Global assessment of how averaging over spatial heterogeneity in precipitation and potential evapotranspiration affects modeled evapotranspiration rates

Elham Rouholahnejad Freund[1,2], Ying Fan[3], James W. Kirchner[2,4,5]

[1]Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium
[2]Department of Environmental Systems Science, ETH Zurich, 8092, Zurich, Switzerland
[3]Department of Earth and Planetary Sciences, Rutgers University, New Brunswick, NJ, United States
[4]Swiss Federal Research Institute WSL, Birmensdorf, 8903, Switzerland
[5]Dept. of Earth and Planetary Science, University of California, Berkeley, CA 94720, United States

Correspondence to: Elham Rouholahnejad Freund, elham.rouholahnejad@gmail.com

Short summary

Evapotranspiration (ET) rates and the properties that regulate them are spatially heterogeneous. Averaging over spatial heterogeneity in precipitation and potential evapotranspiration as main drivers of ET may lead to biased estimates of energy and water fluxes from the land surface to the atmosphere. Here we show that this bias will be largest in mountainous terrain, in regions with temperate climates and dry summers, and in landscapes where spatial variations in precipitation and potential evapotranspiration are inversely correlated.

Abstract

The major goal of large-scale Earth System Models (ESMs) is to understand and predict global change. However, computational constraints require ESMs to operate on relatively large spatial grids (typically ~1 degree or ~100 km in size), with the result that the heterogeneity in land surface properties and processes at smaller spatial scales cannot be explicitly represented. Averaging over this spatial heterogeneity may lead to biased estimates of energy and water fluxes. Here we estimate how averaging over spatial heterogeneity in precipitation (P) and potential evapotranspiration (PET) may affect grid-cell-averaged evapotranspiration (ET) rates, as seen from the atmosphere over heterogeneous landscapes across the globe. Our goal is to identify where, under what conditions, and at what scales this heterogeneity bias could be most important, but not to quantify its absolute magnitude. We use Budyko curves as simple functions that relate ET to precipitation (P) and potential evapotranspiration (PET). Because the relationships driving ET are nonlinear, averaging over sub-grid heterogeneity in P and PET will lead to biased estimates of average ET. We examine the global distribution of this bias, its scale dependence, and its sensitivity to variations in P versus PET. Our analysis shows that this "heterogeneity bias" is more pronounced in mountainous terrain, in landscapes where spatial variations in P and PET are inversely correlated, and in regions with temperate climates and dry summers. We also show that this heterogeneity bias increases on average, and expands over larger areas, as the grid cell size increases.

1. Introduction

Earth System Models (ESMs) are designed to understand interactions between the land surface, atmosphere, and oceans and to predict global environmental changes. However, the Earth system and its underlying physical processes are highly heterogeneous across orders of magnitude in scale below the scale of typical ESM grids (e.g., 1° by 1°). Despite increasing recognition of the need to mechanistically represent physical processes in ESMs, currently even the most disaggregated large-scale ESMs cannot explicitly represent the spatial heterogeneity of land surface hydrological properties at scales that are important to atmospheric fluxes. Averaging over land surface properties at the scale of ESM model grid cells may have important implications for water and energy flux estimates (Avissar and Pielke, 1989; Giorgi and Avissar, 1997; Ershadi et al., 2013; Lu et al., 2014).

Estimates of evapotranspiration (ET) fluxes have significant implications for future temperature predictions. Smaller ET fluxes imply greater sensible heat fluxes and, therefore, drier and warmer conditions in the context of climate change (Seneviratne et al., 2010). Surface evaporative fluxes (and thus energy partitioning over land surfaces) are nonlinear functions of available water and energy, and thus are coupled to spatially heterogeneous surface characteristics (e.g., soil type, vegetation, topography) and meteorological inputs (e.g., radiative flux, wind, and precipitation; Kalma et al., 2008; Shahraeeni and Or, 2010; Holland et al., 2013). These characteristics are spatially variable on length scales of <1 m to many kilometers, well below typical ESM grid scales of ~100 km. ESMs calculate grid-averaged surface and atmospheric fluxes using parameterizations that correspond to grid-averaged properties of the land surface (Sato et al., 1989; Koster et al., 2006; Santanello and Peters-Lidard, 2011). Thus ET estimates that are derived from spatially-averaged land surface properties do not capture ET variations driven by the underlying surface heterogeneity (McCabe and Wood, 2006). Because the relationships driving ET are nonlinear, the average ET flux from a heterogeneous landscape may be different from an ET estimate calculated from spatially averaged inputs (Rouholahnejad Freund and Kirchner, 2017).

Several studies have quantified the effects of land surface heterogeneity on potential evapotranspiration (PET) and latent heat (LH) fluxes, and have found that averaging over land surface heterogeneity can potentially bias ET estimates either positively or negatively. For example, Boone and Wetzel (1998) studied the effects of soil texture variability within each pixel in the Land-Atmosphere-Cloud Exchange (PLACE) model, which has a spatial resolution of approximately 100 by 100 km. They reported that accounting for sub-grid variability in soil texture reduced global ET by 17%, increased total runoff by 48%, and increased soil wetness by 19%, compared to using a homogenous soil texture to describe the entire grid cell. Kollet (2009) found that heterogeneity in soil hydraulic conductivity had a strong influence on evapotranspiration during the dry months of the year, but not during months with sufficient moisture availability. Hong et al. (2009) reported that aggregating radiance data from 30 m to 60, 120, 250, 500, and 1000 m resolution (input upscaling) and then calculating ET from these aggregated inputs at these grid scales using Surface Energy Balance Algorithm for Land (SEBAL, Bastiaanssen et al., 1998a) yields slightly larger ET estimates as compared to ET calculated with finer resolution inputs and then aggregated at the desired grid scales (output upscaling). The discrepancy between ET estimated with the output upscaling method and the input upscaling method grows as the size of the grid cell increases (the difference between ET calculated from the input and output upscaling methods is ~20% more at a grid scale of 1 km by 1 km compared to a grid scale of 120 m by 120 m). Aminzadeh et al. (2017) investigated the effects of averaging surface heterogeneity and soil moisture availability on potential evaporation from a heterogeneous land surface including bare soil and vegetation patches. They found that if the heterogeneity length scale is smaller than the convective atmospheric boundary layer (ABL) thickness, averaging over heterogeneous land surfaces has only a small effect on average potential evaporation rates. Averaging over larger-scale heterogeneities, however, led to overestimates of potential evaporation.

Heterogeneity biases have also been identified in ET calculation algorithms that use remote sensing data as inputs. McCabe and Wood (2006) found that remote sensing retrievals of ET are larger than the corresponding in-situ flux estimates and characterized the roles of land surface heterogeneity and remote sensing resolution in the retrieval of evaporative flux. McCabe and Wood (2006) used Landsat (60 m), Advanced Space borne Thermal Emission and Reflection Radiometer (ASTER) (90 m), and MODIS (1020 m) independently to estimate ET over the Walnut Creek watershed in Iowa. They compared these remote sensing estimates to eddy covariance flux measurements and reported that Landsat and ASTER ET estimates had a higher degree of consistency with one another and correlated better to the ground measurements (r=0.87 and r=0.81, respectively) than MODIS- based ET estimates did. All three remote sensing products overestimated ET as compared to ground measurements (at 12 out of 14 tower sites). Upon aggregation of Landsat and ASTER retrievals to MODIS scale (1 km), the correlation with the ground measurements decreased to r=0.75 and r=0.63 for Landsat and ASTER, respectively.

Contrary to overestimation bias, many remotely sensed ET estimates that include parameters related to aerodynamic resistance are significantly affected by heterogeneity, and underestimate ET as the scale increases (Ershadi et al., 2013). Because aerodynamic resistance is significantly affected by land surface properties (e.g., vegetation height, roughness length, and displacement height), decreases in aerodynamic resistance at coarser resolutions could lead to smaller estimates of evapotranspiration. Ershadi et al. (2013) showed that input aggregation from 120m to 960 m in Surface Energy Balance System (SEBS, Su, 2002) leads to up to 15 % underestimation of ET at the larger grid resolution in a study area in the south-east of Australia.

