1 **Title: Extending the range of applicability of the semi-empirical ecosystem flux**
2 **model PRELES for varying forest types and climate**

3 **Running head:** Applicability of ecosystem flux models

4 Xianglin Tian^{1,2}, Francesco Minunno², Tianjian Cao¹, Mikko Peltoniemi³, Tuomo Kalliokoski^{2,4},
5 Annikki Mäkelä²

6 ¹College of Forestry, Northwest A&F University, Yangling, 712100, China

7 ²Department of Forest Sciences, University of Helsinki, P.O. Box 27, Helsinki FI-00014, Finland

8 ³Natural Resources Institute Finland, Latokartanonkaari 9, Helsinki FI-00790, Finland

9 ⁴Department of Physics, University of Helsinki, P.O. Box 64, Helsinki FI-00014, Finland

10 Correspondence to: Tianjian Cao (cao@nwafu.edu.cn)

11 **Keywords:** Light-use efficiency, plant functional type, gross primary production, evapotranspiration,
12 geographical variations, inverse modelling, multisite calibration.

13 **Paper type:** Primary Research Articles

14

15 **Abstract.** Applications of ecosystem flux models on large geographical scales are often limited by
16 model complexity and data availability. Here, we calibrated and evaluated a semi-empirical
17 ecosystem flux model, PRELES, for various forest types and climate conditions, based on eddy
18 covariance data from 55 sites. A Bayesian approach was adopted for model calibration and
19 uncertainty quantification. We applied the site-specific calibrations and multisite calibrations to nine
20 plant functional types (PFTs) to obtain the site-specific and PFT specific parameter vectors for
21 PRELES. A systematically designed cross-validation was implemented to evaluate calibration
22 strategies and the risks in extrapolation. The combination of plant physiological traits and climate
23 patterns generated significant variation in vegetation responses and model parameters across but not
24 within PFTs, implying that applying the model without PFT-specific parameters is risky. But within
25 PFT, the multisite calibrations performed as accurately as the site-specific calibrations in predicting
26 gross primary production (GPP) and evapotranspiration (ET). Moreover, the variations among sites
27 within one PFT could be effectively simulated by simply adjusting the parameter of potential light-
28 use efficiency (LUE), implying significant convergence of simulated vegetation processes within
29 PFT. The hierarchical modelling of PRELES provides a compromise between satellite-driven LUE
30 and physiologically oriented approaches for extrapolating the geographical variation of ecosystem
31 productivity. Although measurement errors of eddy covariance and remotely sensed data propagated
32 a substantial proportion of uncertainty or potential biases, the results illustrated that PRELES could
33 reliably capture daily variations of GPP and ET for contrasting forest types on large geographical
34 scales if PFT-specific parameterizations were applied.

35 **1 Introduction**

36 One of the major problems in the applications of ecosystem models and physiological models is the
37 level of complexity (Landsberg & Sands, 2011). Models concerning detailed physiological
38 mechanisms or ecosystem processes can theoretically be extrapolated to new sites or to future
39 climates, but appropriate input data and parameters are often difficult to obtain (Landsberg, 2003;

40 Mäkelä et al., 2000), despite the profound development of physiological measurement equipment
41 during last decades. Simplified models are less data-demanding with fewer parameters, but usually
42 extrapolate poorly and may overlook crucial interactions of the ecosystems (Monserud, 2003;
43 Weiskittel, Hann, Kershaw Jr, & Vanclay, 2011). Therefore, in applying models on a larger
44 geographical scale or under changing environmental conditions, it is always necessary to recalibrate
45 the models or test their applicability. Due to improved measurement techniques and automated data-
46 recording systems, numerous databases such as eddy flux, soil property and species distribution are
47 becoming available to fulfil the need for detailed information on stand characteristics or dynamics.
48 For instance, remotely sensed estimates such as canopy light interception, measured as the fraction
49 of absorbed photosynthetically active radiation (f_{APAR}), could determine the spatial variation of input
50 for ecosystem models (Waring, Coops, & Landsberg, 2010). Meanwhile, inverse modelling
51 approaches, such as Bayesian calibration (BC), adjust model parameters and processes according to
52 their ability to reproduce stand-level field observations, which bridges the gap between complex
53 models and various databases (e.g. Hartig et al., 2012; Van Oijen, Rougier, & Smith, 2005). By
54 combining these data and modelling approaches, it is possible to test or extend the applicable ranges
55 of ecosystem models that were originally developed for small-scale regions.

56 Gross primary production (GPP), the sum of the net photosynthesis by all photosynthetic
57 tissue measured at the ecosystem scale, is a key factor in the ecosystem carbon balance. It is the
58 original carbon source for all the forest ecosystem carbon fluxes. Measurements and simulations of
59 GPP help us to understand the development of forest ecosystem and its interactions with climate.
60 Benefited from the rapid development of the eddy covariance network during recent decades
61 (Aubinet, Vesala, & Papale, 2012), both empirical and semi-empirical canopy GPP models can be
62 calibrated and validated sufficiently. Empirical ecosystem flux models have been applied to explain
63 distinctions in productivity between sites or vegetation types (e.g. Falge et al., 2002). Furthermore,
64 satellite-driven LUE approaches have been frequently used for monitoring geographical variation of

ecosystem productivity (e.g. Potter et al., 1993; Sims et al., 2008; Yuan et al., 2007). The trade-off of model simplicity is that much of the ecosystem variation remains unexplained (Yuan et al. 2014; Zheng et al., 2018), although the data requirement of those models can be globally fulfilled. Semi-empirical canopy GPP models have commonly been used as a sub-module of process-based models, such as the photosynthesis modules of 3-PG (Landsberg & Waring, 1997), PnET-II (Aber & Federer, 1992), and FOREST-BGC (Running & Coughlan 1988). Instead of reducing the data requirement, those models rely on the adequacy of the underlying physiological assumptions, extending the applicability of the model to all stands where the physiological parameters can be evaluated.

PRELES (PREdict Light-use efficiency, Evapotranspiration and Soil water) is a semi-empirical ecosystem flux model that predicts daily gross primary production, evapotranspiration (ET) and soil water (Peltoniemi et al., 2015a). The model requires soil characteristics, daily f_{APAR} and meteorological observations as inputs. The GPP predictions are based on a reformulation of the light-use efficiency (LUE) model of Mäkelä et al. (2008). PRELES has been calibrated and validated in the boreal region mainly for coniferous forests (Minunno et al., 2016; Peltoniemi et al., 2012; Peltoniemi et al., 2015a). When national inventory and map data were available, PRELES predicted GPP estimates in Finland similar to those of the model JSBACH and MODIS GPP product (MOD17), although the input data sources differed (Peltoniemi et al., 2015b). Linked with downscaled global circulation model projections, PRELES has been used to predicted boreal forest productivity under climate change scenarios, and its parametric uncertainty is marginal when compared with other sources of uncertainty (Kalliokoski, Mäkelä, Fronzek, Minunno, & Peltoniemi, 2018). Furthermore, PRELES has been linked with a process-based carbon allocation model CROBAS (Mäkelä, 1997; Valentine & Mäkelä 2005) in simulating forest variables with a country-generic calibration in Finland (Minunno et al., 2019). Mäkelä et al. (2008) showed that daily temperature, vapour-pressure deficit (VPD) and absorbed PPFD (photosynthetic photon flux density) accounted for most of the daily variation in GPP in the model, but unexplained variation remained in the site-specific maximum LUE,

90 which correlated linearly with canopy nitrogen (Peltoniemi et al. 2012). When the model was fitted
91 to data, differences between sites could be explained by potential LUE, leaf area and environmental
92 conditions. For wider applications, the ability of the model to extrapolate to conditions outside the
93 original modelling sites must be evaluated. Minunno et al. (2016) tested the applicability of PRELES
94 for 10 boreal coniferous forests in Fennoscandia and obtained a generic vector of model parameters
95 by multisite calibration. Based on a comparison between site-specific and multisite calibration, the
96 generic parameter vector from multisite calibration can be reliably used at the regional scale for boreal
97 coniferous forests. However, in this multisite calibration, all the sites were coniferous forests and
98 shared the same parameters, thus omitting the differences in potential LUE by site fertility or species
99 range. Incorporating the processes of light saturation, temperature acclimation, VPD stress, and soil
100 water dynamics, PRELES is theoretically qualified for monitoring and predicting ecosystem
101 productivity of various forest-climate types, but this wide range of applicability has not been tested
102 in warmer climate types, broad-leaved forests or very fertile soils.

103 The objectives of the study were: 1) to test, with additional modules of seasonality and water
104 dynamics incorporated, whether the LUE approach could sufficiently explain geographical variations
105 of GPP and ET, with respect to contrasting environmental conditions and distinctive forest
106 ecosystems; 2) to propose a generic parameter vector for each plant functional type (PFT) and to
107 hierarchically quantify the differences among sites while fitting the model with pooled data and 3) to
108 quantify the uncertainty in extrapolating to conditions outside the original sites.

109 **2 Material and methods**

110 **2.1 Framework of PRELES**

111 PRELES is a semi-empirical ecosystem flux model that predicts daily GPP (P , $\text{g C m}^{-2} \text{ d}^{-1}$), ET (E ,
112 mm d^{-1}) and soil water (mm). The requirements of site-specific inputs include the soil depth exploited
113 by the roots (mm), field capacity (mm) and wilting point (mm) of the soil, f_{APAR} , and daily
114 meteorological observations that include the PPFD ($\text{mol m}^{-2} \text{ d}^{-1}$) above the canopy, air temperature

115 (°C), VPD (kPa) and precipitation (mm d⁻¹). A detailed description of PRELES can be found in
 116 Peltoniemi et al. (2015a), and the code applied in this study is provided in the GitHub repository
 117 (<https://github.com/checcomi/Rprebas>). Here, we introduce a brief framework of the P and E
 118 submodels. Daily photosynthetic production during day k , P_k , is predicted as follows:

$$119 \quad P_k = \beta \cdot \phi_k \cdot f_{APAR,k} \cdot f_{L,k} \cdot f_{S,k} \cdot \min(f_{D,k}, f_{W,P,k}) \cdot f_{CO_2,P,k} \quad (1)$$

120 where β is the potential LUE (g C mol⁻¹), ϕ_k the PPFD (mol m⁻² d⁻¹) and $f_{APAR,k}$ the fraction of ϕ_k
 121 absorbed by the canopy during day k . The $f_{L,k}$, $f_{S,k}$, $f_{D,k}$ and $f_{W,P,k}$, constrained between 0 and 1, are
 122 respectively the modifiers that account for the suboptimal conditions in light L, temperature
 123 acclimation S, VPD ($f_{D,k}$), and soil-water stress W ($f_{W,P,k}$). The explanations of these modifiers can
 124 be found in Mäkelä et al. (2008). The modifier $f_{CO_2,P,k}$ accounts for the impact of ambient CO₂
 125 concentration on photosynthesis (for details see Kalliokoski et al. 2018). The daily ET during day k ,
 126 E_k , is simulated as follows:

$$127 \quad E_k = \alpha \cdot P_k \cdot f_{W,P,k}^v \cdot D_k^{1-\lambda} + \chi \cdot (1 - f_{APAR,k}) \cdot \phi_k \cdot f_{W,E,k} \cdot \frac{s_{DS,k}}{s_{DS,k} + p_{psychrom}} \quad (2)$$

128 where α is a transpiration parameter, χ an evaporation parameter and λ an adjustment parameter for
 129 the effect of D on transpiration during day k . The $f_{W,P,k}$ is raised to the power v , since the response
 130 of E_k to soil-water stress may differ from that of P_k . Another modifier, $f_{CO_2,E,k}$, was adopted to
 131 replace $f_{CO_2,P,k}$ in Eq. (1) when calculating the impact of the CO₂ concentration on transpiration
 132 (Kalliokoski et al. 2018). The modifier $f_{W,E,k}$ accounts for the suboptimal condition of evaporation
 133 due to soil water, while $s_{DS,k}$ is the slope of the relationship between the saturation vapour pressure
 134 (kPa) and air temperature (°C), and $p_{psychrom}$ is the psychrometric constant (Campbell, 1977) that
 135 relates the partial pressure of water in air to the air temperature (kPa °C⁻¹).

