



Is a generic
calibration
sufficient?

F. Minunno et al.

This discussion paper is/has been under review for the journal Geoscientific Model Development (GMD). Please refer to the corresponding final paper in GMD if available.

A simplified gross primary production and evapotranspiration model for boreal coniferous forests – is a generic calibration sufficient?

F. Minunno¹, M. Peltoniemi², S. Launiainen², M. Aurela³, A. Lindroth⁴, A. Lohila³, I. Mammarella⁵, K. Minkkinen¹, and A. Mäkelä¹

¹Department of Forest Sciences, University of Helsinki, P.O. Box 27, Helsinki 00014, Finland

²Natural Resources Institute Finland (Luke), Jokiniemenkuja 1, 01301 Vantaa, Finland

³Finnish Meteorological Institute, 00560 Helsinki, Finland

⁴Lund University, Department of Physical Geography and Ecosystem Science, Lund, Sweden

⁵Department of Physics, P.O. Box 48, 00014, University of Helsinki, Finland

Received: 16 May 2015 – Accepted: 08 June 2015 – Published: 02 July 2015

Correspondence to: F. Minunno (francesco.minunno@helsinki.fi)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

The problem of model complexity has been lively debated in environmental sciences as well as in the forest modelling community. Simple models are less input demanding and their calibration involves a lower number of parameters, but they might be suitable only at local scale.

In this work we calibrated a simplified ecosystem process model (PRELES) to data from multiple sites and we tested if PRELES can be used at regional scale to estimate the carbon and water fluxes of Boreal conifer forests. We compared a multi-site (M-S) with site-specific (S-S) calibrations. Model calibrations and evaluations were carried out by the means of the Bayesian method; Bayesian calibration (BC) and Bayesian model comparison (BMC) were used to quantify the uncertainty in model parameters and model structure. To evaluate model performances BMC results were combined with more classical analysis of model-data mismatch (M-DM). Evapotranspiration (ET) and gross primary production (GPP) measurements collected in 10 sites of Finland and Sweden were used in the study.

Calibration results showed that similar estimates were obtained for the parameters at which model outputs are most sensitive. No significant differences were encountered in the predictions of the multi-site and site-specific versions of PRELES with exception of a site with agricultural history (Alkkia).

Although PRELES predicted GPP better than evapotranspiration, we concluded that the model can be reliably used at regional scale to simulate carbon and water fluxes of Boreal forests.

Our analyses underlined also the importance of using long and carefully collected flux datasets in model calibration. In fact, even a single site can provide model calibrations that can be applied at a wider spatial scale, since it covers a wide range of variability in climatic conditions.

GMDD

8, 5089–5137, 2015

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



1 Introduction

Biogeochemical flux models quantify the material and energy exchanges between atmosphere, biosphere and soil as a function of soil and vegetation characteristics and weather variables (Baldocchi and Meyers, 1998). Flux models are focal components of forest growth models and dynamic vegetation models (Friend et al., 2014) that describe the interactions and long-term feedbacks between the vegetation cover, soils and the atmosphere. Information about flux rates is also useful for monitoring the current carbon and water balances, such as in national greenhouse gas inventories (Peltoniemi et al., 2015b). Although the physical and physiological processes related to biogeochemical fluxes are theoretically fairly well understood (Farquhar et al., 1980; Monteith, 1981), their reliable quantification in the large geographical scale still remains a challenge. This has been demonstrated by several model comparison studies providing vastly variable predictions (e.g. Medlyn et al., 2011a). For example, a recent comparison of seven dynamic vegetation models concluded that although the net primary productivity (NPP) predictions were very similar, the related vegetation biomass predictions varied vastly, implying that the models also differed in their descriptions of photosynthesis and/or respiration rates for a given vegetation type and biomass (Friend et al., 2014).

The models of vegetation ecosystem carbon and water exchange range from complex descriptions of canopy structure accompanied with short sub-daily time steps (Juang et al., 2008; Launiainen et al., 2011; Leuning et al., 1995; Meyers and Baldocchi, 1988; Ogée et al., 2003; Olchev et al., 2008), to big-leaf models operating often also at lower temporal resolution (Kimball et al., 1997; Liu et al., 1997). On one hand the more complex mechanistic models reproduce in detail the processes of ecosystems, potentially covering a variety of responses and interactions, but also dependent on a large number of inputs with relatively high uncertainty (van Oijen et al., 2013). The more simple summary type models, on the other hand, are less input demanding, involve a lower number of parameters, and could more easily be incorporated in

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



larger-scale vegetation models and other applications. However, because of the simplifications, some of the mechanistic interactions generating site-specific differences may have been excluded, establishing a need for site-specific calibration.

The light-use-efficiency (LUE) approach provides a simple model for describing vegetation carbon fluxes and has already been applied in regional scale in the MODIS algorithm, where the gross-primary productivity (GPP) and NPP are estimated from daily weather data and leaf area index retrieved from remote sensing images (Heinrich et al., 2006). The LUE approach was further developed by Mäkelä et al. (2008) to be suited particularly for boreal and temperate conifers, and the resulting model was found to describe daily GPP rather generally and independently of site (Mäkelä et al., 2008; Peltoniemi et al., 2012). In a recent study, Peltoniemi et al. (2015a) extended the approach to include evapotranspiration (E) through its coupling to photosynthesis by assuming that GPP is a good proxy of transpiration of coniferous forests that are aerodynamically well-coupled to the atmosphere (Brümmer et al., 2012). They calibrated the resulting model, PRELES, by means of Bayesian analysis applied to eddy-covariance (EC) flux and soil moisture data at two Scots pine-dominated boreal sites. In a separate study Peltoniemi et al. (2015b) also demonstrated that the GPP predicted by PRELES across Finland, using field-based leaf area measurements as structural input, was similar to predictions by the JSBACH dynamic vegetation model (Raddatz et al., 2007) calibrated for Finland. Both predicted much lower GPP values than the standard MODIS algorithm, possibly due to leaf area index input data differences.

In model development, model calibration represents a crucial step that strongly affects the reliability of predictions (Minunno et al., 2013b). Process-based models need parameters that are directly related to physiological, functional and structural properties of the system. While detailed process-based ecosystem models that upscale processes from canopy element level to a stand scale, can mostly be calibrated based on scale-appropriate measurements or literature values (i.e. leaf gas-exchange data, soil properties etc.), simpler semi-empirical models often require calibration against ecosystem level data. The calibration is required especially for parameters for which

Is a generic calibration sufficient?

F. Minunno et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



direct measurements are difficult or impossible and must thus be estimated inversely, comparing model outputs with observed data (Hartig et al., 2012; van Oijen et al., 2005). In environmental sciences large amounts of data (e.g., EC-fluxes, national forest inventory data, remote-sensing data, and physiological measurements) are becoming available for model calibration and validation purposes. At the same time, developments in computational techniques allow to efficiently quantify model uncertainties, analyse model structure and evaluate prediction accuracy and reliability (Minunno et al., 2013a, b; van Oijen et al., 2011). The EC flux-tower network (Baldocchi, 2008), already providing more than a decade of continuous measurements, offers a good opportunity to test and calibrate models of carbon and water fluxes by providing model input variables as well as stand and site characteristics.

For the development of a generally applicable, calibrated model with explicitly expressed uncertainty bounds, systematic methods of parameter estimation from data are useful. In ecological models the parameters can usually be assigned a plausible range of variability that should be taken into account in the calibration, rather than finding the over-all best statistical fit of the model to data. Bayesian calibration offers a good method for taking into account such prior distributions which can be modified so as to reduce the uncertainty by systematic comparisons of model predictions with available data (Green et al., 2000; van Oijen et al., 2005). Recently, calculation methods have been developed to the use of Bayesian methods in combination with sensitivity analysis, error propagation and uncertainty estimates (Minunno et al., 2013a; van Oijen et al., 2011).

The objective of this study was to assess if the PRELES model can be used as a tool to estimate the carbon and water fluxes of boreal coniferous forests in Fennoscandia. Firstly, we prepared a comprehensive sensitivity analysis of PRELES and then used the Bayesian framework to calibrate and evaluate the model to data from multiple boreal coniferous sites in Fennoscandia. Using these analyses as basis, we sought answers to three questions: (1) Can we find a generic set of model parameters that adequately performs at all sites? (2) Under what conditions – if any – should the multi-site calibra-

tion be used in favour of the site-specific calibration, if both exist for a site? (3) How should data be selected for model calibration to extend model predictions of GPP and ET to a site with no prior data?

2 Materials and methods

2.1 PRELES model

PRELES (PREdict with LESs – or – PREdict Light-use efficiency, Evapotranspiration and Soil water) is an ecosystem model of intermediate complexity developed by Peltoniemi et al. (2015a), in which the dependent variables, GPP (P , $\text{gC m}^{-2} \text{day}^{-1}$), evapotranspiration, ET (E , mm) and soil water (θ , mm), are interlinked. The model works at daily time-step and requires minimal input data. The climatic driving variables are daily mean temperature (T , $^{\circ}\text{C}$), vapour pressure deficit (D , kPa), precipitation (R , mm) and photosynthetic photon flux density (PPFD, Φ , $\mu\text{mol m}^{-2} \text{day}^{-1}$). The only stand structural information is the fraction of absorbed PPFD (f_{aPPFD}), estimated using the Beer–Lambert law as f_{aPPFD}

$$f_{\text{aPPFD}} = 1 - \exp^{-kL} \quad (1)$$

where L is the leaf area index ($\text{m}^2 \text{m}^{-2}$) and k the extinction coefficient.