Rouholahnejad Freund and Kirchner (2017) quantified the impact of sub-grid heterogeneity on grid-average ET using a simple Budyko curve (Turc, 1954; Mezentsev, 1955) in which long-term average ET is a non-linear function of long-term averages of precipitation (P) and potential evaporation (PET). They showed mathematically that averaging over spatially heterogeneous P and PET results in overestimation of ET within the Budyko framework (Fig. 1). Their analysis implies that large-scale ESMs that overlook land surface heterogeneity will also yield biased evapotranspiration estimates due to the inherent nonlinearity in ET processes. They did not, however, determine where around the globe, and under what conditions, this heterogeneity bias is likely to be most important.

The recognition that spatial averaging can potentially lead to biased flux estimates has prompted methods for representing sub-grid-scale heterogeneities and processes within large scale land surface models and ESMs. Accounting for land surface heterogeneity in large-scale ESMs is not merely constrained by limitations in both computational power (Baker et al. 2017) and the availability of high-resolution forcing data, but also by the fact that the atmospheric and land surface components of some ESMs operate at different resolutions. There have been several attempts to integrate sub-grid heterogeneity in ESMs while keeping the computational costs affordable. In “mosaic” approaches, the model is run separately for each surface type in a grid cell, and then the surface-specific fluxes are area-weighted to calculate the grid-cell average fluxes (e.g., Avissar and Pielke, 1989; Koster and Suarez, 1992). The “effective parameter” approach (e.g., Wood and Mason, 1991; Mahrt et al., 1992), by contrast, seeks to estimate effective parameter values at the grid cell scale that subsume the effects of sub-grid heterogeneity. Estimating these effective parameters can be challenging because the relevant land-surface processes typically depend nonlinearly on multiple interacting parameters, and land-surface signals at different scales are propagated and diffused differently in the atmosphere. Alternatively, the "correction factor" approach (e.g., Maayar and Chen, 2006) uses sub-grid information on spatially heterogeneous land-surface processes and properties to estimate multiplicative correction factors for fluxes that are originally calculated from spatially averaged inputs at the grid-cell scale. All three approaches try to reduce the heterogeneous problem to a homogeneous one that has equivalent effects on the atmosphere at the grid-cell scale.

There is a growing need to understand how sub-grid heterogeneity (and the atmosphere’s integration of it) affect grid-scale water and energy fluxes, and to develop effective methods to incorporate these effects in ESMs (Clark et al., 2015, Fan et al., 2019). In a previous study, we proposed a general framework for quantifying systematic biases in ET estimates due to averaging over heterogeneities (Rouholahnejad Freund and Kirchner, 2017). We used the Budyko framework as a simple estimator of ET, and demonstrated theoretically how averaging over heterogeneous precipitation and potential evapotranspiration can lead to systematic overestimation of long-term average ET fluxes from heterogeneous landscapes. In the present study, we apply this analysis across the globe and highlight the locations where the heterogeneity bias is largest. Our hypotheses, derived from the Budyko framework as summarized in Eq. (4) below, are that (1) strongly heterogeneous landscapes, such as mountainous terrain, will exhibit greater heterogeneity bias, (2) this bias will be larger in climates where P and PET are inversely correlated in space, and (3) heterogeneity bias will decrease as the spatial scales of averaging decrease.

2. Effects of sub-grid heterogeneity on ET estimates in the Budyko framework

Budyko (1974) showed that long-term annual average evapotranspiration is a function of both the supply of water (precipitation, P) and the evaporative demand (potential evapotranspiration, PET) under steady-state conditions and in catchments with negligible changes in storage (Eq. 1; Turc, 1954; Mezentsev, 1955):

$$ET = f(P, PET) = \frac{P}{\left(\left(\frac{P}{PET}\right)^n + 1\right)^{1/n}}. \tag{1}$$

where ET is actual evapotranspiration, P is precipitation, PET is potential evaporation, and n (dimensionless) is a catchment-specific parameter that modifies the partitioning of P between ET and discharge.

Evapotranspiration rates are inherently bounded by energy and water limits. Under arid conditions ET is limited by the available supply of water (the water limit line in Fig. 1b), while under humid conditions ET is limited by atmospheric demand (PET) and converges toward PET (the energy limit line in Fig. 1b). Budyko showed that over a long period and under steady-state conditions, hydrological systems function close to their energy or water limits. These intrinsic water and energy constraints make the Budyko curve downward-curving.

In a heterogeneous landscape, like the simple example of two model columns in Fig. 1a, P and PET vary spatially. The two columns with heterogeneous P and PET are represented by the two solid black circles on the Budyko curve in Fig. 1b. In this hypothetical two-column example, the true average of ET values calculated from individual heterogeneous inputs (the solid black circles) lies below the curve (the grey circle, labeled “true average”). However, if we aggregate the two columns and consider the system as one column with average properties, the function of average inputs (averaged P and PET over the two columns) lies on the Budyko curve (the open circle) which is larger than the true average of the two columns. In short, in any downward curving function, the function of the average inputs (the open circle) will always be larger than the average of the individual function values (the true average; grey circle). The difference between the two can be termed the "heterogeneity bias".

In a previous study (Rouholahnejad Freund and Kirchner, 2017) we showed that when nonlinear underlying relationships are used to predict average behaviour from averaged properties, the magnitude of the resulting heterogeneity bias can be estimated from the degree of the curvature in the underlying function and the range spanned by the individual data being averaged. Here we summarize theses findings as building blocks of the current study. The second-order, second-moment Taylor expansion of the ET function f(P,PET) (Eq. 1) around its mean directly yields:

$$\bar{f}(P, PET) = \overline{ET} \approx f(\bar{P}, \overline{PET}) + \frac{1}{2}\frac{\partial^2 f}{\partial P^2}\, var(P) + \frac{1}{2}\frac{\partial^2 f}{\partial PET^2}\, var(PET) + \frac{\partial^2 f}{\partial P\, \partial PET} cov(P, PET) \quad , \tag{2}$$

where $\bar{f}(P, PET)$ is the true average of the spatially heterogeneous ET function, $f(\bar{P}, \overline{PET})$ is the ET function evaluated at its average inputs $\bar{P}$ and $\overline{PET}$, and the derivatives are calculated at $\bar{P}$ and $\overline{PET}$. Evaluating the derivatives using Eq. (1) and reshuffling the terms, Rouholahnejad Freund and Kirchner (2017) obtained the following expression for the heterogeneity bias, the difference between the average ET, $\bar{f}(P, PET)$, and the ET function evaluated at the mean of its inputs, $f(\bar{P}, \overline{PET})$:

$$f(\bar{P}, \overline{PET}) - \bar{f}(P, PET) \approx (n+1)\frac{\bar{P}^{n+1}\overline{PET}^{n+1}}{(\bar{P}^n + \overline{PET}^n)^{2+1/n}}\left[\frac{1}{2}\frac{var(P)}{\bar{P}^2} + \frac{1}{2}\frac{var(PET)}{\overline{PET}^2} - \frac{cov(P, PET)}{\bar{P}\,\overline{PET}}\right]. \qquad (3)$$

To more clearly show the effects of variations in P and PET, Eq. (3) can be reformulated as follows:

$$f(\bar{P}, \overline{PET}) - \bar{f}(P, PET) \approx$$
$$(n+1)\frac{\bar{P}^{n+1}\overline{PET}^{n+1}}{(\bar{P}^n + \overline{PET}^n)^{2+1/n}}\left[\frac{1}{2}\left(\frac{SD(P)}{\bar{P}}\right)^2 + \frac{1}{2}\left(\frac{SD(PET)}{\overline{PET}}\right)^2 - r_{P,PET}\left(\frac{SD(P)}{\bar{P}}\right)\left(\frac{SD(PET)}{\overline{PET}}\right)\right] \quad . \qquad (4)$$

Equation (4) shows that the heterogeneity bias depends on only four quantities: the fractional variation (i.e., the coefficient of variation) in precipitation $\left(\frac{SD(P)}{\bar{P}}\right)$ and in potential ET $\left(\frac{SD(PET)}{\overline{PET}}\right)$, the correlation between precipitation and potential ET ($r_{P,PET}$), and the function $(n+1)\frac{\bar{P}^{n+1}\overline{PET}^{n+1}}{(\bar{P}^n+\overline{PET}^n)^{2+1/n}}$, which quantifies the curvature in the ET function in Budyko space. As shown by Fig. 1b and Eq. (2), the discrepancy between average of the ET function and the ET function of the average inputs (the heterogeneity bias) is proportional to both the degree of nonlinearity in the function, as defined by its second derivatives, and the variability of P and PET. Equation (4) allows one to estimate how much the curvature of the ET function and the fractional variability (standard deviation divided by mean) of P and PET will affect estimates of ET. However, to the best of our knowledge, the consequences of these nonlinearities for global evaporative flux estimates have not previously been quantified.