136 PRELES has 20 parameters (Table 1) and only two state variables (Peltoniemi et al., 2015a).
 137 One state variable is soil water content, and the other is the state of temperature acclimation. The soil
 138 water balance module simulates the ecosystem as a bucket being filled by precipitation and emptied

139 by drainage and evapotranspiration. The state of temperature acclimation considers adaptive
140 strategies of plants by simulating the slow response of photosynthesis to changes in ambient
141 temperature. A table listing the symbols with their units and meanings is given in Table S2, including
142 model input, output, estimated variables, parameters, and mathematical symbols.

143 **2.2 Data**

144 2.2.1 Eddy covariance data

145 The meteorological and eddy covariance data were maintained and shared by the FLUXNET
146 community. Daily meteorological and flux records of 399 site-years from 55 sites (Fig. 1) were
147 selected and downloaded from the ‘FLUXNET2015 dataset’, in which half-hourly observations were
148 gap-filled, aggregated and transformed to daily records by a standard methodology (Papale et al.,
149 2006; Reichstein et al., 2005). Records of GPP were not directly measured but inferred from the net
150 ecosystem exchange (NEE) of CO₂ using nighttime data-based partitioning method. The partitioning
151 methodology has been validated for various climate and plant functional types (Reichstein et al.,
152 2005) and implemented in FLUXNET following a standard protocol. Further detailed information on
153 the FLUXNET sites used in model calibration is given in Table S1. Various forest and climate types
154 were considered in our study, ranging from tropical broad-leaved forests to cold continental
155 coniferous forests.

F 1: Study sites. DBF = deciduous broad-leaved forest, EBF = evergreen broad-leaved forest, ENF = evergreen needle-leaved forest, MF = mixed forest.

The daily meteorological observations of the dataset constituted the input variables for PRELES. In addition, the daily eddy covariance records of GPP and ET were used for comparing with the model outputs. The daily records were originally generated uniformly from half-hourly observations. A quality flag, constrained between 0 and 1, was assigned to each day to indicate the proportion of measured (nongapfilled) and good quality gap-filled half-hourly data used to calculate the daily value. For the calibration and analysis conducted in this study, we used only data with a quality flag higher than 0.7.

2.2.2 MODIS f_{APAR} data

The daily time series of f_{APAR} throughout the growing season were collected from remotely sensed data products from the Moderate Resolution Imaging Spectroradiometer (MODIS) collections (ORNL DAAC 2008, ORNL DAAC 2017). The product MOD15A2 is an 8-day 1-km-resolution product on a sinusoidal grid (Myneni, Knyazikhin, & Park, 2015a), and the product MOD15A2H is an 8-day composite dataset with 500-m pixel size (Myneni, Knyazikhin, & Park, 2015b). We chose data from Terra (MOD) instead of Aqua (MYD) or the combined product (MCD), since the time of

the Terra overpass (about 10:30 A.M.) is a better approximation of the daily integrated black sky (i.e. assuming only direct radiation from the sun) f_{APAR} (Martínez, Camacho, Verger, García-Haro, & Gilabert, 2013). A simple harmonic model was constructed to simulate the temporal dynamics of f_{APAR} (Kozlov, Kozlova, & Skorik, 2016).

$$f(t) = a_0 + \sum_{j=1}^n b_j \cos 2\pi j t + \sum_{j=2}^n c_j (\sin 2\pi j t - j \sin 2\pi t) \quad (3)$$

where $f(t)$ is the f_{APAR} at time t , t is the time in percentage normalized within the growing season, $a_0, b_1, \dots, b_n, c_2, \dots, c_n$ are coefficients, n represents a particular number of harmonics and j is the index of summation.

2.2.3 Soil information and climate classification

For each site, water-holding capacity information, including soil field capacity and soil wilting point, was collected from the Global Gridded Surfaces of Selected Soil Characteristics dataset (within a global 5-arcminute grid), which was developed by the Global Soil Data Task Group (2000) of the International Geosphere-Biosphere Programme (IGBP) Data and Information System (DIS). For soil depth, we gathered information combined from two datasets, one being the Global 1-km Gridded Thickness of Soil, Regolith and Sedimentary Deposit Layers dataset (Pelletier et al., 2016), which provides high-resolution estimates of the thickness of the permeable layers above the bedrock within a global 30-arcsecond grid. Another dataset is the International Satellite Land-Surface Climatology Project Initiative II (ISLSCP II) Ecosystem Rooting Depths (Schenk & Jackson, 2009), which provides mean ecosystem rooting depths for 1-degree by 1-degree grid cells. Climate classification for all 55 sites was based on an updated world map of Köppen-Geiger climate classification within a global 0.1-degree grid (Peel, Finlayson, & McMahon, 2007). The climate classification was a crucial criterion for grouping of the sites in multisite calibration as explained in the following section.

2.3 Methods

2.3.1 Site-specific calibration and multisite calibration

Statistical calibration of the PRELES model parameters was accomplished in a Bayesian framework by inferring the joint posterior probability density distribution of parameters conditioned on observations (Van Oijen et al., 2005). We implemented two types of calibration: site-specific calibration and multisite calibration. The site-specific calibration included 17 parameters and was applied to each site independently (Table 1). The five parameters concerning local soil, canopy or terrain information were included in site-specific calibration but excluded in multisite calibration (Table 1). For instance, the soil depth parameter was calibrated within a $\pm 15\%$ range, because the soil information came from a dataset with low resolution (Section 2.2.3), and soil depth varies largely with terrain attributes in reality. The records of ecosystem rooting depths were set as medians in prior settings. A higher range, e.g. $\pm 30\%$, was set when the record of the rooting depths largely differed from the soil depth data. The updated soil depth information from site-specific calibrations, the MAP (maximum a posteriori probability estimate), was directly used as inputs in the multisite calibrations and simulations.

Table 1: Parameters in PRELES. Note: The 12 parameters in multisite calibration are ordered by their sensitivity to the model outputs (Peltoniemi et al., 2015a). The minimum and maximum values of X_0 and S_{max} are adjusted, based on the seasonal temperature ranges at each site or of each plant functional type. The ranges of prior for soil-related parameters were set separately for each site based on information from global datasets. The reasons for exclusion of p_{GPP} and p_{ET} from calibration are given in Section S3. Coefficient m was set as a constant according to Kuusisto (1984). VPD = vapour-pressure deficit, GPP = gross primary production, PPFD = photosynthetic photon flux density.

Symbol	Meaning	Units	Prior minimum	Prior maximum	Included in site-specific calibration?	Included in multisite calibration?
χ	Evaporation parameter	$\text{dm}^3 \text{ mol}^{-1}$	0	2.5	Yes	Yes
γ	Light modifier parameter for saturation with irradiance	$\text{mol}^{-1} \text{ m}^{-2}$	1.03e-4	0.503	Yes	Yes
α	Transpiration parameter	mm	1e-6	10	Yes	Yes

$(\text{g C m}^{-2}\text{kPa}^{1-\lambda})^{-1}$						
X_0	Threshold for state of acclimation change	$^{\circ}\text{C}$	-	-	Yes	Yes
β	Potential light-use efficiency	gC mol^{-1}	0.2	2.5	Yes	Yes
$S_{\max}$	Threshold above which the acclimation modifier reaches its maximum	$^{\circ}\text{C}$	-	-	Yes	Yes
λ	Parameter adjusting water-use efficiency with vapour-pressure deficit	-	1e-4	0.999	Yes	Yes
ρ_P	Threshold for the effect of soil-water stress on photosynthesis	-	0	0.999	Yes	Yes
ν	Parameter adjusting water-use efficiency whether soil water limits gross primary production	-	1e-4	2.5	Yes	Yes
κ	Sensitivity parameter for vapour-pressure deficit response	kPa^{-1}	-1	-1e-3	Yes	Yes
ρ_E	Threshold for the effect of soil-water stress on evaporation	-	0	0.999	Yes	Yes
τ	Delay parameter for ambient temperature response	-	1	25	Yes	Yes
D_{soil}	Effective depth of soil that excludes stones and can be explored by plant roots	mm	-	-	Yes	No
θ_{FC}	Effective field capacity	mm	-	-	Yes	No
θ_{WP}	Effective wilting point	mm	-	-	Yes	No
$\theta_{\text{surf,max}}$	Maximum of the water storage on canopy surface	mm	0.5	10	Yes	No
τ_F	Delay parameter of drainage	-	1	5	Yes	No
p_{GPP}	Parameter adjusting the effect of ambient CO_2 concentration on photosynthesis	-	-	-	No	No
p_{ET}	Parameter adjusting the effect of ambient CO_2 concentration on transpiration	-	-	-	No	No
m	Coefficient for temperature dependence of snowmelt rate	$^{\circ}\text{C}^{-1}\text{d}^{-1}$	-	-	No	No

218

219 For multisite calibration, we selected 50 from 55 sites and divided them into nine PFTs, based
 220 on the forest types and Köppen-Geiger climate classification (Table 2). The division excluded five
 221 sites because they belong to either mixed forests or unique climate types (Table S1) and thus could

not be classified into any group in Table 2. A generic parameter vector for each cluster was calibrated, using the Bayesian hierarchical modelling method.

Table 2: Plant functional types for multisite calibration. Note: Detailed meanings of the letters in forest-climate classification are explained in Table S1.

Forest-climate cluster	Description	No. of sites	FLUXNET ID
DBF_Cf	Temperate deciduous broad-leaved forests (without dry season)	2	FR-Fon, IT-PT1
DBF_Cs	Mediterranean deciduous broad-leaved forests	3	IT-CA1, IT-Col, IT-Ro2
DBF_Df	Boreal deciduous broad-leaved forests (without dry season)	7	DE-Hai, JP-MBF, US-Ha1, etc.
EBF_Am	Tropical monsoon evergreen broad-leaved forests	3	AU-Rob, BR-Sa3, GF-Guy, etc.
EBF_Cf	Temperate evergreen broad-leaved forests (without dry season)	6	AU-Cum, AU-Whr, CN-Din, etc.
EBF_Cs	Mediterranean evergreen broad-leaved forests	3	FR-Pue, IT-Cp2, IT-Cpz
ENF_Cf	Temperate evergreen needle-leaved forests (without dry season)	4	AR-Vir, CN-Qia, NL-Loo, etc
ENF_Cs	Mediterranean evergreen needle-leaved forests	5	IT-SR2, US-Blo, US-Me2, etc.
ENF_Df	Boreal evergreen needle-leaved forests (without dry season)	17	CA-NS1, CH-Dav, FI-Hyy, etc.

