

Dear Editor,

Thank you for your comments. Below please find our point by point response to your comments and questions. The Editor's comments are in regular font and our response is in bold. **The line numbers correspond to the revised manuscript (rev2) in "all mark up" view setting.**

Editor Decision: Publish subject to minor revisions (review by editor) (25 Aug 2020) by Anke Hildebrandt
Comments to the Author:

Dear authors,

thank you for your revisions and new version of the manuscript. I have some few requests that should be quick to accomodate, before the manuscript is ready for publication.

Some of the questions of the reviewers were only answered in the response, but no additional information was added to the manuscript. Could you please add this information?

For example, both reviewers requested which procedure was used for averaging. Could you please enter this information at a suitable place into the manuscript?

We added the statements to the manuscript: lines 118-120, 228-230

Similarly, both reviewers were wondering, whether soil moisture variance only was a strong driver on those days selected for presentation.

We now added the statements that was initially only provided to the reviewers to the manuscript: lines 289-291

Myself, I am wondering about the presentation of the „randomly selected day“. How did you select them „randomly“? Can you please add this information to the manuscript.

The two selected days are days 150 and 200 of Julian day calendar of year 2004 and were selected arbitrarily. We added this information to the manuscript, changed “randomly” to “arbitrarily” as it conveys better the way we selected the days, and we corrected for the actual date of these two days. (The codes are counting day 150 and 200 of the year 2004, so no miscalculation there). Lines 247-249, and throughout the text and figures.

Finally, reviewer #2 requested, whether you are showing volumetric soil moisture (the more common term is volumetric soil water content, which ranges from 0 to porosity) or soil moisture saturation (which ranges from 0 to 1). In your response you state that you are presenting soil volumetric soil moisture which ranges from 0 to 1. This must be an error somewhere. From Fig 1 it appears you are presenting soil moisture saturation. Would you please have another look at this?

Thank you for picking on this again. Indeed you are right. The data we have used and mapped is soil moisture saturation. We corrected for this in the manuscript (Lines 132-134) and in Figure 1, and S2, S3, S5, and S6.

I am looking forward to the new version of you manuscript,

Thank you very much for taking time.

Anke Hildebrandt

1 **Averaging over spatiotemporal heterogeneity substantially biases evapotranspiration rates in a mechanistic**
2 **large-scale land evaporation model**

3 Elham Rouholahnejad Freund^{1,2,3}, Massimiliano Zappa⁴, James W. Kirchner^{3,4,5}

4

5 ¹Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium

6 ²Chair of Hydrology, Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany

7 ³Department of Environmental Systems Science, ETH Zurich, CH-8092 Zürich, Switzerland

8 ⁴Swiss Federal Research Institute WSL, CH-8903 Birmensdorf, Switzerland

9 ⁵Department of Earth and Planetary Science, University of California, Berkeley, CA 94720 USA

10

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

12

13 **Abstract**

14 Evapotranspiration (ET) influences land-climate interactions, regulates the hydrological cycle, and contributes
15 to the Earth's energy balance. Due to its feedback to large-scale hydrological processes and its impact on
16 atmospheric dynamics, ET is one of the drivers of droughts and heatwaves. Existing land surface models differ
17 substantially, both in their estimates of current ET fluxes and in their projections of how ET will evolve in the
18 future. Any bias in estimated ET fluxes will affect the partitioning between sensible and latent heat, and thus
19 alter model predictions of temperature and precipitation. One potential source of bias is the so-called
20 "aggregation bias" that arises whenever nonlinear processes, such as those that regulate ET fluxes, are
21 modeled using averages of heterogeneous inputs. Here we demonstrate a general mathematical approach to
22 quantifying and correcting for this aggregation bias, using the GLEAM land evaporation model as a relatively
23 simple example. We demonstrate that this aggregation bias can lead to substantial overestimates in ET fluxes
24 in a typical large-scale land surface model when sub-grid heterogeneities in land surface properties are
25 averaged out. Using Switzerland as a test case, we examine the scale-dependence of this aggregation bias and
26 show that it can lead to an average overestimation of daily ET fluxes by as much as 10% across the whole
27 country (calculated as the median of the daily bias over the growing season). We show how our approach can
28 be used to identify the dominant drivers of aggregation bias, and to estimate sub-grid closure relationships
29 that can correct for aggregation biases in ET estimates, without explicitly representing sub-grid heterogeneities
30 in large-scale land surface models.

31 **Plain Language Summary**

32 Evapotranspiration (ET) is the largest flux from the land to the atmosphere and thus contributes to Earth's
33 energy and water balance. Due to its impact on atmospheric dynamics, ET is a key driver of droughts and
34 heatwaves. In this paper, we demonstrate how averaging over land surface heterogeneity contributes to
35 substantial overestimates of ET fluxes. We also demonstrate how one can correct for the effects of small-scale
36 heterogeneity without explicitly representing it in land surface models.

37

38 1. Introduction

39 Earth's surface and subsurface are characterized by spatial heterogeneity spanning wide ranges of scales,
40 including scales that cannot be explicitly resolved by large-scale Earth System Models (ESMs), which are
41 typically run at resolutions of 10-100 kilometers. Averaging over this finer-scale heterogeneity can bias model
42 estimates of water and energy fluxes and hence alter future temperature predictions. Earth system model
43 estimates of global terrestrial evaporation differ substantially from atmospheric reanalyses based on in-situ
44 and satellite remote sensing observations (Mueller et al., 2013), but it is unclear how much of these
45 differences could be attributed to errors in capturing sub-grid heterogeneity.

46

47 Several recent studies (e.g., Fan et al., 2019; Shrestha et al., 2018) have emphasized the need to account for
48 land surface heterogeneity in large-scale ESMs. Despite recent community efforts in refining ESMs' spatial
49 resolution (Huang et al., 2016; Rauscher et al., 2010; Ringler et al., 2008; Skamarock et al., 2012; Zarzycki et al.,
50 2014), the grid resolution of present-day ESMs is still too coarse to explicitly capture important effects of
51 surface heterogeneity. Whether the solution lies in hyper-resolution large-scale land surface modeling remains
52 an open question, because heterogeneities that are important to land-atmosphere fluxes will not be fully
53 resolved even at scales of 100 m (Beven and Cloke, 2012).

54

55 The effects of aggregating over spatial heterogeneity in land surface models have been assessed using several
56 approaches. Most of these approaches compare grid-cell-averaged energy and water fluxes with flux estimates
57 for finer-resolution grids, or for grid cells that are subdivided into mosaics of several surface types which
58 separately exchange momentum, energy, and water vapor with the overlying atmosphere (e.g., Giorgi, 1997).
59 Several studies have reported increases in average evapotranspiration (ET) (e.g., Kuo et al., 1999; Boone and
60 Wetzel, 1998; Hong et al., 2009; McCabe and Wood, 2006; El Maayar and Chen, 2006), and at least one has
61 reported decreases in grid-cell average ET (Ershadi et al., 2013), as model grids are coarsened and less spatial
62 heterogeneity is accounted for. Shrestha et al. (2018) studied the effects of horizontal grid resolution on ET
63 partitioning in the TerrSysMP Earth system model and found that the aggregation of topography decreases
64 average slope gradients and obscures small-scale convergence and divergence zones, directly impacting
65 surface and subsurface flow. They observed 5 and 8 percent decreases in the transpiration/evapotranspiration
66 ratio for a dry and a wet year, respectively, when their model grid cells were coarsened from 120 m to 960 m.
67 All these studies calculate the effects of land surface heterogeneity on ET fluxes using numerical experiments
68 that refine the model's spatial resolution, either directly or through the use of land-surface mosaics.

69

70 Quantifying the effect of sub-grid scale heterogeneity on grid-cell-averaged fluxes is especially important when
71 highly nonlinear processes are involved. Regardless of scale, the main challenge is not to explicitly represent
72 the heterogeneity in all its details, but instead to define an appropriate scale-dependent sub-grid closure
73 relationship that recognizes the important heterogeneities within the grid elements and the nonlinearities in

74 the processes (Beven, 2006). Such a sub-grid closure scheme would capture the effects of sub-grid
75 heterogeneity in large-scale land surface models without forcing them to run at finer spatial resolutions.
76

77 We have recently proposed a general theoretical framework, based on Taylor series expansions, that
78 quantifies the "aggregation bias" that results from averaging over sub-grid heterogeneity when grid-cell-
79 averaged ET is estimated (Rouholahnejad Freund and Kirchner, 2017; Rouholahnejad Freund et al., 2019). In
80 contrast to the numerical experiments described above, this theoretical framework does not depend on a
81 particular evapotranspiration model or grid scale. Our previous work demonstrated this framework using
82 Budyko curves as a see-through "toy" model, leaving open the question of how strongly ET estimates would be
83 affected by sub-grid heterogeneity in a more typical mechanistic evapotranspiration model. Here we use the
84 mechanistic evapotranspiration model GLEAM to quantify how aggregation biases vary across a range of
85 scales, using Switzerland as a case study. We show how our Taylor expansion framework can be used to
86 quantify the sensitivity of ET fluxes to heterogeneity in their individual drivers. We further demonstrate how
87 this framework can be used to estimate correction factors (i.e., sub-grid closure relationships) that account for
88 the effects of sub-grid heterogeneity without explicitly modeling it, and show how these correction factors can
89 be used to improve grid-scale ET estimates. Because our framework is not model-specific, the analysis
90 presented here could also be applied to many other evapotranspiration algorithms.

91

92 **2. Methods and results**

93 **2.1. A common mechanistic framework for predicting evapotranspiration**

94 Most large-scale land surface models calculate ET as a function of available water and energy at daily time
95 steps. They typically multiply an estimate of potential evapotranspiration (PET) by a conversion factor to
96 calculate actual evapotranspiration. PET is generally understood as the maximum rate of evapotranspiration
97 from a large area (to avoid the effect of local advection) covered completely and uniformly by actively growing
98 vegetation with adequate moisture at all times (Brutseart, 1984). Models typically estimate PET using the
99 Penman equation (Penman, 1948; intended for open water surfaces), the Penman-Monteith equation
100 (Monteith, 1965, Monteith and Unsworth, 1990; intended for reference crop evapotranspiration by adding
101 atmospheric transport processes and stomatal resistance to Penman's open water evaporation), or the
102 Priestley-Taylor equation (Priestley and Taylor, 1972; intended for open water and water-saturated crops and
103 grasslands). The conversion factor that is used to estimate ET from PET typically depends on plant physiology
104 and on the water that is available for evaporation.

105

106 Here, we employ an ET algorithm that is used by several land surface models (i.e., Global Land-surface
107 Evaporation: The Amsterdam Methodology (GLEAM); Miralles et al., 2011; Martens et al., 2017), in which
108 actual ET is calculated as a fraction of PET. This fraction is expressed as a multiplicative factor, often called a
109 stress factor, which ranges between 0 and 1 and thus limits ET rates. Under wet conditions, ET can equal PET
110 (stress factor equals one) while under dry conditions, PET is multiplied by a stress factor smaller than one
111 depending on the degree of water stress. This approach is employed by the GLEAM model, among others.

112 GLEAM is a diagnostic satellite-data-driven method that is used to estimate global land evaporation fluxes.
 113 GLEAM uses the Priestley-Taylor formula and remotely sensed datasets of radiation and temperature to
 114 calculate PET. In GLEAM, actual ET is calculated by constraining PET estimates by a stress factor that is based
 115 on estimates of root-zone soil moisture. The root zone soil moisture is derived from a multi-layer water
 116 balance module that describes the infiltration of precipitation through the vertical soil profile. ET estimates
 117 from GLEAM have been applied in many studies (e.g., Miralles et al., 2013; Miralles et al., 2014; Greve et al.,
 118 2014; Jasechko et al., 2013). GLEAM operates on daily time steps at 0.25-degree spatial resolution. 0.25
 119 degrees is about 27.6 km in the north-south direction and 18.9 km in the east-west direction at the latitude of
 120 Switzerland. To the best of our knowledge, there are no prior studies quantifying the aggregation bias in ET
 121 estimates from GLEAM or other models with similar ET formulations.

122

123 GLEAM calculates ET as an explicit function of the stress factor and potential evaporation:

$$124 \quad ET = S \cdot PET + (1 - \beta) I, \quad (1)$$

125 where ET is actual evapotranspiration (mm d^{-1}), S is the evaporative stress factor (-) that accounts for
 126 environmental conditions that reduce actual ET relative to potential ET, I is interception loss (mm d^{-1}), and β is
 127 a constant ($\beta = 0.07$ – Gash and Stewart, 1977) that avoids double-counting of interception losses during hours
 128 with wet canopy. The stress factor (S) depends on the soil moisture conditions, and is parametrized separately
 129 for tall canopy, short vegetation, and bare soil. GLEAM uses the following soil-moisture-based
 130 parameterization to calculate the stress factor (Miralles et al., 2011; Martens et al., 2017):

$$131 \quad S = 1 - \left(\frac{w_c - w_w}{w_c - w_{wp}} \right)^2, \quad (2)$$

132 where S is the stress factor (-) for tall canopy, w_w is ~~soil moisture saturation~~ soil moisture saturation the volumetric soil moisture at any
 133 given time ($\text{m}^3 \text{m}^{-3}$), and w_c and w_{wp} are the critical soil moisture saturation level and ~~soil moisture at~~ wilting
 134 point. For soil moisture saturation values below the wilting point w_{wp} , the stress is maximal (stress factor
 135 equals 0), causing ET to sharply decline to zero. For values above the critical moisture level w_c , there is no
 136 water stress (stress factor equals 1) and ET equals PET. Between w_{wp} and w_c the stress increases as soil
 137 moisture decreases following a parabolic function (Eq. 2). In the analysis presented below, we set the critical
 138 soil moisture level (w_c) and ~~soil moisture at~~ wilting point (w_{wp}) to 0.6 and 0.1, $\text{m}^3 \text{m}^{-3}$ respectively. To simplify
 139 the analysis presented below, we have used the tall-canopy stress factor (Eq. 2) for all of Switzerland, even
 140 though the short-canopy or bare-soil formulations may be better suited to some locations.

