heatwave risks (Miralles et al., 2014a; Mueller and Seneviratne, 2012) and alter precipitation patterns (Koster et al., 2004).

35   However, expensive costs and operational difficulties make ET_a observation networks (e.g., the FLUXNET; Baldocchi et al., 2001) subject to limited spatial extent, short data lengths, and questionable data quality. Hence, modeling approaches are inevitable for a regional- or a global-scale ET_a analysis, and usually based on physical theories (e.g., Zhang et al., 2016), machine-learning techniques (e.g., Jung et al., 2011), and conceptual land surface schemes (e.g., Haverd et al., 2018) that inherently have numerous error sources. For instance, Jung et al. (2019) produced the global mean ET_a from 557 to 668
40   mm a^{-1} for 2001-2005 with physical and machine learning models. The previous global mean ET_a estimates have varied even in a larger range of 417-650 mm a^{-1} for a part or the whole period of 1982-2011 (Pan et al., 2020). The large discrepancies in the global means imply that modeling prescriptions for ET_a have diverse uncertainty sources, such as forcing errors, ill-posed parameterizations, structural deficiencies, and insufficient training, and thus necessitate intercomparison studies to assess the associated limitations and uncertainties (e.g., Pan et al., 2020).

45   When it depends on a physical equation, such as the Penman-Monteith (Monteith, 1965) or the Priestley-Taylor (Priestley and Taylor, 1972) equations, an ET_a model assumes typically that ET_a under water deficiency is proportional to the atmospheric evaporative demand (ET_p). In the Global Land Evaporation Amsterdam Model (GLEAM; Martens et al., 2017), for example, the Priestley-Taylor equation is multiplied by a stress module to predict ET_a under water limited conditions. Similar approaches are easily found with physical and land surface models that adjust the surface roughness length of the
50   Penman-Monteith equation to represent water stress (e.g., Zhang et al., 2016; Pan et al., 2015). Note that adjusting the surface roughness length is mathematically equivalent to multiplying a coefficient to ET_p (Seneviratne et al., 2010). Though it has reliably predicted ET_a at multiple scales (e.g., Martens et al., 2017; Fisher et al., 2008), the proportionality assumption unavoidably requires soil moisture information to quantify the degree of water stress, giving rise to practical difficulties such as data unavailability, computational inefficiency, and delayed data dissemination. Importantly, the assumption of the positive
55   relation between ET_a and ET_p could be rejected by observational evidence that supports negative correlations between the two (e.g., Brutsaert, 2006; Ramírez et al., 2005; Hobbins et al., 2004). Han et al. (2014) emphasized that the correlation between ET_a and ET_p depends mainly on water availability rather than being always positive.

The drawbacks of the proportionality assumption can be remedied at least in part by employing the complementary relationship (CR) of evapotranspiration. Bouchet (1963) found that the pan evaporation rate over a small wet patch surrounded
60   by water-limited areas is higher than when the same surroundings are entirely wet. Since the small wet patch hardly influences the overlying atmosphere, ET_p over the wet surface is raised by blending with the drier and hotter surroundings. This "oasis effect", by contrast, is negligible in the case that the surrounding areas are entirely wet and large enough to transform the overpassing atmosphere. In other words, even under the same surface radiation and wind speed conditions, ET_p responds to changes in regional water availability. Hence, one can predict water-limited ET_a by gauging how much ET_p is raised from the

65   hypothetical evaporation rate that should occur under the full wetness (referred to as the wet-environment ET; ET_w). Since a

higher adjustment in ET_p indicate a lower water availability and thus ET_a, the CR supports inverse correlations between ET_a

and ET_p (i.e., complementarity). In practice, the complementarity allows users to predict ET_a with no surface information,

because ET_p and ET_w are all obtainable from meteorological data.

After Brutsaert and Parlange (1998) who used the CR to interpret the globally declining pan evaporation rates, various

70   CR methods have been formulated, e.g., Anayah and Kaluarachchi (2014), Crago and Qualls (2013), Huntington et al. (2011),

Kahler and Brutsaert (2006), Crago and Crowley (2005), Hobins et al. (2004) among others. While those CR methods have

been deemed mere heuristic methods with limited reliability (Shuttleworth et al., 2009), the non-dimensional derivation of

Brutsaert (2015) and the following modifications (Szilagyi et al., 2017; Crago et al., 2016) have suggested the generality and

definitiveness of the CR principle. The non-dimensional CR formulations have shown outstanding performance in predicting

75   water-limited ET_a at local, regional, and global scales (e.g., Brutsaert et al., 2020; Crago and Qualls, 2018; Brutsaert et al.,

2017), and applications are extended to drought assessments (Kyatengerwa et al., 2020; Kim et al., 2019b) and even used for

predicting the crop coefficient under the proportionality assumption (Kim et al., 2019a).

Though the non-dimensional CR formulations are still under improvement based on the thermodynamic foundations

(Szilagyi, 2021; Qualls and Crago, 2020), they mostly require any ET_a observations to identify required parameters. Szilagyi

80   et al. (2017) is the only calibration-free CR method that analytically determines the parameter for ET_w with no requirement of

ET_a data. By transferring the parameter analytically obtained in highly humid locations to the entire region of interest, the

calibration-free CR formulation showed superior performance in predicting ET_a to typical land surface and machine-learning

models in the conterminous United States and China where climates are very diverse (Ma et al., 2019; Ma and Szilagyi, 2019).

However, the same approach has not been examined in a continent where only small parts are under humid climates, and thus

85   it is questionable whether the parameter transferring from humid locations is valid.

In this work, therefore, we applied the calibration-free CR formulation for the Australian continent where land

surfaces are mostly under (semi-)arid climates, and its predictive performance was compared with a bunch of physical,

machine-learning, and land surface models. Here, we addressed that the use of a single parameter for the entire continent could

lead the CR method to low performance in preserving spatial coherence, interannual variability, and decadal trends of water-

90   balance ET_a. We also provided some perspectives for improving the non-dimensional CR formulations.

2 Methodology and data

2.1 The generalized complementary relationship approach

The CR principle explains the feedback response of ET_p to regional water deficiency using the three evaporation

rates, namely, ET_a, ET_p, and ET_w. Again, ET_a is the actual water flux from a homogeneous land surface to the atmosphere, and

95   ET_p is the atmospheric capacity to receive water vapor that responds actively to water availability on the surface. ET_w is the

hypothetical ET_a rate that would take place under the same atmospheric conditions but ample water. ET_p under regional water

deficiency would be far higher than the hypothetical ET_w, because high atmospheric vapor pressure deficits (VPD) co-exist with low soil moisture across the globe (Zhou et al., 2019).

In the CR formulation by Szilagyi et al. (2017), the two dimensionless variables, $x \equiv ET_w/ET_p$ and $y \equiv ET_a/ET_p$, are defined, and they are linked with four boundary conditions. If water is ample on the surface, ET_a reaches ET_w that should be equal to ET_p owing to no water deficiency; thus, the first boundary condition is (i) $y = 1$ for $x = 1$. When the surface is entirely desiccated, ET_a must be nil (i.e., $y = 0$), and by energy balance, the surface radiation should be fully transformed to the sensible heat flux that fully amplifies VPD. In other words, under the given radiation and wind speed, ET_p is maximized when $ET_a = 0$, providing another zero-order boundary condition: (ii) $y = 0$ for $x = x_{min} \equiv ET_w/E_{pmax}$, where E_{pmax} is the maximized ET_p. When $x = 1$ (i.e., with ample water), ET_a would change as much as changes in ET_w, yielding a first-order boundary condition: (iii) $dy/dx = 1$ for $x = 1$. On the other boundary (i.e., $x = 0$), ET_a should be constant irrespective of any changes in ET_w; thus, another zero-order boundary condition is (iv) $dy/dx = 0$ for $x = 0$. The simplest polynomial equation satisfying the four boundary conditions is:

$$y = 2X^2 - X^3, \tag{1a}$$

with X defined as:

$$X \equiv \frac{x - x_{min}}{1 - x_{min}} = \frac{E_{pmax} - ET_p}{E_{pmax} - ET_w} \frac{ET_w}{ET_p}. \tag{1b}$$

Since ET_p, ET_w, and E_{pmax} could be all obtainable from a set of net radiation, air temperature and humidity, and wind speed data, Eqs. (1a) and (1b) allow users to estimate ET_a with no direct soil moisture information (e.g., remote-sensing soil moisture products). Szilagyi et al. (2017) used the Penman (1948) equation for ET_p:

$$ET_p = \frac{\Delta_{avg}}{\Delta_{avg} + \gamma} \frac{R_n}{\lambda_v} + \frac{\gamma}{\Delta_{avg} + \gamma} f_u VPD, \tag{2}$$

where, Δ_{avg} is the slope of the saturation vapor pressure curve (kPa °C⁻¹) at the mean air temperature T_{avg} (°C), γ is the psychometric constant (kPa °C⁻¹), R_n is the surface net radiation less the soil heat flux (MJ m⁻² d⁻¹), λ_v is the latent heat of vaporization (MJ kg⁻¹) (here we quantified it by $\lambda_v = 2.501 - 0.00236 T_{avg}$), $f_u = 2.6(1 + 0.54 u_2)$ is the Rome wind function, where u_2 is the 2-m wind speed (m s⁻¹), and VPD is $e_s(T_{avg})$ minus e_a, where $e_s(T_{avg})$ is the saturation vapor pressure at T_{avg} and e_a is the actual vapor pressure, respectively.