2.3.2 Likelihood based on assumption of measurement uncertainty

Using eddy covariance measurements, three main characteristics were included in our likelihood function. Firstly, the measurement error followed a double-exponential (or Laplace) distribution instead of Gaussian (Hollinger & Richardson, 2005). Although daily records were aggregated from half-hourly measurements, the processes of gap-filling and aggregating could cause the Lindeberg's condition of the central limit theorem not to be satisfied. In our experiment, the distributions of the residuals from most sites also more closely followed the double-exponential distributions instead of a normal distribution. Secondly, the standard deviation of the random measurement uncertainty increased with the magnitude of the measurements (Richardson et al., 2008). This relationship can be approximated linearly, and the intercept has a wider range of variation compared with the slope (Aubinet, Vesala, & Papale, 2012). Thirdly, both GPP and ET measurements were considered

simultaneously during the calibration, but each followed its own error distribution separately. Eventually, the likelihood was written as the probability of the observation, conditional on the model output being the true value, which means that the residuals include both measurement error and model structure error (Van Oijen, 2017). The likelihood of the site-specific calibration was as follows:

$$\begin{aligned}
 p(\mathbf{Y}|\boldsymbol{\theta}) &= p(\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{M}(\boldsymbol{\theta})) \\
 &= \prod_{j=1}^2 \prod_{i=1}^{N_j} \frac{1}{2} \text{Exp}\left(|\varepsilon_{j,i}|; \frac{1}{a_j + b_j M(\boldsymbol{\theta})_{j,i}}\right) \\
 &= \prod_{j=1}^2 \prod_{i=1}^{N_j} \frac{1}{2(a_j + b_j M(\boldsymbol{\theta})_{j,i})} \exp\left(\frac{-|\varepsilon_{j,i}|}{a_j + b_j M(\boldsymbol{\theta})_{j,i}}\right)
 \end{aligned} \tag{4}$$

where $\mathbf{Y}$ represents the observations, $\boldsymbol{\theta}$ the parameters of the PRELES model, $M(\boldsymbol{\theta})$ the outputs of model, ε the measurement error and an unknown model structural error. $\text{Exp}(\cdot; \cdot)$ is the probability density function of the exponential distribution, and $\frac{1}{a_j + b_j M(\boldsymbol{\theta})_{j,i}}$ is its rate parameter. The j -subscripts index the two types of output variable, which are GPP and ET; the i -subscripts index the data and N_j is the total number of valid observations for variable j . Parameters a and b were calibrated simultaneously with $\boldsymbol{\theta}$ to approximate the relationship between rate parameter and measurement uncertainty.

For each forest-climate cluster, we proposed a generic vector of parameters by multisite calibration within a Bayesian hierarchical modelling approach (Fig. S8, Section S5). For each PFT (Table 2), data from different sites were combined in BC. The sites within one PFT shared the same generic parameters, which means eventually nine vectors of generic parameters were obtained respectively for the nine PFTs. To explain the variation within one PFT, two parameters, potential LUE (β) and measurement uncertainty intercept (a), were considered ‘site-specific’ and generated from distributions that represented random effects. Then the joint posterior distribution of parameters $p(\boldsymbol{\theta}|\mathbf{Y})$ is written as

$$p(\boldsymbol{\theta}, \mathbf{c}, \mathbf{d}, \mathbf{g}, \mathbf{h}|\mathbf{Y}) \propto \prod_{j=1}^2 \prod_{s=1}^S \prod_{i=1}^{N_{j,s}} \text{Exp}\left(|\varepsilon_{j,s,i}|; \frac{1}{a_j + b_{j,s} M(\boldsymbol{\theta})_{j,s,i}}\right) \prod_{j=1}^2 \prod_{s=1}^S \Gamma(a_{j,s}; c_j, d_j) \prod_{s=1}^S \Gamma(\beta_s; g, h) \tag{5}$$

262 where the s -subscripts index the site and S is the total number of sites in one cluster, $\Gamma(.; .)$ represents
263 the probability density function of the gamma distribution that describes the heterogeneity of potential
264 LUE and measurement uncertainty, c and g are the shape parameters of the gamma distributions, and
265 d and h are the rate parameters. The gamma distribution was chosen because we assumed that β and
266 a were nonnegative and followed right-skewed distributions, based on the results of site-specific
267 calibrations. The priors and hyperpriors were ignored in Eq. (5) because they were assumed as
268 independent uniform distributions. The ranges of uniform distributions for parameters in PRELES
269 were given in Table 1. Detailed explanations for the structural distinctions of site-specific calibration
270 and multisite calibration are given in Section S5.

271 2.3.3 MCMC sampler and convergence diagnostic

272 Markov chain Monte Carlo (MCMC) sampling techniques were used (Hastings, 1970; Metropolis,
273 Rosenbluth, Rosenbluth, Teller, & Teller, 1953) since the posterior distribution was nonanalytical.
274 The MCMC was simulated, using differential evolution adaptive Metropolis with snooker updating
275 (DREAMzs), which runs a few chains in parallel and explores the parameter space in an efficient way
276 (Laloy & Vrugt, 2012; Vrugt et al., 2009). We used the DREAMzs algorithm implemented in the R
277 package BayesianTools (Hartig, Minunno, & Paul, 2017).

278 The MCMC convergence diagnostic (Brooks & Gelman, 1998; Gelman & Rubin, 1992) was
279 used to monitor the convergence in the MCMC output. The multivariate potential scale reduction
280 factor (MPSRF) was calculated, based on two MCMC runs, each of which has three internal chains.
281 A large MPSRF means that the output from all chains is distinguishable and a notable difference
282 exists between variance and intrachain variance. In our study, convergence was diagnosed when the
283 MPSRF was below 1.05, which is a relatively strict criterion (Brooks & Gelman, 1998).

284 2.3.4 Model evaluation

285 Model performance was evaluated using a systematically designed cross validation procedure.

286 Calibration strategies were designed separately for six different cases of applying PRELES at a given
287 site:

288 (1) S-S: Data from a site are available for model calibration. This leads to *site-specific calibration*.

289 (2) M-S: Data from the subject site and from other sites in the same PFT are available for calibration.

290 This yields *multisite calibration*.

291 (3) S.in: No data are available for the subject site. Predictions are made with and S-S calibration of
292 another site in the same PFT.

293 (4) M.in: No data are available for the subject site. Predictions are made with and M-S calibration of
294 other sites in the same PFT.

295 (5) S.out: No data are available for the subject site. Predictions are made with and S-S calibration of
296 a site in a different PFT.

297 (6) M.out: No data are available for the subject site. Predictions are made with and M-S calibration
298 of other sites in a different PFT.

299 A two-fold validation strategy was applied to calculate the model-data mismatches. For each
300 time, half of the GPP and ET observations from the site were randomly selected for the model
301 calibration, and the remaining observations were used for the validation. In cases (3) to (6) the data
302 from the subject site was excluded in calibration but was used for validation. Eventually, the
303 reliability and stability of both the PRELES model and calibration strategies were evaluated for
304 each site independently. A comparison of cases (1) and (2) informed us about the applicability of a
305 generic parameter vector of the PFT. Case (3) to (6) were designed to find out what calibration
306 strategy to adopt when applying PRELES to new sites without any flux data or LUE related
307 information. Additionally, the model reliability in the extrapolation during the drought event in
308 2018 was assessed using another eddy covariance dataset ‘Drought 2018’ (Drought 2018 Team &

309 ICOS Ecosystem Thematic Centre, 2019) and MODIS GPP product MOD17A2H (Running &
310 Zhao, 2015; ORNL DAAC, 2018) in Section S6.

311 Besides the root mean squared error (RMSE), we also used the partitioning of the mean
312 squared error (MSE) that provides both statistical and graphical analysis for model performance
313 (Theil, 1966). Kobayashi and Salam (2000) demonstrated that the MSE could be divided into three
314 components by comparing the measurements and predictions: squared bias, squared difference
315 between standard deviations and lack of correlation weighted by the standard deviation. Gauch,
316 Hwang, & Fick (2003) suggested a slightly different partitioning of MSE: squared bias (SB), nonunity
317 (NU) slope and lack of correlation (LC). These three MSE components are distinct and additive and
318 relate transparently to correlation and linear regression parameters. The quality and difference of
319 these approaches were commented on in an exchange of letters by the authors (Gauch Jr, Hwang, &
320 Fick, 2004; Kobayashi, 2004). Here, we adopted Gauch's method, and the statistics were calculated
321 as follows:

$$322 \quad \text{MSE} = \frac{\sum_{i=1}^n (X_n - Y_n)^2}{N} = \text{SB} + \text{NU} + \text{LC} \quad (6)$$

$$323 \quad \text{SB} = (\bar{X} - \bar{Y})^2 \quad (7)$$

$$324 \quad \text{NU} = (1 - s_l)^2 \left(\sum \frac{(X_n - \bar{X})^2}{N} \right) \quad (8)$$

$$325 \quad \text{LC} = (1 - r^2) \left(\sum \frac{(Y_n - \bar{Y})^2}{N} \right) \quad (9)$$

326 where $\bar{X}$ and $\bar{Y}$ are, respectively, the means of the model predictions (X) and observations (Y); s_l is
327 the slope of the least-squares regression of Y on X ; r^2 the square of the correlation coefficient and N
328 the number of observations. SB represents the translation, which is the mean squared distance
329 between the simulations and measurements. NU represents the rotation away from the 1:1 line of
330 equality and LC the scatter that practically represents the random errors. In other words, SB represents
331 the annual overestimation or underestimation of PRELES; NU shows whether the model is equally

reliable in both low and high predictions; and LC is the random error that was not considered or explained in PRELES.

3 Results

3.1 PFT differences in the posteriori parameters

The posterior of the parameters differed with the sites or PFTs (Fig. 2). Distinctions between the PFTs were revealed by comparisons of the multisite calibrations. The tropical EBFs (EBF-Am) showed the highest evaporation parameter χ and transpiration parameter α . Except for the tropical cluster, DBFs needed higher temperatures to start the temperature acclimation (higher X_0) than evergreen forests. The delay parameters for the ambient temperature response (τ) in deciduous forests were longer than in evergreen forests. The EBFs were more strongly affected by light saturation (higher γ) than DBFs.

342

Figure 2: Marginal posterior distribution of parameters for multisite calibrations of nine plant functional types (Table 2) and the summary of maximum a posteriori parameter vectors of the site-specific calibrations (S-MAP). DBF = deciduous broad-leaved forest, EBF = evergreen broad-leaved forest, ENF = evergreen needle-leaved forest.