141

142 GLEAM uses the Priestley-Taylor approach to calculate PET (Priestley and Taylor, 1972):

$$143 \quad PET = \frac{\alpha}{\lambda} \frac{\Delta}{\Delta + \gamma} (R_n - G), \quad (3)$$

144 where PET is potential evapotranspiration (mm d^{-1}), α is a dimensionless coefficient that parametrizes the
 145 resistance to evaporation and is set to 0.8 for tall canopy in GLEAM (Miralles et al., 2011), $\lambda = 2.26$ (MJ kg^{-1}) is
 146 the latent heat of vaporization, R_n is net radiation ($\text{MJ m}^{-2} \text{d}^{-1}$), G is the ground heat flux, approximated as
 147 $G = 0.05 R_n$ ($\text{MJ m}^{-2} \text{d}^{-1}$) for tall canopy in GLEAM, T is temperature ($^{\circ}\text{C}$), and Δ is the slope of the

148 temperature/saturated vapor pressure curve ($\text{kPa}^\circ\text{C}^{-1}$), which is functionally related to temperature (Tetens,
149 1930; Murray, 1967; Stanghellini, 1987):

$$150 \quad \Delta = ae^{bT}, \quad (4)$$

151 where $a = 0.04145$ ($\text{kPa}^\circ\text{C}^{-1}$), $b = 0.06088$ ($^\circ\text{C}^{-1}$), and γ is the psychrometric constant ($\text{kPa}^\circ\text{C}^{-1}$) which can be
152 calculated as (Brunt, 1952):

$$153 \quad \gamma = \frac{C_{p_{air}} * P}{\lambda * MW_{ratio}}, \quad (5)$$

154 where $C_{p_{air}} = 0.001013$ ($\text{MJ kg}^{-1}\text{C}^{-1}$) is the specific heat of air at constant pressure, $P = 101.3$ (KPa) is
155 atmospheric pressure, and $MW_{ratio} = 0.622$ (-) is the molecular weight ratio of $\text{H}_2\text{O}/\text{air}$. Substituting the
156 aforementioned constants in Eq. 5 yields $\gamma = 0.073$ ($\text{kPa}^\circ\text{C}^{-1}$). Expanding Eq. 1 using Eqs. 2-5 yields the ET
157 function as calculated by GLEAM:

$$158 \quad \begin{aligned} ET_{[mmd^{-1}]} &= \left[-4w_w[m^3m^{-3}]^2 + 4.8w_w[m^3m^{-3}] - 0.44 \right] * \frac{\alpha_{[]}}{\lambda_{[MJ kg^{-1}]} * \Delta_{[kPa^\circ C^{-1}]} + \gamma_{[kPa^\circ C^{-1}]}} \\ &\quad * 0.95 * \frac{86400}{1000000} * R_n[w_m^{-2}] + (1 - \beta) I_{[mmd^{-1}]} \\ &= \left[-4w_w^2 + 4.8w_w - 0.44 \right] * 0.02905 * \frac{a e^{bT}}{a e^{bT} + 0.073} R_n + (1 - 0.07) I_{[mmd^{-1}]} \end{aligned} \quad (6)$$

159
160 In the analysis below, we use the GLEAM evapotranspiration algorithm to demonstrate how aggregation biases
161 can be estimated in land surface modeling schemes. We chose GLEAM because its governing equations are
162 amenable to the analytical solutions derived below. Here we make no particular claim for the accuracy or
163 validity of GLEAM as an evapotranspiration model, nor is our analysis intended to test this. Likewise our
164 analysis should not be interpreted as implying that GLEAM is any more, or less, susceptible to aggregation bias
165 than other evapotranspiration schemes, because this question is beyond the scope of the current paper.

166

167 **2.2. Mathematical framework for predicting aggregation bias**

168 *Nonlinear averaging using second-order Taylor expansions*

169 ET is a nonlinear function of its drivers. An intrinsic property of any nonlinear function is that the average of
170 the function will not equal the function evaluated at the average inputs (e.g., Rastetter et al., 1992; Giorgi and
171 Avissar, 1997; Rouholahnejad Freund and Kirchner, 2017). Thus averaging over sub-grid heterogeneity in ET
172 drivers, as large-scale land surface models do, would be expected to lead to biased ET estimates, even if the
173 underlying equations were exactly correct. For an ET function of three variables, namely R_n , w_w , and T , the
174 mean of the ET function, in terms of the function's value at the mean of its inputs, can be approximated by the
175 second-order Taylor series expansion of the ET function (Eq. 6):

$$176 \quad \begin{aligned} \overline{ET} &\approx \widehat{ET} + \frac{1}{2} \left[\frac{\partial^2 ET}{\partial R_n^2} \text{Var}(R_n) + \frac{\partial^2 ET}{\partial w_w^2} \text{Var}(w_w) + \frac{\partial^2 ET}{\partial T^2} \text{Var}(T) \right] \\ &\quad + \frac{\partial^2 ET}{\partial R_n \partial T} \text{Cov}(R_n, T) + \frac{\partial^2 ET}{\partial R_n \partial w_w} \text{Cov}(R_n, w_w) + \frac{\partial^2 ET}{\partial w_w \partial T} \text{Cov}(w_w, T), \end{aligned} \quad (7)$$

177 where \overline{ET} is the estimate of the true average of the nonlinear ET function over its variable inputs, \widehat{ET} is the ET
178 function evaluated at its mean inputs, and the derivatives are understood to be evaluated at the mean values
179 of the variables ($\overline{R_n}, \overline{w_w}, \overline{T}$) and multiplied by the corresponding variances and covariances among finer-

180 resolution input data. For the specific case of the GLEAM model, the ET function is evaluated at its mean inputs
 181 (\overline{ET}) and these derivatives are derived analytically from the ET function described by Eq. 6, directly yielding the
 182 following expressions:

$$183 \quad \widehat{ET} = [-4\overline{w}_w^2 + 4.8\overline{w}_w - 0.44] * 0.02905 * \frac{a e^{b\overline{T}}}{a e^{b\overline{T}} + 0.073} \overline{R}_n, \quad (8)$$

$$184 \quad \frac{\partial^2 ET}{\partial R_n^2} = 0, \quad (9)$$

$$185 \quad \frac{\partial^2 ET}{\partial w_w^2} = [-8] * 0.02905 * \frac{\Delta}{\Delta + \gamma} R_n \quad (w_{wp} \leq w_w \leq w_c), \quad (10a)$$

$$186 \quad \frac{\partial^2 ET}{\partial w_w^2} = 0 \quad (w_w < w_{wp}, \quad w_w > w_c), \quad (10b)$$

$$187 \quad \frac{\partial^2 ET}{\partial T^2} = [-4w_w^2 + 4.8w_w - 0.44] * 0.02905 * R_n * b^2 * \frac{\gamma^2 \Delta - \gamma \Delta^2}{(\gamma + \Delta)^3}, \quad (11)$$

$$188 \quad \frac{\partial^2 ET}{\partial R_n \partial T} = [-4w_w^2 + 4.8w_w - 0.44] * 0.02905 * \frac{\Delta}{\Delta + \gamma} * \frac{b\gamma}{\Delta + \gamma}, \quad (12)$$

$$189 \quad \frac{\partial^2 ET}{\partial R_n \partial w_w} = [-8w_w + 4.8] * 0.02905 * \frac{\Delta}{\Delta + \gamma} \quad (w_{wp} \leq w_w \leq w_c), \quad (13a)$$

$$190 \quad \frac{\partial^2 ET}{\partial R_n \partial w_w} = 0 \quad (w_w < w_{wp}, \quad w_w > w_c), \quad (13b)$$

$$191 \quad \frac{\partial^2 ET}{\partial w_w \partial T} = [-8w_w + 4.8] * 0.02905 * \frac{\Delta}{\Delta + \gamma} * \frac{b\gamma}{\Delta + \gamma} * R_n \quad (w_{wp} \leq w_w \leq w_c), \quad \text{and} \quad (14a)$$

$$192 \quad \frac{\partial^2 ET}{\partial w_w \partial T} = 0 \quad (w_w < w_{wp}, \quad w_w > w_c), \quad (14b)$$

193 where Δ depends on temperature as described in Eq. (4). The difference between the average of the functions
 194 (\overline{ET}) and the function of the averages (\widehat{ET}), or, equivalently, the sum of all the other terms in Eq. (7),
 195 represents the aggregation bias. The magnitude of this bias can be calculated by combining Eqs. 7-14 with
 196 estimates of the variances and covariances of the input variables. Note that the interception term in equation
 197 6 is dropped out from the derivatives as the interception loss in GLEAM is a linear function of amount of
 198 rainfall necessary to saturate the canopy and therefore has negligible effect when averaged.

199
 200 The approach outlined in Eq. (7) is general and could be extended to other land surface modeling schemes.
 201 The partial derivatives in Eqs. (8-14), of course, are specific to the GLEAM equations; for other models they
 202 would differ. More complex land surface model algorithms may not have such simple analytical derivatives; in
 203 that case, the derivatives can be evaluated numerically.

204

205 **2.3. Sub-grid heterogeneity and aggregation bias in ET estimates across Switzerland**

206 Drivers of ET (i.e., soil moisture, net radiation, and temperature) can be highly heterogeneous within the grid
 207 cells of typical ESMs. Soil moisture can show pronounced spatial variability, especially in areas where surface
 208 roughness, porosity, and permeability vary by orders of magnitude across a variety of length scales (Giorgi and
 209 Avissar, 1997). Temperature and incoming radiation vary significantly with season, elevation, altitude, and

210 albedo. Switzerland, for example, shows strong local variations in average annual temperature, soil moisture
211 content, net radiation, and albedo (Fig. 1; albedo values in Fig. S1).

212

213 We quantified how averaging over spatial (and temporal) heterogeneities of ET drivers affects estimated ET at
214 several grid scales across Switzerland, as an example case for which high-resolution data are available. Our
215 analysis is based on 500-m input data of temperature (interpolation of MeteoSwiss data after Viviroli et al.,
216 2009), net radiation (Viviroli et al., 2009), and soil moisture (simulations from the hydrological model PREVAH,
217 Brunner et al., 2019; Speich et al., 2015; Orth et al., 2015; Zappa et al., 2003) at daily time steps for the 2004
218 growing season. Although our soil moisture data are derived from model simulations whose accuracy is
219 difficult to assess due to the scarcity of real-world soil moisture measurements, for our purposes all that is
220 necessary is that the simulated values exhibit realistically complex spatial variability.

221

222 We used the GLEAM equations, as outlined in Sect. 2, to calculate ET for each day at the 500-m resolution of
223 these input data. We use these 500-m ET estimates as virtual "truth" for the purpose of our analysis, because
224 our goal is not to determine whether GLEAM estimates of ET are accurate (compared to direct measurements,
225 for example), but rather to quantify how spatial aggregation affects them.

226

227 To quantify how spatial aggregation affects model estimates of ET, we calculated ET over larger spatial scales
228 in two different ways. ~~First, we~~ We First calculated the arithmetic average of averaged the 500-m ET estimates
229 over 1/32, 1/16, 1/8, 0.25, 0.5, 0.75, 1, and 2-degree grid cells across Switzerland, to represent the "true"
230 average ET at those grid scales. ~~Second, we averaged~~ Next we calculated the arithmetic average of the 500-m
231 input data (of temperature, soil moisture, and net radiation) over the same grid cells, and then used these
232 grid-cell-averaged input data in the GLEAM equations to calculate the modeled coarse-resolution ET at each
233 grid scale. The deviation of the modeled coarse-resolution ET from the "true" average ET measures the
234 aggregation bias. Because this numerical experiment uses the same model equations, based on the same
235 underlying data, for the ET calculations at each spatial resolution, it isolates spatial aggregation as the only
236 possible cause of the difference between the "true" average ET (\overline{ET} in Eq. 7) and the coarse-resolution
237 modeled ET (\widehat{ET} in Eq. 7) at each grid scale.

238

239 Figure 2a shows that the ET aggregation bias varies considerably across Switzerland, and also varies
240 considerably with grid scale. The average aggregation bias is higher at coarser grid scales, averaging 10% at 2-
241 and 1-degree grid resolution across all of Switzerland (calculated as the median of the daily aggregation biases
242 over the growing season; Fig. 2a). Smaller grid scales typically exhibit smaller aggregation biases (averaging 4%
243 at 1/16-degree grid resolution across all of Switzerland calculated as the median of the daily aggregation
244 biases over the growing season) because they typically average over less spatial heterogeneity, but even at the
245 smallest grid scales, aggregation biases can locally reach 40% as indicated by the scatter plot in Fig. 3. These
246 figures are medians of the daily aggregation biases over the entire growing season of 2004; the aggregation
247 biases of two ~~randomly arbitrarily~~ selected days (May ~~31st~~-29th and July ~~21st~~-18th, 2004) at several spatial scales

lead to much larger overestimation of ET in parts of southern Switzerland (Figs. S2, S3). [The two selected days are days 150 and 200 of Julian day calendar of year 2004.](#)

Using our 500-m input data, we can test how well Eq. (7) estimates the difference between the "true" average ET and the coarse-resolution modeled ET at each grid scale. We used Eqs. (8-14) to calculate the partial derivatives of the GLEAM equations for each grid cell and time step, using the grid-cell averaged values of the input data. We then multiplied these derivatives by the corresponding variances and covariances among the 500-m input data to obtain bias estimates via Eq. (15) for each grid cell and time step:

$$\text{Bias} = \widehat{\text{ET}} - \overline{\text{ET}} \approx -\frac{1}{2} \left[\frac{\partial^2 \text{ET}}{\partial R_n^2} \text{Var}(R_n) + \frac{\partial^2 \text{ET}}{\partial w_w^2} \text{Var}(w_w) + \frac{\partial^2 \text{ET}}{\partial T^2} \text{Var}(T) \right] - \frac{\partial^2 \text{ET}}{\partial R_n \partial T} \text{Cov}(R_n, T) - \frac{\partial^2 \text{ET}}{\partial R_n \partial w_w} \text{Cov}(R_n, w_w) - \frac{\partial^2 \text{ET}}{\partial w_w \partial T} \text{Cov}(w_w, T), \quad (15)$$

where $\overline{\text{ET}}$ is the true average ET at some grid resolution, $\widehat{\text{ET}}$ is the modeled coarse-resolution ET at the same spatial scale, the right-hand side is the Taylor expansion estimate of the aggregation bias. We then compared these estimated biases against the "true" aggregation biases (the difference between the "true" average ET and the coarse-resolution modeled ET) in the numerical experiment described above. The true bias, in other words, is $\widehat{\text{ET}} - \overline{\text{ET}}$ in Eq. (15), and the estimated bias is the Taylor approximation on the right-hand side.

Figure 2b shows that the aggregation bias estimated by Eq. (15) is generally similar, in both overall magnitude and spatial distribution, to the "true" aggregation biases calculated by the numerical experiment. This comparison is shown more explicitly in Fig. 3, in which the estimated aggregation bias is compared with the "true" aggregation bias for each grid cell at each grid scale. Figures 2 and 3 show that Eq. (15) is generally a good predictor of aggregation bias. Both the estimated aggregation biases (Fig. 2) and the "true" aggregation biases are markedly higher in regions of greater topographic complexity (Fig. S4).