As it is parameterized under wet surface conditions, ET_w could be quantified by the Priestly and Taylor (1972) equation:

$$ET_w = \alpha_e \frac{\Delta_{ws}}{\Delta_{ws} + \gamma} \frac{R_n}{\lambda_v}, \tag{3}$$

where, α_e is a coefficient typically ranging between 1.10 and 1.32 (Szilagyi et al., 2017), and Δ_{ws} is the slope of the vapor pressure curve (kPa °C⁻¹) at the wet surface temperature T_{ws} for which Szilagyi (2014) used the two methods based on the Bowen ratio and the derivation of Monteith (1980). We used the latter:

$$T_{ws} = T_{wb} + \frac{\gamma R_n VPD}{(\Delta_{wb} + \gamma)(c_1 R_n + c_2 f_u VPD)}, \tag{4a}$$

$$c_1 = \frac{\Delta_{avg}(\Delta_{wb}+\gamma)-\Delta_{wb}(\Delta_{avg}+\gamma)}{\Delta_{avg}+\gamma}, \tag{4b}$$

$$c_2 = \frac{\lambda_v\gamma(\Delta_{wb}+\gamma)}{\Delta_{avg}+\gamma}, \tag{4c}$$

130 where, T_{wb} is the wet-bulb temperature that can be estimated by the Bowen ratio for an adiabatic change:

$$\gamma\frac{T_{wb}-T_{avg}}{e_s(T_{wb})-e_a} = -1. \tag{5}$$

E_{pmax} is calculated with the same Penman equation but with different temperature and humidity conditions. The air overpassing a desiccated surface is likely devoid of humidity; thus, e_a would become negligible:

$$E_{pmax} = \frac{\Delta_{dry}}{\Delta_{dry}+\gamma}\frac{R_n}{\lambda_v} + \frac{\gamma}{\Delta_{dry}+\gamma}f_u e_s(T_{dry}), \tag{6}$$

135 where, Δ_{dry} is the slope of the vapor pressure curve (kPa °C^{-1}) at the dry air temperature T_{dry} (°C). T_{dry} is the hypothetical air temperature that would adiabatically reach when the latent heat flux is nil:

$$T_{dry} = T_{wb} + \frac{e_s(T_{wb})}{\gamma} = T_{avg} + \frac{e_s(T_{avg})}{\gamma}. \tag{7}$$

The coefficient α_e in Eq. (3) is the only parameter for the CR method. Szilagyi et al. (2017) proposed to use the mean value of α_e values analytically obtained at humid locations for a region of interest, and we achieved $\alpha=1.09$ by the same

140 approach (the details are given in the Appendix). Although it is smaller than the typical Priestley-Taylor coefficient (1.26), the obtained α_e was still higher than the physical lower limit (i.e., unity). It should be noted that the α_e incorporated in the CR method is a model parameter analogous to the Priestley-Taylor coefficient rather than having the same physical meaning (Ma et al., 2019; Brutsaert et al., 2017). In prior continental-scale studies, the optimal α_e values for continental-scale applications were often lower than the typical Priestley-Taylor coefficient (1.26) (e.g., Ma et al, 2020; Kim et al., 2019b; Ma et al., 2019;

145 Ma and Szilagyi, 2019).

2.2 Data used for ET_a estimation and performance evaluation

2.2.1 Atmospheric forcing and evaluation datasets

The study area was the Australian continent lying within [10°- 45° S, 113°- 155° E], and the atmospheric forcing data for the CR method were collected from the ERA Interim reanalysis archive (Dee et al., 2011) of the European Centre for

150 Medium-Range Weather Forecasts (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim; last access on mmm-dd/2020). The monthly averages of surface net solar radiation and net thermal radiation, 2-m air temperature, 2-m dew-point temperature, and 10-m U and V wind speed components were downloaded at the 0.5°×0.5° grid resolution for 1979-2018. R_n was estimated simply by summing the two net radiation data (i.e., the soil heat flux was assumed to be negligible). T_{avg} and VPD were directly quantified by the air temperature and the dew-point temperature datasets, while the 10-m U and V

155 components were converted to u_2 values using the logarithmic wind speed profile.

The CR ET_a estimates from Eq. (1a) were evaluated with the water-balance ET_a (ET_{wb}) estimates at the same 0.5° grid resolution. To achieve the grid-scale ET_{wb}, some syntheses (e.g., spatial and temporal gap filling and/or conceptual modelling) are inevitable due to the non-uniformity and unavailability of in-situ precipitation, streamflow, and terrestrial water storage (TWS) observations. We collected the global precipitation (P) product v.2018 of the Global Precipitation Climate

160   Centre (GPCC) together with the grid runoff (Q) products by Ghiggi et al. (2019) and Hobeichi et al. (2019) and the TWS anomalies reconstructed by Humphrey and Gudmundsson (2019). The GPCC monthly P data are readily available for 1891-present from https://psl.noaa.gov/data/gridded/data.gpcc.html (last access on Jun-01/2020). The monthly GRUN was produced by Ghiggi et al. (2019) at the 0.5° resolution for 1902-2014 using in-situ streamflow observations and a machine learning algorithm, and cross-validated by independent discharge data in major river basins across the world

165   (https://doi.org/10.6084/m9.figshare.9228176; lass access on May-20/2020). The Linear Optimal Runoff Aggregate (LORA) of Hobeichi et al. (2016) merged the syntheses of eleven land surface models using an optimal weighting approach for 1980-2012 (https://geonetwork.nci.org.au/geonetwork/srv/eng/catalog.search#/metadata/f9617_9854_8096_5291m; last access on Dec-23/2020). Hobeichi et al. (2016) validated the LORA Q data with published discharge observations at many river basins. Since we believed that Q values from multiple sources are more reliable than a single model synthesis, the Q value at each

170   grid was determined by simple averages of the GRUN and the LORA data. The TWS data, namely the GRACE-REC, extend the Gravity Recovery and Climate Experiments (GRACE) land water-equivalent-thickness data given for 2003-2015 to the year of 1901 using a statistical methods (https://doi.org/10.6084/m9.figshare.7670849; last access on Jan-4/2021). We calculated annual TWS changes (δS) at each grid by differencing the December TWS estimates of two consecutive years. Then, the annual ET_{wb} in each year was calculated by the water balance equation, i.e., $\overline{ET_{wb}} = \overline{P} - \overline{Q} - \delta S$, where $\overline{ET_{wb}}$, $\overline{P}$, and

175    $\overline{Q}$ are the annual averages of ET_{wb}, P, and Q, respectively.

Figure 1 displays the distribution of the wetness index (the long-term-average ratio of P to ET_p) for 2002-2012 calculated with the GPCC P and the ET_p from the ERA-Interim forcing. Typically, the wetness index categorizes hyper-arid, arid, semi-arid, semi-humid, and humid climates with the thresholds of 0.05, 0.2, 0.5, and 0.65, respectively (Barrow, 1992). The range of the wetness index was 0.07-5.65 in Australia, indicating that 81% of the land surfaces was under (semi-)arid

180   climates though there are no hyper-arid areas. Humid climates are found in northern, southwestern, and southeastern coastal regions, where major cities and agricultural areas have developed.

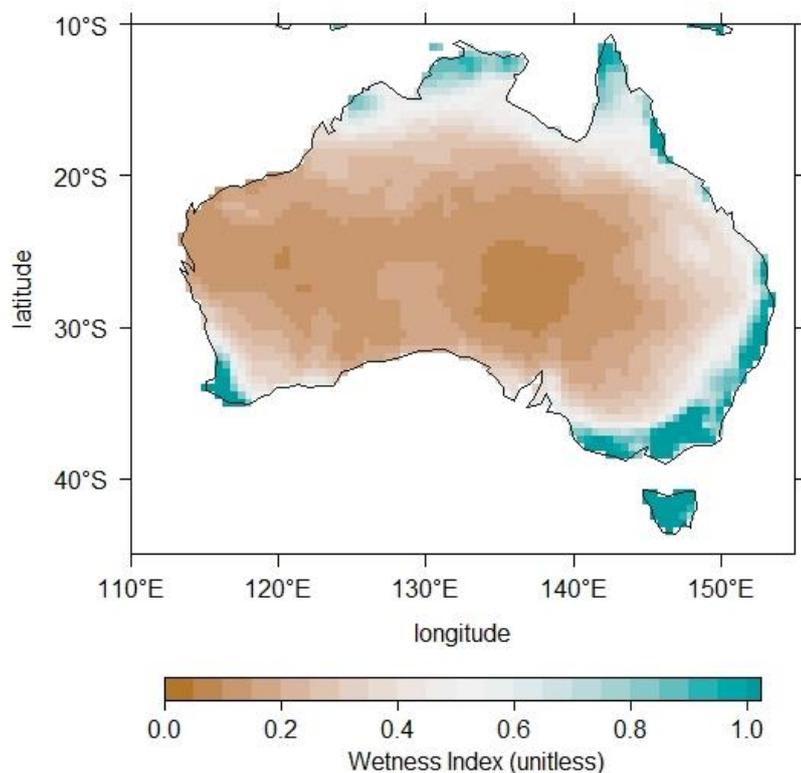



Figure 1: Distribution of the wetness index for 2002-2012 in Australia.

2.2.2 ET$_a$ products for comparative evaluation

185   The performance of the CR method was compared with products from two remote-sensing-based physical models, a machine-learning model, and four land surface schemes. The physical models include the Global Land Evaporation Amsterdam Model (GLEAM) v3.2 (Martens et al., 2017; https://www.gleam.eu; last access on Jun-03/2020) and the Priestley-Taylor Jet Propulsion Laboratory Model (PT-JPL) (Fisher et al., 2008; http://josh.yosh.org/datamodels.htm; last access on Dec-03/2020). The GLEAM uses a vertically-stacked bucket model in combination with the Priestley-Taylor equation

190   constrained by microwave-derived soil moisture, surface temperature, and vegetation optical depth. It synthesizes canopy transpiration, bare-soil evaporation, and interception losses under water- and energy-limited conditions separately. On the other hand, the PT-JPL employs the same Priestley-Taylor equation multiplied with the relative surface wetness from remote-sensing vegetation indices. The GLEAM and the PT-JPL ET$_a$ products were shown to have high coherence with eddy-covariance latent heat flux observations (Martens et al., 2017; Fisher et al., 2008).

195   In addition, the FluxCom was selected as the machine-learning ET$_a$ product (Jung et al., 2019; http://www.fluxcom.org/; last access Mar-18/2019). This dataset was produced by upscaling the point flux-tower observations with 11 machine learning algorithms with remote-sensing vegetation indices. Jung et al. (2019) checked the cross-consistency

of the FluxCom data via comparison against the model-tree ensemble (Jung et al., 2011), the GLEAM v3.1, and the LandFlux-Eval datasets (Mueller et al. 2013). The FluxCom provides several variations; among them, we chose the dataset forced by the

200   CRUNCEPv8 with the length from 1950 to 2016.

The land surface models include the Australian Water Resources Assessment Landscape Model v.5 (AWRA-L; Frost et al., 2016), the Noah Land Surface Model v.3.3 (Noah3.3; Ek et al., 2003), the Catchment Land Surface Model v. Fortuna 2.5 (CLSMF2.5; Koster et al., 2000), and the ERA-Interim land surface scheme (ERA-Interim; Balsamo, 2008). The AWRA-L is a one-dimensional grid water balance model developed for monitoring water stresses in Australian soil, underground, and

205   land surfaces (http://www.bom.gov.au/water/landscape; last access on Dec-30/2020). The AWRA-L ET_a is the sum of interception, soil evaporation, groundwater evaporation, and transpiration from the root zone and deep groundwater generated by a conceptual tank model combining the Penman-Monteith equation. The Noah3.3 and the CLSMF2.5 ET_a products were generated via the Land Information System (LIS; Kumar et al., 2006) of the National Aeronautics and Space Administration (NASA), which supports application of multiple community land surface models. The Noah3.3 is the operational land surface

210   scheme of the National Center for Atmospheric Research (NCAR) discretizing the surfaces with the finite difference method and solving the governing equations associated with the soil-vegetation-snowpack continuum. On the other hand, the CLSMF2.5 was developed by the NASA Global Modeling and Assimilation Office, subdividing the irregular shapes of catchments into saturated, sub-saturated, and wilting fractions, and simulating simulates water fluxes in each fraction of catchments that evolves over time. For our comparative evaluation, both the Noah 3.3 and CLSMF2.5 models were forced by

215   the National Centers for Environmental Prediction Global Data Assimilation System (NCEP GDAS) forcing dataset (https://portal.nccs.nasa.gov). The simulation time step and spin-up period were 15 min and 76 years (four times 19 years (2000–2018), respectively. The Hydrologogicval Modeling and Analysis Platform (Getirana et al. 2012) routing scheme was applied for routed streamflow estimates. Lastly, the ERA-Interim uses the improved land surface scheme formulated by Viterbo and Beljaars (1995) to simulate the evolution of heat and water storages in soil and snow layers. It classifies a land

220   surface using satellite data and ancillary information. The downward water fluxes in a land pixel are generated by the governing equations, and the latent heat flux to the lowest atmospheric level is computed with the Obukhov length.