Distinctions of parameters between the PFTs could also be shown in site-specific calibrations
 once the parameter correlations were considered (Fig. 3). The correlations resulted from the mutual
 effects of the parameters and partially compensated for the distinctions between sites. Although the
 correlations differed among the sites, a general pattern was found in the 55 site-specific calibrations
 (Fig. S7). The highest parameter correlation was between the potential LUE (β) and the light
 saturation parameter (γ). The second highest correlation occurred between two parameters in the
 temperature modifier, which were the beginning (X_0) and the maximum (S_{max}) of the acclimation
 state. Moreover, the third pair comprised the transpiration coefficient (α) and evaporation coefficient
 (χ). The correlations of the parameters occurred not only in the posterior distributions for each site-
 specific calibration, but also on a global scale. A strongly negative correlation ($r_{Pearson} = -0.6$) was
 found between two threshold parameters, X_0 and S_{max} , by summarizing the calibrations of various
 sites (Fig. 3a). The DBFs acquired higher X_0 than did the ENFs. The uncertainty in the parameters
 greatly differed among the sites. Another distinction between forest types was revealed by parameters
 β and γ (Fig. 3b). The EBFs acquired higher γ than did the DBFs. The sites with larger β also contained
 higher levels of uncertainty, e.g. AU-Rob and AU-Wac. The distinctions of temperature acclimation
 between forest types were not only revealed by the parameters, but also by the temperature modifier
 (Fig. 4). The mean values of f_s for boreal ENF sites were about 0.4, whereas for the EBFs of tropical
 sites the values were around 0.9. DBFs showed lower f_s than evergreen forests, even with the same
 mean annual temperature.

367 **Figure 3:** (a) Thresholds of start (X_0) and maximum (S_{max}) of the temperature acclimation modifier.
 368 (b) Potential light-use efficiency (β) and light saturation parameter (γ) of the light modifier. Note:
 369 The range bars represent the uncertainty in the parameters, which is a 95% Bayesian Credible Interval.
 370 The dashed lines are from linear regressions. DBF = deciduous broad-leaved forest, ENF = evergreen
 371 needle-leaved forest, EBF = evergreen broad-leaved forest, f_{APAR} = fraction of absorbed
 372 photosynthetically active radiation.

373 **Figure 4:** Mean annual value of temperature acclimation modifier for various forest types. DBF =
 374 deciduous broad-leaved forest, EBF = evergreen broad-leaved forest, ENF = evergreen needle-leaved
 375 forest, MF = mixed forest. The f_s is a modifier that accounts for temperature acclimation (in Eq. 1).
 376

377 3.2 Site-specific calibration vs. multisite calibration

378 The ranges of the parameters varied widely between sites in the site-specific calibration (S-MAP in
 379 Fig. 2), whereas the multisite calibration strictly constrained the ranges of the parameters. In
 380 comparison to the observations, the calibrated PRELES model effectively simulated the seasonal
 381 variations within sites for most PFTs (Fig. 5). This means that different parameter vectors can lead to
 382 the similar model performance. By adding measurement errors that consider residual distributions,
 383 the predictive uncertainty describes the ranges of eddy covariance observations that could possibly
 384 occur. For the two Mediterranean climate clusters, the declines in GPP and ET during the dry summer
 385 were captured in model simulations. The prediction uncertainty also covered the variation in daily
 386 measurements. It was difficult to judge the tropical sites, since there was no seasonal pattern, and the
 387 main environmental driver for the daily variation was unclear. Even though precipitation seems to

388 relate with the daily GPP, the Pearson correlation was always lower than 0.2 on both weekly and
389 monthly step. Models based on multisite calibration and site-specific calibration performed similarly
390 in the simulations of both GPP (Fig. 5a) and ET (Fig. 5b). Distinctions only occurred occasionally in
391 a few site-year cases, and it was difficult to judge which calibration was better based on observations,
392 because one fitted the higher observations and the other the lower observations (e.g. IT-Col in Fig.
393 5a, CN-Qia in Fig. 5b).

396 **Figure 5:** (a) Daily gross primary production (GPP) and (b) daily evapotranspiration (ET) for nine
397 plant functional types (PFTs). Note: One site and one year were randomly selected from each PFT.
398 The circles represent observations of eddy covariance measurements. The orange areas represent the
399 uncertainty in the multisite calibrated model in the case M-S (Section 3.3.4). The dark orange area is
400 the parametric uncertainty. The light orange area represents the predictive uncertainty given by the
401 parametric uncertainty and measurement error. The dark blue solid line is generated by site-specific
402 calibrations with MAP (maximum a posteriori parameter vector) in the case S-S. The dashed lines
403 represent the ranges of predictive uncertainty based on the site-specific calibration. DBF = deciduous
404 broad-leaved forest, EBF = evergreen broad-leaved forest, ENF = evergreen needle-leaved forest,
405 DoY = day of year.

406 The accuracy of the predictions varied markedly among the sites (Fig. 6). For the average
407 data-model mismatches in multisite calibration, the proportion of random error (LC) in the MSE of
408 GPP was 93%. For the mismatch in seasonal variation (NU) and mean bias of annual prediction (SB),
409 the proportions were respectively 3% and 4%. The accuracy of the ET predictions was lower, since
410 17% of the MSE was SB, and for most sites the ET biases were due to underestimation. The main
411 component of the deviation was LC for both site-specific and multisite calibrations. In comparison to
412 MSE with site-specific calibrations, multisite calibrations showed 12% higher MSE for GPP and 14%
413 higher for ET on average. The accuracy differences between site-specific and multisite calibrations
414 were generally negligible, but noticeable for a few sites (Fig. 6).

Figure 6: Decomposed mean squared error (MSE) for prediction of gross primary production (GPP, $\text{g C m}^{-2} \text{d}^{-1}$) and evapotranspiration (ET, mm d^{-1}) based on site-specific calibrations (S) in the case S-S (Section 2.3.4) and multisite calibrations (M) in the case M-S. SB = squared bias, NU = nonunity, LC = lack of correlation.

Although the multisite calibrations showed higher model-data mismatches, they could be more reliable in certain cases, especially for site-years with inadequate data (Fig. 7). The MSE of multisite calibration for site-year IT-Cpz_2001 was 13% higher than that of the site-specific calibration. Only 20% of daily flux observations was deleted based on the data quality during six years. However, for the data gap during the dry season in 2001, the site-specific calibration described the daily GPP as oscillating unrealistically, whereas the multisite calibration showed a reasonable pattern of decreasing productivity.

Figure 7: Comparison of site-specific calibration and multisite calibration for the site with inadequate data (IT-Cpz, 2001). The orange areas represent the uncertainty in the multisite calibrated model in the case M-S (Section 3.3.4). The dark orange area is the parametric uncertainty. The light orange area represents the predictive uncertainty given by the parametric uncertainty and measurement error. The blue solid line is generated by site-specific calibrations with MAP (maximum a posteriori parameter vector) in the case S-S. The dashed lines represent the ranges of predictive uncertainty based on the site-specific calibration. GPP = gross primary production, DOY = day of year.

3.3 Extrapolations and site random effects

PRELES integrates simplified ecosystem processes associated with GPP and evapotranspiration. However, the reliability of extrapolation beyond the ranges of calibration datasets depends on how different the plant traits and environment conditions have been changed. To evaluate our estimates of GPP and ET, four kinds of parameter vectors representing cases (3) – (6) were calculated (Fig. 8). The S-S calibration provided a baseline for this. In the out-of-sample testing, S.out shows the highest risks in the extrapolations, while the M.out calibrations were more reliable. For the EBF_Cs and EBF_Am forests, using data from the same PFT in calibration could distinctly reduce the errors in extrapolation.

Figure 8: Root-mean-squared error (RMSE) of daily (a) gross primary production (GPP, $\text{g c m}^{-2} \text{ d}^{-1}$) and (b) evapotranspiration (ET) in the out-of-sample testing. The RMSE values were grouped based on the PFT of validation sites. S.out = site-specific calibration of sites from other PFTs, M.out = multisite calibration of other PFTs, S.in = site-specific calibration of other sites in same PFT, M.in = multisite calibration of the same PFT but excluding data from the validation site, S-S = interpolations while using half of data for calibration and other half for validation.

The mean of a gamma distribution $\Gamma(\beta_s; g, h)$ in Eq.5 was adopted as the value of β in the evaluation when multisite calibration was used because the tree species and site fertility were assumed unknown for the new sites. For each PFT, this value was calculated as g/h in Eq.5. The performance of PRELES in the case M.in can be largely improved by only adjusting the parameter

456 β (Fig. 9). The performance of cluster EBF_Cs was distinctively better than the others in the
457 validation, because all three sites represented the same tree species (*Quercus ilex*, Table S1).

Figure 9: (a) Daily gross primary production (GPP) and (b) daily evapotranspiration (ET) for nine plant functional types (PFTs). Note: One site and one year were randomly selected from each PFT. The circles represent observations of eddy covariance measurements. The orange areas represent the uncertainty in the multisite calibrated model (case M.in). The dark orange area is the parametric uncertainty. The light orange area represents the predictive uncertainty given by the parametric uncertainty and measurement error. The red solid line is generated by the MAP (maximum a posteriori parameter vector) in the case M.in while using the mean of the parameter β in the cluster (calculated as g/h in Eq.5). The purple solid line is generated by the MAP in the case M.in while adjusting the parameter β using fluxes data in the validation site. DBF = deciduous broad-leaved forest, EBF = evergreen broad-leaved forest, ENF = evergreen needle-leaved forest, DoY = day of year.

Compared with the extrapolation within one PFT, the extrapolation beyond the PFT might lead to a higher risk. The site CN-Cha, which is a mixed forest (MF) with dry springs, was not included in any cluster of the multisite calibrations. The composition of tree species makes this site neither ENF nor DBF, and the climate of this site could be classified as either Df or Dw (Table A1). We simulated the GPP and ET of this site, using parameters respectively calibrated from the two

476 spatially closest sites (CN-Qia and JP-MBF, Fig. 1) and two climate-similar PFTs (ENF-Df and DBF-
 477 Df). The simulation from the CN-Cha site-specific calibration accurately matched the observations,
 478 because it was originally calibrated with data from this site, while the simulations from the other four
 479 calibrations showed biases and a large degree of uncertainty (Fig. 10). The CN-Qia version of the
 480 calibration failed to simulate the spring GPP of the colder site CN-Cha, because their temperature
 481 acclimation processes were very different (Table S1). Meanwhile, the evaporation was highly
 482 overestimated in spring. The JP-MBF site was similar to the CN-Cha site for coldness, but was more
 483 humid with higher precipitation, which made the JP-MBF version fail in the simulation of the late
 484 spring drought at the CN-Cha site. In comparison to the site-specific versions, the two multisite
 485 calibrations performed better in both GPP and ET simulations. The higher prediction uncertainty
 486 covered the variations more thoroughly in the GPP simulations and more efficiently in the ET
 487 simulations. Nevertheless, the random effect of β introduced a large degree of parametric uncertainty
 488 into the simulations. For the DBF-Df in Fig. 10, on average 16% of the predictive uncertainty of GPP
 489 and 13% of the uncertainty in ET were due to parametric uncertainty. For site-specific calibrations,
 490 the average proportions of parametric uncertainty were only about 5% (Fig. S1).