2.4. Correcting for aggregation bias

2.4.1. Identifying drivers of aggregation bias

The Taylor expansion in Eq. (15) not only allows one to quantify the aggregation bias; it also allows one to quantify the relative importance of the three input variables (net radiation, soil moisture, and temperature) as drivers of that bias. Each of the terms in Eq. (15) combines a variance or covariance that expresses how variable the input data are, and a second derivative that expresses how sensitive the average ET is to that variability. Each of these terms – a derivative multiplied by a variance or covariance – has the same units as ET, and thus they can be directly compared to one another.

Table 1 shows each of the aggregation bias terms, calculated over all of Switzerland for the two

[arbitrarily randomly](#) chosen days mentioned in Sect. 2.3 (May [31st29th](#) and July [21st18th](#), 2004). For these two example days, the aggregation bias is clearly dominated by a single term, associated with the variance of soil moisture. The variance in net radiation (R_n) creates no aggregation bias, because GLEAM ET is a linear function

283 of R_n ; thus positive and negative deviations from average R_n will increase and decrease ET by exactly
284 offsetting amounts. Similarly, the variance in temperature (T) also results in little aggregation bias, because
285 GLEAM ET increases nearly linearly with T across a wide range of temperature. The covariance terms similarly
286 lead to little aggregation bias. By contrast, the strong curvature in the quadratic dependence of ET on soil
287 moisture (Eq. 6) implies that positive and negative deviations from mean soil moisture will not have offsetting
288 ET effects, and thus that spatial heterogeneity in soil moisture can significantly alter average ET. On most of
289 the days of the year 2004, the soil moisture variance term is the dominant driver of the aggregation bias.
290 However, there are some days in which other factors such as the T and R_n covariance term are the dominant
291 factors.

292

293 2.4.2. Correcting for aggregation bias using sub-grid closure relationships

294 The Taylor expansion framework in Eq. (7) can be used not only to diagnose aggregation bias, but also to
295 estimate sub-grid closure relationships that correct for the effects of small-scale heterogeneity. The variance
296 and covariance terms in Eq. (7) express how sub-grid heterogeneity affects average ET at the grid scale,
297 implying that these aggregation bias estimates could be used to improve grid-scale ET estimates, without
298 explicitly modeling ET at high resolutions. This approach could be particularly useful in land surface algorithms
299 that are part of coarser-resolution Earth system models; in such cases it may be much more efficient to
300 evaluate Eqs. 7-14 at the coarse grid resolution than to directly evaluate the underlying ET model, Eq. 6, at
301 high resolution. The Taylor expansion approach could also be attractive where we lack spatially explicit high-
302 resolution maps of the ET drivers, but where their variances and covariances can nonetheless be estimated
303 from other sources (such as from the variability of topography, mapped soil units, remote sensing data, etc.).

304

305 It is beyond our scope here to construct such variance and covariance estimates, but we can illustrate how
306 they could potentially be used. The solid red symbols in Fig. 4 show the relationships between "true" average
307 ET and modeled grid-cell-averaged ET, for each grid cell (and one example day, May 31st, 2004) at several
308 different grid scales. For comparison, the open grey symbols in Fig. 4 show average ET estimated by the Taylor
309 expansion approach of Eq. (7), which corrects for sub-grid heterogeneity effects using only grid-cell-averaged
310 estimates of the ET drivers and their small-scale variances and covariances.

311

312 The heterogeneity-corrected ET estimates shown by the open symbols in Fig. 4 cluster much closer to the 1:1
313 line than the modeled grid-cell-averaged ET values shown by the solid red symbols, suggesting that the Taylor
314 expansion approach may substantially improve estimates of grid-cell-averaged ET. Real-world results may be
315 less clear than those shown in Fig. 4, because the heterogeneity-corrected ET estimates (the open symbols in
316 Fig. 4) are calculated using exact values for the variances and covariances of the ET drivers within each grid
317 cell, and in real-world cases these variances and covariances will not be known precisely. Figure 4 nonetheless

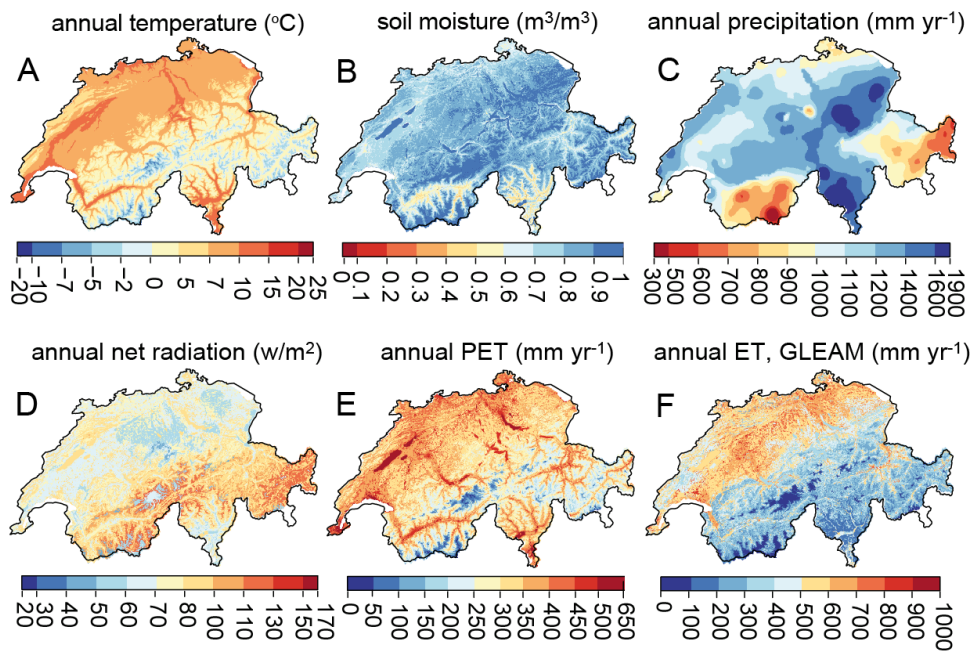
318 demonstrates the potential value of knowing, or being able to estimate, those variances and covariances.
 319 Efforts to determine those variances and covariances can be focused on the terms that matter the most, if one
 320 can identify the main drivers of aggregation bias using the methods described in Sect. 2.2 above.

321
 322
 323
 324
 325
 326
 327
 328
 329

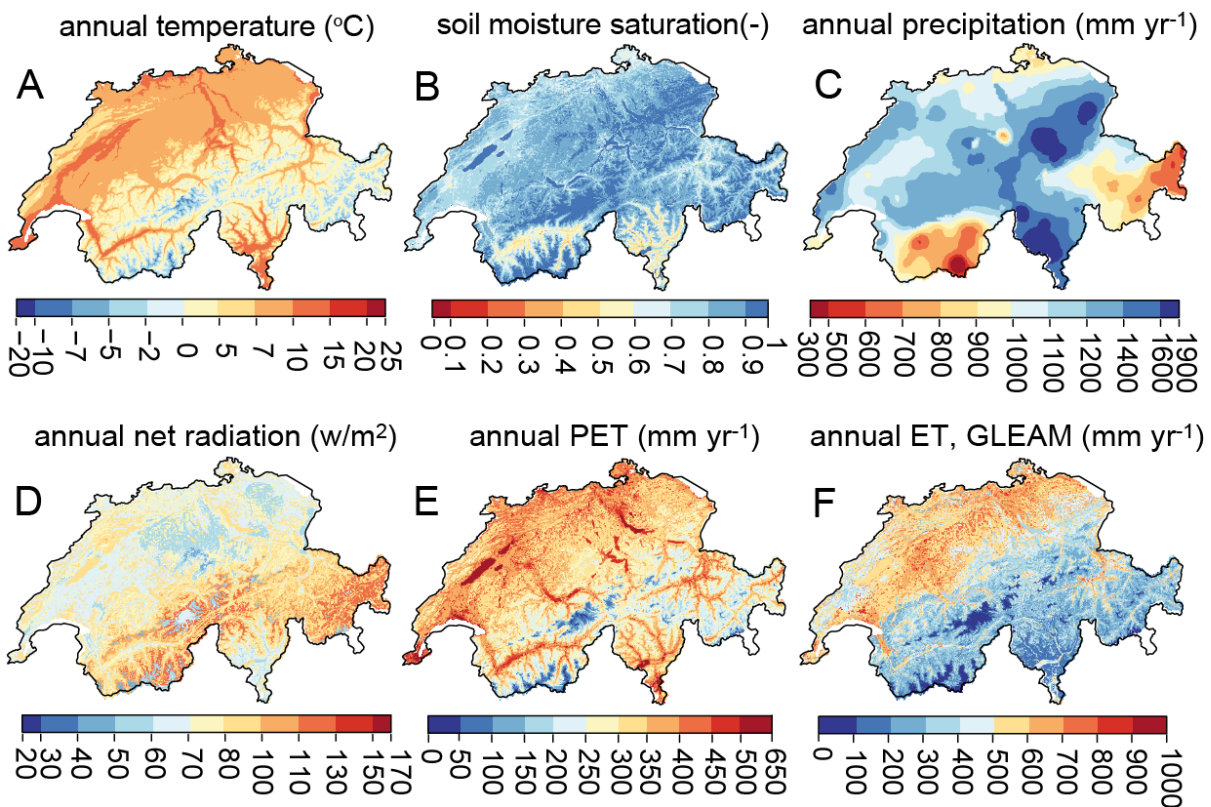
330 Table 1. Relative importance of different ET drivers in aggregation bias estimates (different terms in Eq. 15). Va
 331 lues are calculated for all of Switzerland for the two ~~arbitrarily randomly~~ chosen days (May ~~31st-29th~~
 332 ~~st-18th~~, 2004). The aggregation bias is dominated by the term associated with the variance of soil moisture for t
 333 hese two example days.

	\widehat{ET} mm d ⁻¹	\overline{ET} mm d ⁻¹	Bias %	Contribution of $Var(R_n)$ term in % aggregation bias (%)	Contribution of $Var(w_w)$ term in % aggregation bias (%)	Contribution of $Var(T)$ term in % aggregation bias (%)	Contribution of $Cov(R_n, T)$ term in % aggregation bias (%)	Contribution of $Cov(R_n, w_w)$ term in % aggregation bias (%)	Contribution of $Cov(R_n, w_w)$ term in % aggregation bias (%)
Calculation	(Eq. 8)	(Eq. 7)	(Eq. 15)	$\frac{1}{2} \frac{\partial^2 ET}{\partial R_n^2} Var(R_n)$ $(\widehat{ET}.Bias)$	$\frac{1}{2} \frac{\partial^2 ET}{\partial w_w^2} Var(w_w)$ $(\widehat{ET}.Bias)$	$\frac{1}{2} \frac{\partial^2 ET}{\partial T^2} Var(T)$ $(\widehat{ET}.Bias)$	$\frac{\partial^2 ET}{\partial R_n \partial T} Cov(R_n, T)$ $(\widehat{ET}.Bias)$	$\frac{\partial^2 ET}{\partial R_n \partial w_w} Cov(R_n, w_w)$ $(\widehat{ET}.Bias)$	$\frac{\partial^2 ET}{\partial w_w \partial T} Cov(w_w, T)$ $(\widehat{ET}.Bias)$
31.05.20	2.3	1.89	21.7	0	81.65	0.90	1.05	2.80	14.41
21.07.20	2.11	1.84	14.84	0	83.35	2.34	6.56	1.84	6.01

334
 335



338

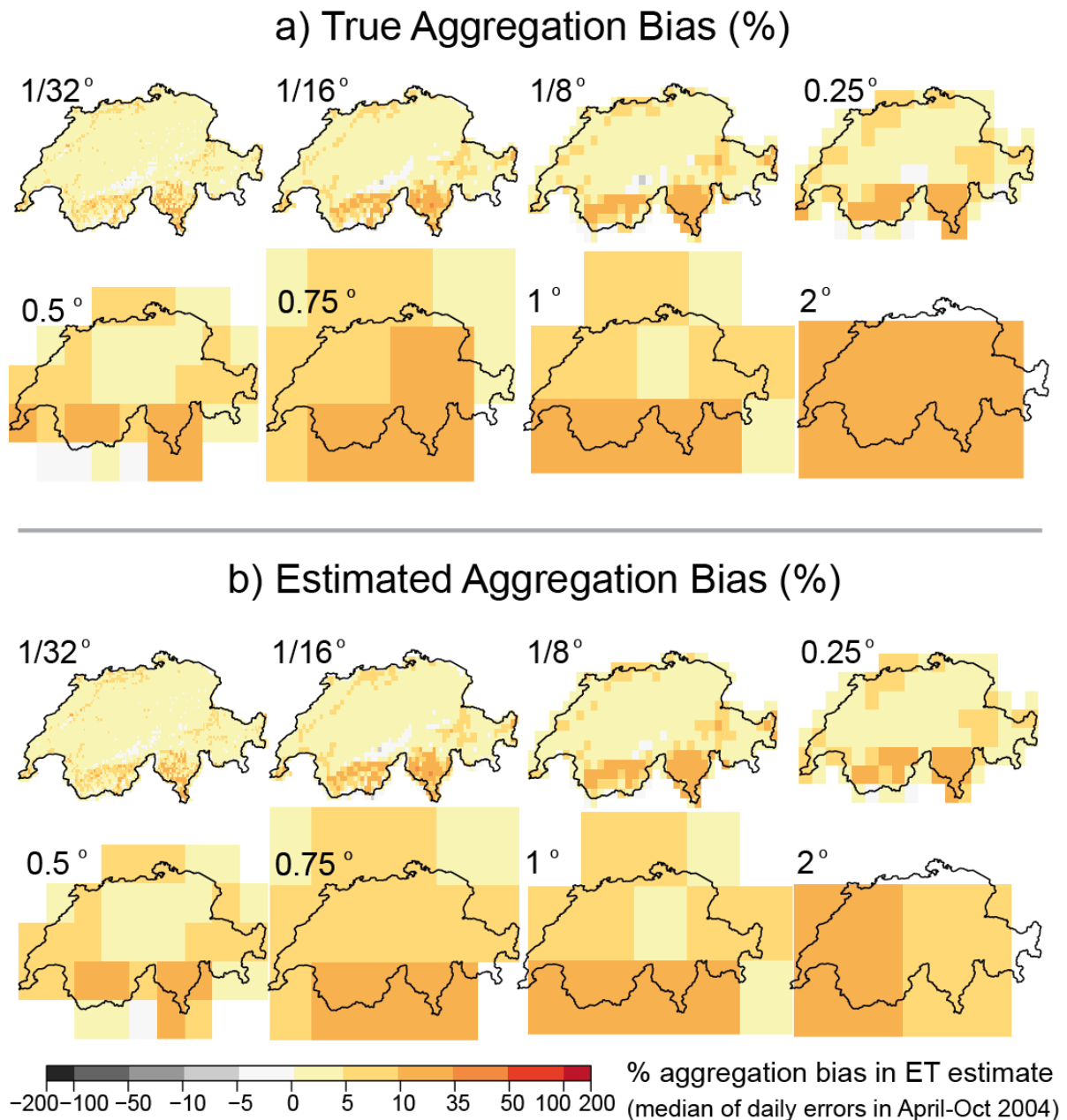


339

340 Figure 1. Spatial distribution of input data for the year 2004 at 500-m resolution: Annual mean (A) temperature
 341 (°C), (B) soil moisture saturation ($m^3 m^{-3}$), simulated by the PREVAH hydrological model), (C) precipitation (mm
 342 yr^{-1}), (D) net radiation ($W m^{-2}$), (E) potential evapotranspiration (PET, $mm yr^{-1}$) using the Priestley-Taylor
 343 equation (Eq. 3), and (F) evapotranspiration (ET, $mm yr^{-1}$) using the approach used in the GLEAM model (Eq. 1).
 344 See Table. S1 for references.