For comparison between the nine ETa products, the different spatial resolutions were bilinearly unified to the standard 0.5° grid of the ERA-Interim forcing data. We compared all the ET_a products to the grid-scale ET_{wb}, and evaluated the reproducibility in spatial variation of long-term mean ET_a and the continental means together with the grid-scale interannual

225   variability and linear trends.

3 Results

3.1 Evaluation of spatial variations and continental means

Figure 2a illustrates the distribution of long-term ET_{wb} means for 2002-2012. The spatial variation of the mean ET_{wb} was consistent with the analysis of Zhang et al. (2010) for which a typical Budyko function was adopted to produce the mean

230   ET_a at 0.05° resolution across Australia for 2000-2005. Despite the different data length and the spatial resolution, Zhang et al. (2010) suggested that that the mean ET_a in Australia is the lowest in the east-central part receiving very small precipitation. The region with AI < 0.10 in Figure 1 corresponds approximately to where the mean ET_a is very low in Zhang et al. (2010) and Figure 2a. In the arid Australian continent, the precipitation pattern mostly determines the spatial variation of mean ET_a, because about 90% of precipitation returns to the atmosphere (Glenn et al., 2011). Figure 2a, too, depicts that the means of

235   ET_{wb} were small in the arid east-central part, increasing towards northern and eastern coasts where precipitation is abundant due to monsoonal and easterly winds. The continental mean ET_{wb} for 2002-2012 was 431 mm a^{-1} approximately close to the value (439 mm a^{-1}) given by the global assessment of Zhang et al. (2016). The consistency to prior studies led us believe that the annual ET_{wb} product from the reanalysis precipitation and the reconstructed runoff and TWS data could become an acceptable evaluation reference.

240   As expected, the mean ET_{wb} tended to increase with AI. In arid regions (AI < 0.2), the mean ET_{wb} was 260 ± 71.9 mm a^{-1} (mean ± standard deviation), while semi-arid regions (0.2 ≤ AI < 0.5) had a range of 489 ± 152 mm a^{-1}. Under semi-humid (0.5 ≤ AI < 0.65) and humid (AI ≥ 0.65) climates, it was within 761 ± 168 mm a^{-1} and 797 ± 273 mm a^{-1}, respectively. Compared to ET_{wb}, the CR ET_a had positive biases. The mean ET_a from the CR method for 2002-2012 ranged in 221 ± 105 mm a^{-1}, 564 ± 231 mm a^{-1}, 976 ± 235 mm a^{-1}, and 1,057 ± 297 mm a^{-1} from arid to humid regions, respectively (Figure 2b).

245   Though the pattern correlation between the mean ET_{wb} and CR ET_a was fairly high (Pearson r was 0.87), the CR method overestimated ET_a in coastal areas, while underrating it in the central-western part under arid climates.

        The two physical models, on the other hand, were biased negatively. The GLEAM produced smaller ET_a in the wet northern coastal, and the (semi-)arid central and southwestern parts than ET_{wb}, while the PT-JPL seemed to generally underestimated it across the continent (Figure 3c and d). On the contrary, the FLUXCOM product was positively biased in

250   (semi-)arid inland areas with suppressed spatial variation (Figure 3e). The land surface scheme of AWRA-L generated the unexpected dry hotspots in the mid-western part, which was not found from the water balance (Figure 3f). The two LIS land surface models, the Noah3.3 and the CLSMF2.5, relatively well captured the spatial variation of mean ET_{wb}, although there were some underestimations by the Noah3.3 in the northern part (Figure 3g and h). The ERA-Interim ET_a was of largely biased in coastal areas (Figure 3i).

255   The continental means ET_a for 2002-2012 from the eight models provide a consistent indication (Figure 3a). The CR method generated a positive bias of +58.2 mm a^{-1} (+14%) relative to the water-balance ET_{wb}. Among the eight models, the CLSMF2.5 produced the minimum bias (+0.5%), whereas the largest bias (+24%) was from the ERA-Interim ET_a. In the Taylor (2001) diagram that measures the standard deviation, and the root mean square error (RMSE) and the pattern correlation to ET_{wb} together, the CR method appeared to be in the 6th rank. The CLSMF2.5 ET_a was the closest to ET_{wb}, and followed in

260   order by the PT-JPL, the AWRA-L, the FluxCom, the Noah3.3, the CR method, the GLEAM, and the ERA-Interim.

        The continental annual means of CR ET_a were considerably higher than those from ET_{wb} owing to the positive biases in coastal areas, and their interannual variability was smaller than that of ET_{wb} (Figure 4). The GLEAM well captured the interannual variation of the continental mean ET_{wb} through 1981-2012 with slight underestimation; however, the PT-JPL ET_a

were of considerable negative biases. As did the kindred machine learning model of Jung et al. (2011), the FluxCom suppressed
265 the temporal variation of the annual means. Though the AWRA-L, the Noah3.3, and the CLSMF2.5 appeared to have small
biases to the annual means of ET_{wb}, the temporal correlations of the two LIS land surface models to ET_{wb} were lower than the
other models (0.7 and 0.75 for the Noah3.3 and the CLSMF2.5 vs. 0.86-0.98 for the other models). As expected, the ERA-
Interim land surface scheme produced even higher continental means than the CR method.

Overall, though the calibration-free CR have showed outstanding performance with the simple mathematical formulations,
270 e.g., in the conterminous United States (Ma and Szilagyi, 2019; Kim et al., 2019b) and in China (Ma et al., 2019), it was
unlikely to outperform the other models in the arid Australian continent. The CR ET_a estimates were considerably biased in
wet coastal areas, their spatial coherence with ET_{wb} was weaker than the other ET_a products, and their variation of continental
means was suppressed.

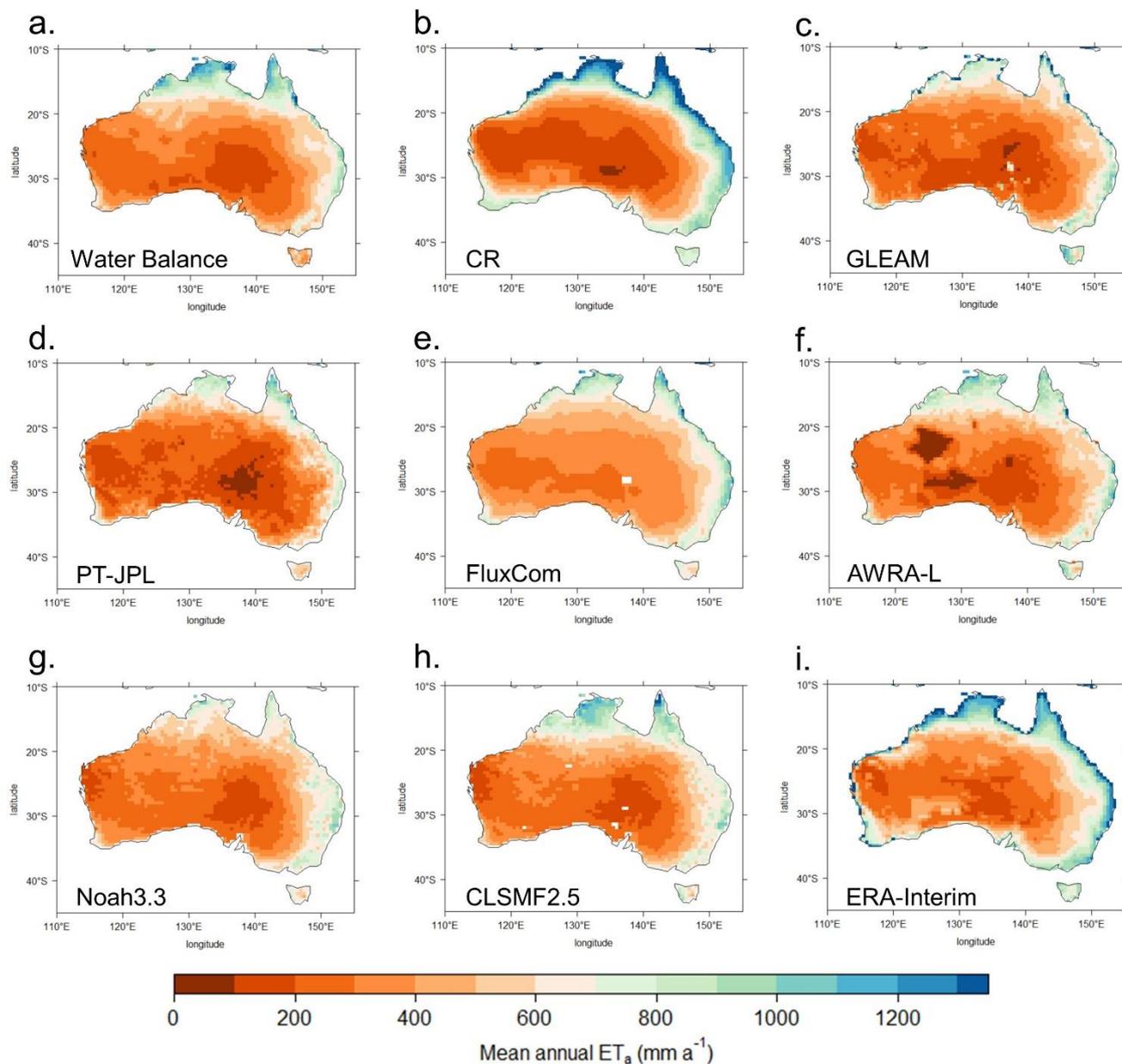



Figure 2: Distributions of the mean annual ET$_a$ for 2002-2012 produced by the (a) water balance, (b) CR method, (c) GLEAM, (d) PT-JPL, (e) FluxCom, (f) AWRA-L, (g) Noah3.3, (h) CLSMF2.5, and (i) ERA-Interim.

a.

b.