491 **Figure 10:** Validation of different calibrations of PRELES with observations at the dry spring site
 492 CN-Cha. The circles represent observations of eddy covariance measurements at site CN-Cha. The
 493 orange areas are model simulations based on calibrations from different sites. The dark orange area
 494

495 is the parametric uncertainty. The light orange area represents the predictive uncertainty given by the
496 parametric uncertainty and measurement error. GPP = gross primary production, ET =
497 evapotranspiration, DoY = day of year.

498 **4 Discussion**

499 The model calibrations and validations demonstrated that PRELES could accurately simulate GPP
500 and ET on a large geographical scale. The simulations were reliable even for extremely contrasting
501 environmental conditions and distinctive forest ecosystems when given sufficient data. The multisite
502 calibrations were as accurate as the site-specific calibrations in the interpolations, but were more
503 reliable in the extrapolations. Based on the hierarchical quantification of the random effects among
504 sites, the predictive uncertainty was extensive for extrapolations to new sites with unknown tree
505 species and site fertility.

506 **4.1 A generic parameter vector**

507 Minunno et al. (2016) examined a generic calibration of PRELES for the boreal coniferous forests in
508 Fennoscandia and showed that the multisite calibration and the site-specific calibration performed
509 similarly. In this study, we extended the applications of PRELES to a larger regional/global scale,
510 using a Bayesian hierarchical modelling approach. PRELES assumes that the actual LUE changes
511 with weather conditions, including the intensity of light, temperature, VPD and soil water. The
512 generality of parameters in LUE models depends on the complexity of model structures and the
513 accuracy of input data. On the one hand, a universal set of parameters can be sufficient enough for
514 satellite driven LUE models across biomes and geographic regions (Yuan et al. 2014). On the other
515 hand, various studies have illustrated that many other external factors also affect the LUE, including
516 age of trees (Saldarriaga & Luxmoore, 1991), fertilization treatment (Leuning, Cromer, & Rance,
517 1991), specific leaf nitrogen (Hammer & Wright, 1994; Kergoat, Lafont, Arneth, Le Dantec, &
518 Saugier, 2008; Peltoniemi et al., 2012), and tree species (Ahl et al., 2004). Since these factors were
519 not considered in the calibrations when combining the data, we assumed that the potential LUE β was
520 different among sites (Fig. S8). Thus, the crucial assumption became that the differences among sites

521 within a single cluster could be simulated by simply adjusting the potential LUE, which was
522 confirmed and illustrated in Fig. 9 and Fig. S10. The performance of the site-specific and multisite
523 calibrations was similar (Fig. 5), and the differences between them in the Decomposed MSE tests
524 (Fig. 6) were almost negligible, which also corroborated this assumption.

525 The site-specific calibration assumes that the sites are completely unrelated. The boreal-region
526 generic calibration in the study of Minunno et al. (2016) ignored all site-to-site variability. The
527 challenge in our global data analysis and forecasting is to correctly partition different sources of
528 variability. Our multisite calibration represents the continuum between treating data sets independent
529 versus treating them identical. As a result, we partitioned process variability between the different
530 levels of the hierarchy (Section 5). Using a Bayesian hierarchical modelling approach, the random
531 effect among sites was quantified not only for the potential LUE β but also for the measurement
532 uncertainty parameter a (Eq. 5, Fig. S8). The intercept a was chosen, due to its wider range of
533 variation compared with the slope (Richardson et al., 2008). This pattern was blurred with the results
534 of the 55 site-specific calibrations. The intercept a varied among the sites, with values from 0.10 to
535 2.42 g C m⁻² d⁻¹ for GPP and 0.004 to 0.99 mm d⁻¹ for ET. By comparison, the slope b was confined
536 from 0.0007 to 0.36 for GPP and 0.0001 to 0.70 for ET.

537 The main motivations for applying the hierarchical Bayesian framework in this study include
538 combining datasets with different measurement errors, integrating the random effects for each site
539 and quantifying the uncertainty. The Bayesian framework consistently provided natural structures for
540 achieving these purposes by treating all terms in the model calibrations and predictions as probability
541 distributions (Clark, 2007; Dietze, 2017). Nevertheless, it is also possible to achieve a generic
542 parameter vector by other mathematical methods. Combinations of multisource data could be
543 considered as having multiple likelihoods or weighted objectives (Marler & Arora, 2010). Random
544 effects could be characterized by multilevel mixed models (Bijleveld & van der Kamp, 1998; Ware
545 & Liang, 1996). Uncertainty quantification could be achieved by the bootstrap method (Efron, 1979).

546 4.2 Interpolation vs. extrapolation

547 Based on the site-specific and multisite calibrations, three different vectors of PRELES parameters
548 were optional for applications: the site-specific calibrated version, the multisite calibrated version
549 with a ‘site-specific’ LUE parameter β and ‘site-specific’ measurement uncertainty a , and the
550 multisite calibrated version with unknown values (random effects) of β and a . The latter two
551 parameter vectors were only two different strategies for using the multisite calibration. This is not to
552 say that one of them is generally better or always more reliable than the other; instead, the choice of
553 method is dependent on the objectives of the model used. When the analysis is based on a local scale
554 or a region of the same site condition and a comprehensive and complete dataset is available
555 (Minunno et al., 2016), site-specific calibration would be the best option. In forestry practice,
556 however, it is common that a dataset with various possible local weather conditions is unavailable or
557 difficult to access. Moreover, the model applications often involve a wider variability in terms of
558 climate and forest structure. In that case, the multisite calibration with site-specific β and a would be
559 more reliable than the site-specific calibration (e.g. IT-Cpz in Fig. 7 and Fig. S1c). When the model
560 is extrapolated to new situations with unknown tree species and site fertility, multisite calibration of
561 the same PFT should be the best option, and site-specific calibration of other sites in same PFT should
562 be the next-best option (Fig. 8). The choices of parameter vectors should depend on the similarity of
563 PFTs instead of geographical distances. For instance, when we validated several calibrations for the
564 site CN-Cha, which was not included in the multisite calibration (Fig. 10), the site-specific potential
565 LUE parameter β and measurement uncertainty parameter a were not available from the original
566 calibration. Thus, we generated these two parameters from the gamma distributions calibrated in the
567 hierarchical Bayesian modelling approach (Eq. 5). The random effects in multisite calibration reflect
568 the actual predictive uncertainty when extrapolating entirely outside the original sites. If more
569 information were available about β , possibly based on tree species and site fertility (canopy nitrogen

570 concentration), we could also have decreased the uncertainty by constraining the value of β (the
571 purple lines in Fig. 9 and Fig. S10).

572 **4.3 The role of data quality**

573 The input of soil information is crucial for simulations of the soil-water content. We collected the
574 information from three global gridded datasets, which were inaccurate and may have affected the
575 simulations of drought events. The field water capacity and wilting point are determined by the
576 physical properties of the soil (Kirkham, 2014). Both soil texture and soil depth might vary widely
577 with the terrain. The strong correlations between soil parameters allowed only one parameter to be
578 adjusted. When calibrating all soil-related parameters simultaneously, the marginal posterior
579 distribution simply converged to the prior distribution, which means that the uncertainty in this
580 parameter was entirely dependent on the prior information (similar with the case of CO₂ module,
581 Section S3). When comparing those soil datasets with field measurements (literature in Table S1),
582 larger mismatches were found in soil depth than in soil texture. We chose to calibrate only the soil
583 depth for each site in the site-specific calibration by using the information from global datasets as the
584 prior. Eventually, the adjustments improved the simulations of those sites with drought events or dry
585 seasons.

586 The f_{APAR} is another important input for PRELES, and it interfered with the estimation of β in
587 the calibrations. We exercised particular care in interpreting the f_{APAR} data. We filtered the f_{APAR} data
588 and fitted the harmonic model, using only the observations during the growing seasons. Even so,
589 large random errors and biases could still be contained in the simulated curves of f_{APAR} (Fig. S2).
590 Although it was theoretically possible to compare the maximum LUEs of all the different tree species
591 after calibration, the error propagated from f_{APAR} obscured any relevant interpretations.

592 The global scale evaluation of the model is dependent not only on the applicability of the
593 model itself but also on the quantity and quality of the data. We filtered the eddy covariance data,
594 based on the quality flag, but outliers still occurred, which widened the mismatches. For sites with

few quality-acceptable observations, the outliers resulted in higher NU or SB (e.g. ET performance of site CA-SF2). The outliers were one of the main reasons that residuals followed double-exponential distributions instead of normal distributions. The heavy-tailed distributions likely weakened the impact of erratic observations and outliers (Sivia & Skilling, 2006). The differences between the S-S and M-S calibrations in data-model mismatch were imperceptible (Fig. 6), but the real performance could differ noticeably between the calibrations for certain gap periods (e.g. the dry season of IT-Cpz and the spring of US-Wi3 in Fig. S1c). This suggests that the information lost in gaps could have been useful for the calibrations.

Considering that most sites in our study were from boreal and temperate forests in Europe and North America, extrapolation to forests of Asia, South America and Africa could be problematic, especially for the tropical forests. No seasonal or monsoon pattern was revealed by PRELES for the tropical sites (Fig. S1a). Gebremichael and Barros (2006) found that the MODIS GPP products showed large degrees of uncertainty and were biased in the tropical monsoon regions when validated with flux tower observations. Yuan et al. (2014) compared seven LUE models on a global scale and illustrated that most models performed better in capturing the temporal changes and magnitude of GPP in DBFs and MFs than in the EBFs. Although the model-data mismatch increased with mean annual temperature (Fig. S5), it is still difficult to interpret which PFT was not suitable for PRELES. For example, the site GF-Guy showed the highest model-data mismatch for predicting GPP (Fig. 6), which was actually caused by its extremely high tree species richness and productivity (Bonal et al., 2008). The measurement errors, stand structure and silviculture treatments varied immensely within single PFTs, which obscured the distinctions among PFTs.

4.4 Biological interpretation of parameters

Instead of using direct physiological measurements of the parameters, this study applied BC and eddy covariance data to adjusting parameter values at the level of the whole system. One common concern about this approach is whether the parameters still have a biological meaning. An inadequate dataset

620 may lead to overfitting (e.g. outliers of MAPs in site-specific calibrations in Fig. 2). Since the inverse
621 modelling approach to model calibration is based on statistical analysis instead of detailed
622 physiological measurements, the MAPs may easily have deviated from physiologically meaningful
623 parameter values if the uncertainty ranges were not efficiently constrained by the data. Similarly, the
624 correlations between parameters may have led to wide uncertainty ranges (e.g. IT-SR2 in Fig. 3a). In
625 these cases, different combinations of parameters could have led to the same predictions, implying
626 that the data used in the calibration were not sufficient to reduce the parametric uncertainty. A dataset
627 from tropical or subtropical sites may not effectively constrain the parameters of the temperature
628 modifier, which was the reason for setting the priors of S_{max} and X_0 respectively for each site or PFT
629 based on the local temperature ranges (Table 1). The multisite calibration resulted in more accurate
630 estimations of parameters with a lower probability of overfitting by assimilating information from a
631 wider range of weather conditions (Fig. 2). With almost the same performance, the multisite
632 calibration contained less parametric uncertainty with more reasonable MAPs. However, the risk in
633 multisite calibrations lies in assuming that forests from different sites respond to environmental
634 factors in exactly the same pattern. Thus, instead of one global calibration, we adopted nine multisite
635 calibrations respectively designed for nine PFTs.