A detailed description of the PRELES can be found in Peltoniemi et al. (2015a); herein we briefly outline model structure and provide all the equations.

Water storage (θ) consists of three pools: intercepted water (θ_{surf}) (mainly on canopy surfaces), snow/ice (θ_{snow}) and soil water storage (θ_{soil}). All components are described

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



by simple bucket models.

$$\theta = \theta_{\text{surf}} + \theta_{\text{snow}} + \theta_{\text{soil}} \quad (2)$$

$$\theta_{\text{surf}} = \theta_{\text{surf}} + R^1 - E^{\text{surf}} - F_{\text{surf}} \quad (3)$$

$$\theta_{\text{snow}} = \theta_{\text{snow}} + R^0 - E^{\text{snow}} - M \quad (4)$$

$$5 \quad \theta_{\text{soil}} = \theta_{\text{soil}} + F_{\text{surf}} + M - F - E^{\text{soil}} \quad (5)$$

where R^1 is rainfall and R^0 is snowfall. Precipitation is assumed to be snow when air temperature is below 0°C . M is snowmelt, F_{surf} is drainage from canopy surface and F is drainage from the soil. Ecosystem evapotranspiration is given by the sum of evaporation and transpiration from the three water storage components, i.e. canopy, snow and soil (Eq. 6).

$$10 \quad E = E^{\text{surf}} + E^{\text{snow}} + E^{\text{soil}} \quad (6)$$

We assume that the canopy intercepts precipitation up to a maximum, $\theta_{\text{surf, max}}$.

$$F_{\text{surf}} = \min(0, \theta_{\text{surf, max}} - \theta_{\text{surf}}) \quad (7)$$

When precipitation exceeds this limit the additional water reaches the soil and accumulates in the soil water storage (θ_{soil}) up to the field capacity of soil. Additional water drains away from the system with a fix time constant (τ_F). We used $\tau_F = 3$ days derived from soil water measurements at a boreal forest site on mineral soil (Peltoniemi et al., 2015a).

$$15 \quad F = \frac{\theta_{\text{soil}} - \theta_{\text{FC}}}{\tau_F} \quad (8)$$

20 Snow water storage θ_{snow} accumulates when $T < 0^\circ\text{C}$ and melts at a rate m when $T > 0^\circ\text{C}$ (Kuusisto, 1984).

$$M = \begin{cases} mT, & T > 0 \\ 0, & T < 0 \end{cases} \quad (9)$$

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The GPP-submodel is a modification of Prelued (Mäkelä et al., 2008). The photosynthetic production is related to the light use efficiency of the stand (β) and the absorbed photosynthetic photon flux density.

$$P = \beta \Phi f_{aPPFD} \quad (10)$$

5 Light use efficiency is given by the potential light use efficiency (β_P) multiplied by an array of modifiers (f_i) varying between 0 and 1 that accounts for the environmental conditions.

$$\beta = \beta_P \prod f_i \quad (11)$$

10 The saturation of photosynthetic production with high PPFD is expressed by the light modifier f_L , that follows the rectangular-hyperbola photosynthesis model (Mäkelä et al., 2008).

$$f_L = (\gamma \Phi + 1)^{-1} \quad (12)$$

Temperature impacts photosynthesis using a modifier for temperature acclimation (f_S) (Mäkelä et al., 2004, 2008).

$$15 \quad f_S = \min(S/S_{\max}, 1) \quad (13)$$

$$S_k = \max(X - X_0, 0) \quad (14)$$

$$X = X_{t-1} + \frac{1}{\tau}(T - X_{t-1}), \text{ where } X_1 = T_1 \quad (15)$$

20 S_{\max} (°C) is the minimum temperature threshold at which canopy photosynthesis is not limited by temperature (i.e., $f_S = 1$ for $T \geq S_{\max}$). S (°C) is the state of acclimation that depends on the limit (X_0) above which f_S is higher than 0 and on the a priori estimate for the state of acclimation (X). X is calculated using a first-order dynamic delay model influenced by the daily air temperature T (°C) and the value of X during the previous

Is a generic calibration sufficient?

F. Minunno et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



day (X_{t-1}). The parameter τ , expressed in days, represents the speed of response of the current acclimation status to changes in T .

Plant water stress ($f_{DW,P}$) reduces photosynthesis and it can be caused by vapour pressure deficit of the atmosphere (f_D) and soil water availability ($f_{W,P}$). We assumed that only the most limiting factor between f_D and $f_{W,P}$ reduces photosynthesis (Landsberg and Waring, 1997).

$$f_{DW,P} = \min(f_D, f_{W,P}) \quad (16)$$

where f_D affects GPP through an exponential relationships:

$$f_D = e^{\kappa D} \quad (17)$$

$f_{W,P}$ depends on the relative extractable water W

$$f_{WP,k} = \min(1, W_k / \rho_P) \quad (18)$$

$$W = \frac{\theta - \theta_{WP}}{\theta_{FC} - \theta_{WP}} \quad (19)$$

where θ_{soil} is water stored in the soil, θ_{WP} is the wilting point and θ_{FC} is the field capacity, and ρ_P is the threshold of W (relative extractable water) below which P is reduced linearly.

Evapotranspiration is calculated by means of a simple empirical equation that requires minimal input data, but links the predicting variables P , E and θ_{soil} .

$$E = \alpha \frac{P}{D^\lambda} f_{W,P}^\nu D + \chi(1 - f_{aPPFD}) \varphi f_{W,E} \quad (20)$$

where α and χ are empirical parameters; λ is a parameter that relates evapotranspiration and vapour pressure deficit (Medlyn et al., 2011b; Peltoniemi et al., 2015a). E is influenced by the soil water modifier, but $f_{W,P}$ is raised to the power ν since the response of P and E to drought is different. The E is also affected by soil drought through the $f_{W,E}$ modifier; $f_{W,E}$ follows the same equation of $f_{W,P}$ (Eq. 19) but it has its own threshold ρ_E .

GMDD

8, 5089–5137, 2015

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



2.2 Carbon and water flux data

Stand-scale net ecosystem exchange (NEE) of CO₂, evapotranspiration and meteorological data from ten boreal coniferous forest sites located in Finland and Sweden were used in this study (Table 2). The sites cover a latitudinal band from 60 to 67° N with annual mean temperatures ranging from 0.8 to 7.1 °C, and precipitation from ~ 550 to ~ 850 mm. Leaf area index (LAI) at each site was treated as one lumped LAI, i.e. all the canopy layers were included in one unique layer. The total (all-sided) LAI varies between ~ 3.8 and ~ 12 m² m⁻² offering a good possibility to address both climatic and LAI controls on forest GPP and ET. A brief summary of the sites is provided in Table 2, and complete descriptions can be found in the respective references.

The NEE and ET were measured above the forest canopies by the eddy-covariance method and the 1/2 h fluxes computed according to common practices (Aubinet et al., 2012). Gaps in data caused by instrumental failures or methodological issues, such as insufficient turbulent mixing, were gap-filled, and NEE was partitioned into component fluxes before the 1/2 h data was aggregated into daily averages or sums. The gap-filling of NEE was done using a combination of look-up tables and mean diurnal variability according to Reichstein et al. (2005). The gaps in meteorological data were filled either by linear interpolation or by the mean diurnal variability determined in a 14 day moving window.

The GPP was separated from the measured NEE as $GPP = -NEE + R_e$, where the ecosystem respiration R_e is (Kolari et al., 2009; Reichstein et al., 2005)

$$R_e = R_{10} Q_{10}^{[(T-10)/10]} \quad (21)$$

The R_{10} ($\mu\text{mol m}^{-2} \text{s}^{-1}$) represents the temporally varying base respiration rate at 10 °C temperature (T) and Q_{10} (unitless) represents the short-term temperature sensitivity, which is assumed constant in time but can vary among the sites. The R_e model parameters were determined for each site by a non-linear least squares fit of Eq. (22) to nighttime NEE measured in turbulent conditions (friction velocity μ^* exceeding an em-

GMDD

8, 5089–5137, 2015

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



pirically defined site-specific threshold) using measured soil or air temperature as an independent variable. The Q_{10} was first computed by pooling all available growing season data, defined here as May–September (June–August in Sodankylä due to northern location). Secondly, Q_{10} was fixed and the temporal variability of R_{10} was determined by fitting Eq. (22) to data in four-day non-overlapping windows, and linearly interpolating between the window centres. Finally, R_e was computed by extrapolating the obtained regression model to daytime temperatures, allowing the GPP to be approximated for each 1/2 h period.

After gap-filling, the fluxes and meteorological variables were aggregated at daily time step. A quality flag (F) varying between 0 and 1 was assigned to each day to represent the fraction of gap-filled data used to compute the daily value, and used in later analysis to weight the observations error (see “*Model calibration and model comparison*” section).

2.3 Overview of the model analyses

Bayesian calibration (BC) and Bayesian model comparison (BMC) were used to quantify the uncertainty in model parameters and model structure. For a comprehensive understanding of model behaviour, the Bayesian analyses were combined with a model-data mismatch analysis and a global sensitivity analysis (i.e., Morris method, Morris, 1991) following the framework proposed by van Oijen et al. (2011) and improved by Minunno et al. (2013a).