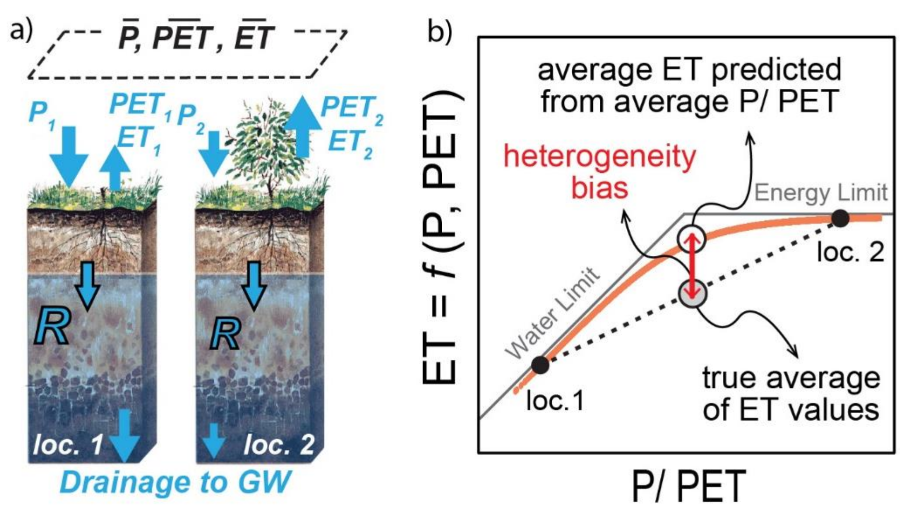

Figure 1. Heterogeneity bias in a hypothetical two-column model in the Budyko framework. The true average ET of the columns (gray circle) lies below the curve and is less than the average ET estimated from the average P/PET of the two columns (open circle). The heterogeneity bias depends on the curvature of the function and the spread of its inputs. Both panels are adapted from Rouholahnejad Freund and Kirchner (2017).

3. Effects of sub-grid heterogeneity on ET estimates at 1° by 1° grid scale across the globe

Across a landscape of similar size to a typical ESM grid cell (1° by 1°), soil moisture, atmospheric demand (PET) and precipitation (P) will vary with topographic position; hillslopes will typically be drier, and riparian regions will be wetter. To map the spatial pattern in the heterogeneity bias that results from averaging over this land surface heterogeneity, we applied the approach outlined in section 2 to the global land surface area at 1° by 1° grid scale. Within each 1° by 1° grid cell, we used 30 arc-second values of P (WorldClim; Hijmans et al., 2005) and PET (WorldClim; Hijmans et al., 2005) to examine the variations in small-scale climatic drivers of ET. Because 30 arc-seconds is nearly 1 km, hereafter we refer to the 30 arc-second data as 1km values for simplicity. The spatial distribution of long-term annual averages (1960-1990) of P and PET values at 1 km resolution, along with 1km values of the aridity index (AI=P/PET), are shown in Fig 2a-c. ET values calculated from these 1km P and PET values using Eq. (1) are then averaged at 1° by 1° scale ("true average", Fig. 2e). We also averaged the 1km values of P and PET within each grid cell and then modeled ET using the Budyko curve (Eq. 1) applied to these averaged input values. The difference between these two ET estimates is the heterogeneity bias.

We also calculated the heterogeneity bias using Eq. (4), which describes how the nonlinearity in the governing equation and the heterogeneity in P and PET jointly contribute to the heterogeneity bias. The heterogeneity bias estimates obtained by Eq. (4) were functionally equivalent (R^2=0.97, root mean square error of 0.17%) to those obtained by direct calculation using Eq. (1) as described above.

Fig. 3a-d illustrates the variability (quantified by standard deviation) of 1km values of P, PET, aridity index, and altitude at the 1° by 1° grid scale. The heterogeneity bias in long-term average ET fluxes at the 1° by 1° grid scale (Fig. 3e) highlights regions around the globe where ET fluxes are likely to be systematically overestimated. The spatial distribution of the heterogeneity bias calculated using Eq. 4 (Fig. 3e) closely coincides with locations where the aridity index is highly variable (Fig. 3c), which is driven in turn by topographic variability (Fig. 3d). Strongly heterogeneous landscapes exhibit significant heterogeneity biases in long-term average ET fluxes. Although the global average heterogeneity bias is small (<1%), physically based ET calculations may exhibit larger heterogeneity biases than the modest values we calculate here, because the Budyko approach already subsumes spatial heterogeneity effects at the catchment scale (and also temporal heterogeneity effects due to its steady-state assumptions). The heterogeneity biases in ET estimates shown in Fig. 3e correspond to long-term average ET estimates. Given the fact that P and PET can vary temporally (i.e., seasonality), the actual bias could be much larger, particularly where P and PET are inversely correlated (see the last term of Eq. 4).

Our results show that the topographic gradient, and hence the variability in the aridity index across a given grid scale, drives consistent, predictable patterns of heterogeneity bias in evapotranspiration estimates at that scale. Equation 4 shows that this bias is equally sensitive to fractional variability in P and PET (standard deviation divided by mean). However, because P is typically more variable (in percentage terms) than PET across landscapes, the variability in P will usually make a larger contribution to the heterogeneity bias.