345

346



347

348

349

350

351

352

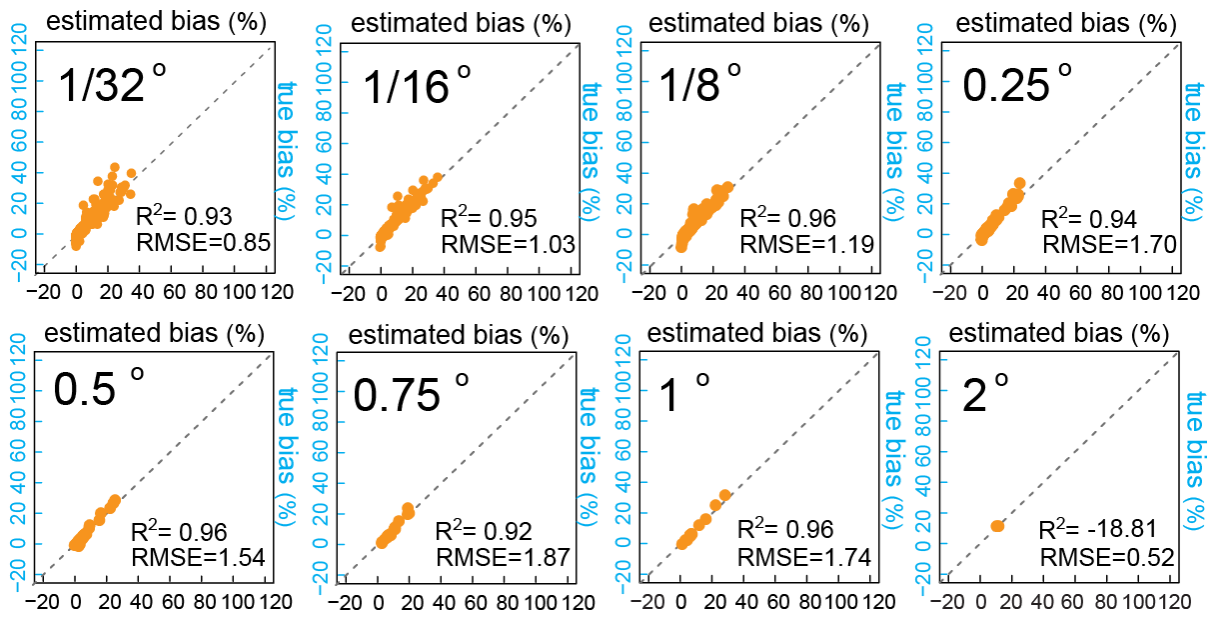
353

354

355

356

Figure 2. a) “True” aggregation bias in ET, as calculated by averaging the 500-m resolution ET estimates using fine-resolution input data in Eq. 6, over 1/32, 1/16, 1/8, 0.25, 0.5, 0.75, 1, and 2-degree grid cells across Switzerland. b) Aggregation bias in ET, as estimated by Eq. 7 from grid-cell averaged temperature ($^{\circ}\text{C}$), soil moisture (w_w), net radiation (R_n), their variances at each grid scale, and the covariances of all pairs of variables using the 500-m input data. At finer grid scales, the aggregation bias is more localized, and smaller on average. Across Switzerland as a whole, average aggregation bias becomes smaller as grid scales become finer, but never disappears completely.

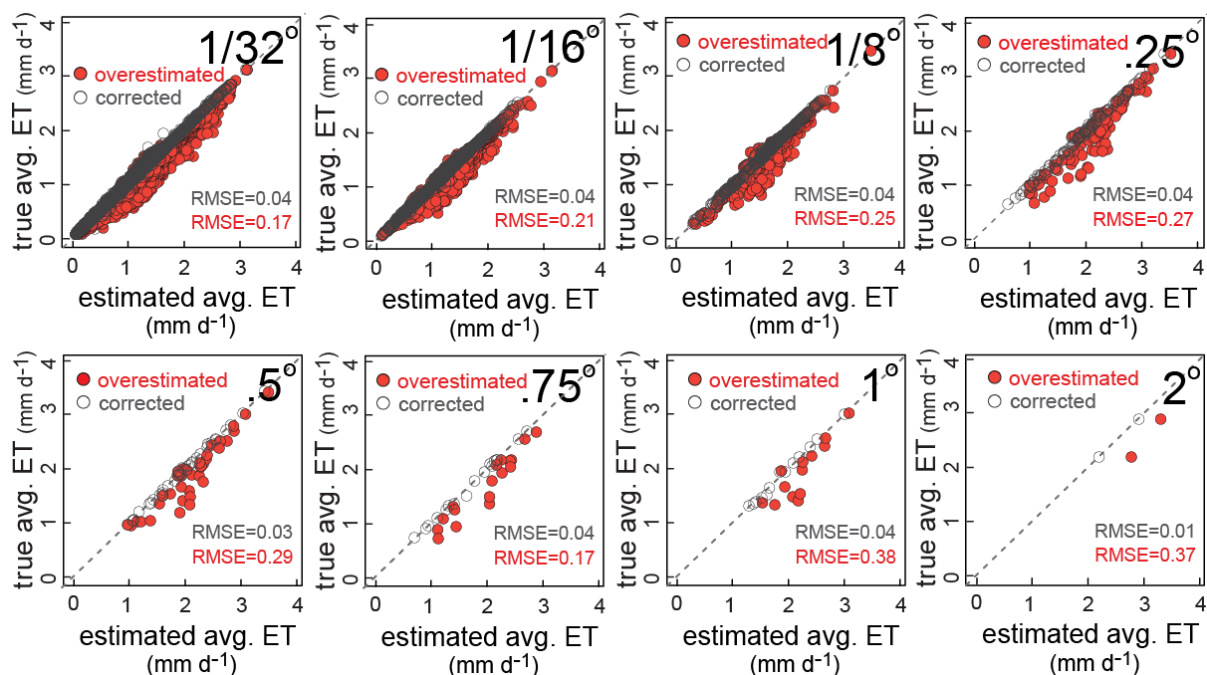


357

358 Figure 3. Daily estimated aggregation bias in ET estimates (% , median of daily biases in Apr.-Oct. 2004) versus
 359 daily true aggregation bias in ET estimates (% , median of daily biases in Apr.-Oct. 2004) at several spatial
 360 scales. Estimated aggregation biases are calculated using Eq. 7. True aggregation biases are calculated as
 361 differences between the finer resolution ET estimates from finer resolution input data, averaged over several
 362 spatial scales (average of functions) and ET values calculated from average inputs at each spatial scale
 363 (function of averages). The coefficients of determination (R^2) between the true and estimated aggregation
 364 biases verify the reliability of the Taylor expansion method and Eq. 7 as estimates of the aggregation bias.

365

366



368

369 Figure 4. Daily estimated ET rates versus “true” average ET at each grid cell at several different grid scales
 370 (example day, May 31st, 2004). The solid red symbols demonstrate the relationships between “true” average
 371 ET calculated using fine-resolution data at each grid cell and modeled grid-cell-averaged ET using grid-cell-
 372 averaged inputs in Eq.8, for each grid cell at several different grid scales (overestimated). For comparison, the
 373 open symbols show true average versus average ET estimated by the Taylor expansion approach of Eq. (7),
 374 which corrects for sub-grid heterogeneity effects using only grid-cell-averaged estimates of the ET drivers and
 375 their small-scale variances and covariances (heterogeneity-corrected ET estimates, corrected).

376

377 3. Discussion

378 Averaging over spatially heterogeneous ET drivers leads to substantial aggregation biases in ET flux estimates
 379 from a typical mechanistic large-scale land surface model. This aggregation bias arises from the inherent
 380 nonlinearities in evapotranspiration processes, coupled with the inherent spatial heterogeneity in the driving
 381 factors. The joint effects of these nonlinearities and heterogeneities can be estimated using second-order
 382 Taylor expansions of the governing equations. Using Switzerland as a test case, we have shown that median
 383 aggregation biases of 10-35% are common, even at grid scales substantially smaller than those typically used in
 384 land surface models (Fig. 2). These biases can be much larger for individual days (Figs. S2 and S3) and
 385 potentially have substantial consequences for water and energy flux estimates in land surface models and
 386 consequently for temperature predictions in coupled models. The overestimated evaporative fluxes would
 387 lead to overestimated latent heat fluxes and underestimated sensible heat fluxes, and thus potentially to
 388 underestimates of expected temperature increases in a changing climate. Unrealistically high evaporation
 389 estimates lead to cooler modeled temperatures and wetter modeled climates. Correcting for the aggregation

390 bias in ET fluxes would lead to reduced evaporative cooling and increased atmospheric heating via sensible
391 heat flux.

392

393 In coupled Earth system models, ET fluxes influence how surface temperature, net radiation, and soil moisture
394 evolve through time, and thus influence future values of ET. The analyses shown in Figs. 2-4 are based on static
395 values for each day, and thus do not account for the propagation of aggregation biases forward through time.
396 Estimating the consequences of aggregation biases for dynamic modeling would require fully coupled Earth
397 system model simulations rather than the single ET algorithm analyzed here. In a dynamic model, the Taylor
398 expansion approach can potentially be used to correct for aggregation biases in each time step, using
399 statistical models for the variances and covariances of the ET drivers. Thus, estimating aggregation biases in a
400 dynamic model would not require explicitly simulating sub-grid heterogeneity at every time step. Correcting
401 for aggregation biases at each modeling time step would prevent them from propagating further into future
402 time steps, or into the partitioning of future water and energy fluxes at the land surface. The present paper
403 does not illustrate this dynamic correction for aggregation biases, but establishes the theoretical framework
404 for it.

405

406 The purpose of our analysis was to demonstrate how aggregation bias due to spatial heterogeneity can be
407 quantified (Sects. 2.2-2.3), how its dominant drivers can be identified (Sect. 2.4.1), and how its effects can be
408 efficiently corrected for, using sub-grid closure relationships (Sect. 2.4.2). For this demonstration, we chose
409 GLEAM as an illustrative example, and Switzerland as a topographically complex case study where high-
410 resolution data on the ET drivers are available. Applications of this approach to more complex land surface
411 models may require calculating the necessary derivatives (see Eq. 7) numerically rather than analytically, and
412 applications where high-resolution data are unavailable may require statistically estimating the variances and
413 covariances among the drivers of ET, based on their relationships with topography, soil types, land cover, etc.
414 Using the approach outlined here, one can account for the effects of sub-grid heterogeneity without explicitly
415 modeling ET at fine spatial resolution, which could be impractical due to computational costs, or impossible
416 due to a lack of fine-resolution input data.

417

418 In our analysis, spatial heterogeneity in soil moisture emerged as the dominant driver of aggregation bias in ET
419 estimates. Particularly if this result can also be confirmed in other regions and climates, it points to the
420 importance of improving our understanding of spatial patterns of soil moisture and what controls them. The
421 lower topographic curvature of coarsely gridded landscapes can lead models to predict higher soil moisture at
422 coarser grid scales (Kuo et al., 1999); higher soil moisture at larger grid scales would lead to even higher
423 modeled values of ET, beyond the effects of the aggregation biases analyzed here. Soil moisture may also be
424 substantially influenced by lateral subsurface transfers of water, which are ignored in our analysis and are also

425 ignored by many land surface models. Overlooking lateral transfers could potentially bias ET estimates in large-
426 scale land surface models (Fan et al., 2019), but this is beyond the scope of the present study.

427

428

429 **Acknowledgements**

430 We thank Prof. Ying Fan Reinfelder for numerous insightful discussions and for helpful comments on the
431 manuscript. E.R.F. acknowledges support from the Swiss National Science Foundation (SNSF) under Grant No.
432 P2EZP2_162279.

433 **Data Availability Statement**

434 We will upload the source data for this study to a FAIR repository and provide the URL with the final version of
435 the paper.