The work consisted of three analyses where we compared multi-site (M-S) and site-specific (S-S) calibrations. M-S has the advantage that the data involved in the calibration cover a wider variability in terms of climate and forest structure since they come from different sites, including measurement and other errors which may or may not partially cancel out when all data are used in parameter inference. In contrast, S-S could provide good correspondence to local data, but may not be spatially generalizable, firstly because the processes may not be generic, and secondly because the risk of

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



bias increases with less measurements. More specifically, we conducted the following comparative analyses:

2.3.1 Analysis #1 – “Global or local” – Is PRELES generic enough to be applied at regional scale using one generic calibration?

5 We compared M-S and S-S calibrations, in order to test if PRELES is a model of general applicability, and to test how well one calibration can predict ecosystem fluxes. In total, 11 BCs were performed; the model was independently calibrated for each of the ten sites (S-S calibrations) and a multi-site calibration was achieved using data from all the sites in one BC. Parameter estimates and model outputs from the M-S and S-Ss were
10 compared in order to detect any significant differences between the calibrations.

2.3.2 Analysis #2 – “Forward prediction” – Is a multi-site calibration better than site-specific in predicting fluxes for a site for which data are already available?

15 The aim of this exercise was to compare M-S and S-S calibrations in predicting future carbon and water fluxes of a site for which data are already available. For this analysis the datasets of each site were split in two parts, the first half was used for model calibrations (calibration dataset) and the second half for the comparison (comparison dataset). Similarly to Analysis #1, but using just the shorter calibration datasets, 11 BCs were performed. In addition, 10 model comparisons were carried out, one for each
20 site, using the comparison dataset and outputs from the M-S and the S-S versions of PRELES.

2.3.3 Analysis #3 – “New site” – To predict GPP and ET for a new site, should a single site calibration or a multi-site calibration be used?

25 In this analysis we compared the M-S and S-S to test which calibration is more suitable for predicting ET and GPP for a site where the model has not been calibrated before. In

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



calculated as:

$$e_i(\mathbf{X}) = \frac{[y(x_1, \dots, x_{j-1}, x_j + \Delta, x_{j+1}, \dots, x_k) - y(\mathbf{X})]}{\Delta} \quad (22)$$

The OAT experiments are designed in order to uniformly cover the whole parameter space; r trajectories are generated, with each trajectory having a different starting point randomly selected. So, r EEs are computed; the mean (μ^*) (Campolongo et al., 2007) and the standard deviation (σ), from the distributions of the absolute values of the EEs represent the sensitivity measures. μ^* gives the overall importance of a parameter, while σ describes non-linear effects and interactions between parameters. A complete description of the method can be found in Morris (1991) and Campolongo (2007). Morris sensitivity analysis has been used for several process-based forest models (Minunno et al., 2013a; Song et al., 2013, 2012; van Oijen et al., 2011).

Sensitivity analyses of PRELES were carried out for the parameter space defined by the minimum and maximum values (Table 1), corresponding to the parameter space of the prior distribution of the Bayesian calibrations. The prior parameter space was chosen because it helps us to understand model behaviour in relation to the dataset used in the calibration process, thus helping us to interpret the results of the Bayesian calibrations. Morris method was applied to each site in order to test if the sensitivity results change with the climatic conditions in Fennoscandia.

2.5 Model output sensitivity to LAI

The fraction of absorbed PPFD (f_{aPPFD}) is the only stand structural variable used to drive PRELES. In this work, f_{aPPFD} of each site was estimated using LAI (including both the main canopy and understory), by means of Beer's law (Eq. 1). In order to quantify model output response to variations in LAI we conducted a sensitivity analysis using the Hyytiälä dataset. Average annual GPP and ET were calculated running PRELES with different LAI values ranging from 0 to 16. Considering that total (all-sided) stand LAI at Hyytiälä was about 8, for this sensitivity analysis LAI was varied $\pm 100\%$. Model runs

were conducted using the parameter estimates achieved by the multi-site calibration (i.e., all data from all sites were included in the calibration; details are provided in the next section).

2.6 Model calibration and comparison

Bayesian calibration (BC) provides an updated joint probability distribution of the parameters (*posterior* distribution) combining existing parameter knowledge (*prior* distribution) and new information enclosed in the data (*likelihood*).

The 12 most influential parameters on PRELES outputs (Peltoniemi et al., 2015a) were included in the BC (Table 1); the priors were uniformly distributed between the minimum and maximum values reported (Table 1).

A variable number of GPP and ET data points were available for the BC at each site (Table 2). The data were considered to be normally distributed so the likelihood was a Gaussian distribution and the standard deviation (s_i) of each data point i was assumed to be proportional to the number of gap-filled data in a day (see the quality flag F defined above) and was calculated using the following equation:

$$s_i = a_j + b_j F_i \quad (23)$$

where F_i is a the quality flag and the parameters a_j and b_j were specific for each data type j (i.e., GPP and E) and were included in the BCs. We only used data with a quality flag lower than 0.7 in the calibrations.

2.7 Model-data mismatch

The Bayesian approach jointly uses the prior and all the data to calculate the posterior distribution, but it provides little information about the strengths and weaknesses of a model. On the contrary, the more classical analyses of the mismatch between the simulated and the observed data give useful insights about model behaviour. In particular the decomposition of the mean squared error (MSE) provides indications about the

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



accuracy and the precision of the predictions (Minunno et al., 2013a; van Oijen et al., 2011).

MSE can be decomposed into three components: the bias error, the variance error and the correlation error (Kobayashi and Salam, 2000) (Eq. 24).

$$5 \text{ MSE} = \overline{(S - O)^2} = (\bar{S} - \bar{O})^2 + (\sigma_S - \sigma_O)^2 + 2(\sigma_S \sigma_O)(1 - r) \quad (24)$$

where O are the observations, S the model predictions, σ_O and σ_S are the standard deviation of the observed and simulated data respectively and r is the correlation between the O and S .

10 The bias error quantifies the distance of the predictions from the data; the variance error expresses if the model is able to catch data variability; the correlation error indicates if the model is able to reproduce the pattern of data fluctuations. The latter component expresses the lack of positive correlation between the observed and simulated data and is weighted with standard deviations (Eq. 3), therefore there is an overlap between the variance and the correlation error (Kobayashi and Salam, 2000).
15 In practice this MSE component seems to capture all types of error, such as random errors, left after accounting for the bias and the differences in the variances. MSEs were calculated for both GPP and ET and for each site.

3 Results

3.1 Sensitivity analyses

20 The Morris sensitivity metrics, μ and σ , for GPP indicate that X_0 (temperature threshold), χ (non-canopy evapotranspiration), α (canopy evapotranspiration), γ (light saturation), and β (potential GPP) were the most influential parameters (highest μ) with respect to the modelled photosynthetic activity in Hyytiälä (Fig. 2a). The results for Hyytiälä are representative of all sites except Norunda where the sensitivity of GPP to

**Is a generic
calibration
sufficient?**

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(Fig. 5a), whereas differences occurred in estimates of the parameters at which model outputs are less sensitive (Fig. 5b). In the CAge12yr site-specific calibration γ , X_0 and β marginal posterior distributions were quite different from the rest of the calibrations.

Bayesian calibration provides a joint posterior distribution of model parameters, i.e. BC also considers the interactions between parameters. This kind of information is not derivable from the marginal posterior distributions (Fig. 5), but it can be expressed through the correlations (r) between parameters calculated from the posterior sampled by the MCMC. To simplify the result presentation we briefly summarize the results below. In the M-S calibration the highest correlations were between β and γ ($r = 0.83$) and between X_0 and S_{\max} ($r = -0.78$). Significant correlations were also found between α and χ ($r = -0.67$), and between α and λ ($r = -0.64$). For the Kalevansuo, Flakaliden and Norunda S-S calibrations parameter correlations were similar to the M-S calibration; while for the other sites some differences in the parameter correlations of the posterior distributions were found. For instance β was positively correlated to S_{\max} in the calibrations for Knottåsen ($r = 0.84$), CAge75yr ($r = 0.71$), Hyytiälä ($r = 0.66$), CAge12yr ($r = 0.63$) and Alkkia ($r = 0.63$).

We evaluated model performances in terms of R^2 and the slopes of the simulated vs. observed data, calculated for each calibration and each model output (i.e., GPP and ET) at daily time step (Table 3). The predictions were generated using the maximum a posteriori (MAP, i.e. the modal parameter vector of the posterior distribution) parameter vectors of M-S and S-S (Fig. 6). The variance explained by the model was higher for GPP than for ET, both being in most of the cases higher than 70% (R^2 of Table 3); however the model tended to underestimate carbon and water fluxes (slopes lower than 1) (Table 3). Model fit to the Flakaliden data was generally rather poor. Furthermore, the multi-site calibration significantly underestimated evapotranspiration at Alkkia site (slope = 0.62). In general, after BC, model outputs were characterized by low uncertainty (not shown in the plots).

In model predictions of GPP and ET for each site, the differences between the multi-site and site-specific calibrations were small in most of the cases (Fig. 6). The exception

GMDD

8, 5089–5137, 2015

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



is the evapotranspiration at the Alkkia site, where ET by the multi-site calibration was clearly different from the site-specific calibration prediction (Fig. 6a).

Mean squared errors were calculated for GPP and ET and decomposed to the bias, variance and correlation errors for the multi-site and site-specific calibrations (Fig. 7).