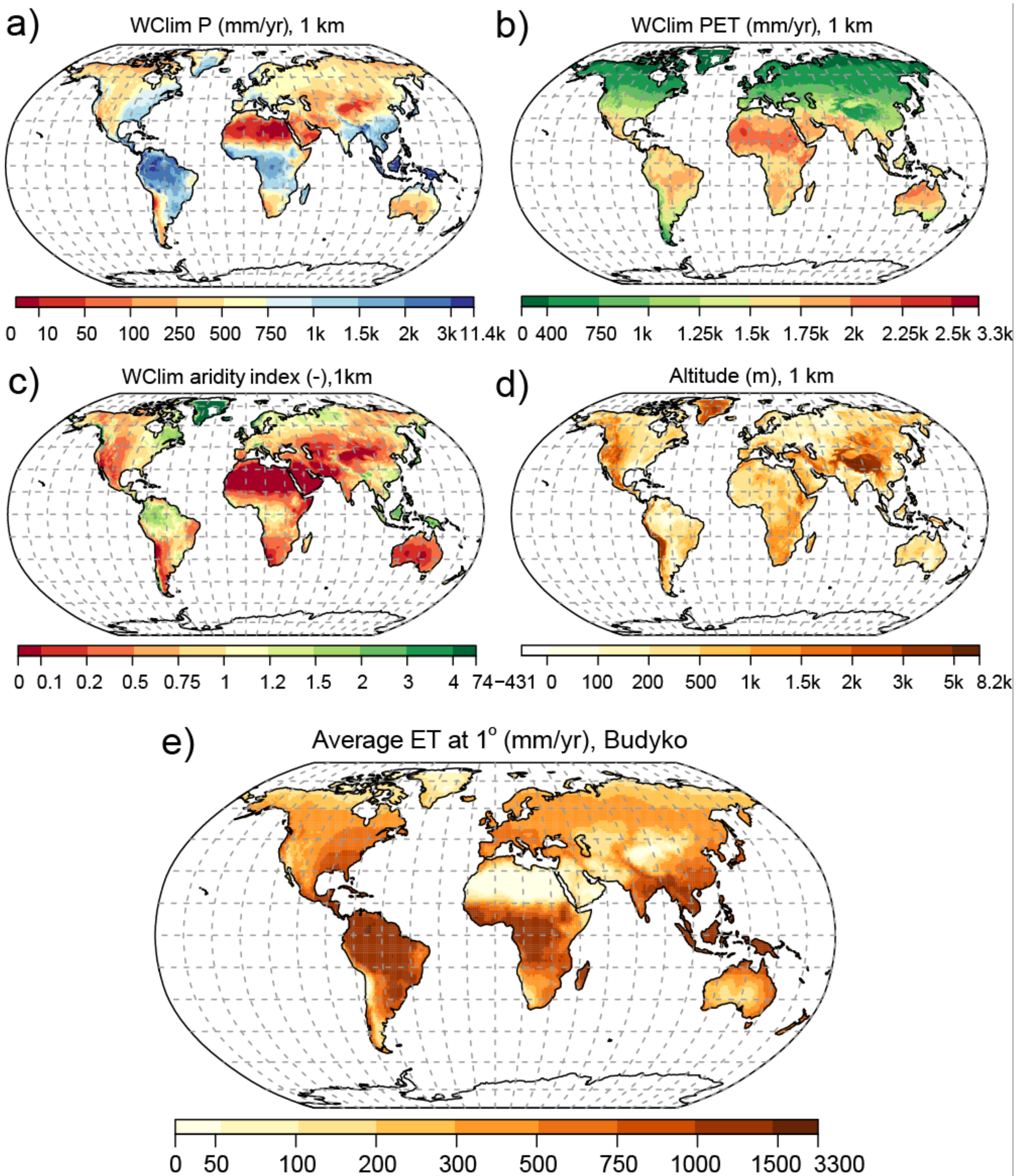

Figure 2. Global distribution of one-kilometer resolution annual mean precipitation (a: P; WorldClim; Hijmans et al., 2005), potential evapotranspiration (b: PET; WorldClim; Hijmans et al., 2005), aridity index (c: AI=P/PET; WorldClim; Hijmans et al., 2005), and topography (d: SRTM; Jarvis et al., 2008), along with (e) evapotranspiration (ET) at 1° by 1° scale by averaging 1km values of ET calculated using the Budyko function (Eq. 1).

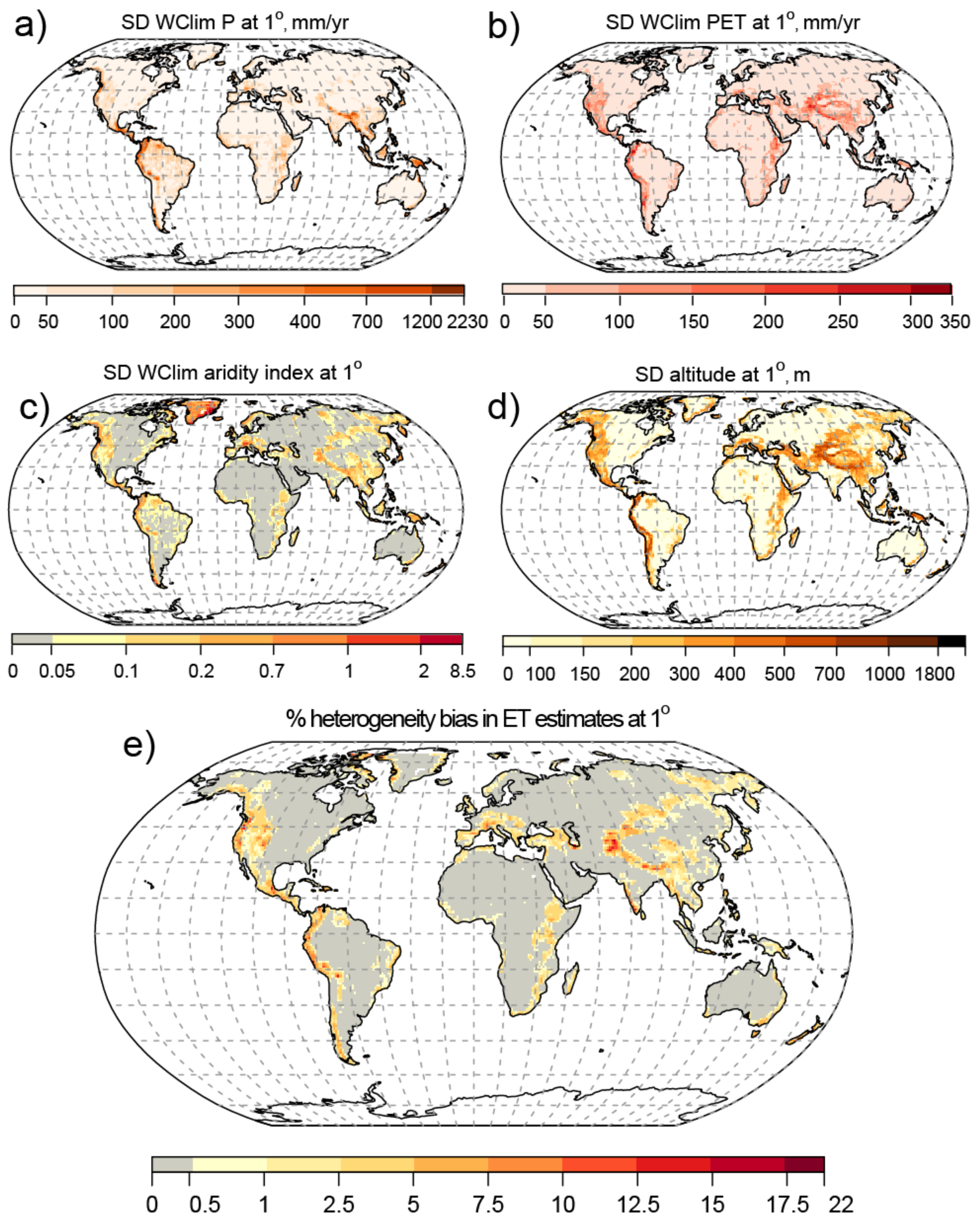

Figure 3. Global spatial distribution of variability (standard deviation) of one-kilometer values of a) precipitation (P), b) potential evapotranspiration (PET), c) aridity index (AI=P/PET), and d) altitude at 1° by 1° grid cell. The

heterogeneity bias in ET estimates (e) is calculated using Eq. (4). Grid cells with larger standard deviation in altitude and aridity index have larger heterogeneity bias.

4. Variation in heterogeneity bias across climate zones, data sources, and grid scales

With increased availability of spatial data, it is becoming standard practice to assess input data uncertainties and their propagated impacts on water and energy flux estimates in land surface models. To quantify how choices among alternative input data products could affect the heterogeneity bias in ET estimates, we calculated the heterogeneity bias at 1 ° by 1° grid cell resolution across the contiguous US using four different pairs of P and PET data products. Two precipitation data sets, Prism (http://prism.oregonstate.edu) and WorldClim (Hijmans et al., 2005), along with two PET data sets, MODIS (Mu et al., 2007) and WorldClim (Hijmans et al., 2005). As Prism precipitation data is available at 4 km resolution, all other data sets were aggregated to 4 km. Two P products and two PET products were combined in all possible pairs. The WorldClim PET dataset (Hijmans et al., 2005) is based on the Hargreaves method (Hargreaves and Samani 1985) while the MODIS PET product (Mu et al, 2007) is based on the Penman–Monteith equation (Monteith, 1965). The heterogeneity bias in ET estimates (Eq. 4), as outlined in Sect. 2, was evaluated from 4km values of P, PET, and the estimated average ET using the Budyko relationship (Eq. 1) for each of the four input data pairs. Figure 4a-e compares the spatial distributions of heterogeneity bias across the contiguous US for the four pairs of P and PET data products. The heterogeneity bias in ET estimates reached as high as 36 % in the western US using Prism P and WorldClim PET as input to the ET model (Fig. 4b). A visual comparison of Figs. 4b and Fig. 4d shows that the choice of P data source (Prism vs. WorldClim) had a bigger effect on the heterogeneity bias than the choice of PET data source (MODIS vs. WorldClim), meaning that the fractional variability in P is the dominant variable. In all cases, data sources that were more variable in relation to their means (Prism for P and WorldClim for PET; Fig. 4b) led to larger heterogeneity biases, as expected from Eq. (4). Thus we infer that we would have obtained larger heterogeneity biases if we had conducted our global analysis (Fig. 3) with Prism P and either WorldClim or MODIS PET, but we cannot show that result explicitly at global scale because Prism P is not freely available globally.