436

437

438 **References**

- 439 Beven, K. J., and H. L. Cloke: Comment on “Hyperresolution global land surface modeling: Meeting a grand
440 challenge for monitoring Earth's terrestrial water” by Eric F. Wood et al., *Water Resour. Res.*, 48, W01801,
441 <https://doi.org/10.1029/2011WR010982>, 2012.
- 442 Beven, K. J.: The holy grail of scientific hydrology: $Q_t=H(SR)A$ as closure, *Hydrol. Earth Syst. Sci.*, 10, 609–618,
443 <https://doi.org/10.5194/hess-10-609-2006>, 2006.
- 444 Boone, A., and O. J. Wetzel: A simple scheme for modeling sub-grid soil texture variability for use in an
445 atmospheric climate model, *Journal of the Meteorological Society of Japan*, 77(1), 317–333,
446 https://doi.org/10.2151/jmsj1965.77.1B_317, 1998.
- 447 Brunner, M. I., K. Liechti, and M. Zappa: Extremeness of recent drought events in Switzerland: dependence on
448 variable and return period choice, *Natural Hazards and Earth System Science*, 19(10), 2311–2323,
449 <https://doi.org/10.5194/nhess-19-2311-2019>, 2019.
- 450 Brunt, D.: *Physical and dynamical meteorology*, 2nd ed., Univ. Press, Cambridge. 428 pp, 1952.
- 451 Brutsaert, W.: *Evaporation into the atmosphere*, ISBN 978-90-481-8365-4, [https://doi.org/10.1007/978-94-](https://doi.org/10.1007/978-94-017-1497-6)
452 [017-1497-6](https://doi.org/10.1007/978-94-017-1497-6), 1984.
- 453 BFS, *Die Bodennutzung der Schweiz: Arealstatistik 1979/85*, Bundesamt fuer Statistik, Bern, 1995.
- 454 Budyko, M. I.: *Climate and life*, Academic, New York, 1974.
- 455 Bundesamt für Landestopographie: *Digitales Höhenmodell RIMINI*, Wabern,
456 [https://www.bfs.admin.ch/bfs/de/home/dienstleistungen/geostat/geodaten-](https://www.bfs.admin.ch/bfs/de/home/dienstleistungen/geostat/geodaten-bundesstatistik/topografie.assetdetail.230215.html)
457 [bundesstatistik/topografie.assetdetail.230215.html](https://www.bfs.admin.ch/bfs/de/home/dienstleistungen/geostat/geodaten-bundesstatistik/topografie.assetdetail.230215.html), 1991.
- 458 El Maayar, M. , and J. M. Chen: Spatial scaling of evapotranspiration as affected by heterogeneities in
459 vegetation, topography, and soil texture, *Remote Sensing of Environment*, 102, 33–51,
460 <https://doi.org/10.1016/j.rse.2006.01.017>, 2006.
- 461 Ershadi A., M. F. McCabe, J. P. Evans, and J. P. Walker: Effects of spatial aggregation on the multi-scale
462 estimation of evapotranspiration, *Remote Sensing of Environment*, 131, 51–62,
463 <http://dx.doi.org/10.1016/j.rse.2012.12.007>, 2013.
- 464 Fan, Y., M. Clark, D. M. Lawrence, S. Swenson, L. E. Band, S. L. Brantley, P. D. Brooks, W. E. Dietrich, A. Flores,
465 G. Grant, J. W. Kirchner, D. S. Mackay, J. J. McDonnell, P. C. D. Milly, P. L. Sullivan, C. Tague, H. Ajami, N.
466 Chaney, A. Hartmann, P. Hazenberg, J. McNamara, J. Pelletier, J. Perket, E. Rouholahnejad-Freund, T. Wagener,
467 X. Zeng, E. Beighley, J. Buzan, M. Huang, B. Livneh, B. P. Mohanty, B. Nijssen, M. Safeeq, C. Shen, W. van
468 Verseveld, and J. Volk, D. Yamazaki: Hillslope hydrology in global change research and Earth system modeling,
469 *Water Resources Research*, 55, <https://doi.org/10.1029/2018WR023903>, 2019.
- 470 Gash, J. H. C.: an analytical model of rainfall interception by forests, *Q. J. R. Meteorol. Soc.* 105 (433), 43–55,
471 <https://doi.org/10.1002/qj.49710544304>, 1979.

472 Giorgi, F.: An Approach for the Representation of Surface Heterogeneity in Land Surface Models. Part I:
473 Theoretical Framework, *Mon. Wea. Rev.*, 125, 1885–1899, <https://doi.org/10.1175/1520->
474 [0493\(1997\)125<1885:AAFTRO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1885:AAFTRO>2.0.CO;2), 1997.

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

477 Greve, P., B. Orlowsky, B. Mueller, J. Sheffield, M. Reichstein, and S. I. Seneviratne: Global assessment of
478 trends in wetting and drying over land, *Nature Geoscience*, 7: 716, 2014.

479 Hong, S. H., J. M. H. Hendrickx, and B. Borchers: Up-scaling of SEBAL derived evapotranspiration maps from
480 Landsat (30 m) to MODIS (250 m) scale, *Journal of Hydrology*, 370, 122–138,
481 <https://doi.org/10.1016/j.jhydrol.2009.03.002>, 2009.

482 Huang, X., A. M. Rhoades, P. A. Ullrich, and C. M. Zarzycki: An evaluation of the variable-resolution- CESM for
483 modeling California’s climate, *J. Adv. Model. Earth Syst.*, 8, 345–369, doi:10.1002/2015MS000559, 2016.

484 Jasechko, S., Z. D. Sharp, J. J. Gibson, S. J. Birks, Y. Yi, and P. J. Fawcett: Terrestrial water fluxes dominated by
485 transpiration, *Nature*, 496: 347, <https://doi.org/10.1890/ES13-00391.1>, 2013.

486 Kuo, W. L., T. S. Steenhuis, C. E. McCulloch, C. L. Mohler, D. A. Weinstein, S. D. DeGloria, and D. P. Swaney:
487 Effect of grid size on runoff and soil moisture for a variable-source-area hydrology model, *Water Resour. Res.*,
488 35(11), 3419–3428, <https://doi.org/10.1029/1999WR900183>, 1999.

489 Martens, B., D. G. Miralles, H. Lievens, R. van der Schalie, R. A. M. de Jeu, D. Fernández-Prieto, H. E. Beck, W. A.
490 Dorigo, and N. E. C. Verhoest: GLEAM v3: satellite-based land evaporation and root-zone soil moisture, *Geosci.*
491 *Model Dev.* 10(5): 1903–1925, <https://doi.org/10.5194/gmd-10-1903-2017>, 2017.

492 McCabe M., and E. Wood: Scale influences on the remote estimation of evapotranspiration using multiple
493 satellite sensors, *Remote Sensing of Environment* 105, 271–285, <https://doi.org/10.1016/j.rse.2006.07.006>,
494 2006.

495 Miralles, D. G., T. R. H. Holmes, R. A. M. De Jeu, J. H. Gash, A. G. C. A. Meesters, and A. J. Dolman: Global land-
496 surface evaporation estimated from satellite-based observations, *Hydrol. Earth Syst. Sci.* 15(2): 453–469,
497 <https://doi.org/10.5194/hess-15-453-2011>, 2011.

498 Miralles, D. G., A. J. Teuling, C. C. van Heerwaarden, and J. Vilà-Guerau de Arellano: Mega-heatwave
499 temperatures due to combined soil desiccation and atmospheric heat accumulation, *Nature Geosci.* 7(5):
500 345–349, <https://doi.org/10.1038/NGEO2141>, 2014.

501 Miralles, D. G., M. J. van den Berg, J. H. Gash, R. M. Parinussa, R. A. M. de Jeu, H. E. Beck, T. R. H. Holmes, C.
502 Jiménez, N. E. C. Verhoest, W. A. Dorigo, A. J. Teuling, and A. Johannes Dolman: El Niño–La Niña cycle and
503 recent trends in continental evaporation, *Nature Climate Change*, 4: 122,
504 <https://doi.org/10.1038/NCLIMATE2068>, 2013.

505 Monteith, J. L., and M. H. Unsworth: *Principles of Environmental Physics*, Edward Arnold, London, 1990.

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

508 Mueller, B., M. Hirschi, C. Jimenez, P. Ciais, P. A. Dirmeyer, A. J. Dolman, J. B. Fisher, M. Jung, F. Ludwig, F.
509 Maignan, D. G. Miralles, M. F. McCabe, M. Reichstein, J. Sheffield, K. Wang, E. F. Wood, Y. Zhang, and S. I.
510 Seneviratne: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, *Hydrol.*
511 *Earth Syst. Sci.* 17(10): 3707–3720, <https://doi.org/10.5194/hess-17-3707-2013>, 2013.

512 Murray, F. W.: On the computation of saturation vapor pressure, *J. Appl. Meteor.* 6: 203–204,
513 [https://doi.org/10.1175/1520-0450\(1967\)006<0203:OTCOSV>2.0.CO;2](https://doi.org/10.1175/1520-0450(1967)006<0203:OTCOSV>2.0.CO;2), 1967.

514 Orth, R., M. Staudinger, S. I. Seneviratne, J. Seibert, and M. Zappa: Does model performance improve with
515 complexity? A case study with three hydrological models, *Journal of Hydrology*, 523, 147–159,
516 <https://doi.org/10.1016/j.jhydrol.2015.01.044>, 2015.

517 Penman, H. L.: Natural evaporation from open water, bare soil, and grass, *Proc. Roy. Soc. London A*193,
518 120–146, 1948.

519 Priestley, C. H. B., and R. J. Taylor: On the assessment of surface heat flux and evaporation using large-scale
520 parameters, *Monthly Weather Review*, 100, 81–92, [https://doi.org/10.1175/1520-0493\(1972\)100<0081:otaosh>2.3.co;2](https://doi.org/10.1175/1520-0493(1972)100<0081:otaosh>2.3.co;2), 1972.

522 Rauscher, S. A., E. Coppola, C. Piani, and F. Giorgi: Resolution effects on regional climate model simulations of
523 seasonal precipitation over Europe, *Clim. Dyn.*, 35(4), 685–711, <https://doi.org/10.1007/s00382-009-0607-7>,
524 2010.

525 Ringler, T., L. Ju, and M. Gunzburger: A multiresolution method for climate system modeling: Application of
526 spherical centroidal Voronoi tessellations, *Ocean Dyn.*, 58(5–6), 475–498, <https://doi.org/10.1007/s10236-008-0157-2>, 2008.

528 Rouholahnejad Freund, E., and J. W. Kirchner: A Budyko framework for estimating how spatial heterogeneity
529 and lateral moisture redistribution affect average evapotranspiration rates as seen from the atmosphere,
530 *Hydrology and Earth System Sciences*, 21(1), 217–233, <https://doi.org/10.5194/hess-21-217-2017>, 2017.

531 Rouholahnejad Freund, E., Y. Fan, and J. W. Kirchner: Global assessment of how averaging over spatial
532 heterogeneity in precipitation and potential evapotranspiration affects modeled evapotranspiration rates,
533 *Hydrol. Earth Syst. Sci.*, 24, 1927–1938, <https://doi.org/10.5194/hess-24-1927-2020>, 2020.

534 Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B., Orlowsky, and A. J. Teuling:
535 Investigating soil moisture–climate interactions in a changing climate: A review, *Earth-Science Reviews* 99(3–
536 4): 125–161, <https://doi.org/10.1016/j.earscirev.2010.02.004>, 2010.

537 Shrestha, P., M. Sulis, C. Simmer, and S. Kollet: Impacts of grid resolution on surface energy fluxes simulated
538 with an integrated surface-groundwater flow model, *Hydrol. Earth Syst. Sci.* 19(10): 4317–4326,
539 <https://doi.org/10.5194/hess-19-4317-2015>, 2015.

540 Shrestha, P., M. Sulis, C. Simmer, and S. Kollet: Effects of horizontal grid resolution on evapotranspiration
541 partitioning using TerrSysMP, *Journal of Hydrology*, 557: 910–915,
542 <https://doi.org/10.1016/j.jhydrol.2018.01.024>, 2018.

543 Skamarock, W. C., J. B. Klemp, M. G. Duda, L. D. Fowler, S.-H. Park, and T. D. Ringler, A multiscale
544 nonhydrostatic atmospheric model using centroidal Voronoi tessellations and C-grid staggering, *Mon. Weather*
545 *Rev.*, 140(9), 3090–3105, <https://doi.org/10.1175/MWR-D-11-00215.1>, 2012.

546 Speich, M. J. R., L. Bernhard, A. J. Teuling, and M. Zappa: Application of bivariate mapping for hydrological
547 classification and analysis of temporal change and scale effects in Switzerland, *Journal of Hydrology*, 523, 804–
548 821, <https://doi.org/10.1016/j.jhydrol.2015.01.086>, 2015.

549 Stanghellini, C., *Transpiration of Greenhouse Crops*. PhD thesis, Wageningen University, Wageningen, The
550 Netherlands, 1987.

551 Tetens, O., Über einige meteorologische Begriffe. *z. Geophys.* 6:297–309, 1930.

552 Viviroli, D., M. Zappa, J. Gurtz, and R. Weingartner: An introduction to the hydrological modelling system
553 PREVAH and its pre- and post-processing-tools, *Environmental Modelling and Software*, 24(10), 1209–1222,
554 <https://doi.org/10.1016/j.envsoft.2009.04.001>, 2009.

555 Zappa, M. and J. Gurtz: Simulation of soil moisture and evapotranspiration in a soil profile during the 1999
556 MAP-Riviera Campaign, *Hydrol. Earth Syst. Sci.*, 7, 903–919, <https://doi.org/10.5194/hess-7-903-2003>, 2003.

557 Zarzycki, C. M., M. N. Levy, C. Jablonowski, J. R. Overfelt, M. A. Taylor, and P. A. Ullrich: Aquaplanet
558 experiments using CAM’s variable-resolution dynamical core, *J. Clim.*, 27(14), 5481–5503,
559 <https://doi.org/10.1175/JCLI-D-14-00004.1>, 2014.