For the site-specific calibrations the main component of model error was the correlation error, while the other two components were negligible (Fig. 7). Also the MSEs of the multi-site calibration predictions were mainly constituted from the correlation error (Fig. 7); however the other two error components were significant at some sites, varying between 10 and 30 % of MSE. ET predictions at the Alkkia site for the multi-site calibration were characterized by the highest bias error (i.e. 40 % of MSE, Fig. 7b).

Both M-S and S-S calibrations showed robust performances in predicting the photosynthetic activity of boreal forests also at annual time step (Figs. 8a and 9a); while the model was less accurate in reproducing the annual evapotranspiration (Figs. 8b and 9b). Note that to compute the “annual” fluxes, the daily fluxes were summed only if the quality flag was lower than 0.7 (i.e. at maximum 30 % of 1/2 h fluxes were missing and gap-filled for that particular day), and the numbers in Figs. 8 and 9 are not representative for true annual balances. PRELES was able to catch the pattern of GPP inter-annual fluctuations for the sites with the long-term datasets (i.e., Hyytiälä, Sodankylä, Flakaliden, Norunda and Kalevansuo) (Fig. 9a), but at Flakaliden for some years (i.e., 1997, 2002) the relative difference between the observed and modelled GPP was around 50 %. The M-S and S-S annual prediction were really similar (Figs. 8 and 9), apart from the GPP at Flakaliden (Fig. 9a).

3.3 Analysis #2 – “Forward predictions” –

M-S and S-S calibrations were evaluated at each site using the validation dataset and considering both output fluxes (i.e., ET and GPP) at the same time. In six sites S-S had 100 % probability of being the best model version, while in the other sites BMCs supported the M-S calibration (Table 4). In general the NRMSEs calculated with the two types of calibration did not differ substantially (Fig. 10). At Hyytiälä and Skyttorp

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the GPP NRMSE of S-S was 10–20 % higher than the GPP NRMSE of M-S, while at Flakaliden and CAge75yr the M-S calibration error was significantly higher (Fig. 10a). The NRMSEs of the evapotranspiration M-S were always higher than those of the S-S calibration, except for Skyttorp (Fig. 10b). At Alkkia the M-S had a NRMSE of about 60, 20 % higher than the error of the S-S calibration; on the contrary, at Skyttorp the NRMSE of S-S was 20 % higher than the M-S NRMSE. For the rest of the sites the ET NRMSEs of the two calibrations differed less than 5 %.

3.4 Analysis #3 – “New site” –

According to the BMC, the Hyytiälä S-S calibration had 100 % probability of being better than the M-S calibration on every site; on the contrary, the M-S calibrations were always better than the other S-S calibrations (Table 5). The normalised root mean squared errors calculated for Analysis #3 are consistent with the BMC probabilities (Fig. 11, Table 4). The ET and GPP NRMSEs of the M-S calibration were slightly higher than those of Hyytiälä S-S calibration, but the differences are negligible (Fig. 11). The NRMSEs of the other S-S calibrations were always higher than the errors generated from the M-S calibration (Fig. 11).

4 Discussion

Evaluating model performances in the light of site-specific calibrations and a multi-site calibration gives useful information about the general applicability of the model. PRELES is a simple model, with a strong empirical component, however calibration results showed that a generic calibration can be used to estimate the gross primary production and the evapotranspiration of all the sites considered in this study. In fact, model performances obtained using the multi-site calibration were similar to those achieved by the site-specific versions at both daily and annual time steps (Table 3, Figs. 6–9), with exception of a site with agricultural history (Alkkia). Although errors in the data can-

GMDD

8, 5089–5137, 2015

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Is a generic calibration sufficient?F. Minunno et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

not be excluded as potential reasons, the most likely explanation is that Alkkia forest is located on a peatland drained for agriculture in 1930's and subsequently afforested about 35 years ago. The agricultural history of the site is seen as high nutrient contents in the soil (due to use of fertilizers) that are reflected in the amount and species composition of the understory vegetation (Lohila et al., 2007). The vigorously growing understory at Alkkia is composed of deciduous species that have less conservative water use strategies than Scots pine and Norway spruce that dominate the LAI at other sites. Kalevansuo is also a drained peatland forest, but for this site M-S and S-S ET predictions were similar. Contrary to Alkkia, Kalevansuo has no agricultural history and its understory consists of dwarf shrubs and mosses similar to the mineral soil sites part of this study (Table 2). The failure of the M-S calibration to predict ET at Alkkia could be partially related also to an improper representation of LAI. We used a lumped LAI for trees and understory, but they have different seasonal dynamics and different physiology. The problem can also be due to uncertainty in soil hydraulic characteristics (e.g., field capacity) as well as to simplistic representation of the soil water balance in PRELES. The water storages are described by small superficial water storage and a simple bucket model with a pre-defined fixed drainage coefficient and has no explicit description of lateral flows such as drainage to ditches. In the S-S calibration at Alkkia, the model structural deficiency may have been compensated by different parameter estimates (see parameter χ of Fig. 5a).

PRELES predicts GPP better (Table 3) than evapotranspiration, the total water flux from several sources. While transpiration is highly correlated with GPP through linkage between stomatal conductance and assimilation rate (Katul et al., 2010; Medlyn et al., 2011b), are other water sources constrained more by stand characteristics, microclimate and soil properties. Modelling highly dynamic processes such as interception and evaporation at daily time step and neglecting the layered structure of forest ecosystems could be one reason for poorer ET predictions. Also, EC-based evapotranspiration estimates have in most cases higher uncertainty than carbon fluxes due to unclosed

energy balance (Foken, 2008) and technical problems measuring water vapour at high air humidity (Mammarella et al., 2009).

The analysis on the MSE decomposition (Fig. 7) allowed us to better understand model behaviour. The main component of the MSE was usually the correlation error, probably due to the summer peaks that occur in particular environmental conditions and that the model is not able to reproduce. Furthermore, in this study we were using fixed annual values of LAI; a better representation of seasonal LAI dynamics might help improve the predictions of rainfall interception and thereby the water flux partitioning between transpiration and evaporation. However, the seasonal cycle modifier f_s , (Eqs. 14–16) partially accounts for both the annual cycle of photosynthetic capacity and LAI but it influences only GPP and transpiration predictions and has an upper limit of 1 that is reached typically before midsummer. Earth observed data would allow to integrate the intra annual LAI variability in the model. In spite of these reservations, we found that the mean squared error was low for most of the sites and the deviation from the annual aggregated data was lower than 10% in most of the cases, confirming that the model can be considered reliable tool to predict the carbon and water fluxes of boreal forests.

The sensitivity analysis carried out through the prior parameter space allowed us to identify the parameters that a priori were most influential on the outputs. The differences between the sensitivity results at Norunda and the rest of the sites were mainly due to the differences in leaf area index, LAI being much higher at Norunda. LAI has a strong impact on PRELES outputs, especially on GPP. The sensitivity of photosynthesis to LAI follows the exponential curve of Eq. (1), since GPP is linearly related to the fraction of absorbed PPFD (Eq. 10). It is important to have accurate estimates of LAI, especially for stands with low foliage biomass (e.g, young stands, low productive sites), because small errors in LAI strongly affect GPP calculations. Nowadays, thanks to remote sensing techniques, it is possible to obtain inputs for ecosystem models at high spatial and temporal scale, making possible the application of process-based

GMDD

8, 5089–5137, 2015

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



models in practice (Härkönen et al., 2011). The relative weak response of forest ET to LAI (Fig. 4) is in line with EC-based observations in boreal Canada (Amiro et al., 2006).

Combining the sensitivity of model output to the parameters and the uncertainty analysis it was possible to extract useful information about PRELES behaviour and its general applicability. The uncertainty of the most influential parameters was strongly reduced by the data and the parameter estimates were quite similar for the different versions of the model (Fig. 5a). Model output was not strongly sensitive to soil-water related parameters (ρ_P , ρ_E , ν) and the posterior distribution of those parameters remained quite uncertain. The reason for this is that the boreal forests are not often water stressed so there is little information to estimate these parameters. However, those parameters could become crucial if the model is applied to more xeric sites. In the CAge12yr site-specific calibration, some of the most important parameters (i.e., γ , X_0 and β) were quite different from the rest of the calibrations. This could be due to the understory that accounted for almost 50% of the LAI of this site. Also, in young regenerating stands the contribution of deciduous tree species is more abundant than at older sites. The physiological differences of forest plant species might influence the stand level carbon and water fluxes. In the future it must be investigated if modelling stand layers separately as well as describing the soil water balance in more detail will improve model performance.

No significant differences were encountered in the parameter estimates and model outputs of the Scots pine and Norway spruce dominated stands. The delay parameter for ambient temperature response (τ , Fig. 5b) was the only parameter for which the Norway spruce dominated sites (i.e., Flakaliden, Norunda and Knottåsen) had similar marginal posterior distributions, while τ estimates for the Scots pine dominated sites were different. Consistently with recent results by Linkosalo et al. (2014), the photosynthetic activity of Norway spruce starts earlier than Scots pine, explaining the lower values of τ . Flakaliden was the site where the model showed the worst performance in predicting GPP. Nevertheless, since the model performed well at the other two spruce sites, we believe that there is no need for a species specific calibration of PRELES,

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



which also speaks for the generality of model calibration, at least given the uncertainties involved.