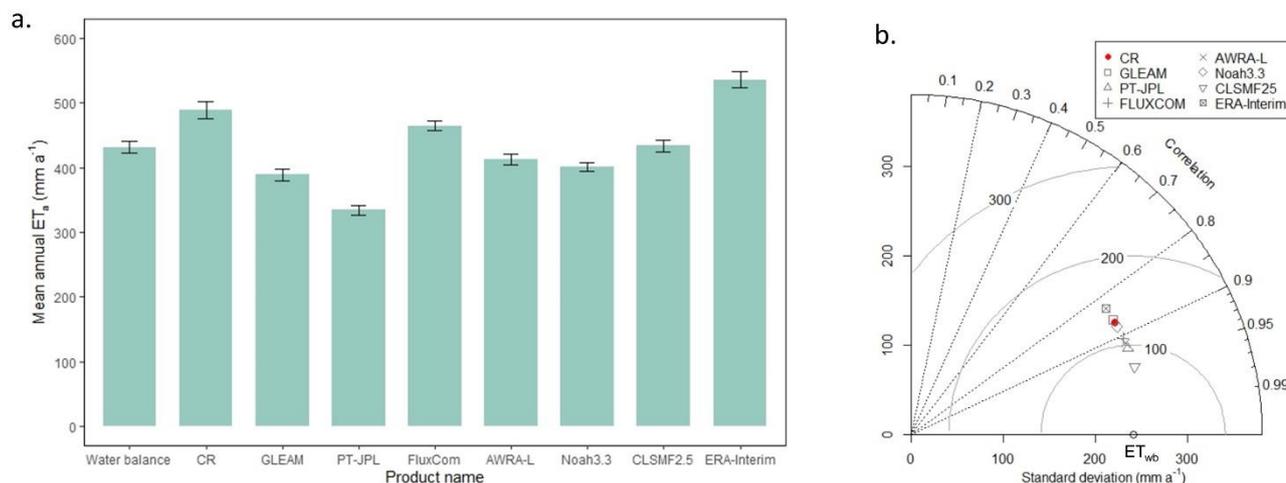



Figure 3: (a) Comparison of the continental means of the ET$_a$ estimates for 2002-2012, and (b) the Taylor diagram comparing the standard deviation, the root mean square error, and the pattern correlation between the modeled mean annual ET$_a$ and ET$_{wb}$. The whiskers indicate the two standard deviations of the continental means.

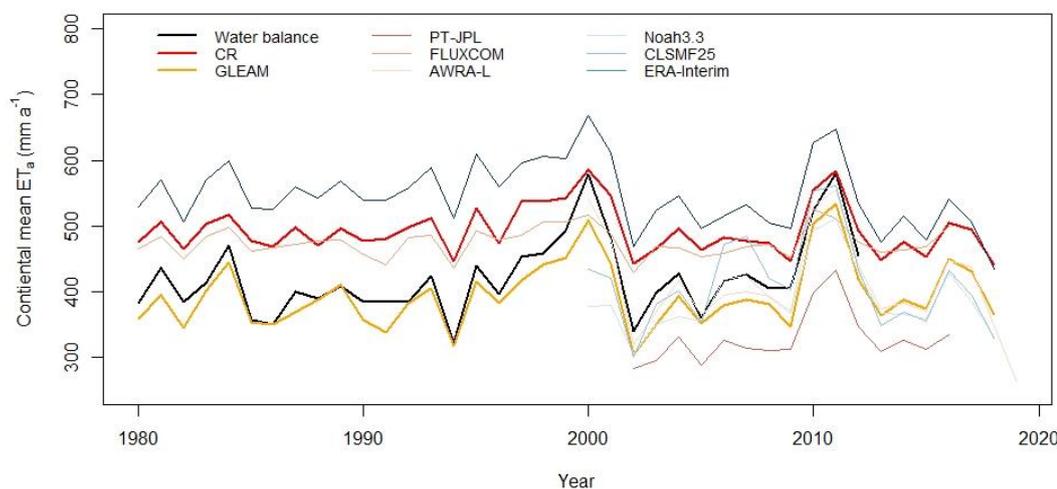



Figure 4: Interannual variations of the continental annual means for 1980-2019 estimated by the nine methods. Note that the AWRA-L, the Noah3.3, and the CLSMF2.5 data are given from 2000, while the PT-JPL data were produced for 2002-2016.

3.2 Evaluation of interannual variability and temporal trends at the grid scale

Figure 5 displays spatial patterns of the standard deviations of annual ET$_a$ for 2002-2012 from the water balance and the eight models. The ET$_{wb}$ was highly variable particularly in the northern and the eastern parts relative to other regions. The high ET$_a$ variability in the northern and the eastern parts of Australia was illustrated by Pan et al. (2020) for which state-of-the-art physical, machine-learning, and land surface models were compared together. The interannual variability of CR ET$_a$ product, however, had much weaker than ET$_{wb}$ particularly in the northern and the eastern coastlines. It was smaller than that

generated by the physical (GLEAM and JPL) and the land surface models (AWRL-A, Noah3.3, CLSMF2.5, and ERA-Interim). While the vast majority of models in Pan et al. (2020) suggested the largest interannual ET_a in regions with annual precipitation between 700 and 1,000mm a^{-1}, the calibration-free CR method was unable to pronounce such variations in (semi-)humid locations.

In the global-scale evaluation of Pan et al. (2020), land surface models pronounced higher ET_a variability relative to physical and machine-learning models. Figure 5 provide a consistent indication that the annual ET_a averages from AWRA-L, Noah3.3, CLSMF2.5, and ERA-Interim had interannual variability much higher than those from the PT-JPL and the machine-learning FluxCom. Although physically-based, the GLEAM showed the similar ET_a variations to the land surface models. The machine-learning FluxCom pronounced the suppressed interannual variability across the continent, as did in Pan et al. (2020) and Ma et al. (2020).

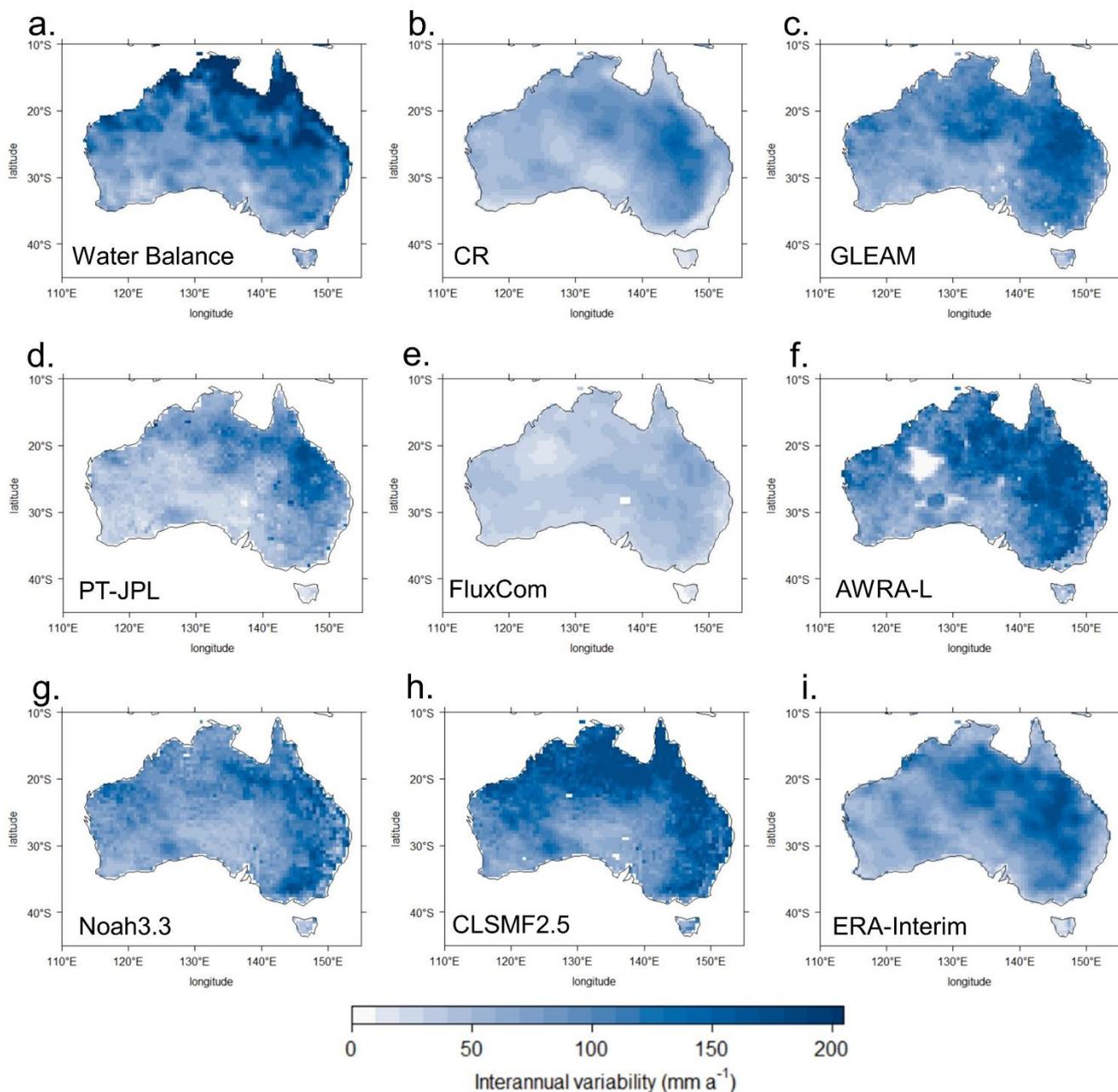
The maps in Figure 6 are distributions of the median trends in annual ET_a time series over 2002-2012. Given the transition from the weak El-Nino in 2002 to the strong La-Nina in 2010-2012 (Miralles et al., 2014b), steep upward trends (> 40 mm a^{-2}) in ET_{wb} were found around the northeastern and the eastern parts of Australia. Relative to ET_{wb}, the CR ET_a estimates gradually increased in the same region, and in the coastal areas, they rather declined unexpectedly. On the contrary, despites discrepancies, the ET_a products from the GLEAM, the AWRA-L, the Noah3.3, and CLSMF2.5 well reflected the variations expected from changes in sea surface temperatures. The PT-JPL and the ERA-Interim provided smaller areas with the strong ET_a trends. Although significant, the rising trends of the FluxCom were the smallest among the eight models with the unexpected declining trends in the eastern coast.

310

Figure 5: As in Figure 2, but for interannual variability for 2002-2012.