If we separate the heterogeneity biases shown in Fig. 4 according to Köppen-Geiger climate zones (Peel et al., 2007; Fig. 5a), we see that they are distinctly higher in particular climate-terrain combinations. Heterogeneity biases are higher in regions with temperate climates and dry summers (climate zone Cs) and in regions with cold, dry summers (climate zone Ds), most likely due to the sharp spatial gradient in their water and energy sources for evapotranspiration (Fig. 5b). These areas typically have high topographic relief, combined with seasonal climate. The heterogeneity effects on ET estimates in these regions are expected to be even larger when a mechanistic model of ET is used. We expect that averaging over temporal variations of drivers of ET, especially in places with strong seasonality, could substantially bias the ET estimates, but this cannot be quantified in the Budyko framework due to its underlying steady-state assumptions. Figure 5b also illustrates the relative magnitudes of the heterogeneity biases obtained with the four pairs of P and PET data sources. The heterogeneity bias is the highest when the Prism P and WorldClim PET datasets are used, followed by the combination of Prism P and MODIS PET, which resulted in the second-highest heterogeneity bias across different climate zones. Wilcoxon signed-rank tests was performed to evaluate the statistical significance of the differences between heterogeneity bias in ET estimates using all pairs of climate zones and data sources that are shown in Fig. 5b (Table S1). These analysis show that while the difference between heterogeneity biases estimated in Cs and Ds climate zones are not statistically significant across all four combinations of datasets, the difference between estimated heterogeneity bias in Cs versus Cf, Ds versus Cf, as well as Cs versus Bs climate zones are significant across all four data combinations (highlighted in Table S1 of the supplementary material).

Equation 4 shows that heterogeneity biases in Budyko estimates of ET are equally sensitive to the same percentage variability in P and PET. Thus the degree of sensitivity, per se, to P and PET variations expressed in percentage terms is the same. Although Figs. 5c and 5d give the visual impression that PET is more variable than P across climate zones and between data sources, Fig. 5e shows that the fractional variability in P is systematically higher than PET, and it also varies more across the climate zones and between the two data sets. Because P is typically more variable than PET (in percentage terms) across landscapes, the variability in P will make a larger contribution to the heterogeneity bias (Fig. 5e) in the Budyko approach. Whether this is true for more physically based ET estimates remains to be seen. Analysis of percent variability of P and PET products shows that percent variabilities of precipitation products are in general larger than PET products and hence contribute more to heterogeneity (Fig 5e). While the percent variabilities of the two PET products are in the same range, the percent variability in Prism precipitation is slightly larger than in WorldClim precipitation, in regions with dry summers (Cs and Ds climate zones in Fig. 5a).

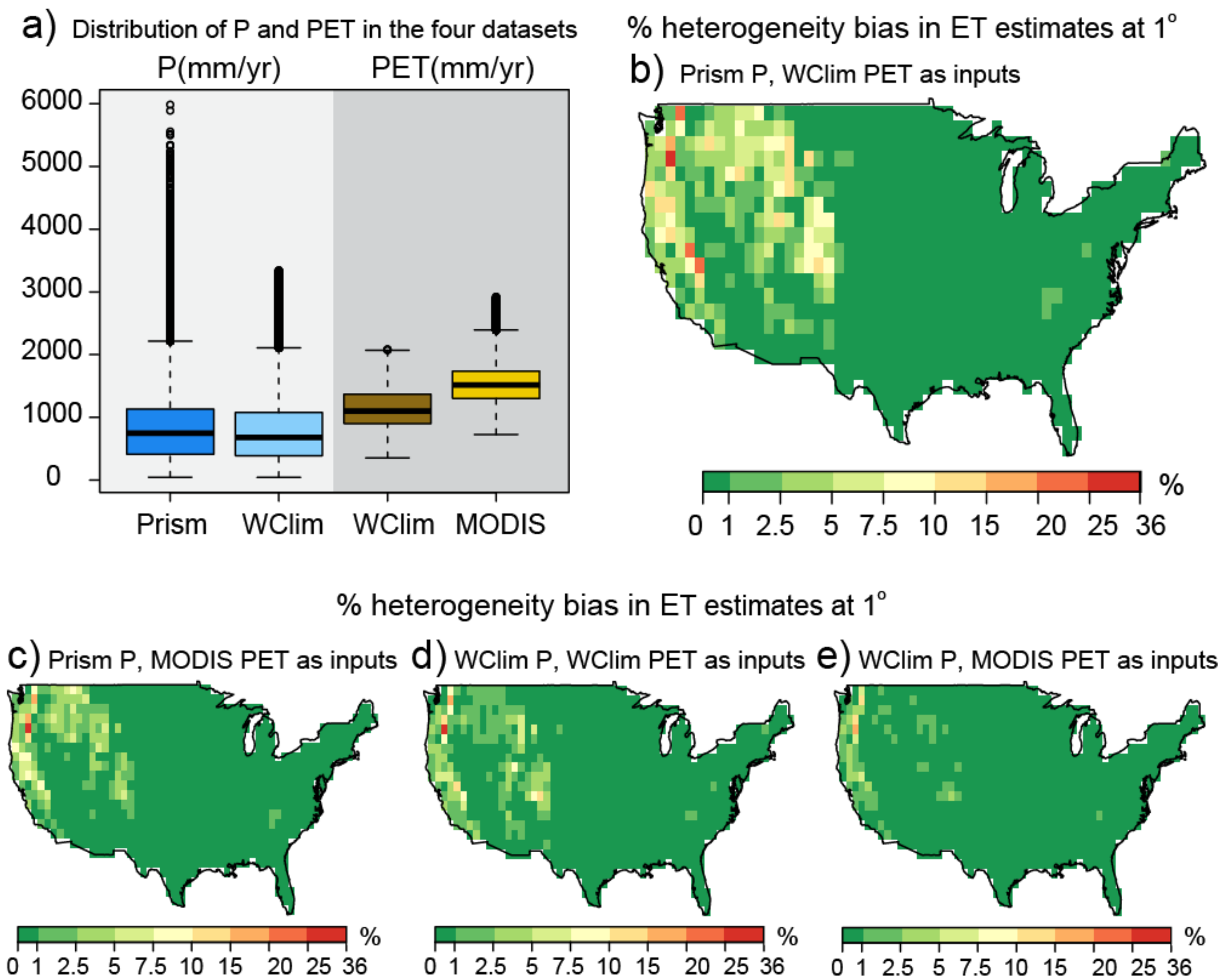

Figure 4. The distribution of P and PET in the four datasets is shown in a). Estimated heterogeneity bias (Eq. 4) across the contiguous US using four-kilometer values of b) Prism P and WorldClim PET c) Prism P and MODIS PET d) WorldClim P and WorldClim PET, and e) WorldClim P and MODIS PET as inputs.