Parameter uncertainty was strongly reduced by the BC since thousands of data points were involved in the calibrations; for this reason also the uncertainties of model predictions were low. As expected M-S resulted in more accurate parameter estimates; however, PRELES parameter uncertainty was low also for S-S of the sites with the long-term datasets (i.e., Hyytiälä, Sodankylä, Flakaliden, Norunda and Kalevansuo). Furthermore the sites with the long-term datasets were the most influential on M-S, having more weight on the likelihood. In particular Hyytiälä data had a strong influence on the M-S calibration. Instead of giving a weight to each site dataset according to the number of data points we preferred to use all the available information to calibrate and test the model.

The Analysis #2 and #3 tested the reliability of the model. Results from BMC can look quite severe (Tables 4 and 5) but in reality, while BMC tells which model is more likely to be the best, this does not mean that the worst model gives completely wrong predictions. Combining BMC with more classical model error quantifications provided a more complete picture about the models under evaluation. In our analyses BMC was assigning always near to 100 % probability of being correct to one version of the model because thousands of data points were used in the comparisons. However the NRM-SEs (Figs. 10 and 11) showed that the differences between M-S and S-S calibrations in predicting carbon and water fluxes are quite low for most of the sites.

Eddy-covariance network is expanding and flux data is currently available for hundreds of sites (Baldocchi, 2008). At some sites measurements have already been collected for more than a decade and it is likely that the inter-annual variability is well represented by the measurements. On the contrary, for other sites the fluxes have been measured just for a few years. Analysis #2 gave insights into which model version (i.e. M-S or S-S) is more appropriate to predict carbon and water fluxes for a site for which data are available. The multi-site calibration showed robust performances in predicting carbon and water fluxes when compared to site-specific versions. M-S was the best

GMDD

8, 5089–5137, 2015

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



calibration for 4 sites over 10 (Table 4) and the NRMSE between M-S and S-Ss were not significantly different in most of the cases. Except for Skyttorp, the evapotranspiration NRMSEs of the multi-site version were always higher than S-S NRMSE (Fig. 10b). As suggested by Analysis #1, the ET module in PRELES seems to be too simplistic, rendering the S-S calibration with better performance. Likely there are site-specific differences in flux-environmental driver relationships that could have been compensated by site-specific parameter estimates. Comparison results are more significant when a high number of data are involved and the measurements cover different years. We concluded that that both versions of the model (the S-S and the M-S) can be used to predict GPP and ET of sites for which flux data are available. The M-S calibration might be preferred for sites with short-term data series.

Results from Analysis #3 provided key indications about the regional applicability of PRELES. The performances of the M-S calibration and the Hyytiälä site-specific calibration were more reliable in predicting ET and GPP compared to the other site-specific versions. Hyytiälä S-S provided the most robust performances and was slightly better than the M-S. This is probably due to the fact that Hyytiälä was the most comprehensive dataset in terms of site years and data quality. This underlines the importance of long and carefully collected flux datasets, even a single site can provide model calibrations that can be applied at a wider spatial scale, since it covers a wide range of variability in climatic conditions. However, the good fit to Hyytiälä data can stem also from the development history and structure of PRELES (Mäkelä et al., 2008; Peltoniemi et al., 2015a) that was partly based on the understanding gained from the Hyytiälä measurements. Therefore, using a multi-site calibrated model for regional analysis represents a more conservative choice in terms of spatial representativeness, and the fact that not all required site conditions can be extrapolated by the modeller. From carbon modelling perspective, use of a few aerially representative sites with long and high quality records would likely be optimal.

Some ecosystem processes, such as photosynthesis, have been found to be sufficiently well understood and generalizable. Our analyses involved just one model and

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



these results could surely not to be expected with all models unless models are generic and calibrated with high quality data. PRELES has a simplified structure but as a computationally efficient model that requires easily available input data it is suitable for applications at regional scale. A useful application of PRELES could be in estimating the impact of climate change on Boreal forests. On one hand, we tested the model with data that covered a wide range of climatic conditions (i.e., D , PPF, R , T); but, on the other hand, we used fixed CO_2 atmosphere concentrations. Nevertheless, according to future climate scenario simulations (IPCC, 2007), CO_2 concentration is expected to increase dramatically in the future, having a strong impact on Boreal forests (Kalliokoski et al., 2015). The CO_2 effects on forest photosynthetic activity has been already estimated by other studies (Kolari et al., 2009); implementing this information in PRELES structure (Kalliokoski et al., 2015) will allow us to use PRELES to make estimates under future climatic scenarios.

Acknowledgements. We thank Pasi Kolari for his general advice in acquiring the data.

The study was supported by the European project “Enabling Intelligent GMES Services for Carbon and Water Balance Modeling of Northern Forest Ecosystems” and the LIFE+ financial instrument of the European Union (LIFE12 ENV/FI/000409 Monimet, LIFE09 ENV/FI/000571 Climforisk).

The authors thank their colleagues for continuing support and discussion around the coffee breaks. The editor thanks X. Y. Furore and another referee for assisting in evaluating this paper.

References

- Amiro, B. D., Orchansky, A. L., Barr, A. G., Black, T. A., Chambers, S. D., Chapin III, F. S., Goulden, M. L., Litvak, M., Liu, H. P., McCaughey, J. H., McMillan, A., and Randerson, J. T.: The effect of post-fire stand age on the boreal forest energy balance, *Agr. Forest Meteorol.*, 140, 41–50, 2006.
- Aubinet, M., Vesala, T., and Papale, D.: *Eddy Covariance: a Practical Guide to Measurement and Data Analysis*, Springer Science & Business Media, 460 pp., London, UK, 2012.

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Baldocchi, D.: TURNER REVIEW No. 15. “Breathing” of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems, *Aust. J. Bot.*, 56, 1–26, 2008.

Brümmer, C., Black, T. A., Jassal, R. S., Grant, N. J., Spittlehouse, D. L., Chen, B., Nestic, Z., Amiro, B. D., Arain, M. A., Barr, A. G., Bourque, C. P.-A., Coursolle, C., Dunn, A. L., Flanagan, L. B., Humphreys, E. R., Lafleur, P. M., Margolis, H. A., McCaughey, J. H., and Wofsy, S. C.: How climate and vegetation type influence evapotranspiration and water use efficiency in Canadian forest, peatland and grassland ecosystems, *Agr. Forest Meteorol.*, 153, 14–30, 2012.

Campolongo, F., Cariboni, J., and Saltelli, A.: An effective screening design for sensitivity analysis of large models, *Environ. Modell. Softw.*, 22, 1509–1518, doi:10.1016/j.envsoft.2006.10.004, 2007.

Farquhar, G. D., von Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species, *Planta*, 149, 78–90, 1980.

Foken, T.: The energy balance closure problem: an overview, *Ecol. Appl.*, 18, 1351–1367, 2008.

Friend, A. D., Lucht, W., Rademacher, T. T., Keribin, R., Betts, R., Cadule, P., Ciais, P., Clark, D. B., Dankers, R., Falloon, P. D., Ito, A., Kahana, R., Kleidon, A., Lomas, M. R., Nishina, K., Ostberg, S., Pavlick, R., Peylin, P., Schaphoff, S., Vuichard, N., Warszawski, L., Wiltshire, A., and Woodward, F. I.: Carbon residence time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO₂, *P. Natl. Acad. Sci. USA*, 111, 3280–3285, doi:10.1073/pnas.1222477110, 2014.

Green, E. J., MacFarlane, D. W., and Valentine, H. T.: Bayesian synthesis for quantifying uncertainty in predictions from process models, *Tree Physiol.*, 20, 415–419, doi:10.1093/treephys/20.5-6.415, 2000.

Hari, P. and Kulmala, M.: Station for measuring ecosystem–atmosphere relations, *Boreal Environ. Res.*, 10, 315–322, 2005.

Hartig, F., Dyke, J., Hickler, T., Higgins, S. I., O’Hara, R. B., Scheiter, S., and Huth, A.: Connecting dynamic vegetation models to data – an inverse perspective, *J. Biogeogr.*, 39, 2240–2252, doi:10.1111/j.1365-2699.2012.02745.x, 2012.