636 The parameters in the temperature acclimation modifier were closely associated with the
637 phenology of the growing season, and plausible parameters were obtained for each PFT. In
638 comparison to evergreen coniferous forests, deciduous forests need higher temperature for
639 acclimation (X_0) and longer delays for ambient temperature response, which shows that deciduous
640 trees recover more slowly with the rising temperatures. The delay parameter for ambient temperature
641 response τ in the DBFs was also larger than those of other clusters (Fig. 2). This distinction in spring
642 phenology was closely linked with the adaptive strategies of DBFs and ENFs. To maximize the
643 carbon fixation, it would benefit the DBFs to leaf-out as early as possible in spring. However, the
644 potential risk is damage to the leaves and conducting tissues when a late frost occurs (Bennie, Kubin,

Wiltshire, Huntley, & Baxter, 2010). ENFs adopt a resource-conserving strategy to produce well-defended needles that have a long lifespan, while DBFs adopt a resource-demanding strategy to produce less costly and poorly defended broad leaves (Rahman & Tsukamoto, 2013). Although the leaf-out day in spring was delayed, the DBFs actually had a longer effective growing season lengths, due to the higher recovery speed and delayed recession day of the growing season (Niu, Fu, Gu, & Luo, 2013).

The distinctions of the parameters among the PFTs were affected by both the physiological characteristics of the plants and the climate patterns. The higher value of light saturation parameter γ in the EBFs (Fig. 3b) indicates that larger proportions of intercepted light were not utilized, due to light saturation in comparison to DBFs. This was probably due to EBFs occurring in tropical or subtropical regions, where the light intensity is much higher than that of temperate or boreal regions. Photosynthesis keeps the light saturated for longer durations in low-latitude regions, due to high irradiance, even though low-latitude plants attain photosynthetic light saturation at higher light intensity (Mooney & Billings, 1961). Extremely high light intensity may result in a decline in photosynthesis, due to photo-oxidation of photosynthetic enzymes and pigments (Lambers, Chapin, & Pons, 2008). High levels of light also lead to an increase in leaf temperature or even heat stress. Since the temperature modifier in PRELES only focuses on seasonal acclimation, the negative impacts of unfavourably high temperature are actually explained by the light saturation modifier and VPD modifier. PRELES assumes a homogeneous environment of PPFD and canopy structure to obtain the photosynthesis of the entire ecosystem, which avoids complex structures for modelling the effect of canopy positions (Campbell, Marini, & Birch, 1992) or optimal canopy nitrogen allocation (Field, 1983; Badeck, 1995).

The ET model (Eq. 2) partitions the water fluxes of ecosystems into transpiration and evaporation. These two components were not sharply distinguished in the calibrations, since only total water fluxes were given in the eddy covariance measurements. Thus, higher uncertainty occurred

670 for the ET parameters ν and ρ_E (Fig. 2). Meanwhile, high correlations occurred between transpiration
 671 parameter α and evaporation parameter χ (Fig. S7). The threshold for the effect of soil-water stress
 672 on evaporation, ρ_E , was distinctively low for the cluster DBF-Df (Fig. 2). This may have resulted
 673 from high precipitation but low potential evaporation of its climate. In addition, the increased f_{APAR}
 674 greatly reduced the evaporation, which made the impact of soil water on evaporation negligible at the
 675 beginning of the growing season. Most parameters in PRELES are difficult to obtain in physiological
 676 measurements. Parameter λ indicates the sensitivity of water use efficiency, so the range was defined
 677 as 0 to 1 (Table 1). However, some of the lumped-parameters are even difficult to define the prior.
 678 The parameter ν is related to the sensitivity of water use efficiency to the rooting pattern, and its
 679 possible range was set based on pre-tests of the likelihood and convergence during calibration instead
 680 of the measurements in physiological studies.

681 Beer et al. (2018) found that inherent water use efficiency is higher for deciduous broad-
 682 leaved forests than evergreen needle-leaved forests based on data from 43 flux tower sites across
 683 biomes. Using MODIS and flux data at 28 sites across United States, Lu and Zhuang (2010) found
 684 that evergreen broad-leaf forest has the highest WUE, intermediate at evergreen needle-leaf forest
 685 and lowest at the deciduous needle/broad leaf forest. The parameter α in PRELES was designed with
 686 a similar interpretation with the inverse of intrinsic water use efficiency (Eq. 2). The posterior
 687 distribution of parameter α illustrated that intrinsic water use efficiency is lowest in evergreen broad-
 688 leaved forests, especially in the tropical broad-leaved forests, but no clear distinction was found
 689 between deciduous broad-leaved forests and evergreen needle-leaved forests (Fig.2). This mismatch
 690 between PRELES parameter and previous studies might be due to incorrect partitioning among the
 691 transpiration, bare soil evaporation and water storage on canopy surface after rainy days (Grelle,
 692 Lundberg, Lindroth, Morén, & Cienciala, 1997). The parameters of the evapotranspiration model
 693 might deviate from its physiologically meaningful value in order to match the observations of
 694 ecosystem total water fluxes.

695 **4.5 Uncertainty quantification**

696 Although many LUE models have previously been calibrated and tested against eddy covariance data
697 (e.g. Heinsch et al., 2006; Yuan et al., 2007), the uncertainty has seldom been quantified. Zheng et
698 al. (2018) separately analysed the uncertainty of model structure, parameters, input data and spatial
699 resolution for remote-sensing data-based LUE models, but the contributions of various sources to the
700 final forecasting were not qualified. Bayesian frameworks allow us to treat all terms in the forecast
701 as probability distributions, thus making it easier to quantify uncertainty and partition uncertainties
702 into different sources (Dietze, 2017).

703 The uncertainty analysis divided the predictive uncertainty into three components: parametric,
704 measurement, and model structural uncertainty. Since only one model, PRELES, was considered in
705 the study, the model structural uncertainty was mixed with the other two components. Measurement
706 uncertainty, which often comprised more than 90% of the predictive uncertainty (Fig. 5), represented
707 the measurement error of GPP and ET. However, the records of GPP were not directly measured but
708 inferred from the NEE of CO₂, using partitioning algorithms (Aubinet, Vesala, & Papale, 2012). A
709 certain amount of ‘measurement uncertainty’ of GPP was actually caused by the partitioning methods
710 (Fig. S4).

711 For predictions of GPP in climate change projections, the parametric and structural
712 uncertainty of PRELES was almost marginal in comparison to the uncertainty propagated from
713 emission scenarios and the global circulation model (Kallioikoski et al., 2018). However, the
714 precondition of the low uncertainty was that a sufficient dataset was obtained for the model calibration
715 and validation in the application area that was relatively homogeneous under climate and stand
716 conditions (Minunno et al., 2016). In the case of various forest types, the prediction uncertainty
717 differed greatly from site to site (Fig. S1). When simulations are based on extrapolation instead of
718 interpolation, the uncertainty will be even higher (Fig. 10), resulting from the assumptions of random
719 site effects and the choice of parameters. The uncertainty for forecasting the impact of ambient CO₂

720 concentration on photosynthesis and transpirations could hardly be assessed from model calibrations
721 (Fig. S6).

722 **4.6 Model simplifications for spatial applications**

723 The LUE approach has been applied at various spatial and temporal scales for simulations of GPP.
724 The spatial-scale application of process-based models is feasible, but requires spatially derived
725 climate data, soil survey, and remotely sensed estimates of f_{APAR} (Waring et al., 2010). Model
726 simplifications can largely reduce the data requirements and allow for simulations on a global scale.
727 The satellite driven LUE approach has been widely used in monitoring spatial and temporal dynamics
728 of global terrestrial GPP, relying on extensive remote-sensing data and simplified model structure.
729 For instance, the EC-LUE model proposed by Yuan et al. (2007) was driven by four variables only:
730 normalized difference vegetation index (NDVI), photosynthetically active radiation (PAR), air
731 temperature and the Bowen ratio. These variables can be directly derived from remote-sensing data.
732 Furthermore, Sims et al. (2008) developed a GPP model based solely on the enhanced vegetation
733 index (EVI) and land-surface temperature (LST) from MODIS. Methods of simplification include
734 setting a constant biome-independent potential LUE value (e.g. Potter et al., 1993; Yuan et al., 2007),
735 and ignoring or indirectly describing the soil-water stress (e.g. VPD accounts for drought stress in
736 MODIS-GPP products (Running, Glassy, & Thornton, 1999; Running et al., 2004)). Zheng et al.
737 (2018) quantified the model structure uncertainty in the LUE approach by comparing 36
738 combinations of optional simplified modifiers, then found the most suitable model structure for the
739 study region. The choice of a suitable model depends on both the accuracy requirement and data
740 availability. For instance, both MODIS GPP product MOD17A2H and PRELES captured the changes
741 of GPP during drought events in the 2018 summer (Fig. S10). The cost of accurate predictions from
742 PRELES is the data or knowledge for unbiased estimation of parameter β . Otherwise, the predictions
743 will contain large ranges of uncertainty. These satellite-based LUE models can be conveniently
744 applied on a global scale (Yuan et al., 2014), but the interpretations of future productivity would be

745 problematic, especially under a changing climate. The hierarchical modelling approach maintained
746 the complexity of PRELES, thus avoiding the errors propagated from model oversimplification.
747 Precipitation and soil information will be the most difficult inputs to acquire for the global simulations
748 in PRELES, whereas other meteorological variables and f_{APAR} could be directly derived from remote-
749 sensing products. For evergreen forests, another practical approach of estimating f_{APAR} is to use
750 Lambert-Beer law when annual leaf area index and extinction coefficient can be obtained.

751 PRELES aims at a compromise between predictive accuracy and model complexity. The
752 generalization of ecosystem processes on the one hand makes the model convincing in extrapolating
753 to changing environments, and on the other hand makes it convenient to parameterize and apply on
754 large geographical scales. The model accurately simulated and explained the seasonal and daily GPP
755 variations for most forest-climate types. Thus, PRELES can be a good candidate for mapping forest
756 production and quantifying uncertainty on regional to global scales under the background of climate
757 change. The potential risk in global applications is that we only calibrated parameters, while the
758 optimal model structure should vary as plant traits and environments change. For instance, the
759 modifier of temperature acclimation was crucial for boreal and temperate PFTs, but was impractical
760 for tropical forests. A key development need of PRELES for global application is to generalize and
761 quantify the ecophysiological distinctions of varying biomes. A more reliable global calibration of
762 PRELES should focus on not only adjusting parameters, but also optimizing the PFT-specific model
763 structures.