560

561

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Supporting Information for

Averaging over spatiotemporal heterogeneity substantially biases evapotranspiration rates in a mechanistic large-scale land evaporation model

Elham Rouholahnejad Freund^{1,2,3}, Massimiliano Zappa⁴, James W. Kirchner^{3,4,5}

¹Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium

²Chair of Hydrology, Faculty of Environment and Natural Resources, University of Freiburg, Freiburg, Germany

³Department of Environmental Systems Science, ETH Zurich, CH-8092 Zürich, Switzerland

⁴Swiss Federal Research Institute WSL, CH-8903 Birmensdorf, Switzerland

⁵Department of Earth and Planetary Science, University of California, Berkeley, CA 94720 USA

Content of this file

- Table S1, PREVAH hydrological and meteorological data
- Figure S1, Land cover map of Switzerland with corresponding albedo values
- Figures S2-S6
- Data availability

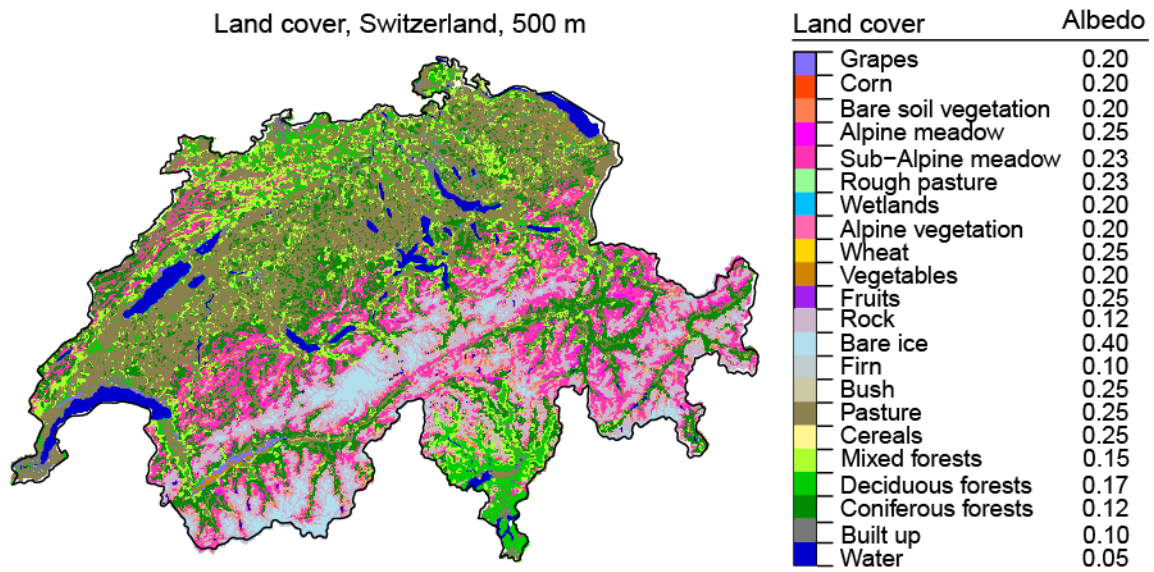
18 Table S1. *PREVAH* hydrological and meteorological data. All data are in gridded format and at 500 m spatial
19 resolution and (if relevant) daily temporal resolution

Data	Source
<i>PREVAH</i> soil moisture (m ³ /m ³)	Simulations from <i>PREVAH</i> hydrological model, Brunner et al., 2019 ; Speich et al., 2015; Orth et al., 2015; Zappa et al., 2003
precipitation (mm d-1)	Interpolation of MeteoSwiss data after Viviroli et al., 2009
radiation (W m-2)	Interpolation of MeteoSwiss data after Viviroli et al., 2009
relative humidity (-)	Interpolation of MeteoSwiss data after Viviroli et al., 2009
sun duration (hr)	Interpolation of MeteoSwiss data after Viviroli et al., 2009
temperature (°C)	Interpolation of MeteoSwiss data after Viviroli et al., 2009
vapor pressure (Pa)	Interpolation of MeteoSwiss data after Viviroli et al., 2009
CH land use	BFS, Swiss Federal Statistical Office, 1995

20

21

22



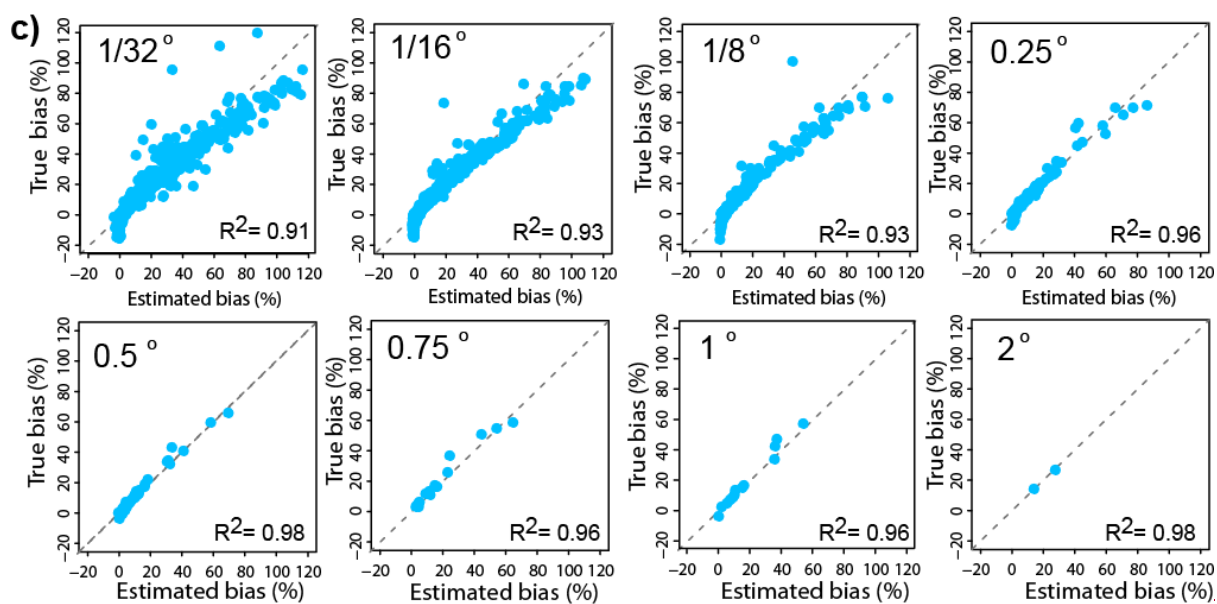
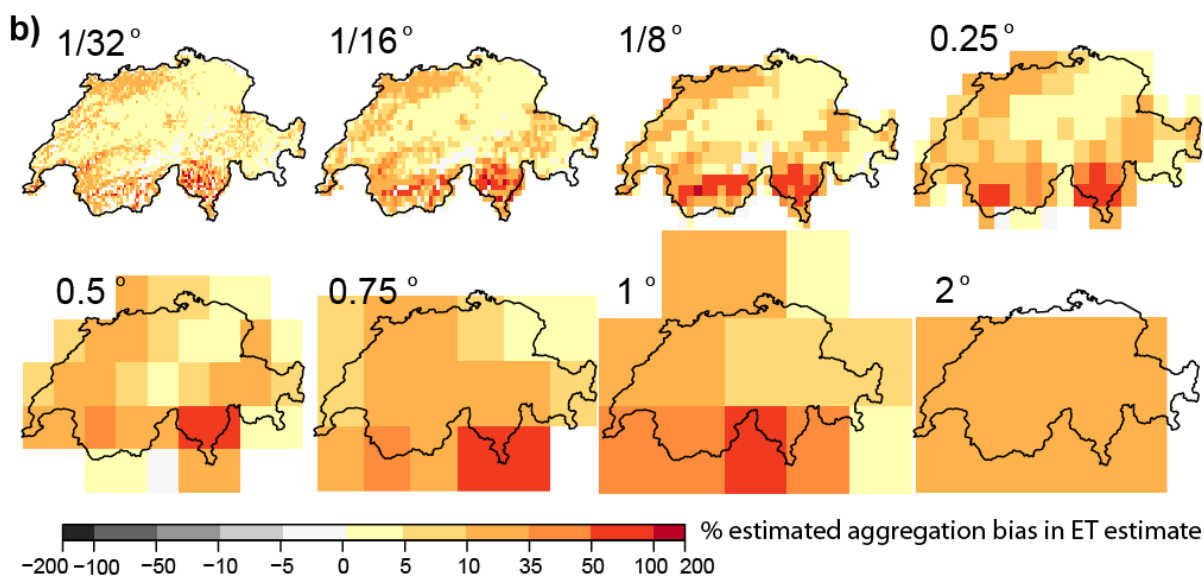
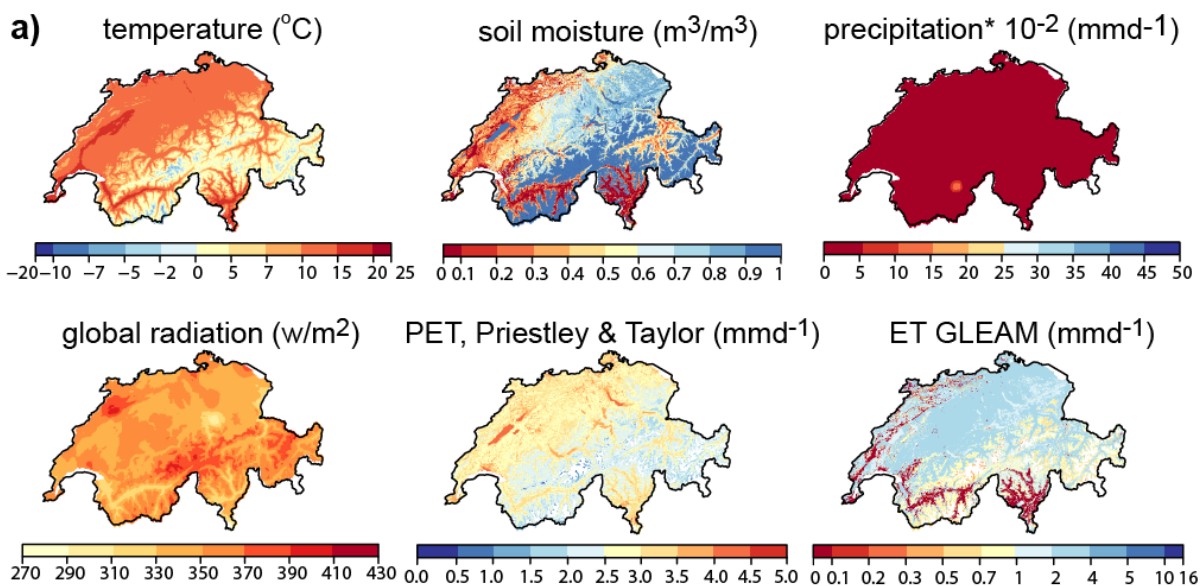
23

24 Figure S1. Land cover map of Switzerland at 500-meter resolution along with the albedo values associated with
 25 each land cover type (BFS, 1995; Viviroli et al., 2009)

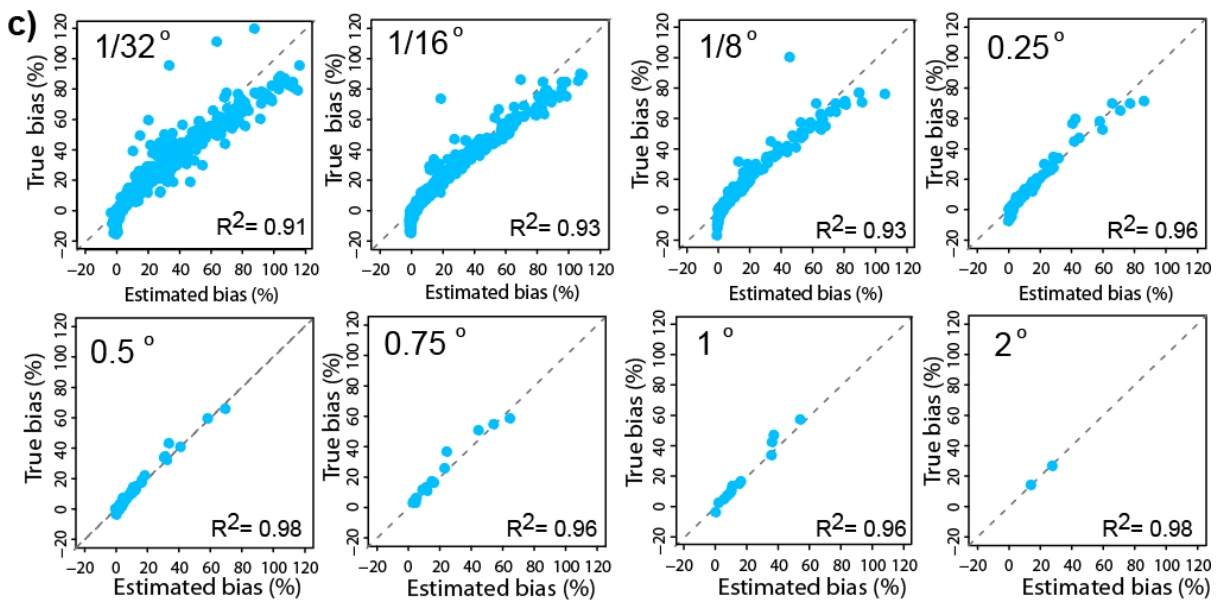
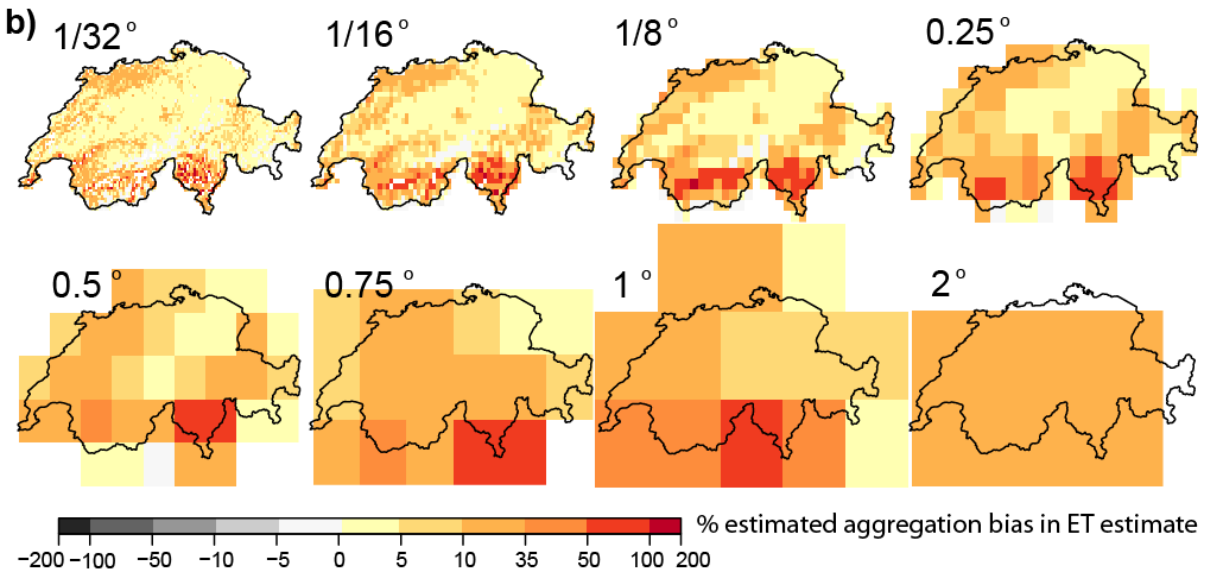
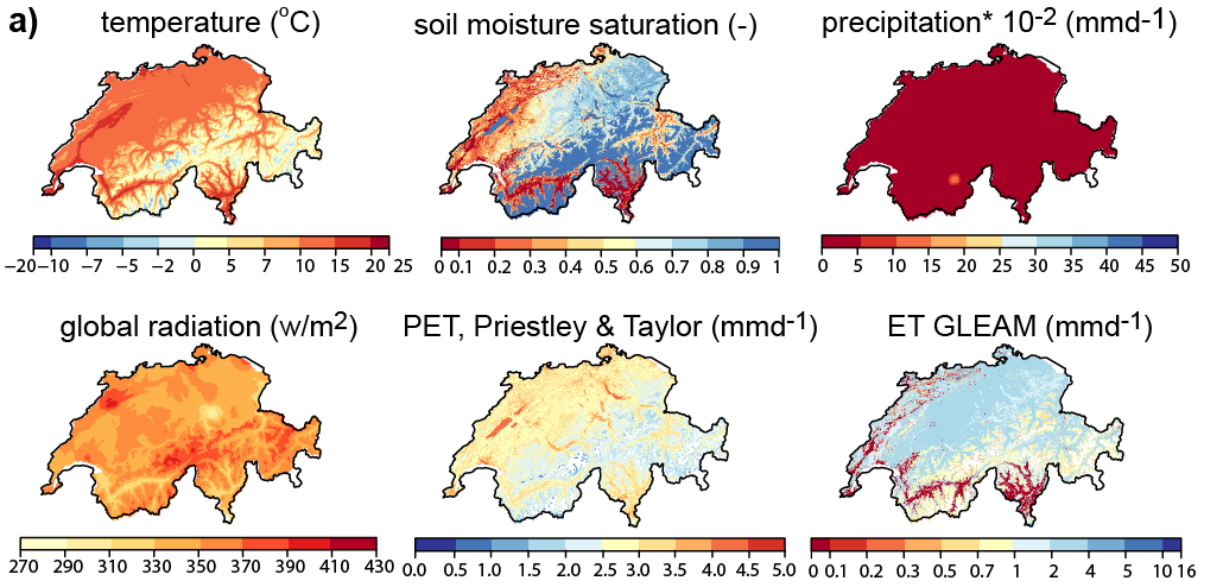
26