000

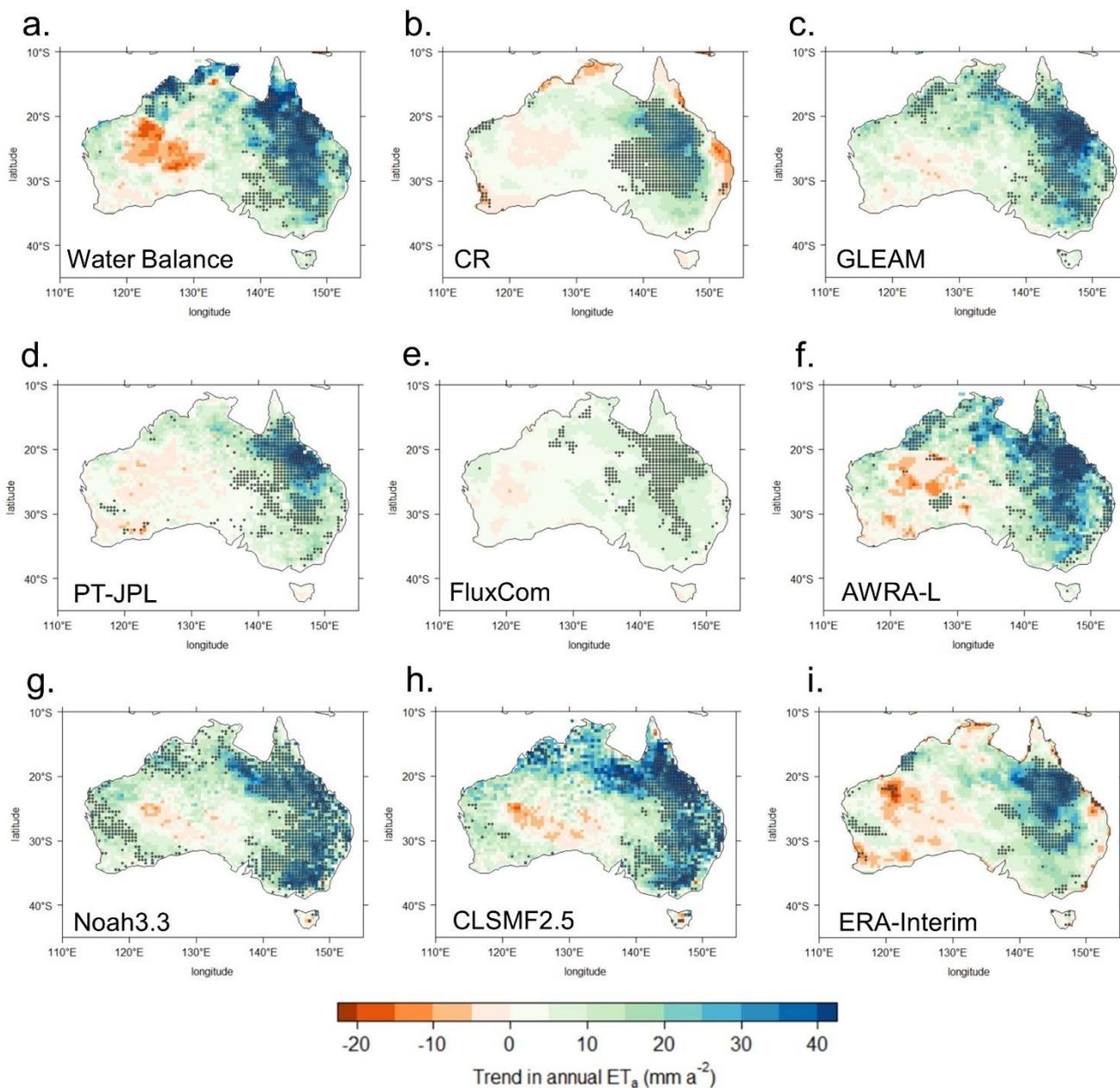



Figure 6: Distributions of the median trends in annual ET$_a$ time series for 2002-2012 given by the nine methods. The dots indicate the statistical significance at 5% level.

4 Discussion and perspectives

4.1 Performance of the CR method in Australia

320   In the continental-scale applications by Ma and Szilagyi (2019), Kim et al. (2019b), and Ma et al. (2019), the calibration-free CR formulation of Szilagyi et al. (2017) was superior to widely-used physical, machine learning, and land surface models. With the slight difference in calculation of T_{ws}, here we applied the same approach for the continent smaller than the conterminous United States. Unexpectedly, the CR method performed worse than the chosen physical and land surface models, possibly because Australian land surfaces are mostly under (semi-)arid climates. In Szilagyi et al. (2017), the long-

325   term sum of CR ET_a was often larger than that of precipitation in some locations in the western United States. Though the extensive irrigation for agriculture partly explains the unrealistically high ET_a in such an area (e.g., Szilagyi and Jozsa, 2018), questions still remain on the relatively poor performance of the CR method in (semi-)arid environments.

We here argue that the inferior performance of the CR method could be led by the parameter α_e=1.09 fixed across the large continent where climatic gradients are steep between dry inland and wet coastal areas. In Brutsaert et al. (2020), the

330   Priestley-Taylor coefficient within a CR formulation is tightly related with climatological aridity. In other words, applying the constant α_e in every location could nullify the influence of climatological variation on the lower bound of ET_p. We hence checked the spatial variation of α_e optimal for the CR method by calibrating it against the long-term mean ET_{wb} for 1991-2001 at each grid. Figure 7 displays the optimal α_e that minimizes the biases between the mean ET_{wb} and CR ET_a, indicating that it tends to increase from coastal areas to western inland locations. In the western part, the calibrated α_e was often greater than

335   the typical range of the Priestley-Taylor coefficient (1.1-1.32). The distribution of the optimal α_e was likely to reduce the positive biases in coastal locations and the underestimation in the western part produced by application of the constant α_e=1.09. When regenerating ET_a for 2002-2012 with the distributed α_e, the CR method became the best among the eight models in reproducing the temporal and spatial variations of mean ET_{wb} and in the Taylor diagram (Figure 8). Note that the optimal α_e was found with the ET_{wb} set separated from the new simulation.

340   The α_e distribution in Figure 7 provides a counteractive indication to Brutsaert et al. (2020) in which the parameter α_e within the CR of Brutsaert (2015) was calibrated against the eddy-covariance observations across the globe. The relationship between climatological aridity and α_e developed by Brutsaert et al. (2020) implicates the tendency of α_e exponentially decreasing with aridity. In contrast, Figure 7 indicates that wet regions have α_e values lower than in (semi-)arid inland locations. The central-eastern part receiving little precipitation has larger α_e values than the western part. The contradictory result from

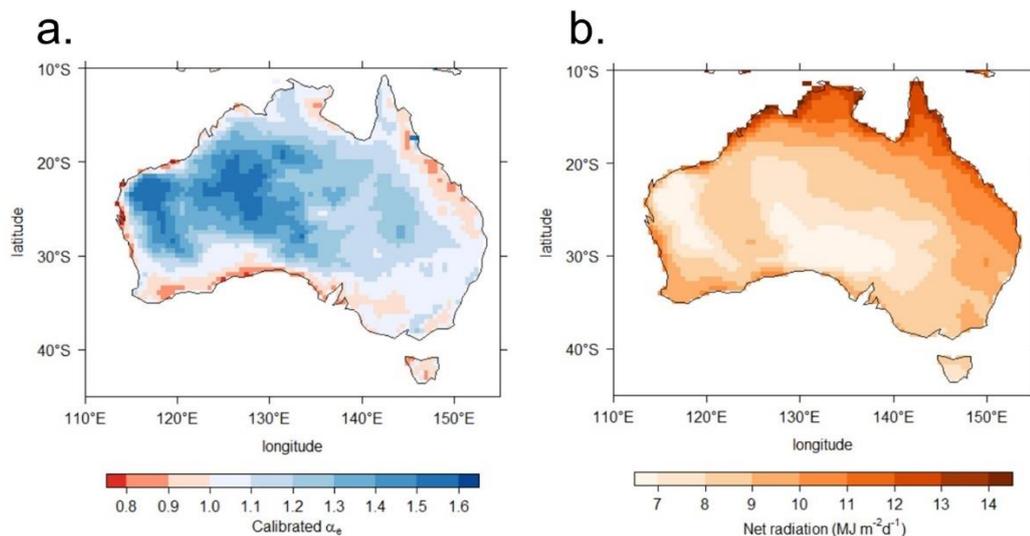
345   our work could be explained by the CR formulation from Brutsaert (2015). In the original CR derivation by Brutsaert (2015), ET_w was forced to become nil for $ET_a = 0$, posing an ill-defined assumption that R_n is always zero even when a desiccated surface is warm (Crago and Qualls, 2018; Szilagyi et al., 2017, Crago et al., 2016), and making the calibrated α_e oversensitively decline with aridity. Since Szilagyi et al. (2017) mended this problem by introducing the upper bound of ET_p (i.e., E_{pmax}), the calibrated α_e in Figure 7a would not hold the same sensitivity to aridity changes.

350

Figure 7: Distributions of (a) the calibrated α_e against the mean ET_{wb} and (b) the mean R_n for 1991-2001.

Instead, one could analytically relate the parameter α_e with climatic variables by equating the Penman and the

Priestley-Taylor equations, because ET_p and ET_w must be equal under ample water conditions, yielding:

$$\alpha_e = 1 + \frac{\gamma}{\Delta_{ws}} \frac{f_u VPD_{ws}}{R_n/\lambda_v}, \tag{8}$$

355   where, VPD_{ws} is the VPD of the atmosphere overpassing the hypothetical wet surface. Although VPD_{ws} would be small owing

to the interacting wet surface (Brutsaert and Stricker, 1979), Eq. (8) implies that the climatic variables, R_n, Δ_{ws}, and u_2, could

amplify the effects of VPD_{ws} on α_e. We conducted simple correlation analyses between the calibrated α_e values and the

corresponding averages of R_n, T_{avg} (i.e., the control of Δ_{ws}), and u_2. The Pearson r values of the α_e to the three variables were -

0.56, 0.44, and 0.29, respectively (significant at 1% levels). The simple regression analyses, in addition, showed that the R_n,

360   T_{avg}, and u_2 explain 32%, 19%, and 9% of the variation of the calibrated α_e values, respectively (significant at 1% levels). In

other words, variation of R_n plays a substantial role in determining the optimal α_e for the CR method, and it was found that

locations with high α_e tend to have low R_n (Figure 7b).

Nevertheless, the calibrated CR provided little improvements in reproducing the interannual variability and trends of

ET_{wb}. Despite the slightly increased interannual variability in central-northern areas, the new maps of interannual variability

365   and trends (Figure 9) were still similar to the outcomes from the fixed $\alpha_e=1.09$. This might be attributable in part to vegetation

dynamics neglected in the CR method. The Penman and Priestley-Taylor equations assume stationary land-surface parameters,

and are unable to capture the plants' behaviors to atmospheric conditions that play a considerable role in the precipitation

partitioning (Jasechko, 2018). Modelling studies showed that the absence of surface responses to CO_2 fertilization has led

offline hydrologic models to runoff projections contradictory to simulations of earth system models (e.g., Milly and Dunne,

370   2016; Swann et al., 2016; Roderick et al., 2015). Yang et al. (2019) mended this problem by incorporating a simple surface

roughness formulation to the elevated CO_2 into the Penman-Monteith equation. Even though the CR method with the Penman

equation has worked well in Australia for the past decades (e.g., Crago and Quall, 2018), the prior studies suggest that surface responses to atmospheric changes could considerably affect temporal changes in ET_p, ET_w and thus ET_a. This necessitates further refinements for the CR method to synthesize the surface behaviors explicitly under non-water-limiting conditions. It is

375   worth noting that Australian carbon sink has enhanced during the 21st century even at increasing wildfire risks owing to the plants' water-use efficiency and productivity increased by CO_2 fertilization (Kelly and Harrison, 2014).

In short, to capture the spatial variation of mean ET_a in Australia, the CR method needs to consider the influence of climatic variables on the parameter α_e. To regenerate the interannual variability and trends, the equations in the CR formulation are seemingly required incorporating dynamic surface parameters. In this case, the operational advantage of the CR method

380   (i.e. no need of surface data) could disappear in return.