764 **Acknowledgements**

765 We acknowledge the contributions of the Academy of Finland project CARB-ARC (No. 286190),
766 the National Natural Science Foundation of China (NSFC 31640646), the Strategic Research
767 Council at the Academy of Finland (IBC-CARBON, decision. #312635, SOMPA 312912), and the
768 Horizon 2020 Research and innovation framework program (Forest Carbon Flux and Storage
769 Mapping Service, proposal #821860). We also appreciate English language editing by James

Thompson. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia and USCCC. The ERA-Interim reanalysis data were provided by the ECMWF and processed by the LSCE. The FLUXNET eddy covariance data processing and harmonization were carried out by the European Fluxes Database Cluster, AmeriFlux Management Project and Fluxdata project of FLUXNET, with the support of the CDIAC and ICOS Ecosystem Thematic Centre, and the OzFlux, ChinaFlux and AsiaFlux offices.

Conflict of Interest

None of the authors has any conflict of interest to declare.

References

Aber, J. D., & Federer, C. A. (1992). A generalized, lumped-parameter model of photosynthesis, evapotranspiration and net primary production in temperate and boreal forest ecosystems. *Oecologia*, 92(4), 463-474.

Ahl, D. E., Gower, S. T., Mackay, D. S., Burrows, S. N., Norman, J. M., & Diak, G. R. (2004). Heterogeneity of light use efficiency in a northern Wisconsin forest: Implications for modeling net primary production with remote sensing. *Remote Sensing of Environment*, 93, 168-178.

Aubinet, M., Vesala, T., & Papale, D. (2012). *Eddy covariance: A practical guide to measurement and data analysis*. Germany: Springer Science & Business Media.

Badeck, F. (1995). Intra-leaf gradient of assimilation rate and optimal allocation of canopy nitrogen: A model on the implications of the use of homogeneous assimilation functions. *Functional Plant Biology*, 22, 425-439.

795 Beer, C., Ciais, P., Reichstein, M., Baldocchi, D., Law, B. E., Papale, D., ... & Gianelle, D. (2009).
 796 Temporal and among-site variability of inherent water use efficiency at the ecosystem level.
 797 Global biogeochemical cycles, 23(2).

798 Bennie, J., Kubin, E., Wiltshire, A., Huntley, B., & Baxter, R. (2010). Predicting spatial and temporal
 799 patterns of bud-burst and spring frost risk in north-west Europe: The implications of local
 800 adaptation to climate. *Global Change Biology*, 16, 1503-1514.

801 Bijleveld, C. C.J.H., van der Kamp, L. J. T., Mooijart, A., van der Kloot, W. A., van der Leeden, R.,
 802 & van der Burg, E. (1998). *Longitudinal data analysis: Designs, models and methods*.
 803 Thousand Oaks. CA: Sage Publications Ltd.

804 Bonal, D., Bosc, A., Ponton, S., Goret, J., Burban, B., Gross, P., Bonnefond, J., Elbers, J., Longdoz,
 805 B., Epron, D., Guehl, J., & Granier, A. (2008). Impact of severe dry season on net ecosystem
 806 exchange in the Neotropical rainforest of French Guiana. *Global Change Biology*, 14, 1917-
 807 1933.

808 Brooks, S. P. & Gelman, A. (1998). General methods for monitoring convergence of iterative
 809 simulations. *Journal of Computational and Graphical Statistics*, 7, 434-455.

810 Campbell, G. S. (1977). *An introduction to environmental biophysics*. New York: Springer-Verlag.

811 Campbell, R. J., Marini, R. P., & Birch, J. B. (1992). Canopy position affects light response curves
 812 for gas exchange characteristics of apple spur leaves. *Journal of the American Society for*
 813 *Horticultural Science*, 117, 467-472.

814 Clark, J. S. (2007). *Models for ecological data: An introduction*. USA: Princeton University Press.

815 Dietze, M. C. (2017). *Ecological forecasting*. USA: Princeton University Press.

816 Drought 2018 Team, & ICOS Ecosystem Thematic Centre. (2019, August 21). Drought-2018
 817 ecosystem eddy covariance flux product in FLUXNET-Archive format - release 2019-1
 818 (Version 1.1). ICOS Carbon Portal. <https://doi.org/10.18160/pzdk-ef78>

819 Efron, B. (1979). Computers and the theory of statistics: Thinking the unthinkable. *SIAM Review*,
820 21, 460-480.

821 Falge, E., Baldocchi, D., Tenhunen, J., Aubinet, M., Bakwin, P., Berbigier, P., ... & Elbers, J. A.
822 (2002). Seasonality of ecosystem respiration and gross primary production as derived from
823 FLUXNET measurements. *Agricultural and Forest Meteorology*, 113(1-4), 53-74.

824 Field, C. J. O. (1983). Allocating leaf nitrogen for the maximization of carbon gain: Leaf age as a
825 control on the allocation program. *Oecologia*, 56, 341-347.

826 Gauch, H. G., Hwang, J. T., & Fick, G. W. (2003). Model evaluation by comparison of model-based
827 predictions and measured values. *Agronomy Journal*, 95, 1442-1446.

828 Gauch Jr, H. G., Hwang, J. T. G., & Fick, G. (2004). Reply: Comments on another way of partitioning
829 mean squared deviation proposed by Gauch et al. (2003) (multiple letters).

830 Gebremichael, M., & Barros, A. P. (2006). Evaluation of MODIS Gross Primary Productivity (GPP)
831 in tropical monsoon regions. *Remote Sensing of Environment*, 100, 150-166.

832 Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences.
833 *Statistical Science*, 7, 457-472.

834 Global Soil Data Task Group (2000). Global Gridded Surfaces of Selected Soil Characteristics
835 (IGBP-DIS). ORNL DAAC, Oak Ridge, Tennessee, USA.
836 <http://dx.doi.org/10.3334/ORNLDAAAC/569>.

837 Grelle, A., Lundberg, A., Lindroth, A., Morén, A. S., & Cienciala, E. (1997). Evaporation
838 components of a boreal forest: variations during the growing season. *Journal of Hydrology*,
839 197(1-4), 70-87.

840 Hammer, G., & Wright, G. (1994). A theoretical analysis of nitrogen and radiation effects on radiation
841 use efficiency in peanut. *Australian Journal of Agricultural Research*, 45, 575-589.

842 Hartig, F., Dyke, J., Hickler, T., Higgins, S. I., O'Hara, R. B., Scheiter, S., & Huth, A. (2012).
843 Connecting dynamic vegetation models to data – an inverse perspective. *Journal of*
844 *Biogeography*, 39, 2240-2252.

845 Hartig, F., Minunno, F., & Paul, S. (2017). BayesianTools: General-purpose MCMC and SMC
846 samplers and tools for Bayesian statistics. R package version 0.1.4. [https://CRAN.R-](https://CRAN.R-project.org/package=BayesianTools)
847 [project.org/package=BayesianTools](https://CRAN.R-project.org/package=BayesianTools).

848 Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.
849 *Biometrika*, 57, 97-109.

850 Heinsch, F. A., Maosheng, Z., Running, S. W., Kimball, J. S., Nemani, R. R., Davis, K. J., ...
851 Flanagan, L. B. (2006). Evaluation of remote sensing based terrestrial productivity from
852 MODIS using regional tower eddy flux network observations. *IEEE Transactions on*
853 *Geoscience and Remote Sensing*, 44, 1908-1925.

854 Hollinger, D. Y., & Richardson, A. D. (2005). Uncertainty in eddy covariance measurements and its
855 application to physiological models. *Tree Physiology*, 25, 873-885.

856 Kalliokoski, T., Mäkelä, A., Fronzek, S., Minunno, F., & Peltoniemi, M. (2018). Decomposing
857 sources of uncertainty in climate change projections of boreal forest primary production.
858 *Agricultural and Forest Meteorology*, 262, 192-205.

859 Kergoat, L., Lafont, S., Arneth, A., Le Dantec, V., & Saugier, B. (2008). Nitrogen controls plant
860 canopy light-use efficiency in temperate and boreal ecosystems. *Journal of Geophysical*
861 *Research: Biogeosciences*, 113(G4).

862 Kirkham, M. B. (2014). Principles of soil and plant water relations. Academic Press.

863 Kobayashi, K. (2004). Comments on another way of partitioning mean squared deviation proposed
864 by Gauch et al. (2003) (multiple letters).

865 Kobayashi, K., & Salam, M. U. (2000). Comparing simulated and measured values using mean
866 squared deviation and its components research partly supported by Core Research for

867 Evolutional Science and Technology (CREST) of Japan Science and Technology Corp. (JST).
 868 Agronomy Journal, 92, 345-352.

869 Kozlov, A., Kozlova, M., & Skorik, N. (2016). A simple harmonic model for FAPAR temporal
 870 dynamics in the wetlands of the Volga-Akhtuba floodplain. Remote Sensing, 8, 762.

871 Kuusisto, E. (1984). Snow accumulation and snowmelt in Finland. Publications of Water Research
 872 Institute, 55, 1–149.

873 Laloy, E., & Vrugt, J. A. (2012). High-dimensional posterior exploration of hydrologic models using
 874 multiple-try DREAM(ZS) and high-performance computing. Water Resources Research, 48.

875 Lambers, H., Chapin, F. S., & Pons, T. L. (2008). Photosynthesis. In: Plant physiological ecology.
 876 Germany: Springer Science & Business Media.

877 Landsberg, J. (2003). Modelling forest ecosystems: state of the art, challenges, and future directions.
 878 Canadian Journal of Forest Research, 33, 385-397.

879 Landsberg, J. J., & Hingston, F. J. (1996). Evaluating a simple radiation/dry matter conversion model
 880 using data from *Eucalyptus globulus* plantations in Western Australia. Tree Physiology, 16,
 881 801-808.

882 Landsberg, J. J., & Sands, P. J. (2011). Physiological ecology of forest production: Principles,
 883 processes and models. Netherlands: Elsevier/Academic Press.

884 Landsberg, J. J., & Waring, R. H. (1997). A generalised model of forest productivity using simplified
 885 concepts of radiation-use efficiency, carbon balance and partitioning. Forest Ecology and
 886 Management, 95, 209-228.

887 Leuning, R., Cromer, R. N., & Rance, S. (1991). Spatial distributions of foliar nitrogen and
 888 phosphorus in crowns of *Eucalyptus grandis*. Oecologia, 88, 504-510.

889 Lu, X., & Zhuang, Q. (2010). Evaluating evapotranspiration and water-use efficiency of terrestrial
 890 ecosystems in the conterminous United States using MODIS and AmeriFlux data. Remote
 891 Sensing of Environment, 114(9), 1924-1939.

892 Mäkelä, A. (1997). A carbon balance model of growth and self-pruning in trees based on structural
 893 relationships. *Forest Science*, 43(1), 7-24.