27

May 31st, 2004

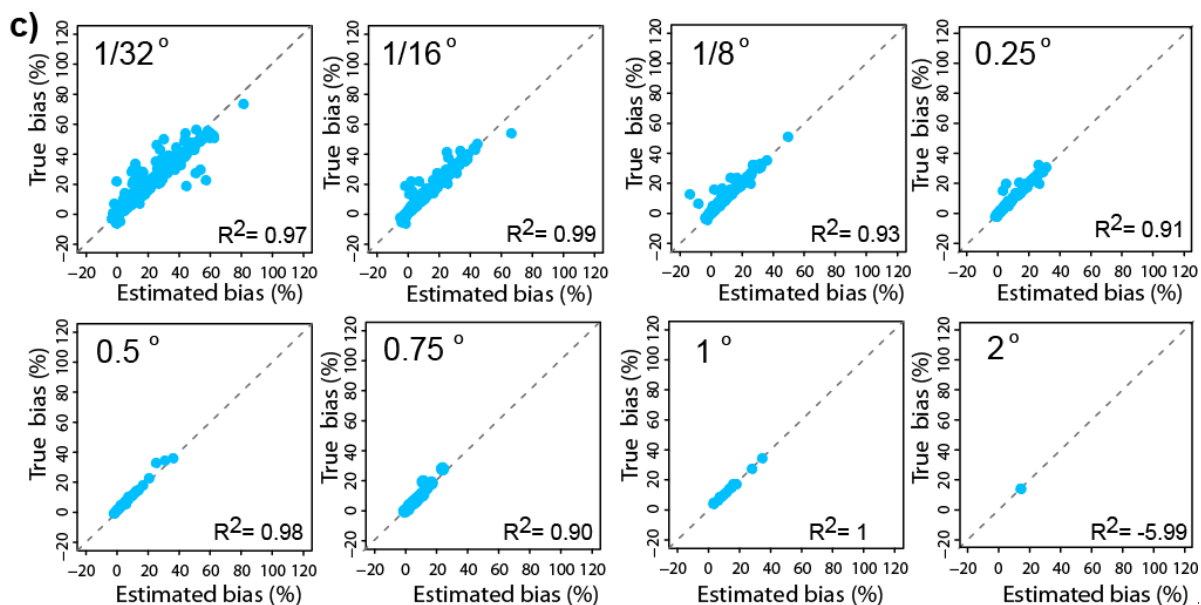
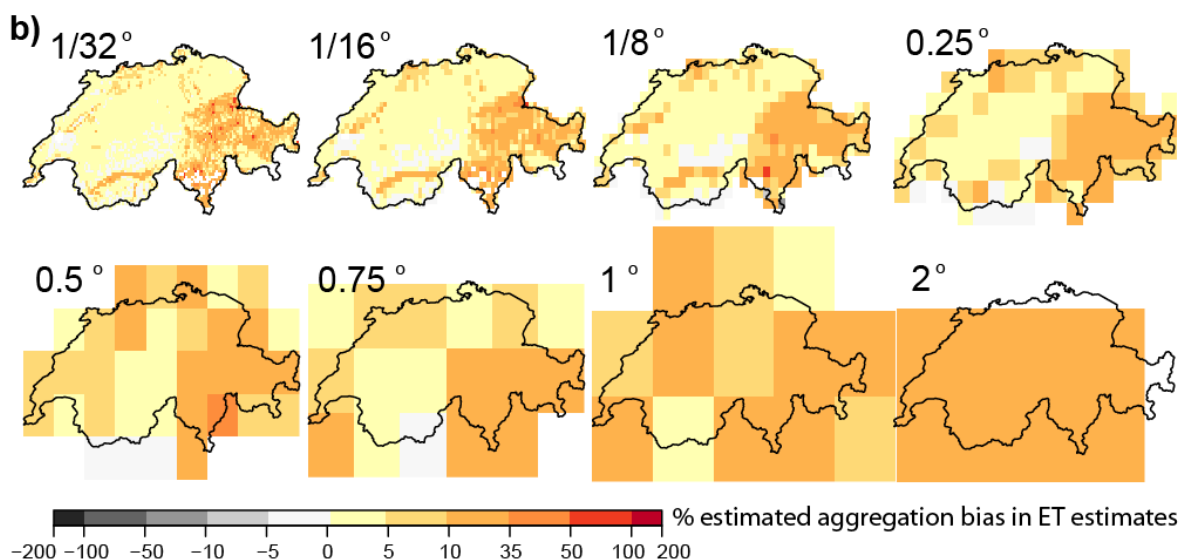
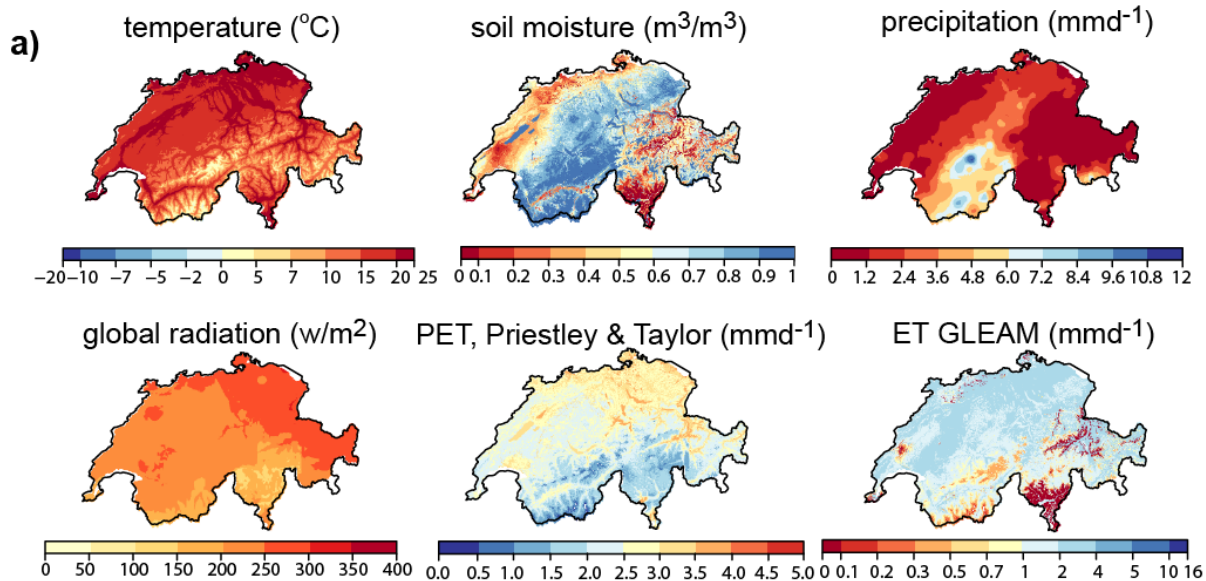


May 29th, 2004

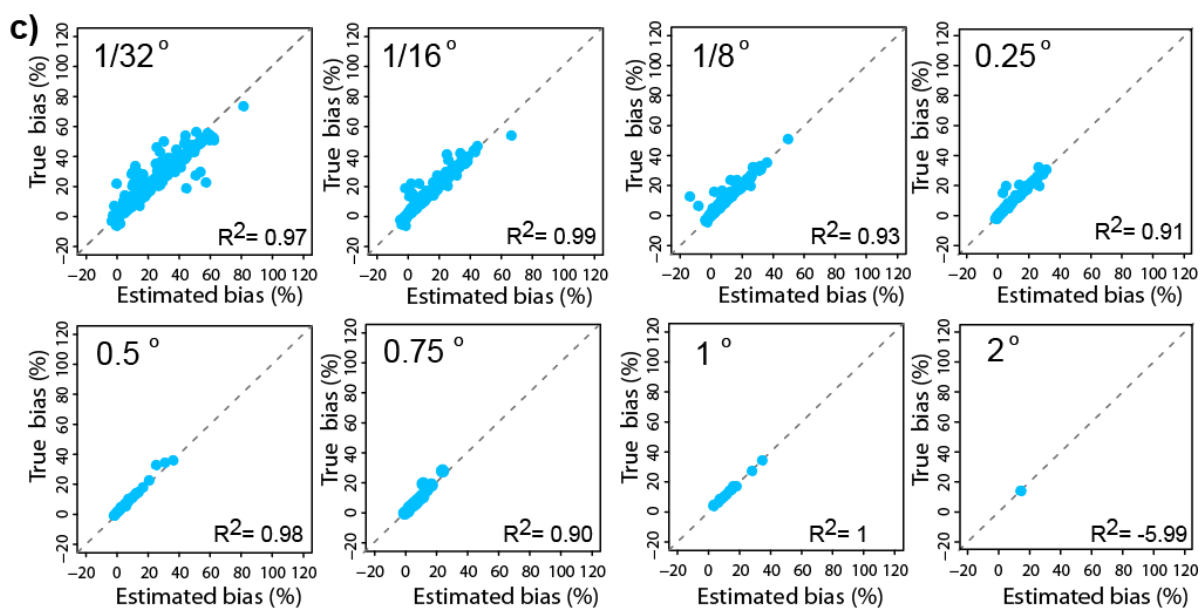
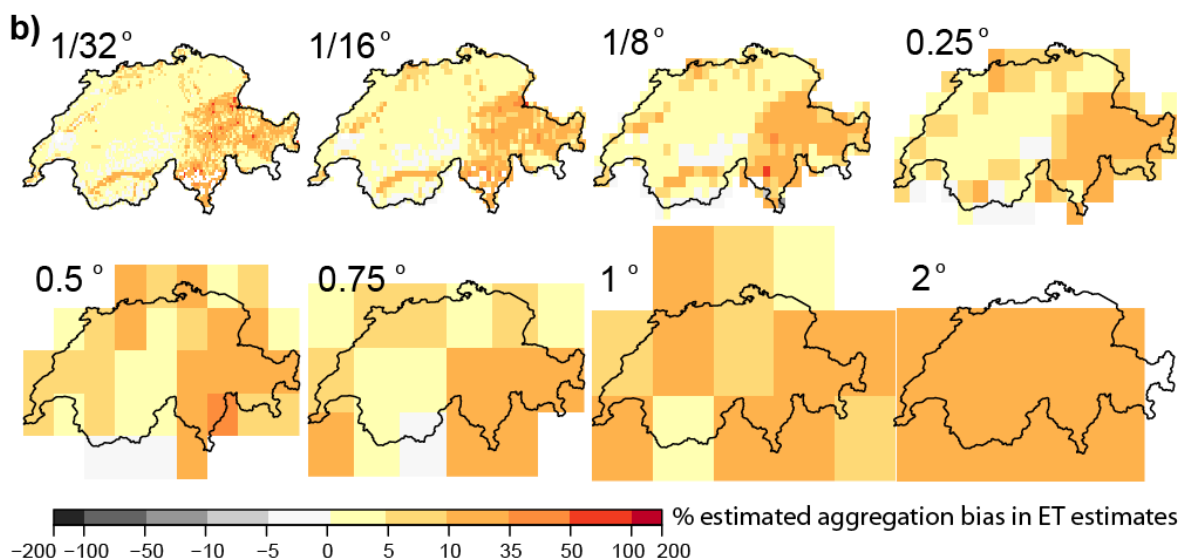
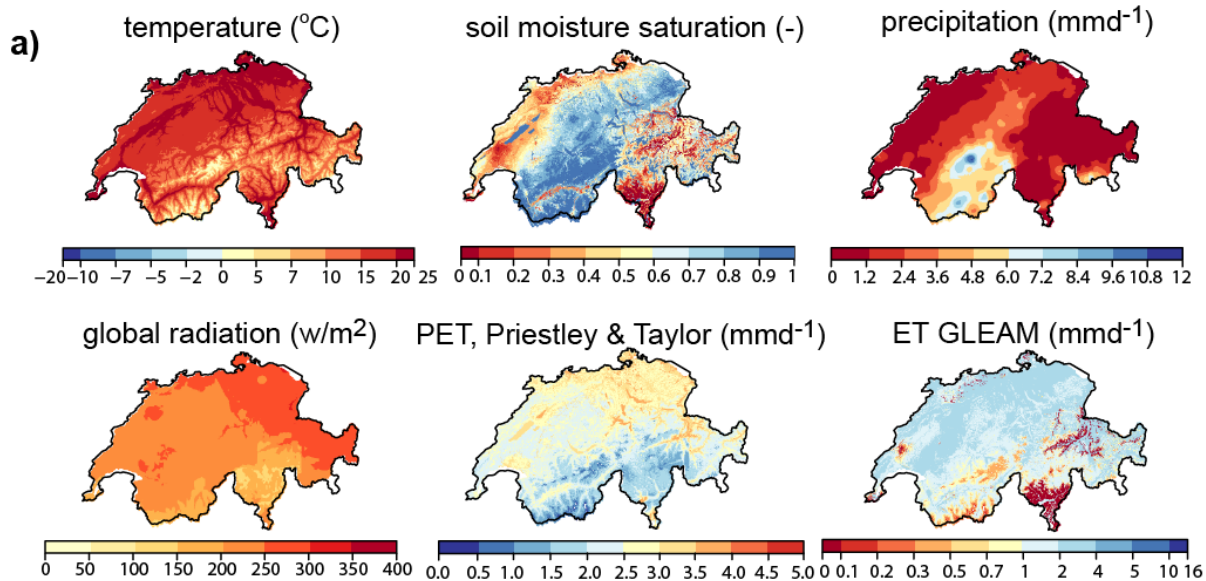


30 Figure S2. a) Spatial distribution of input data at 500 m resolution for a ~~arbitrarily~~randomly selected day
31 (~~31~~29.05.2004) to calculate ET. Potential evapotranspiration (PET, mmyr^{-1}) is calculated using the Priestley-
32 Taylor equation (Eq. 3), and evapotranspiration (ET, mmyr^{-1}) is calculated using the approach used in the
33 GLEAM model (Miralles et al., 2011; Martens et al., 2017; Eq. 1). b) Aggregation bias estimated from 500 m
34 temperature ($^{\circ}\text{C}$), soil moisture (w_w), net radiation (R_n), their variances at each grid scale, and the covariances
35 of all the pairs using Eq. 7. Even at the finest resolutions (1/32 and 1/16 degrees) the aggregation bias rises to
36 50-100 % overestimation in daily ET estimates in South Switzerland. c) Daily approximated aggregation bias in
37 ET estimates versus daily true aggregation bias in ET estimates at several spatial scales for ~~31~~29.05.2004.
38 Approximated aggregation bias is calculated using Eq.7. The true bias is the difference between the finer-
39 resolution ET estimates from finer-resolution input data, averaged over several spatial scales (average of
40 functions) and average ET estimated from average inputs at each spatial scale (function of averages). The
41 coefficient of determination (R^2) between the true and approximated aggregation bias confirms the
42 appropriateness of the proposed method and Eq. 7 for approximating the aggregation bias.