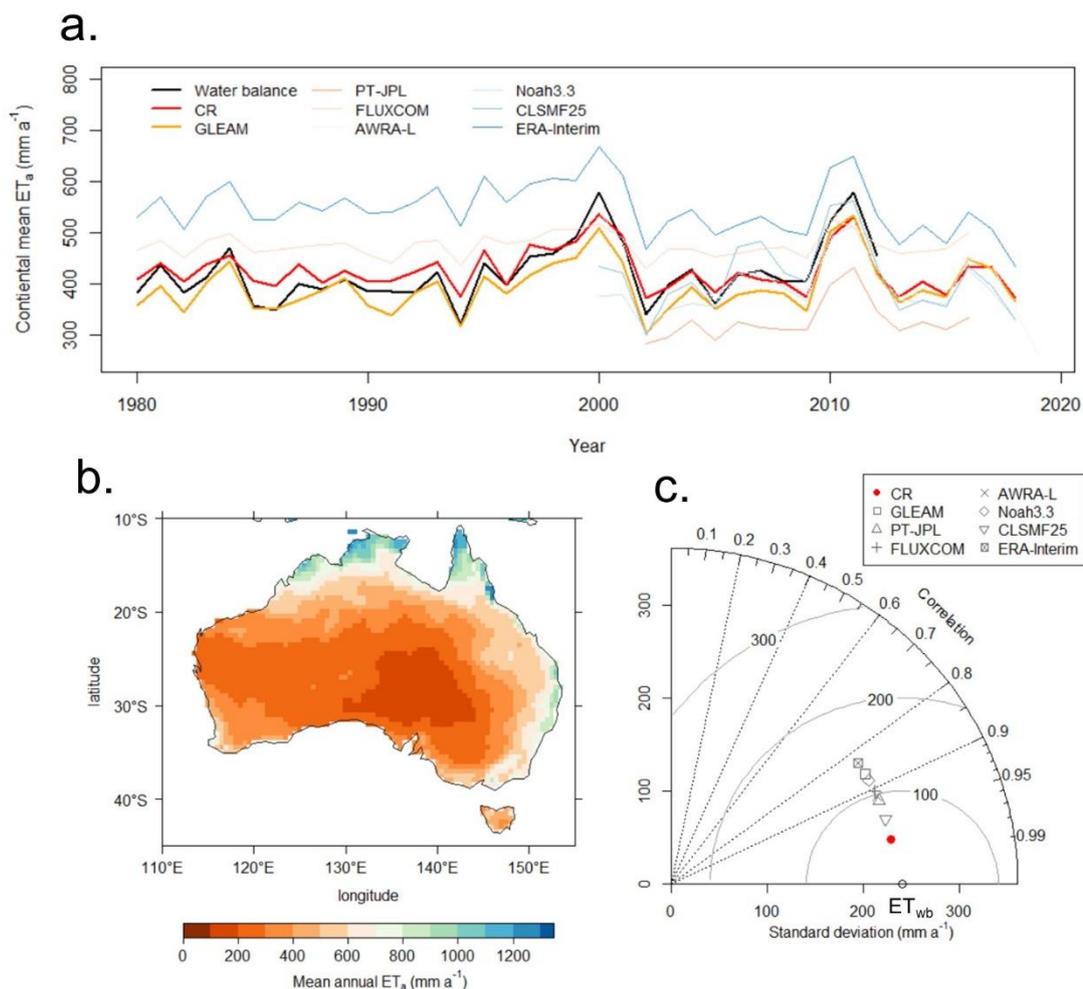



Figure 8: (a) interannual variation of the continental mean annual ET_a, (b) the distribution of mean annual ET_a for 2002-2012 and (c) the Taylor diagram comparing the eight ET_a products against ET_{wb} when the CR method is calibrated.

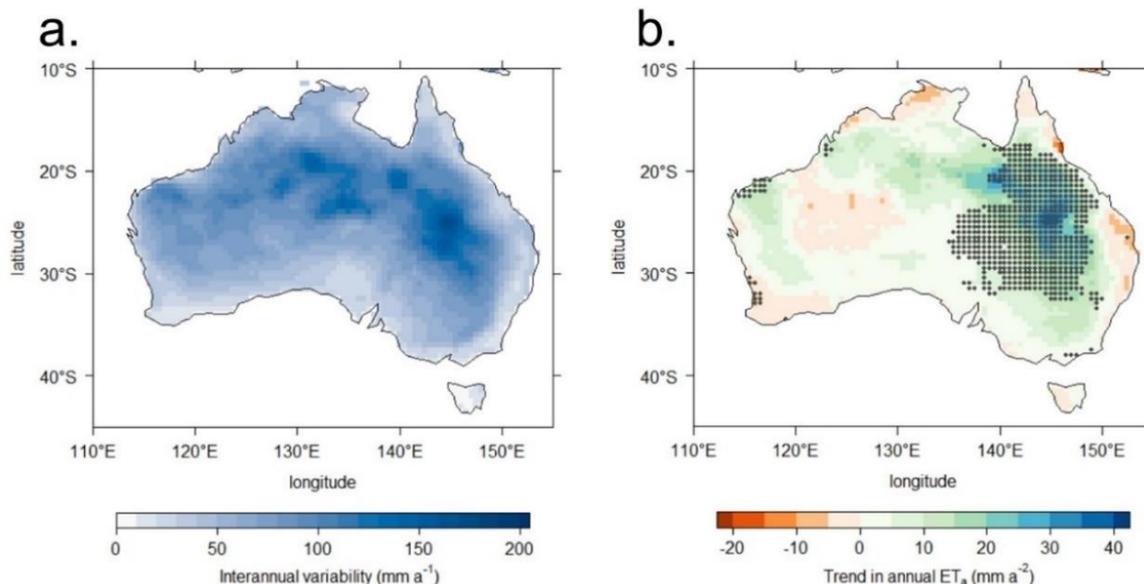



Figure 9: The distributions of (a) interannual variability for 2002-2012 and (b) trends of annual ET$_a$ from the CR method with calibrated α$_e$. The dots indicate the statistical significance at 5% level.

4.2 Intercomparison between ET$_a$ products

Since precipitation is the primary control of ET$_a$ in the dry continent, biases and errors in the GRUN and LORA runoff dataset are unlikely to induce large biases in the grid-scale water balance. As mentioned earlier, the ratio of ET$_a$ to precipitation is approximately 90% in Australia, suggesting that caveats in the GPCC precipitation are major error sources to ET$_{wb}$. Typically, errors in a grid precipitation product are introduced by: (i) the systematic measuring errors from evaporation out of rain gauges and aerodynamic effects, and (ii) the sampling errors from low gauging density. The GPCC precipitation takes an advanced correction and anomaly interpolation methods for reducing the systematic and the sampling errors via a very rigorous quality control framework (Schneider et al., 2014). The precipitation product has well closed the global water budget, becoming a reliable evaluation reference for other grid precipitation products (e.g., Sun et al., 2017). The quality of the GPCC data, hence, was unlikely a major concern.

Compared to ET$_{wb}$, however, the other ET$_a$ models are subject to diverse limitations. The remote-sensing physical models do not account for soil moisture dynamics playing a pivotal role in canopy conductance and bare soil evaporation in (semi-)arid regions (Pan et al., 2020). While the GLEAM takes soil moisture into account in the ET$_a$ synthesis, whereas the PT-JPL does not consider the soil moisture dynamics and underrates ET$_a$ in shrubs and deserts in the southern hemisphere (Miralles et al., 2016). Given that bare soil evaporation introduces the largest error to ET$_a$ estimates (Talsma et al., 2018), the PT-JPL needs any corrections for operational ET$_a$ monitoring in Australia.

On the other hand, the machine-learning FluxCom, which showed the worst performance in this work, has important caveats. Even though it acceptably simulates long-term averages of the surface energy fluxes, the FluxCom carbon fluxes are

405 likely to have too small interannual variations in the energy fluxes (Jung et al., 2019). Recently, Ma et al. (2020) also emphasized the deficiency of the FluxCom in reproducing long-term ET_a trends in the United States. Given that predictive performance of a machine-learning algorithm depends critically on training datasets (Yao et al., 2017), the machine learning ET_a product needs to be further trained by any datasets describing the interannual behaviors.

Pan et al. (2020) showed that interannual variability of ET_a products by 14 land surface models was dominantly

410 controlled by precipitation in most of regions in the Southern Hemisphere. However, they also highlighted the dynamic root parameterization of the ORCHIDEE-MICT model (Guimberteau et al., 2018), which is distinguishable from the other models, suggesting that ET_a changes in Australia could be less sensitive to precipitation changes than indicated by commonly-adopted land surface models. Hence, larger interannual variability than in ET_{wb} could be an indication that land surface parameterization might be oversensitive to precipitation changes (e.g., the simulations in the eastern part by the AWRA-L and the CLSMF2.5).

415 Though it outperformed other models in streamflow generation (Frost et al., 2015), the AWRA-L needs corrections for the unexpected dry hotspots in inland areas. The two LIS land surface models, in addition, provided a consistent indication with the prior application for the Upper Blue Nile River (Jung et al., 2017). The CLSMF2.5 tends to provide higher ET_a and lower streamflow than the Noah3.3, and it better represented the water budget in the sub-humid river basin. The overestimation in the ERA-Interim product was also found by Miralles et al. (2016) and Mueller et al. (2013). Sun et al. (2017) pointed out that

420 the ERA-Interim often prescribed annual precipitation exceeding the GPCC P data.

5 Conclusions

In this work, we evaluated applicability of the calibration-free CR formulation in Australian land surfaces mostly under (semi-)arid climates. The terrestrial evapotranspiration (ET_a) produced by the CR method was compared with a bunch of ET_a products from physical, machine-learning, and land surface models, and their spatial and temporal variations and

425 decadal trends were evaluated against the estimates from water balance. While it could generate the ET_a product strongly correlated with the water-balance estimates, the CR method seemed to introduce considerable biases when comparing to the other models. In Australia mostly under (semi-)arid climates, the approach proposed by Szilagyi et al. (2017) was unlikely to outperform typically-adopted physical, machine-learning, and land surface models, and thus necessitates better parameterization for improvement. We draw the following conclusions worth emphasizing:

430 (1) The optimal coefficient (α_e) for the wet-environment evapotranspiration is unlikely constant. The α_e=1.09 obtained from the calibration-free approach introduced positive biases in (semi-)humid coastal areas while underestimating in arid locations. When calibrating α_e each grid with the independent set of water-balance estimates, α_e seems to respond to climatic variations.

(2) Even with the calibrated α_e, the CR method insufficiently captured the interannual variability and the decadal

435 trends of the water-balance estimates at the grid scale. Since the latent heat flux is not only controlled by water stress but atmospheric conditions (e.g. CO_2 concentration), any formulation that captures land surface behaviors

under non-water-liming conditions would be necessary in quantification of the wet-environment and potential evapotranspiration.

440 (3) The evaluations of the physical, the machine-learning, and the land surface models provided a consistent implication with the prior global-scale studies. A remote-sensing physical model can better represent the surface energy balance by explicit consideration of soil moisture dynamics. The machine-learning depending largely on training datasets can suppress interannual variability and lead to overestimation in arid locations. ET_a products from land surface models could be more sensitive to precipitation variability than physical and machine-learning models.