Heinsch, F. A., Zhao, M., Running, S. W., Kimball, J. S., Nemani, R. R., Davis, K. J., Bolstad, P. V., Cook, B. D., Desai, A. R., Ricciuto, D. M., Law, B. E., Oechel, W. C., Kwon, H., Luo, H., Wofsy, S. C., Dunn, A. L., Munger, J. W., Baldocchi, D. D., Xu, L., Hollinger, D. Y.,

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Launiainen, S., Katul, G. G., Kolari, P., Vesala, T., and Hari, P.: Empirical and optimal stomatal controls on leaf and ecosystem level CO₂ and H₂O exchange rates, *Agr. Forest Meteorol.*, 151, 1672–1689, 2011.
- Leuning, R., Kelliher, F. M., de Pury, D., and Schulze, E.-D.: Leaf nitrogen, photosynthesis, conductance and transpiration: scaling from leaves to canopies, *Plant Cell Environ.*, 18, 1183–1200, 1995.
- Liu, J., Chen, J. M., Cihlar, J., and Park, W. M.: A process-based boreal ecosystem productivity simulator using remote sensing inputs, *Remote Sens. Environ.*, 62, 158–175, 1997.
- Lohila, A., Laurila, T., Aro, L., Aurela, M., Tuovinen, J.-P., Laine, J., Kolari, P., and Minkkinen, K.: Carbon dioxide exchange above a 30-year-old Scots pine plantation established on organic-soil cropland, *Boreal Environ. Res.*, 12, 141–157, 2007.
- Lohila, A., Minkkinen, K., Aurela, M., Tuovinen, J.-P., Penttilä, T., Ojanen, P., and Laurila, T.: Greenhouse gas flux measurements in a forestry-drained peatland indicate a large carbon sink, *Biogeosciences*, 8, 3203–3218, doi:10.5194/bg-8-3203-2011, 2011.
- Lundin, L.-C., Halldin, S., Lindroth, A., Cienciala, E., Grelle, A., Hjelm, P., Kellner, E., Lundberg, A., Mölder, M., Morén, A.-S., Nord, T., Seibert, J., and Stähli, M.: Continuous long-term measurements of soil–plant–atmosphere variables at a forest site, *Agr. Forest Meteorol.*, 98, 53–73, 1999.
- Mäkelä, A., Hari, P., Berninger, F., Hänninen, H., and Nikinmaa, E.: Acclimation of photosynthetic capacity in Scots pine to the annual cycle of temperature, *Tree Physiol.*, 24, 369–376, doi:10.1093/treephys/24.4.369, 2004.
- Mäkelä, A., Kolari, P., Karimäki, J., Nikinmaa, E., Perämäki, M., and Hari, P.: Modelling five years of weather-driven variation of GPP in a boreal forest, *Agr. Forest Meteorol.*, 139, 382–398, doi:10.1016/j.agrformet.2006.08.017, 2006.
- Mäkelä, A., Pulkkinen, M., Kolari, P., Lagergren, F., Berbigier, P., Lindroth, A., Loustau, D., Nikinmaa, E., Vesala, T., and Hari, P.: Developing an empirical model of stand GPP with the LUE approach: analysis of eddy covariance data at five contrasting conifer sites in Europe, *Global Change Biol.*, 14, 92–108, doi:10.1111/j.1365-2486.2007.01463.x, 2008.
- Mammarella, I., Launiainen, S., Gronholm, T., Keronen, P., Pumpanen, J., Rannik, Ü., and Vesala, T.: Relative humidity effect on the high-frequency attenuation of water vapor flux measured by a closed-path eddy covariance system, *J. Atmos. Ocean. Tech.*, 26, 1856–1866, doi:10.1175/2009JTECHA1179.1, 2009.

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Medlyn, B. E., Duursma, R. A., and Zeppel, M. J.: Forest productivity under climate change: a checklist for evaluating model studies, *Wiley Interdiscip. Rev. Clim. Change*, 2, 332–355, 2011a.

Medlyn, B. E., Duursma, R. A., Eamus, D., Ellsworth, D. S., Prentice, I. C., Barton, C. V., Crous, K. Y., de Angelis, P., Freeman, M., and Wingate, L.: Reconciling the optimal and empirical approaches to modelling stomatal conductance, *Global Change Biol.*, 17, 2134–2144, 2011b.

Meyers, T. P. and Baldocchi, D. D.: A comparison of models for deriving dry deposition fluxes of O₃ and SO₂ to a forest canopy, *Tellus B*, 40, 270–284, doi:10.1111/j.1600-0889.1988.tb00297.x, 1988.

Minunno, F., van Oijen, M., Cameron, D. R., Cerasoli, S., Pereira, J. S., and Tomé, M.: Using a Bayesian framework and global sensitivity analysis to identify strengths and weaknesses of two process-based models differing in representation of autotrophic respiration, *Environ. Modell. Softw.*, 42, 99–115, doi:10.1016/j.envsoft.2012.12.010, 2013a.

Minunno, F., van Oijen, M., Cameron, D. R., and Pereira, J. S.: Selecting parameters for bayesian calibration of a process-based model: a methodology based on canonical correlation analysis, *SIAMASA Journal on Uncertainty Quantification*, 1, 370–385, doi:10.1137/120891344, 2013b.

Monteith, J. L.: Evaporation and Environment, in: *The State and Movement of Water in Living Organisms*, edited by: Fogg, C. E., 205–234, Cambridge University Press, Cambridge, UK, 1981.

Morris, M. D.: Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33, 161–174, 1991.

Ogée, J., Brunet, Y., Loustau, D., Berbigier, P., and Delzon, S.: MuSICA, a CO₂, water and energy multilayer, multileaf pine forest model: evaluation from hourly to yearly time scales and sensitivity analysis, *Global Change Biol.*, 9, 697–717, 2003.

Ojanen, P., Minkkinen, K., Lohila, A., Badorek, T., and Penttilä, T.: Chamber measured soil respiration: a useful tool for estimating the carbon balance of peatland forest soils?, *Forest Ecol. Manag.*, 277, 132–140, doi:10.1016/j.foreco.2012.04.027, 2012.

Olchev, A., Ibrom, A., Priess, J., Erasmi, S., Leemhuis, C., Twele, A., Radler, K., Kreilein, H., Panferov, O., and Gravenhorst, G.: Effects of land-use changes on evapotranspiration of tropical rain forest margin area in Central Sulawesi (Indonesia): modelling study with a regional SVAT model, *Ecol. Model.*, 212, 131–137, 2008.

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Peltoniemi, M., Pulkkinen, M., Kolari, P., Duursma, R. A., Montagnani, L., Wharton, S., Lagergren, F., Takagi, K., Verbeeck, H., Christensen, T., Vesala, T., Falk, M., Loustau, D., and Mäkelä, A.: Does canopy mean nitrogen concentration explain variation in canopy light use efficiency across 14 contrasting forest sites?, *Tree Physiol.*, 32, 200–218, doi:10.1093/treephys/tpr140, 2012.

Peltoniemi, M., Pulkkinen, M., Aurela, M., Pumpanen, J., Kolari, P., and Mäkelä A.: A semi-empirical model of boreal forest gross primary production, evapotranspiration, and soil water – calibration and sensitivity analysis, *Boreal Environ. Res.*, 20, 151–171, 2015a.

Peltoniemi, M., Markkanen, T., Härkönen S., Muukkonen. P. T., T., Aalto, T., and Mäkelä A.: Consistent estimates of gross primary production of Finnish forests – comparison of estimates of two process models, *Boreal Environ. Res.*, 20, 196–212, 2015b.

Pihlatie, M. K., Kiese, R., Brüggemann, N., Butterbach-Bahl, K., Kieloaho, A.-J., Laurila, T., Lohila, A., Mammarella, I., Minkkinen, K., Penttilä, T., Schönborn, J., and Vesala, T.: Greenhouse gas fluxes in a drained peatland forest during spring frost-thaw event, *Biogeosciences*, 7, 1715–1727, doi:10.5194/bg-7-1715-2010, 2010.

Raddatz, T. J., Reick, C. H., Knorr, W., Kattge, J., Roeckner, E., Schnur, R., Schnitzler, K.-G., Wetzol, P., and Jungclaus, J.: Will the tropical land biosphere dominate the climate-carbon cycle feedback during the twenty-first century?, *Clim. Dynam.*, 29, 565–574, doi:10.1007/s00382-007-0247-8, 2007.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grunwald, T., Havrankova, K., Ilvesniemi, H., Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global Change Biol.*, 11, 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x, 2005.

Song, X., Bryan, B. A., Paul, K. I., and Zhao, G.: Variance-based sensitivity analysis of a forest growth model, *Ecol. Model.*, 247, 135–143, doi:10.1016/j.ecolmodel.2012.08.005, 2012.

Song, X., Bryan, B. A., Almeida, A. C., Paul, K. I., Zhao, G., and Ren, Y.: Time-dependent sensitivity of a process-based ecological model, *Ecol. Model.*, 265, 114–123, doi:10.1016/j.ecolmodel.2013.06.013, 2013.

Thum, T., Aalto, T., Laurila, T., Aurela, M., Lindroth, A., and Vesala, T.: Assessing seasonality of biochemical CO₂ exchange model parameters from micrometeorological flux observations

at boreal coniferous forest, *Biogeosciences*, 5, 1625–1639, doi:10.5194/bg-5-1625-2008, 2008.

van Oijen, M., Höglind, M., Hanslin, H. M., and Caldwell, N.: Process-based modeling of Timothy regrowth, *Agron. J.*, 97, 1295–1303, doi:10.2134/agronj2004.0251, 2005.

5 van Oijen, M., Cameron, D. R., Butterbach-Bahl, K., Farahbakhshazad, N., Jansson, P.-E., Kiese, R., Rahn, K.-H., Werner, C., and Yeluripati, J. B.: A Bayesian framework for model calibration, comparison and analysis: application to four models for the biogeochemistry of a Norway spruce forest, *Agr. Forest Meteorol.*, 151, 1609–1621, doi:10.1016/j.agrformet.2011.06.017, 2011.

10 van Oijen, M., Reyer, C., Bohn, F. J., Cameron, D. R., Deckmyn, G., Flechsig, M., Härkönen, S., Hartig, F., Huth, A., Kiviste, A., Lasch, P., Mäkelä, A., Mette, T., Minunno, F., and Rammer, W.: Bayesian calibration, comparison and averaging of six forest models, using data from Scots pine stands across Europe, *Forest Ecol. Manag.*, 289, 255–268, doi:10.1016/j.foreco.2012.09.043, 2013.