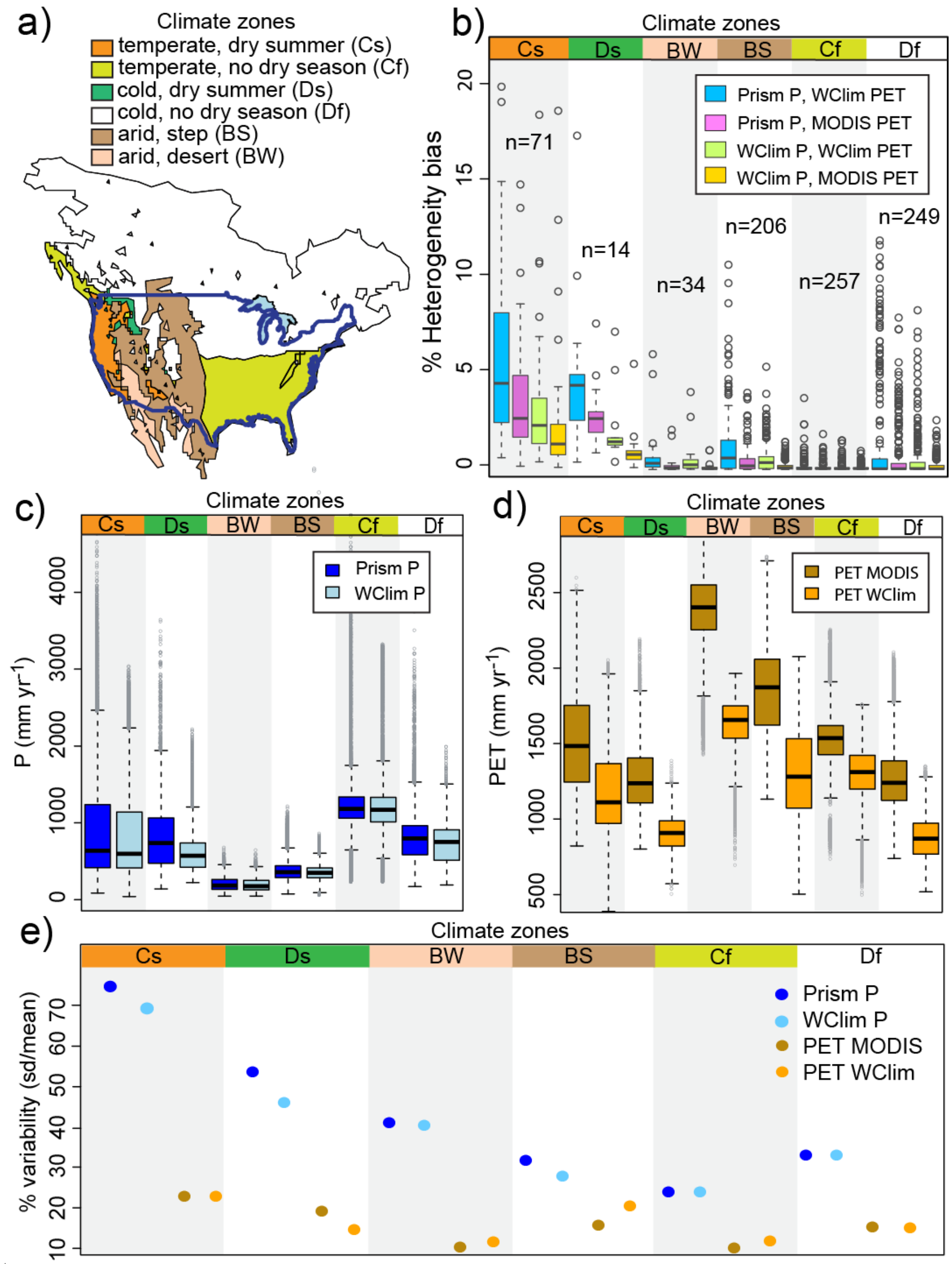

Figure 5. a) Köppen-Geiger climate classification (Peel et al., 2007 in Beck et al. 2013) across the contiguous US, b) the distribution of calculated heterogeneity bias in ET estimates (Eq. 4) at 1° by 1° grid cell in individual climate zones, shown by boxplot (three data points with heterogeneity biases of over 20% are off-scale). The significance of differences between the pairs are presented in Table S1. Panels c and d show the distribution of precipitation products (Prism and WorldClim) and potential evaporation products (MODIS and WorldClim) at individual climate zones, respectively. The color-coded climate zones at the tops of panels b, c, and d correspond to the climate zones mapped in panel a. Panel e compares the percentage variability of the two P and PET data products across climate zones, showing that the percentage variability in P is markedly higher than in PET, and the percentage variability in Prism P is somewhat higher than in WorldClim P, particularly in climate zones with dry summers.

Because future increases in computing power will lead to ESMs with smaller grid cells, it is useful to ask how changes in grid resolution affect the heterogeneity biases that we have estimated in this paper. To quantify the heterogeneity bias in ET estimates as a function of grid scale, we repeated our analysis at various grid resolutions using Switzerland as a test case. We started with high-resolution (500m) maps of long-term average annual precipitation and PET across the Swiss landscape (Fig. 6), and then used Eq. 4 to estimate the heterogeneity bias at grid scales ranging from 1/32° to 2° (~3 km to ~200 km). As Fig. 6 shows, aggregating P and PET over larger scales leads to larger, and more widespread, overestimates in ET. Conversely, at finer grid resolutions, the average heterogeneity bias is smaller, and the locations with large biases are more localized. On average, the heterogeneity bias across Switzerland as a whole grows exponentially as the inputs are averaged over larger grids (as shown in the lower-right panel in Fig. 6).

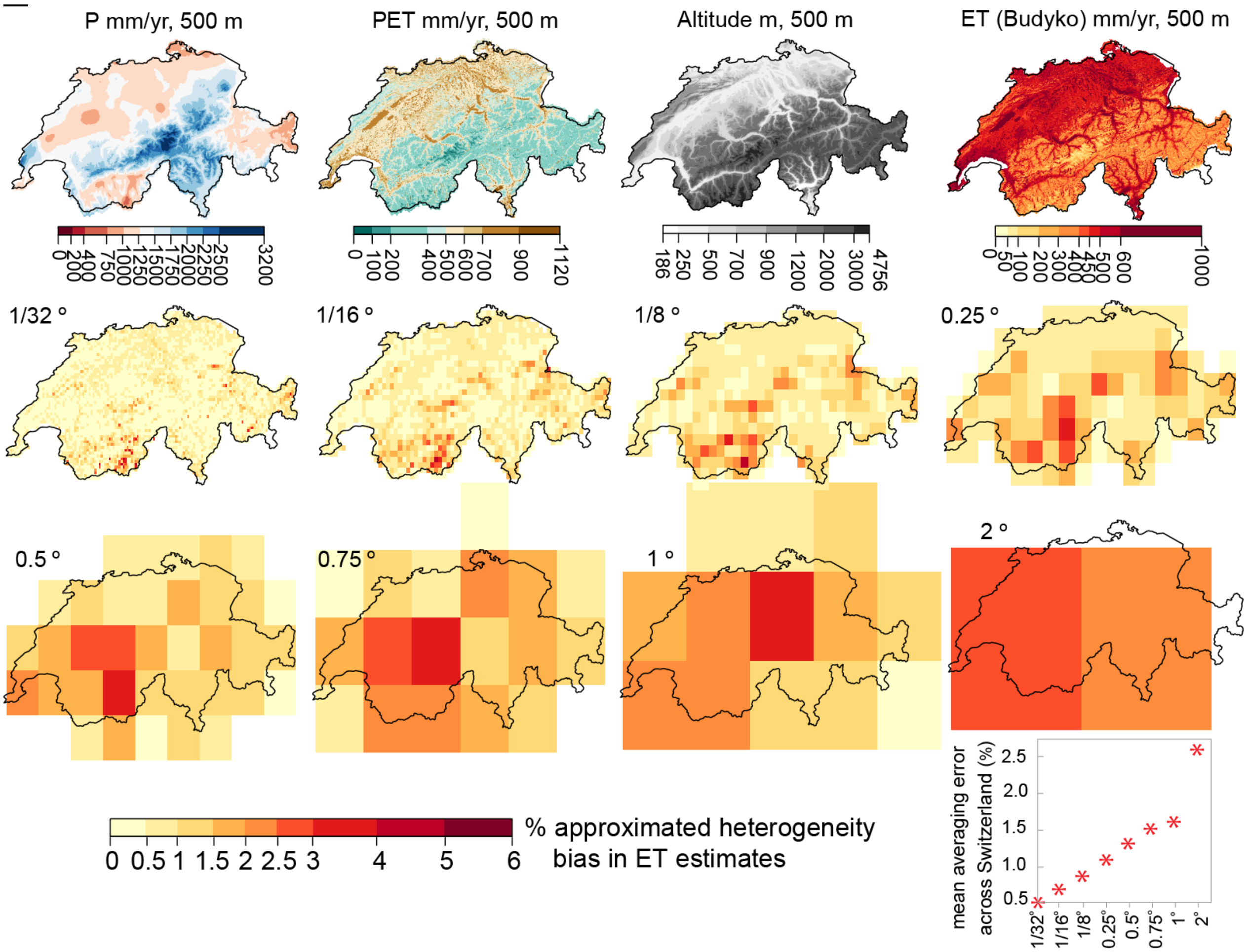

Figure 6. Heterogeneity bias in ET estimates at various scales across Switzerland, estimated from 500m climate data. ET is calculated using the Budyko relationship (Eq. 1). Heterogeneity bias was estimated from 500m precipitation (P) and potential evapotranspiration (PET), and their variances at each grid scale, using Eq. 4. At finer grid resolutions, the heterogeneity bias is more localized, and smaller on average.