Author contributions

445 DK, MC, and JAC designed the study all together. DK simulated terrestrial evapotranspiration with the CR method and drafted the manuscript. JAC run the LIS land surface models and collected the other datasets for comparative evaluation. MC participated in discussion and review of the results and the manuscript.

Competing interests

450 The authors declare no competing interests.

Code availability

The R scripts for the CR method are available upon request from the leading author (daeha.kim@jbnu.ac.kr).

Acknowledgements

This study is supported by the APEC Climate Center. We also acknowledge the support by the National Research Foundation
455 of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2019R1A2B5B01070196).

References

Anayah, F. M. and Kaluarachchi, J. J.: Improving the complementary methods to estimate evapotranspiration under diverse climatic and physical conditions, Hydrol. Earth Syst. Sci., 18, 2049–2064, https://doi.org/10.5194/hess-18-2049-2014, 2014.
Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R.,
460 Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., and Wofsy, S.: FLUXNET: A New Tool to Study the

Temporal and Spatial Variability of EcosystemScale Carbon Dioxide, Water Vapor, and Energy Flux Densities, B. Am. Meteorol. Soc., 82, 2415–2434, doi:10.1175/1520- 0477(2001)0822.3.CO;2, 2001.

Bechtold, P., Köhler, M., Jung, T., Doblas-Reyes, F., Leutbecher, M., Rodwell, M.J., Vitart, F. and Balsamo, G.: Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales, Q. J. R. Meteorol. Soc., 134, 1337–1351, https://doi.org/10.1002/qj.289, 2008.

Barrow, C. J.: World atlas of desertification (United Nations Environment Programme), edited by N. Middleton and D. S. G. Thomas. Edward Arnold, London, 1992. Land Degrad. Dev. 3 (4), 249, https://doi.org/10.1002/ldr.3400030407, 1992.

Bouchet, R. J.: Evapotranspiration re'elle et potentielle, signification climatique, in: General Assembly Berkeley, Int. Assoc. Sci. Hydrol., Gentbrugge, Belgium, Publ. No. 62, 134–142, 1963.

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

Brutsaert, W., and Parlange, M. B.: Hydrologic cycle explains the evaporation paradox, Nature, 396, 30, https://doi.org/10.1038/23845, 1998.

Brutsaert, W., Li, W., Takahashi, A., Hiyama, T., Zhang, L., and Liu, W.: 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 Resour. Res., 53, 270–282, https://doi.org/10.1002/2016WR019472, 2017.

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

Brutsaert, W.: Indications of increasing land surface evaporation during the second half of the 20th century, Geophys. Res. Lett., 33, L20403, https://doi.org/10.1029/2006GL027532, 2006.

Crago, R., and Crowley, R.: Complementary relationships for near-instantaneous evaporation, J. Hydrol., 300(1-4), 199–211, https://doi.org/10.1016/j.jhydrol.2004.06.002, 2005.

Crago, R. D., and Qualls, R. J.: The value of intuitive concepts in evaporation research. Water Resour. Res., 49, 6100–6104. https://doi.org/10.1002/wrcr.20420, 2013.

Crago, R. D., and Qualls, R. J.: Evaluation of the generalized and rescaled complementary evaporation relationships, Water Resour. Res., 54, 8086–8102, https://doi.org/10.1029/2018WR023401, 2018.

Crago, R., Szilagyi, J., Qualls, R., and Huntington, J.: Rescaling the complementary relationship for land surface evaporation. Water Resour. Res., 52, 8461–8471, https://doi.org/10.1002/2016WR019753, 2016.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N.

495   and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale ETa model, J. Geophys. Res., 108(D22), 8851, https://doi.org/10.1029/2002JD003296, 2003.

500   Fisher, J. B., Tu, K. P., and Baldocchi, D. D.: Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites, Remote Sens. Environ., 112, 901–919, https://doi.org/10.1016/j.rse.2007.06.025, 2008.

Frost, A. J., Ramchurn, A., and Smith, A.: The Bureau's Operational AWRA Landscape (AWRA-L) Model. Bureau of Meteorology Technical Report., available at

505   http://www.bom.gov.au/water/landscape/assets/static/publications/Frost__Model_Description_Report.pdf, 2016.

Getirana, A., Boone, A., Yamazaki, D., Decharme, B., Papa, F., and Mognard, N.: The Hydrological Modeling and Analysis Platform (HyMAP): Evaluation in the Amazon basin. J. Hydrometeor., 13, 1641–1665, https://doi.org/10.1175/JHM-D-12-021.1, 2012.

510   Ghiggi, G., Humphrey, V., Seneviratne, S. I., & Gudmundsson, L. GRUN: an observations-based global gridded runoff dataset from 1902 to 2014. Earth Syst. Sci. Data, 11, 1655–1674, https://doi.org/10.5194/essd-11-1655-2019, 2019.

Glenn, E.P., Doody, T.M., Guerschman, J.P., Huete, A.R., King, E.A., McVicar, T.R., Van Dijk, A.I.J.M., Van Niel, T.G., Yebra, M. and Zhang, Y.: Actual evapotranspiration estimation by ground and remote sensing methods: the Australian experience, Hydrol. Process., 25, 4103–4116, https://doi.org/10.1002/hyp.8391, 2011.

515   Guimberteau, M., Zhu, D., Maignan, F., Huang, Y., Yue, C., Dantec-Nédélec, S., Ottlé, C., Jornet-Puig, A., Bastos, A., Laurent, P., Goll, D., Bowring, S., Chang, J., Guenet, B., Tifafi, M., Peng, S., Krinner, G., Ducharne, A., Wang, F., Wang, T., Wang, X., Wang, Y., Yin, Z., Lauerwald, R., Joetzjer, E., Qiu, C., Kim, H., and Ciais, P.: ORCHIDEE-MICT (v8.4.1), a land surface model for the high latitudes: model description and validation, Geosci. Model Dev., 11, 121–163, https://doi.org/10.5194/gmd-11-121-2018, 2018.

520   Han, S., Tian, F., and Hu, H.: Positive or negative correlation between actual and potential evaporation? Evaluating using a nonlinear complementary relationship model, Water Resour. Res., 50, 1322–1336, https://doi.org/10.1002/2013WR014151, 2014.

Haverd, V., Smith, B., Nieradzik, L., Briggs, P. R., Woodgate, W., Trudinger, C. M., Canadell, J. G., and Cuntz, M.: A new version of the CABLE land surface model (Subversion revision r4601) incorporating land use and land cover change, woody

525   vegetation demography, and a novel optimisation-based approach to plant coordination of photosynthesis, Geosci. Model Dev., 11, 2995–3026, https://doi.org/10.5194/gmd-11-2995-2018, 2018.

Hobbins, M. T., Ramírez, J. A., and Brown, T. C.: Trends in pan evaporation and actual evapotranspiration across the conterminous U.S.: Paradoxical or complementary?, Geophys. Res. Lett., 31, L13503, https://doi.org/10.1029/2004GL019846, 2004.

530 Hobeichi, S., Abramowitz, G., Evans, J., and Beck, H. E.: Linear Optimal Runoff Aggregate (LORA): a global gridded synthesis runoff product, Hydrol. Earth Syst. Sci., 23, 851–870, https://doi.org/10.5194/hess-23-851-2019, 2019.

Humphrey, V., and Gudmundsson, L.: 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, 2019.

Huntington, J. L., Szilagyi, J., Tyler, S. W., and Pohll, G. M.: Evaluating the complementary relationship for estimating
535 evapotranspiration from arid shrublands. Water Resour. Res, 47, W05533. https://doi.org/10.1029/2010WR009874, 2011.

Jasechko, S.: Plants turn on the tap, Nature Clim. Change, 8, 562–563, https://doi.org/10.1038/s41558-018-0212-z, 2018.

Jung, H. C., Getirana, A., Policelli, F., McNally, A., Arsenault, K. R., Kumar, S., Tadesse, T., and Peters-Lidard, C. D.: Upper Blue Nile basin water budget from a multi-model perspective, J. Hydrol., 555, 535–546, https://doi.org/10.1016/j.jhydrol.2017.10.040, 2007

540 Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. J. Geophys. Res., 116, G00J07. https://doi.org/10.1029/2010JG001566, 2011.

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

Kim, D., Chun, J. A., and Ko, J.: A hybrid approach combining the FAO-56 method and the complementary principle for predicting daily evapotranspiration on a rainfed crop field, J. Hydrol., 577, 123941, https://doi.org/10.1016/j.jhydrol.2019.123941., 2019a

550 Kim, D., Lee, W.-S., Kim, S. T., and Chun, J. A.: 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, 2019b.

Koster, R.D., Suarez, M.J., Ducharne, A., Stieglitz, M., and Kumar, P.: A catchment based approach to modeling land surface processes in a general circulation model: 1. Model structure, J. Geophys. Res. Atmos., 105, 24809–24822,
555 https://doi.org/10.1029/2000JD900327, 2000.

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C. H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y. C., Taylor, C. M., Verseghy, D., Vasic, R., Xue, Y., and Yamada, T.: Regions of strong coupling between soil moisture and precipitation, Science, 305, 1138–1140, https://doi.org/10.1126/science.1100217, 2004.

560   Kumar, S. V., Peters-Lidard, C. D., Tian, Y., Houser, P. R., Geiger, J., Olden, S., Lighty, L., Eastman, J. L., Doty, B., Dirmeyer, P., Adams, J., Mitchell, K., Wood, E. F., and Sheffield, J.: Land information system: An interoperable framework for high resolution land surface modeling, Environ. Model. Softw., 21, 1402-1415, https://doi.org/10.1016/j.envsoft.2005.07.004, 2006.

Kyatengerwa, C., Kim, D., and Choi, M.: A national-scale drought assessment in Uganda based on evapotranspiration deficits
565   from the Bouchet hypothesis, J. Hydrol., 500, 124348. https://doi.org/10.1016/j.jhydrol.2019.124348, 2020.

Ma, N., and Szilagyi, J.: 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, 2019.

Ma, N., Szilagyi, J., Zhang, Y., and Liu, W: Complementary-relationship-based modeling of terrestrial evapotranspiration across China during 1982–2012: validations and spatiotemporal analyses, J. Geophy. Res. Atmos., 124, 4326–4351,
570   https://doi.org/10.1029/2018JD029850, 2019.

Ma, N., Szilagyim J., Zhang, Y., and Jozsa, J.: Benchmarking large-scale evapotranspiration (ET) estimates: A perspective from a calibration-free complementary relationship (CR) approach and FLUXCOM, J. Hydrol., 125221, https://doi.org/10.1016/j.jhydrol.2020.125221, 2020.

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., Beck, H. E., Dorigo, W.
575   A., and Verhoest, N. E. C.: 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, 2017.

Milly, P. C. D., and Dunne, K. A.: Potential evapotranspiration and continental drying, Nat. Clim. Change., 6, 946–949, https://doi.org/10.1038/nclimate3046, 2016.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan,
580   F., Miralles, D. G., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K., Wood, E. F., Zhang, Y., and Seneviratne, S. I.: 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, 2013.