GMDD

8, 5089–5137, 2015

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Table 1. List of parameters used in the calibration.

Name	Symbol	Units	Minimum	Maximum
Potential light use efficiency	β	gC molPPFD^{-1}	0.2	2.5
Delay parameter for ambient temperature response	τ	–	1	25
Threshold for state of acclimation change	X_0	$^{\circ}\text{C}$	–20	20
Acclimation state maximum	S_{\max}	$^{\circ}\text{C}$	2.3	30
Sensitivity parameter for VPD response	κ	kPa^{-1}	–1	–0.001
Light modifier parameter	γ	$\text{molPPFD}^{-1}\text{m}^{-2}$	1.03×10^{-4}	5.03×10^{-1}
Threshold for linear decrease of $f_{W,P}$	ρ_P	–	0	0.999
Transpiration parameters	α	$\text{mm}(\text{gC m}^{-2}\text{kPa}^{1-\lambda})^{-1}$	1×10^{-6}	10
Parameter adjusting water use efficiency with VPD	λ	–	1×10^{-4}	1.2
Evaporation parameter	χ	mm molPPFD^{-1}	0	2.5
Threshold for linear decrease of $f_{W,E}$	ρ_E	–	0	0.999
Parameter adjusting water use efficiency if soil water is limiting GPP ($f_{W,P}$)	ν	–	1×10^{-4}	5

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Table 2. Continued.

	Annual P (mm)	Annual T (°C)	Years of flux measurements	Ndata GPP	Ndata E	Reference
Hyttiälä	709	2.9	2000–2010	3391	3601	Hari and Kulmala (2005); Kolari et al. (2009)
Sodankylä	527	−0.4	2001–2009	2698	2878	Thum et al. (2008)
Flakaliden	600	2.3	1997, 1998, 2001, 2002, 2007–2009	1414	1653	Berggren et al. (2008)
Norunda	527	5.5	1996–1999, 2003	1476	1499	Lundin et al. (1999); Lindroth et al. (2008)
Kalevansuo	606	4.3	2004–2009	1144	1154	Pihlatie et al. (2010); Lohila et al. (2011); Ojanen et al. (2012)
Knottåsen	613	3.4	2007, 2009	680	699	Berggren et al. (2008)
Alkkia	681	4.1	2002–2004	357	404	Lohila et al. (2007)
Skyttorp	830	7.1	2005	267	282	–
CAge12yr	709	2.9	2002	235	237	Kolari et al. (2004)
CAge75yr	709	2.9	2002	204	198	Kolari et al. (2004)

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Table 3. R^2 and slopes calculated for the multi-site and site-specific calibration of Analysis #1 – “Global or local”.

	GPP				E			
	multi-site R^2	slope	site-specific R^2	slope	multi-site R^2	slope	site-specific R^2	slope
Hyytiälä	0.96	0.98	0.96	0.98	0.89	0.90	0.89	0.92
Sodankylä	0.89	0.82	0.91	0.89	0.75	0.79	0.80	0.80
Flakaliden	0.79	1.09	0.81	0.80	0.68	0.87	0.71	0.77
Norunda	0.89	0.97	0.90	0.92	0.82	0.84	0.85	0.85
Kalevansuo	0.93	0.95	0.95	0.97	0.87	0.85	0.91	0.88
Knottåsen	0.91	0.78	0.91	0.93	0.89	0.74	0.89	0.86
Alkkia	0.89	0.80	0.89	0.88	0.83	0.62	0.84	0.89
Skyttorp	0.80	0.87	0.81	0.85	0.72	0.86	0.72	0.81
CAge12yr	0.73	0.77	0.84	0.87	0.71	0.80	0.75	0.72
CAge75yr	0.93	1.10	0.95	0.97	0.88	0.83	0.92	0.89

GMDD

8, 5089–5137, 2015

Is a generic calibration sufficient?

F. Minunno et al.

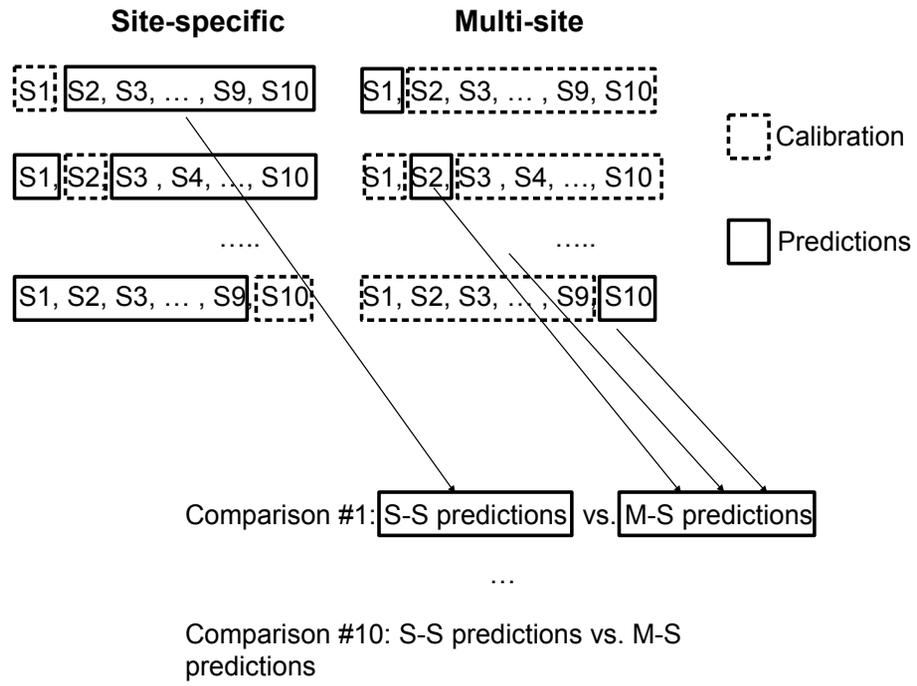


Figure 1. Calibration scheme used in the Analysis #3. 10 site-specific calibrations and 10 multi-site calibrations were performed. In the multi-site calibrations one site was excluded in turn from the calibration.

Title Page

Abstract Introduction
Conclusions References
Tables Figures

⏪ ⏩
◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Is a generic calibration sufficient?

F. Minunno et al.

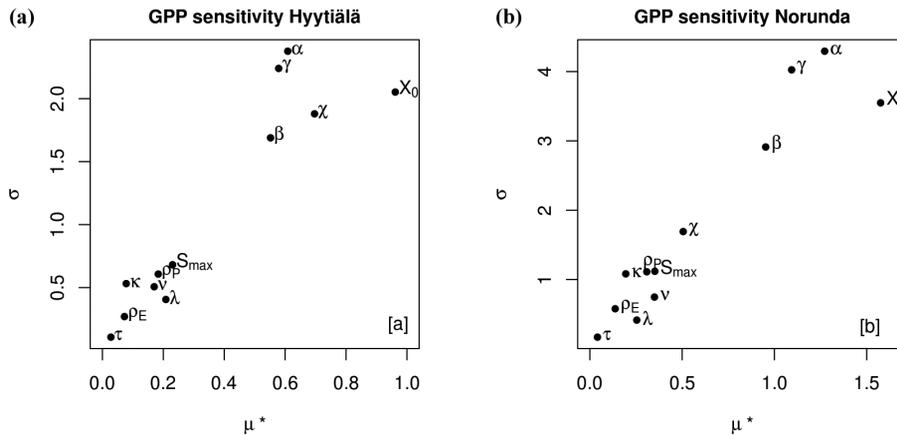


Figure 2. Plots of the sensitivity metrics μ^* and σ calculated for gross primary production (GPP) at Hyytiälä (a) and Norunda (b). Results from Hyytiälä can be considered representative for the remaining sites.

Is a generic calibration sufficient?

F. Minunno et al.

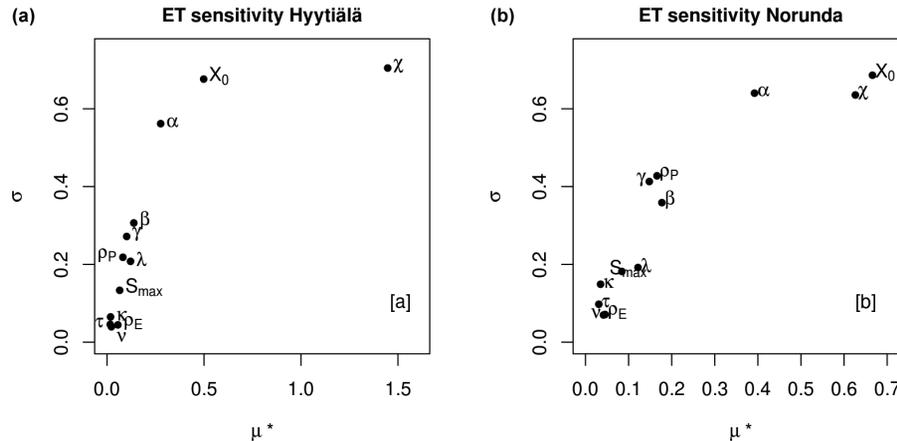


Figure 3. Plots of the sensitivity metrics μ^* and σ calculated for evapotranspiration (ET) at Hyytiälä (a) and Norunda (b). Results from Hyytiälä can be considered representative for the remaining sites.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Is a generic calibration sufficient?