5. Summary and discussion

Because evapotranspiration (ET) processes are inherently bounded by water and energy constraints, over the long term, ET is always a nonlinear function of available water and PET, whether this function is expressed as a Budyko curve or another ET model. These nonlinearities imply that spatial heterogeneity will not simply average out in predictions of land surface water and energy fluxes in ESMs. Overlooking sub-grid spatial heterogeneity in large-scale ESMs could lead to biases in estimated water and energy fluxes (e.g., ET rates). Here we have shown that, across several scales, averaging over spatially heterogeneous land surface properties and processes leads to biases in evapotranspiration estimates. We examined the global distribution of this bias, its scale dependence, and its sensitivity to variations in P versus PET, and showed under what conditions, this heterogeneity bias is likely to be most important. Our analysis does not quantify the heterogeneity biases in ESMs, owing to the many differences between these mechanistic models and the simple empirical Budyko curve. But if the heterogeneity biases in ESMs can be quantified, they can be used as correction factors to improve ESM estimates of surface-atmosphere water and energy fluxes across landscapes. Our paper highlights a general methodology that can be used to estimate heterogeneity biases and to map their spatial patterns, but not to calculate their absolute magnitudes because those will change significantly depending on the ET formulation that is used.

In this study, we used Budyko curves as simple models of ET, in which long-term average ET rates are functionally related to long-term averages of P and PET. We used an approach outlined by Rouholahnejad Freund and Kirchner (2017) to estimate the heterogeneity bias in modeled ET at 1-degree grid scale across the globe (Fig. 3), and also at multiple grid scales across Switzerland (Fig. 6), using finer-resolution P and PET values as drivers of ET. We showed how the heterogeneity effects on ET estimates vary with the nonlinearity in the governing equations and with the variability in land surface properties. Our analysis shows that heterogeneity effects on ET fluxes matter the most in areas with sharp gradients in the aridity index, which are in turn controlled by topographic gradients, and not merely in areas that are either arid or humid (e.g., compare Fig. 3e with Fig. 2c).

According to our analysis, regions within the U.S. that have temperate climates and dry summers exhibit greater heterogeneity bias in ET estimates (Fig. 5). We show that the heterogeneity bias in ET estimates at each grid scale depends on the variance in the drivers of ET at that scale (Fig. 4), and on the choice of data sources used to estimate ET. Heterogeneity bias was significantly larger across the contiguous United States when P and PET data sources with larger variances were used (Fig. 4).

We also explored the magnitude and spatial distribution of heterogeneity bias in ET estimates as a function of the scale at which the climatic drivers of ET are averaged. We found that as heterogeneous climatic variables are aggregated to larger scales, the heterogeneity biases in ET estimates become greater on average, and extend over larger areas (Fig. 6). At smaller grid scales, the heterogeneity bias does not completely disappear, but instead becomes more localized around areas with sharp topographic gradients. Finding an effective scale at which one can average over the heterogeneity of land surface properties and processes has been a longstanding problem in Earth science. Our analysis shows that at smaller resolutions the average heterogeneity bias as seen from the atmosphere becomes smaller, but there is no characteristic scale at which it vanishes entirely (Fig. 6). The magnitude and spatial distribution of this bias depend strongly on the scale of the averaging and degree of the nonlinearity in the underlying processes. The heterogeneity bias concept is general and extendable to any convex or concave function (Rouholahnejad Freund and Kirchner 2017), meaning that in any nonlinear process, averaging over spatial and temporal heterogeneity can potentially lead to bias.

In the analysis presented here, we have assumed a value of 2 for the Budyko parameter n, which approximates the variation of ET/PET with respect to P/PET in MODIS and WorldClim data across continental Europe (Mu et al. 2007; Hijmans et al. 2005; Rouholahnejad Freund & Kirchner, 2017). Although there are suggestions in the literature that n can vary with land use and other landscape properties (e.g., Teuling et al., 2019), here we have assumed that n is spatially and temporally constant in order to focus on the effects of heterogeneity in P and PET. In the supplement we present a sensitivity analysis with values of n ranging from 2 to 5 (Fig. S1). That analysis shows that, as expected from Eqs. 3 and 4, higher values of n lead to larger heterogeneity biases, but the spatial pattern shown in Fig. 3e remains largely unchanged. The Taylor approximation in Eqs. 3 and 4 yields realistic estimates of the heterogeneity bias for all values of n that were tested (Fig. S2). Thus while our numerical estimates of heterogeneity bias depend somewhat on the value of n, our conclusions do not.

One should keep in mind that the true mechanistic equations that determine point-scale ET as a function of point-scale water availability and PET (if such data were available) may be much more nonlinear than Budyko's empirical curves, because these curves already average over significant spatial and temporal heterogeneity. Thus, we expect that the real-world effects of sub-grid heterogeneity are probably larger than those we have estimated in Sects. 3 and 4 of this study. In addition, the 1km P and PET values that are used in our global analysis might be still too coarse to represent small-scale heterogeneity that is important to evapotranspiration processes.

Budyko curves are empirical relationships that functionally relate evaporation processes to the supply of water and energy under steady-state conditions in closed catchments with no changes in storage. Our analysis likewise assumes no changes in storage, nor any lateral transfer between the model grid cells, although both lateral transfers and changes in storage may be important, both in the real world and in models. Unlike the Budyko framework, ET fluxes in most ESMs are often physically based (not merely functions of P and PET) and are calculated at much smaller time steps (seconds to minutes). These models often represent more processes that are important to evapotranspiration (such as storage variations) and include their dynamics to the extent that is computationally feasible. Because these relationships may be much more nonlinear than Budyko curves, there may also be significant heterogeneity biases when complex physically based models are used to estimate ET from spatially aggregated data. Therefore, we are now working to quantify heterogeneity bias in ET fluxes using a more mechanistic land surface model.

Acknowledgements

E.R.F. acknowledges support from the Swiss National Science Foundation (SNSF) under Grant No. P2EZP2_162279. The authors thank Massimiliano Zappa of the Swiss Federal Research Institute WSL for providing the 500m resolution data that enabled the analysis shown in Fig. 6.

References

Aminzadeh M., and D. Or: The complementary relationship between actual and potential evaporation for spatially heterogeneous surfaces, Water Resour. Res., 53, 580–601, doi:10.1002/2016WR019759, 2017.

Avissar, R., R. A. Pielke: A Parameterization of Heterogeneous Land Surfaces for Atmospheric Numerical Models and Its Impact on Regional Meteorology, Monthly Weather Review, vol. 117, issue 10, p. 2113, doi:10.1175/1520-0493(1989)117<2113:APOHLS>2.0.CO;2, 1989.

Baker I. T. , P. J. Sellers , A. S. Denning, I. Medina , P. Kraus, K. D. Haynes , and S. C. Biraud: Closing the scale gap between land surface parameterizations and GCMs with a new scheme, SiB3-Bins, Journal of Advances in Modeling Earth Systems, J. Adv. Model. Earth Syst., 9, 691–711, doi:10.1002/2016MS000764, 2017.

Bastiaanssen, W. G. M., M. Menenti, R. A. Feddes, and A. A. M. Holtslag: A remote sensing surface energy balance algorithm for land (SEBAL): 1. Formulation, Journal of Hydrology, 212-213, 198–212, 1998.