Miralles, D. G., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M. F., Hirschi, M., Martens, B., Dolman, A. J., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F., and Fernández-Prieto, D.: The WACMOS-ET project – Part 2:
585   Evaluation of global terrestrial evaporation data sets, Hydrol. Earth Syst. Sci., 20, 823–842, https://doi.org/10.5194/hess-20-823-2016, 2016.

Miralles, D., G., Teuling, A., J. van Heerwaarden, C. C., and de Arellano, J. V.-G.: Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation, Nature Geosci., 7, 345–349, https://doi.org/10.1038/ngeo2141, 2014a.

590   Miralles, D. G., van den Berg, M. J., Gash, J. H., Parinussa, R. M., de Jeu, R. A. M., Beck, H. E., Holmes, T. R. H., Jiménez, C., Verhoest, N. E. C., Dorigo, W. A., Teuling, A. J., and Dolman, A. J.: El Niño-La Niña cycle and recent trends in continental evaporation, Nature Clim. Change, 4, 122–126, https://doi.org/10.1038/NCLIMATE2068, 2014b.

Monteith, J. L.: Evaporation and surface temperature, Q. J. R. Meteorol. Soc., 107, 1–27. https://doi.org/10.1002/qj.49710745102, 1981.

595   Monteith, J. L.: Evaporation and environment, Symposia of the Society for Exp. Biol., 19, 205–234, 1965.

Mueller, B., and Seneviratne, S. I.: Hot days induced by precipitation deficits at the global scale, P. Natl. Acad. Sci. USA, 109, 12398-12403, https://doi.org/10.1073/pnas.1204330109, 2012

Mueller, B., Hirschi, M., Jiménez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K. C., Wood, E. F., Zhang, Y., and Seneviratne, S. I.:

600   Benchmark products for land evapotranspiration: LandFluxEVAL multi-dataset synthesis, Hydrol. Earth Syst. Sci., 17, 3707–3720, doi:10.5194/hess-17-3707-2013, 2013

Pan, S., Tian, H., Dangal, S. R., Yang, Q., Yang, J., Lu, C., Tao, B., Ren, W., and Ouyang, Z.: Responses of global terrestrial evapotranspiration to climate change and increasing atmospheric CO2 in the 21st century, Earths Future, 3, 15–35, https://doi.org/10.1002/2014EF000263, 2015.

605   Pan, S., Pan, N., Tian, H., Friedlingstein, P., Sitch, S., Shi, H., Arora, V. K., Haverd, V., Jain, A. K., Kato, E., Lienert, S., Lombardozzi, D., Nabel, J. E. M. S., Ottlé, C., Poulter, B., Zaehle, S., and Running, S. W.: Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote sensing, machine learning and land surface modeling, Hydrol. Earth Syst. Sci., 24, 1485–1509, https://doi.org/10.5194/hess-24-1485-2020, 2020.

Pareek, A., Dhankher, O. P., Foyer, C. H.: Mitigating the impact of climate change on plant productivity and ecosystem
610   sustainability, J. Exp. Bot., 71, 451–456, https://doi.org/10.1093/jxb/erz518, 2020.

Penman, H. L.: Natural evaporation from open water, bares soil, and grass, Proc. R. Soc. Lond., Ser. A, 193, 120–146, https://doi.org/10.1098/rspa.1948.0037, 1948.

Priestley, C. H., and Taylor, R. J.: On the assessment of surface heat flux and evaporation using large-scale parameters, Mon. Weather Rev., 100, 81–92. https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

615   Qualls, R. J., and Crago, R. D.: Graphical interpretation of wet surface evaporation equations, Water Resour. Res., 56, e2019WR026766. https://doi.org/10.1029/2019WR026766, 2020.

Ramírez, J. A., Hobbins, M. T., and Brown, T. C.: Observational evidence of the complementary relationship in regional evaporation lends strong support for Bouchet's hypothesis. Geophysical Research Letters, 32, L15401. https://doi.org/10.1029/2005GL023549, 2005.

620   Roderick, M. L., Greve, P. and Farquhar, G. D: On the assessment of aridity with changes in atmospheric CO2, Water Resour. Res., 51, 5450–5463, https://doi.org/10.1002/2015WR017031, 2015.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Ziese, M., and Rudolf, B.: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle, Theor. Appl. Climatol., 115, 15–40, https://doi.org/10.1007/s00704-013-0860-x, 2014.

625   Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture–climate interactions in a changing climate: A review, Earth Sci. Rev., 99, 125–161, 2010.

Shuttleworth, W. J., Serrat-Capdevila, A., Roderick, M. L., and Scott, R.L.: On the theory relating changes in area-average and pan evaporation, Quart. J. Royal Meteor. Soc. 135, 1230–1247, https://doi.org/10.1002/qj.434, 2019.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., and Hsu, K.-L.: A review of global precipitation data sets: Data sources, estimation, and intercomparisons. Rev. Geophys, 56, 79–107. https://doi.org/10.1002/2017RG000574, 2018.

Swann, A. L., Hoffman, F. M., Koven, C. D., and Randerson, J. T.: Plant responses to increasing CO2 reduce estimates of climate impacts on drought severity, P. Natl. Acad. Sci. USA, 113, 10019–10024, https://doi.org/10.1073/pnas.1604581113, 2016.

Szilagyi, J.: 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, 2014.

Szilagyi, J., and Jozsa, J.: 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, 2018.

Szilagyi, J., Crago, R., and Qualls, R: 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, 2017.

Szilagyi, J., Crago, R., and Ma, N.: Dynamic scaling of the generalized complementary relationship (GCR) improves long-term tendency estimates in land evaporation. Adv. Atmos. Sci, 37, 1–12. https://doi.org/10.1007/s00376-020-0079-6, 2020.

Szilagyi, J.: On the thermodynamics foundations of the complementary relationship of evaporation, J. Hydrol., 593, 125916, http://doi.org/10.1016/j.jhydrol.2020.125916, 2021.

Talsma, C. J., Good, S. P., Jimenez, C., Martens, B., Fisher, J. B., Miralles, D. G., McCabe, M. F., and Purdy, A. J.: Partitioning of evapotranspiration in remote sensing-based models, Agr. Forest Meteorol., 131–143, 2018b.

Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106, 7183–7192, https://doi.org/10.1029/2000JD900719, 2001.

Trenberth, K. E., Fasullo, J.. T., Kiehl, J.: Earth's global energy budget, Bull. Am. Meteorol. Soc., 311–324, https://doi.org/10.1175/2008BAMS2634.1, 2009.

Trenberth, K. E., Smith, L., Qian, T., Dai, A., and Fasullo, J.: Estimates of the Global Water Budget and Its Annual Cycle Using Observational and Model Data, J. Hydrometeorol., 8, 758–769, https://doi.org/10.1175/JHM600.1, 2007.

Viterbo, P., and Beljaars, A. C. M.: An improved land surface parameterization scheme in the ECMWF model and its validation, J. Clim., 8, 2716–2748, https://doi.org/10.1175/1520-0442(1995)008<2716:AILSPS>2.0.CO;2, 1995.

Yang, Y., Roderick, M. L., Zhang, S., McVicar, T. R., Donohue, R. J.: Hydrologic implications of vegetation response to elevated CO2 in climate projections. Nature Clim. Change 9, 44–48, https://doi.org/10.1038/s41558-018-0361-0, 2019.

Yao, Y., Liang, S., Li, X., Chen, J., Liu, S., Jia, K., Zhang, X., Xiao, Z., Fisher, J. B., and Mu, Q.: Improving global terrestrial evapotranspiration estimation using support vector machine by integrating three process-based algorithms, Agr. Forest Meteorol., 242, 55–74, https://doi.org/10.1016/j.agrformet.2017.04.011, 2017.

660   Zhang, Y., Leuning, R., Hutley, L. B., Beringer, J., McHugh, I., and Walker, J. P.: Using long-term water balances to parameterize surface conductances and calculate evaporation at 0.05° spatial resolution, Water Resour. Res., 46, W05512, https://doi.org/10.1029/2009WR008716, 2010.

Zhang, Y., Peña-Arancibia, J. L., McVicar, T. R., Chiew, F. H., Vaze, J., Liu, C., Lu, X., Zheng, H., Wang, Y., and Liu, Y. Y.: Multi-decadal trends in global terrestrial evapotranspiration and its components, Sci. Rep., 6, 19124,
665   https://doi.org/10.1038/srep19124, 2016.

Zhou, S., Williams, A. P., Berg, A. M., Cook, B. I., Zhang, Y., Hagemann, S., Lorenz, R., Seneviratne, S. I., and Gentine, P.: Land–atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity, P. Natl. Acad. Sci. USA, 116, 18848–18853, https://doi.org/10.1073/pnas.1904955116, 2019.

670 **Appendix: Calibration-free determination of α_e for the Australian continent.**

In Szilagyi et al. (2017) and Ma et al. (2019), α_e values for ET_w were determined by inserting the Priestley-Taylor equation into the Bowen ratio for a wet environment as:

$$\frac{R_n - ET_w}{ET_w} = \frac{1 - \alpha_e \frac{\Delta_w}{\Delta_w + \gamma}}{\alpha_e \frac{\Delta_w}{\Delta_w + \gamma}} = \gamma \frac{T_{ws} - T_w}{e_s(T_{ws}) - e_a}, \tag{A1}$$

where, Δ_w is the slope of the saturation vapor pressure curve at the wet-environment air temperature (T_w) and the other variables

675 have the same definitions in section 2.1. By rearranging Eq. (A1), α_e could be analytically obtained:

$$\alpha_e = \frac{[\Delta_w + \gamma][e_s(T_{ws}) - e_a]}{\Delta_w \{\gamma [T_{ws} - T_w] + [e_s(T_{ws}) - e_a]\}}. \tag{A2}$$

Szilagyi et al. (2017) identified local wet cells within a large region using sufficiently fine relative humidity (RH) and T_{ws} data from Eq. (4a). The α_e values for wet cells are calculated with Eq. (A2) by inserting the measured air temperature into T_w, and are expected to fall within the theoretical limits of $[1, (\Delta_w + \gamma)/\Delta_w]$ (Priestley and Taylor, 1972).

680 In this work, the wet cells within the Australian continent were identified as locations satisfying the two conditions of $T_{ws} > T_{avg} + 3°C$ and $RH > 90\%$. While Szilagyi et al. (2017) used $T_{ws} > T_{avg} + 2°C$, we considered the fact that T_{ws} estimates from Monteith (1981) is approximately 1°C higher than those estimated by the implicit Bowen ratio (Szilagyi, 2014). The results showed that very few cells (less than 1% of the Australian continent) satisfied the given criteria and their α_e values from Eq. (A2) were within a very narrow range of 1.09 ± 0.01 (mean $\pm$ standard deviation). This calibration-free approach assumes that

685 the mean of the α_e values is applicable for the entire region, thus assuming that a suitable α_e is spatially and temporally constant. More details are found in the Appendix B in Ma and Szilagyi (2019).