894 Mäkelä, A., Landsberg, J., Ek, A. R., Burk, T. E., Ter-Mikaelian, M., Ågren, G. I., ... Puttonen, P.
 895 (2000). Process-based models for forest ecosystem management: current state of the art and
 896 challenges for practical implementation. *Tree Physiology*, 20, 289-298.

897 Mäkelä, A., Pulkkinen, M., Kolari, P., Lagergren, F., Berbigier, P., Lindroth, A., ... Hari, P. (2008).
 898 Developing an empirical model of stand GPP with the LUE approach: analysis of eddy
 899 covariance data at five contrasting conifer sites in Europe. *Global Change Biology*, 14, 92-
 900 108.

901 Marler, R. T., & Arora, J. S. (2010). The weighted sum method for multi-objective optimization: new
 902 insights. *Structural multidisciplinary optimization*, 41, 853-862.

903 Martínez, B., Camacho, F., Verger, A., García-Haro, F. J., & Gilabert, M. A. (2013). Intercomparison
 904 and quality assessment of MERIS, MODIS and SEVIRI FAPAR products over the Iberian
 905 Peninsula. *International Journal of Applied Earth Observation and Geoinformation*, 21, 463-
 906 476.

907 McMurtrie, R. E., Gholz, H. L., Linder, S., & Gower, S. T. (1994). Climatic factors controlling the
 908 productivity of pine stands: A model-based analysis. *Ecological Bulletins*, 1994, 173-188.

909 Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation
 910 of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21, 1087-
 911 1092.

912 Minunno, F., Peltoniemi, M., Launiainen, S., Aurela, M., Lindroth, A., Lohila, A., ... Mäkelä, A.
 913 (2016). Calibration and validation of a semi-empirical flux ecosystem model for coniferous
 914 forests in the Boreal region. *Ecological Modelling*, 341, 37-52.

915 Minunno, F., Peltoniemi, M., Härkönen, S., Kalliokoski, T., Makinen, H., & Mäkelä, A. (2019).
 916 Bayesian calibration of a carbon balance model PREBAS using data from permanent growth
 917 experiments and national forest inventory. *Forest Ecology and Management*, 440, 208-257.
 918 Monserud, R. A. (2003). Evaluating forest models in a sustainable forest management context. *Forest*
 919 *Biometry, Modelling and Information Sciences*, 1, 35-47.
 920 Mooney, H. A., & Billings, W. D. (1961). Comparative physiological ecology of Arctic and Alpine
 921 populations of *Oxyria digyna*. *Ecological Monographs*, 31, 1-29.
 922 Myneni, R., Knyazikhin, Y., & Park, T. (2015a). MOD15A2 MODIS/Terra Leaf Area Index/FPAR
 923 8-Day L4 Global 1km SIN Grid. NASA LP DAAC.
 924 <http://doi.org/10.5067/MODIS/MOD15A2.006>.
 925 Myneni, R., Knyazikhin, Y., & Park, T. (2015b). MOD15A2H MODIS/Terra Leaf Area Index/FPAR
 926 8-Day L4 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC.
 927 <https://doi.org/10.5067/MODIS/MOD15A2H.006>.
 928 Niu, S., Fu, Y., Gu, L., & Luo, Y. (2013). Temperature sensitivity of canopy photosynthesis
 929 phenology in northern ecosystems. In: *Phenology: An Integrative Environmental Science*.
 930 Schwartz, M. D. (Ed.), Springer Netherlands, Dordrecht.
 931 ORNL DAAC (2008). MODIS collection 5 land products global subsetting and visualization tool.
 932 ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/1241>.
 933 ORNL DAAC (2017). MODIS Collection 6 land products global subsetting and visualization tool.
 934 ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/1379>.
 935 ORNL DAAC. (2018). MODIS and VIIRS land products global subsetting and visualization tool.
 936 ORNL DAAC, Oak Ridge, Tennessee, USA.
 937 Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W. L., ... Yakir, D.
 938 (2006). Towards a standardized processing of net ecosystem exchange measured with eddy
 939 covariance technique: Algorithms and uncertainty estimation. *Biogeosciences*, 3, 571-583.

940 Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger
 941 climate classification. *Hydrology and Earth System Sciences Discussions*, 4, 439-473.

942 Pelletier, J. D., Broxton, P. D., Hazenberg, P., Zeng, X., Troch, P. A., Niu, G., ... Gochis, D. (2016).
 943 Global 1-km gridded thickness of soil, regolith, and sedimentary deposit layers. ORNL
 944 DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/1304>.

945 Peltoniemi, M., Pulkkinen, M., Aurela, M., Pumpanen, J., Kolari, P., & Mäkelä, A. (2015a). A semi-
 946 empirical model of boreal-forest gross primary production, evapotranspiration, and soil water
 947 — calibration and sensitivity analysis. *Boreal Environment Research*, 20, 151-171.

948 Peltoniemi, M., Markkanen, T., Härkönen, S., Muukkonen, P., Thum, T., Aalto, T., & Mäkelä, A.
 949 (2015b). Consistent estimates of gross primary production of Finnish forests — comparison
 950 of estimates of two process models. *Boreal Environment Research*, 20, 196-212.

951 Peltoniemi, M., Pulkkinen, M., Kolari, P., Duursma, R. A., Montagnani, L., Wharton, S., ... Mäkelä,
 952 A. (2012). Does canopy mean nitrogen concentration explain variation in canopy light use
 953 efficiency across 14 contrasting forest sites? *Tree Physiology*, 32, 200-218.

954 Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M., Mooney, H. A., &
 955 Klooster, S. A. (1993). Terrestrial ecosystem production: A process model based on global
 956 satellite and surface data. *Global Biogeochemical Cycles*, 7, 811-841.

957 Rahman, M. M., & Tsukamoto, J. (2013). Leaf traits, litter decomposability and forest floor dynamics
 958 in an evergreen- and a deciduous-broadleaved forest in warm temperate Japan. *Forestry: An*
 959 *International Journal of Forest Research*, 86, 441-451.

960 Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., ... Valentini, R.
 961 (2005). On the separation of net ecosystem exchange into assimilation and ecosystem
 962 respiration: Review and improved algorithm. *Global Change Biology*, 11, 1424-1439.

963 Richardson, A. D., Mahecha, M. D., Falge, E., Kattge, J., Moffat, A. M., Papale, D., ... Hollinger, D.
964 Y. (2008). Statistical properties of random CO₂ flux measurement uncertainty inferred from
965 model residuals. *Agricultural and Forest Meteorology*, 148, 38-50.

966 Running, S. W., & Coughlan, J. C. (1988). A general model of forest ecosystem processes for regional
967 applications I. Hydrologic balance, canopy gas exchange and primary production processes.
968 *Ecological modelling*, 42(2), 125-154.

969 Running, S., Glassy, J., & Thornton, P. (1999). MODIS daily photosynthesis (PSN) and annual net
970 primary production (NPP) product (MOD17). Algorithm Theoretical Basis Document.

971 Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H. (2004). A
972 continuous satellite-derived measure of global terrestrial primary production. *BioScience*, 54,
973 547-560.

974 Running, S., Mu, Q., & Zhao M. (2015). MOD17A2H MODIS/Terra Gross Primary Productivity 8-
975 Day L4 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC.
976 <https://doi.org/10.5067/MODIS/MOD17A2H.006>

977 Runyon, J., Waring, R. H., Goward, S. N., & Welles, J. M. (1994). Environmental limits on net
978 primary production and light-use efficiency across the Oregon transect. *Ecological*
979 *Applications*, 4, 226-237.

980 Saldarriaga, J. G., & Luxmoore, R. J. (1991). Solar energy conversion efficiencies during succession
981 of a tropical rain forest in Amazonia. *Journal of Tropical Ecology*, 7, 233-242.

982 Schenk, H.J., & Jackson, R.B. (2009). ISLSCP II Ecosystem rooting depths. ORNL DAAC, Oak
983 Ridge, Tennessee, USA. <http://dx.doi.org/10.3334/ORNLDAAAC/929>.

984 Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, B. Z., Baldocchi, D. D., Bolstad, P. V., ...
985 Xu, L. (2008). A new model of gross primary productivity for North American ecosystems
986 based solely on the enhanced vegetation index and land surface temperature from MODIS.
987 *Remote Sensing of Environment*, 112, 1633-1646.

988 Sivia, D., & Skilling, J. (2006). *Data analysis: A Bayesian tutorial*. UK: Oxford University Press.

989 Theil, H. (1966). Applied economic forecasting. North Holland Publishing Company.

990 Valentine, H. T., & Mäkelä, A. (2005). Bridging process-based and empirical approaches to modeling
 991 tree growth. *Tree Physiology*, 25(7), 769-779.

992 Van Oijen, M. (2017). Bayesian methods for quantifying and reducing uncertainty and error in forest
 993 models. *Current Forestry Reports*, 3, 269-280.

994 Van Oijen, M., Rougier, J., & Smith, R. (2005). Bayesian calibration of process-based forest models:
 995 Bridging the gap between models and data. *Tree Physiology*, 25, 915-927.

996 Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009).
 997 Accelerating Markov chain Monte Carlo simulation by differential evolution with self-
 998 adaptive randomized subspace sampling. In: *International Journal of Nonlinear Sciences and*
 999 *Numerical Simulation*, 10(3), 273-290.

1000 Ware, J., & Liang, K. (1996). The design and analysis of longitudinal studies: A historical
 1001 perspective. *Advances in Biometry*, 1996. 339-362.

1002 Waring, R. H., Coops, N. C., & Landsberg, J. J. (2010). Improving predictions of forest growth using
 1003 the 3-PGS model with observations made by remote sensing. *Forest Ecology and*
 1004 *Management*, 259, 1722-1729.

1005 Weiskittel, A.R., Hann, D.W., Kershaw Jr, J.A., & Vanclay, J.K. (2011). Forest growth and yield
 1006 modeling. USA: John Wiley and Sons, Inc.

1007 Yuan, W., Cai, W., Liu, S., Dong, W., Chen, J., Arain, M. A., ... & Genesio, L. (2014). Vegetation-
 1008 specific model parameters are not required for estimating gross primary production.
 1009 *Ecological modelling*, 292, 1-10.

1010 Yuan, W., Cai, W., Xia, J., Chen, J., Liu, S., Dong, W., ... Wohlfahrt, G. (2014). Global comparison
 1011 of light use efficiency models for simulating terrestrial vegetation gross primary production
 1012 based on the LaThuile database. *Agricultural and Forest Meteorology*, 192-193, 108-120.

1013 Yuan, W., Liu, S., Zhou, G., Zhou, G., Tieszen, L. L., Baldocchi, D., ... Wofsy, S. C. (2007). Deriving
1014 a light use efficiency model from eddy covariance flux data for predicting daily gross primary
1015 production across biomes. *Agricultural and Forest Meteorology*, 143, 189-207.

1016 Zheng, Y., Zhang, L., Xiao, J., Yuan, W., Yan, M., Li, T., & Zhang, Z. (2018). Sources of uncertainty
1017 in gross primary productivity simulated by light use efficiency models: Model structure,
1018 parameters, input data, and spatial resolution. *Agricultural and Forest Meteorology*, 263, 242-
1019 257.