Beck H. E., A. I. J. M. van Dijk, D. G. Miralles, R. A. M. de Jeu, L. A. Bruijnzeel, T. R. McVicar, and J. Schellekens: Global patterns in base flow index and recession based on streamflow observations from 3394 catchments, Water Resour. Res., 49, 7843–7863, doi:10.1002/2013WR013918, 2013.

Boone, A., and O. J. Wetzel: A simple scheme for modeling sub-grid soil texture variability for use in an atmospheric climate model. Journal of the Meteorological Society of Japan, 77(1), 317−333, 1998.

Budyko, M. l.: Climate and life, Academic, New York, 1974.

Clark, M. P., Y. Fan, D. M. Lawrence, J. C. Adam, D. Bolster, D. J. Gochis, R. P. Hooper, M. Kumar, L. R. Leung, D. S. Mackay, R. M. Maxwell, C. Shen, S. C. Swenson, and X. Zeng: Improving the representation of hydrologic processes in Earth System Models, Water Resour. Res., 51, 5929–5956, doi:10.1002/2015WR017096, 2015.

Ershadi A., M. F. McCabe, J. P. Evans, J. P. Walker: Effects of spatial aggregation on the multi-scale estimation of evapotranspiration, Remote Sensing of Environment 131, 51–62, http://dx.doi.org/10.1016/j.rse.2012.12.007, 2013.

Fan, Y., M. Clark, D. M. Lawrence, S. Swenson, L. E. Band, S. L. Brantley, P. D. Brooks, W. E. Dietrich, A. Flores, G. Grant, J. W. Kirchner, D. S. Mackay, J. J. McDonnell, P. C. D. Milly, P. L. Sullivan, C. Tague, H. Ajami, N. Chaney, A. Hartmann, P. Hazenberg, J. McNamara, J. Pelletier, J. Perket, E. Rouholahnejad-Freund, T. Wagener, X. Zeng, E. Beighley, J. Buzan, M. Huang, B. Livneh, B. P. Mohanty, B. Nijssen, M. Safeeq, C. Shen, W. van Verseveld, J. Volk, D. Yamazaki: Hillslope hydrology in global change research and Earth system modeling, Water Resources Research, 55, doi:10.1029/2018WR023903, 2019.

Giorgi, F., and R. Avissar: Representation of heterogeneity effects in Earth system modeling: Experience from land surface modeling, Rev. Geophys., 35, 413–437, doi:10.1029/97RG01754, 1997.

Hargreaves, G. H., and Z. A. Samani: Reference crop evaporation from temperature, Appl. Eng. Agric., 1(2), 96-99, 1985.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis: Very high resolution interpolated climate surfaces for global land areas, Int. J. Climatol., 25, 1965–1978, doi:10.1002/joc.1276, 2005.

Holland, S., J. L. Heitman, A. Howard, T. J. Sauer, W. Giese, A. Ben-Gal, N. Agam, D. Kool, and J. Havlin: Micro Bowen ratio system for measuring evapotranspiration in a vineyard interrow, Agric. For. Meteorol., 177, 93–100, 2013.

Hong, S. H., J. M. H. Hendrickx, and B. Borchers: Up-scaling of SEBAL derived evapotranspiration maps from Landsat (30 m) to MODIS (250 m) scale, Journal of Hydrology, 370, 122–138, 2009.

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E.: Hole-filled SRTM for the globe Version 4, available from the CGIARCSI SRTM 90m Database, http://srtm.csi.cgiar.org (last access: 26 February 2016), 2008.

Kalma, J. D., T. R. McVicar, and M. F. McCabe: Estimating land surface evaporation: A review of methods using remotely sensed surface temperature data, Surv. Geophys., 29, 421–469, doi:10.1007/s10712-008-9037-z, 2008.

Kollet S. J.: Influence of soil heterogeneity on evapotranspiration under shallow water table conditions: transient, stochastic simulations, Environmental Research Letters, 4, 35007, doi:10.1088/1748-9326/4/3/035007, 2009.

Koster R. D. et al.: GLACE: The Global Land– Atmosphere Coupling Experiment. Part I: Overview. J. Hydrometeor., 7, 590–610, 2006.

Koster R. D., and M. Suarez: Modeling the land surface boundary in climate models as a composite of independent vegetation stands, J. Geophysical Research, 97 (D3), 26-97-2715, 1992.

Lu, H., T., Liu, Y. Yang, D. Yao: A hybrid dual-Source model of estimating evapotranspiration over different ecosystems and implications for satellite-based approaches, Remote Sens. 6, 8359–8386, 2014.

Maayar, M. E., J. M. Chen: Spatial scaling of evapotranspiration as affected by heterogeneities in vegetation, topography, and soil texture, Remote Sensing of Environment, 102, 33–51,2006.

Mahrt, L., J. Sun, D. Vickers, J. I. MacPherson, J. R. Perderson, and R. L. Desjardins: Observations of fluxes and inland breezes over a heterogeneous surface, J. Atmos. Sci. 51, 2165e2178, 1992.

McCabe M., and E. Wood: Scale influences on the remote estimation of evapotranspiration using multiple satellite sensors, Remote Sensing of Environment 105 (2006) 271–285, 2006.

Mezentsev, V. S.: More on the calculation of average total evaporation, Meteorol. Gidrol., 5, 24–26, 1955.

Montheith, J. L.: Evaporation and environment, the state of and movement of water in living organisms, Proceeding of Soc. for Exp. Biol., 19, 205-234, doi:10.1002/qj.49710745102, 1965.

Mu, Q., F. A. Heinsch, M. Zhao, and S. W. Running: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sens. Environ., 111, 519–536, doi:10.1016/j.rse.2007.04.015, 2007.

Peel, M. C., B. L. Finlayson, and T. A. McMahon: Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci., 11, 1633-1644, https://doi.org/10.5194/hess-11-1633-2007, 2007.

PRISM Climate Group, Oregon State University, http://prism.oregonstate.edu, created 22 Feb 2017.

Rouholahnejad Freund, E., and J. W. Kirchner: A Budyko framework for estimating how spatial heterogeneity and lateral moisture redistribution affect average evapotranspiration rates as seen from the atmosphere, Hydrology and Earth System Sciences, 21(1), 217-233, 2017.

Santanello J. R., and C. D. Peters-Lidard: Diagnosing the Sensitivity of Local Land–Atmosphere Coupling via the Soil Moisture–Boundary Layer Interaction, J. Hydrometeporology, 12, 766-786, doi: 10.1175/JHM-D-10-05014.1, 2011.

Sato N., P. J. Sellers, D. A. Randall, E. K. Schneider, J. Shukla, J. L. Kinter III, Y. T. Hou, and E. Albertazzi: Effects of Implementing the Simple Biosphere Model in a General Circulation Model, J. Atmospheric Sciences, 46(18), 2757-2782, 1989.

Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling: Investigating soil moisture–climate interactions in a changing climate: A review, Earth-Science Reviews, 99(3–4), 125-161, 2010.

Shahraeeni, E., and D. Or: Thermo-evaporative fluxes from heterogeneous porous surfaces resolved by infrared thermography, Water Resour. Res., 46, W09511, doi:10.1029/2009WR008455, 2010.

Su, Z.: The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrology and Earth System Sciences, 6, 85–100, 2002.

Teuling, A. J. and de Badts, E. A. G. and Jansen, F. A. and Fuchs, R. and Buitink, J. and Hoek van Dijke, A. J. and Sterling, S. M., Climate change, reforestation/afforestation, and urbanization impacts on evapotranspiration and streamflow in Europe, Hydrology and Earth System Sciences, 23, 3631—3652, DOI = {10.5194/hess-23-3631-2019, 2019.Turc, L.: Le bilan d’eau des sols: relation entre la precipitations, l’evaporation et l’ecoulement, Ann. Agron. A, 5, 491–569, 1954.

Wood, N., and P. J. Mason: The influence of static stability on the effective roughness length for momentum and heat transfer, Quart. J. Roy. Meteor. Soc. 117, 1025e1056, 1991.