July 21st, 2004



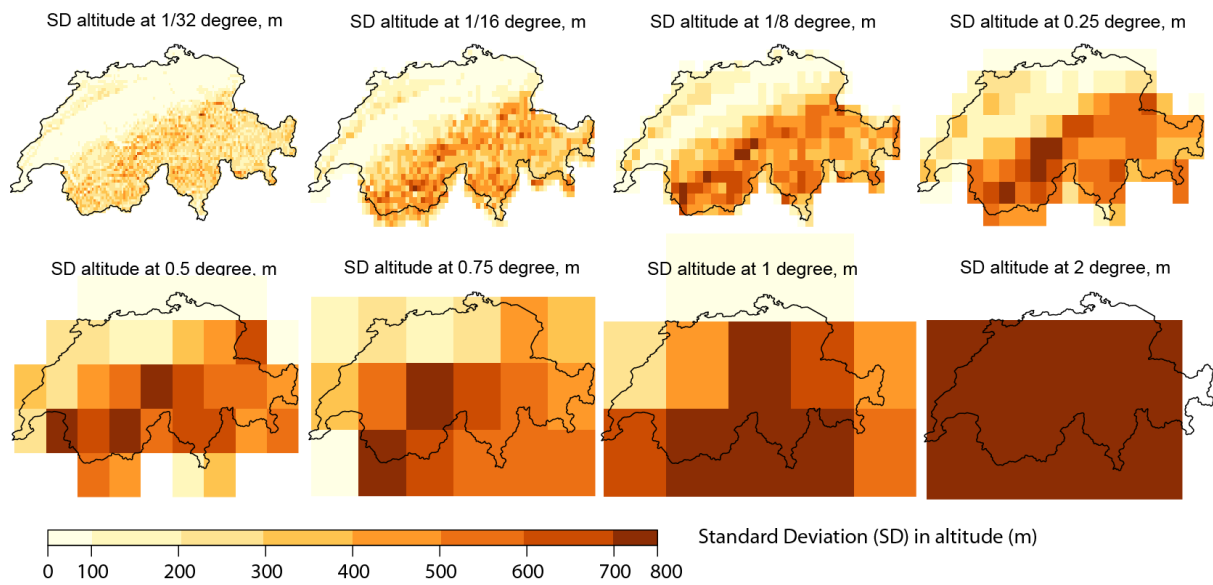
July 18th, 2004



45 Figure S3. Same as in Fig. S2 but for another ~~arbitrarily~~randomly selected day (~~21~~18.07.2004).

46

47

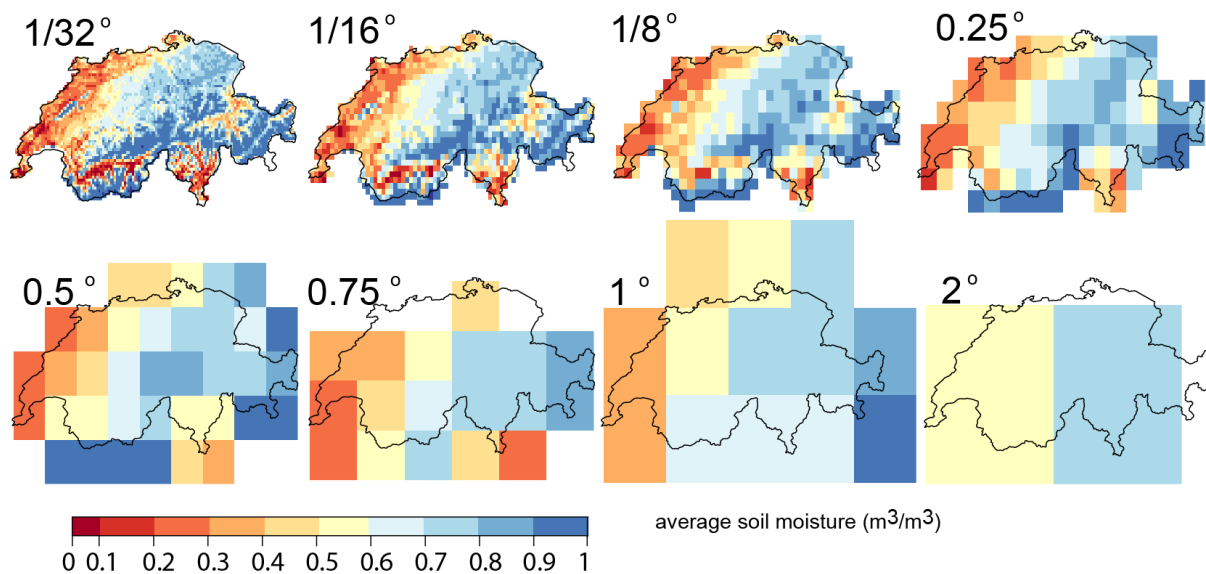


48

49 Figure S4. Standard deviation of altitude across several grid scales calculated from 500 m resolution
 50 topographic data (Bundesamt für Landestopographie, 1991). The spatial patterns of topographic variability at
 51 each grid scale are similar to spatial patterns of the median of daily aggregation biases shown in Fig 2 of the
 52 manuscript.

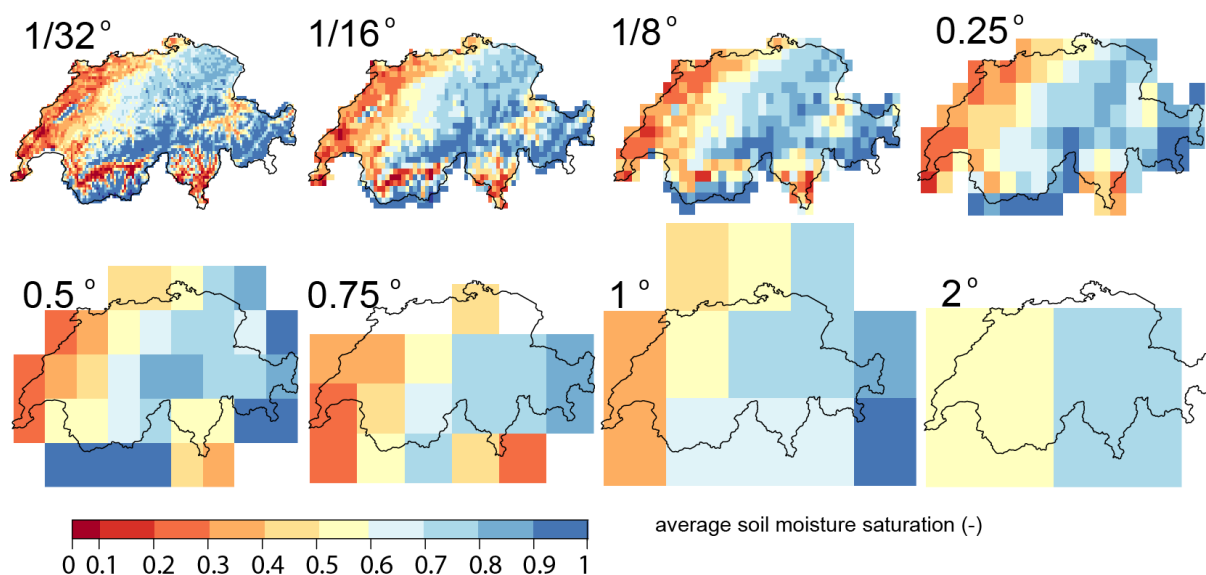
53

May 31st, 2004



54

May 29th, 2004

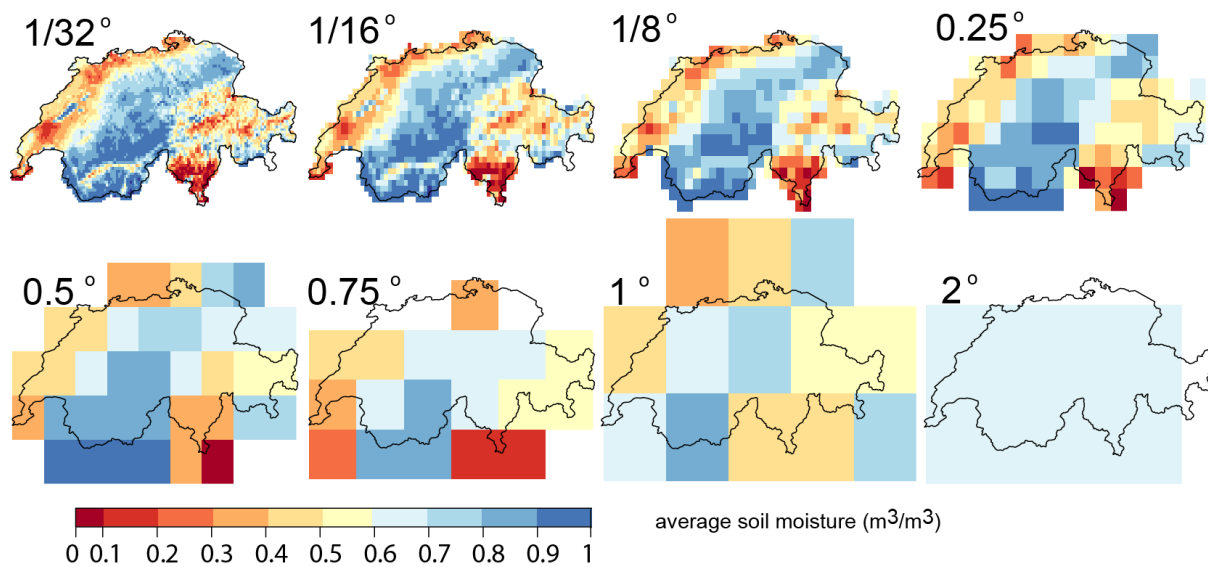


55

56 Figure S5. Soil moisture ($\text{m}^3 \text{m}^{-3}$) averaged over several grid scales calculated from 500 m resolution soil
57 moisture data ($\text{m}^3 \text{m}^{-3}$) for a arbitrarily randomly selected day (3129.05.2004). The 500 m resolution soil
58 moisture data are simulated by the PREVAH hydrological model (Brunner et al., 2019; Speich et al., 2015; Orth
59 et al., 2015; Zappa et al., 2003).

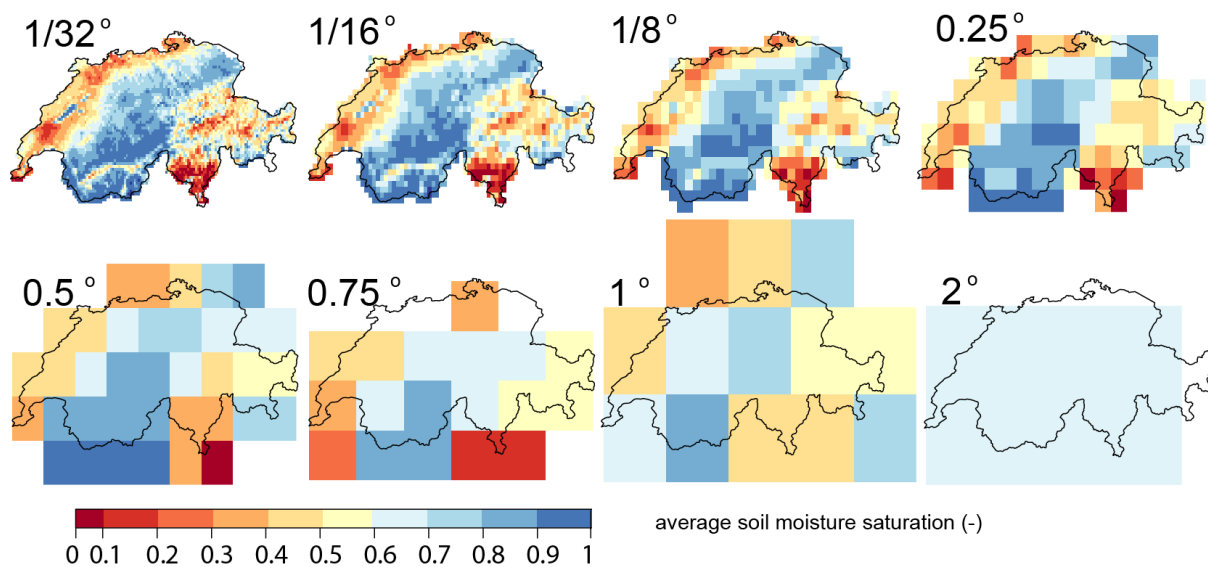
60

July 21st, 2004



61

July 18th, 2004



62

63 Figure S6. Same as in Fig. S5 but for another arbitrarilyrandomly selected day (2118.07.2004).

64

65 **Data availability**

66 We will upload the source data for this study to a FAIR repository and provide the URL with the final version of
67 the paper.

68

69