F. Minunno et al.

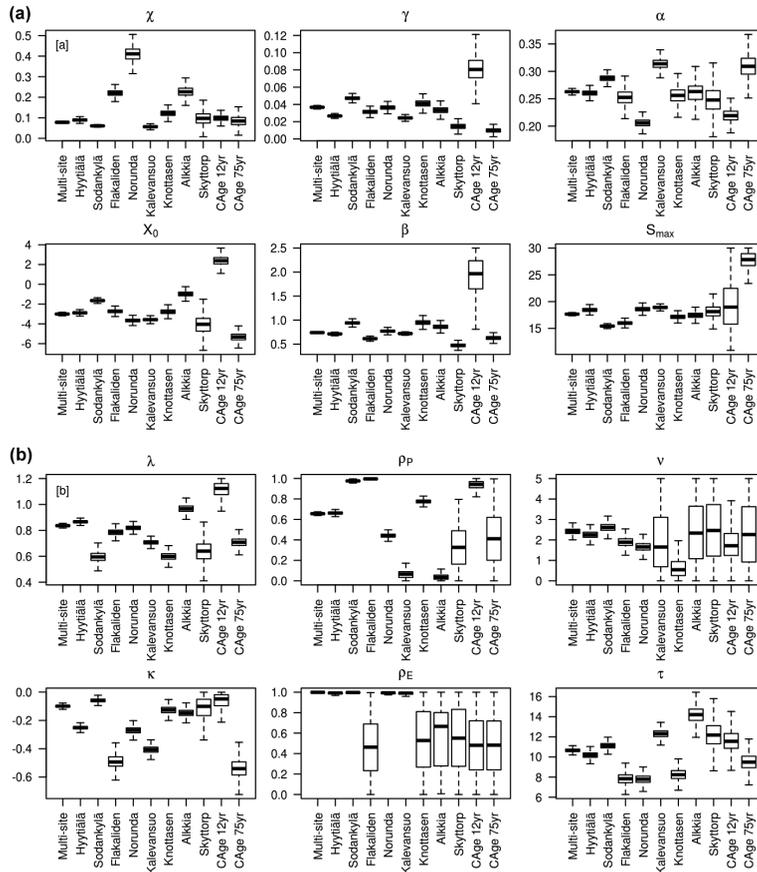


Figure 5. Marginal posterior distributions of PRELES parameters obtained through the multi-site calibration and the site-specific calibrations. **(a)** Parameters of high sensitivity according to the Morris sensitivity analysis, **(b)** parameters of medium and low sensitivity according to the Morris sensitivity analysis.

[Title Page](#)

[Abstract](#) [Introduction](#)

[Conclusions](#) [References](#)

[Tables](#) [Figures](#)

[◀](#) [▶](#)

[◀](#) [▶](#)

[Back](#) [Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

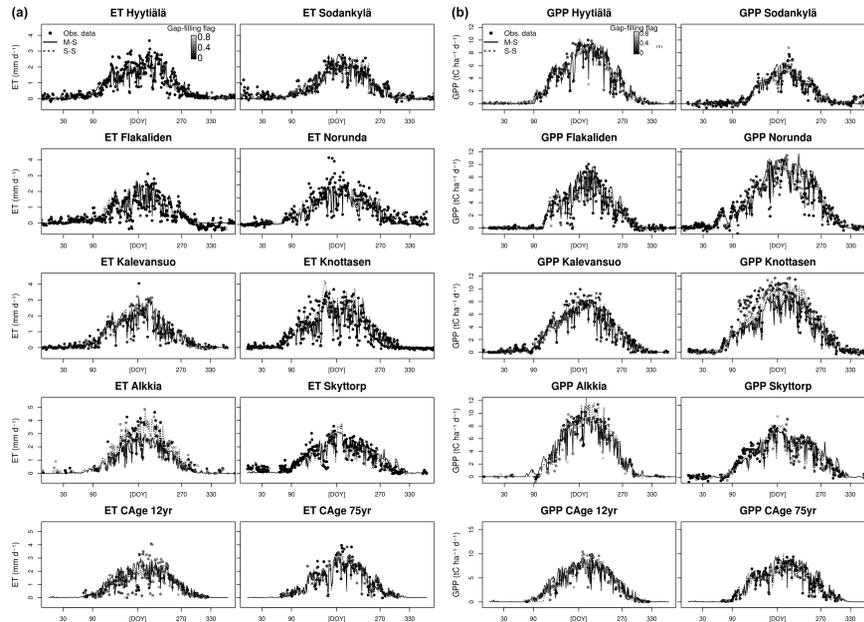


Figure 6. (a) Daily evapotranspiration at each experimental site for a year randomly selected from the dataset. Sites are ordered according to the number of data points available for model calibration. Dots represent the observations and are coloured in grey scale according to the fraction of gap-filled data in a day (i.e., black = all data were observed, white = all data were gap-filled). The lines are PRELES predictions; the dashed line is the output from the site-specific calibrations, while the continuous lines represent the multi-site calibration. (b) Daily gross primary production at each experimental site for a year randomly selected from the dataset. Sites are ordered according to the number of data points available for model calibration. Dots represent the observations and are coloured in grey scale according to the fraction of gap-filled data in a day (i.e., black = all data were observed, white = all data were gap-filled). The lines are PRELES predictions; the dashed line is the output from the site-specific calibrations, while the continuous lines represent the multi-site calibration.

Is a generic calibration sufficient?

F. Minunno et al.

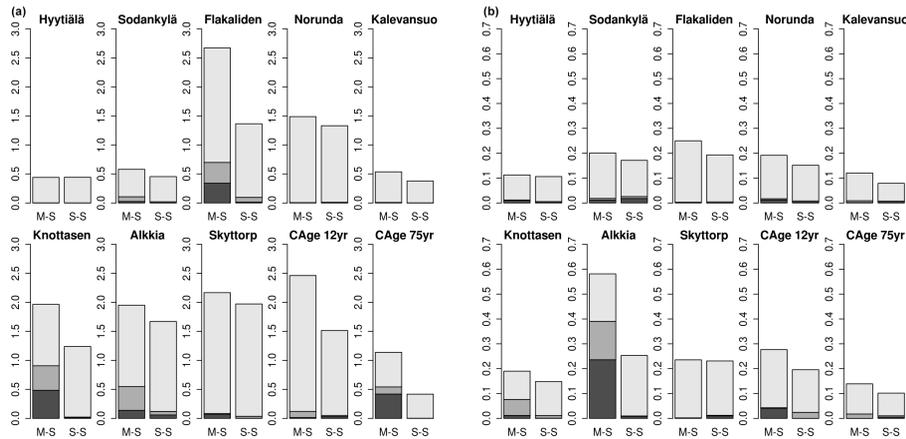


Figure 7. (a) Mean squared error decomposition for GPP. The first bar in the plots is the *MSE* calculated with PRELES outputs generated from the multi-site calibration (M-S), while the second bar is the error of the site-specific calibration (S-S). The maximum values of the y-axes were set to the square of the mean of the GPP observed values of all the sites. (b) Mean squared error decomposition for ET. The first bar in the plots is the *MSE* calculated with PRELES outputs generated from the multi-site calibration (M-S), while the second bar is the error of the site-specific calibration (S-S). The maximum values of the y-axes were set to the square of the mean of the ET observed values of all the sites.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)

[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)


Is a generic calibration sufficient?

F. Minunno et al.

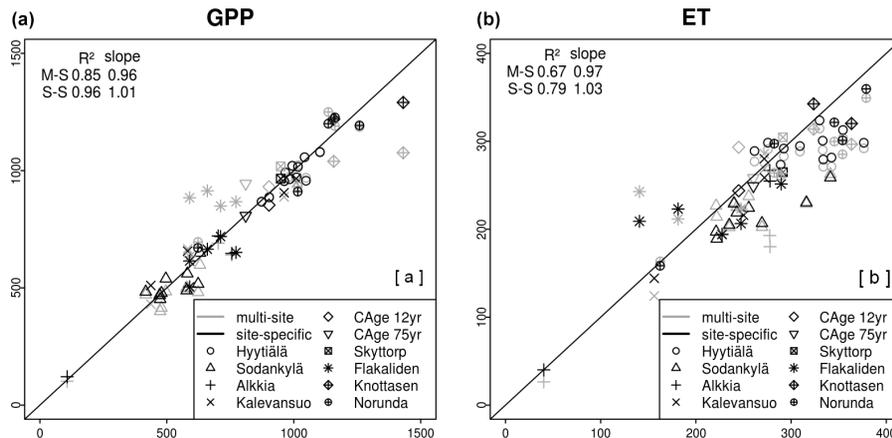


Figure 8. (a) Observed vs. simulated annual gross primary production. Each symbol corresponds to different site; while the colours, grey and black, refer to the multi-site (M-S) and site-specific (S-S) calibration, respectively. The daily observed and simulated data were summed to obtain the annual GPP values in the figure only if the quality flags of the daily GPP measurements were lower than 0.7. In the left up corner of the plotting area a table with the R^2 and slopes are reported for M-S and S-S. (b) Observed vs. simulated annual evapotranspiration. Each symbol corresponds to different site; while the colours, grey and black, refer to the multi-site (M-S) and site-specific (S-S) calibration, respectively. The daily observed and simulated data were summed to obtain the annual ET values in the figure only if the quality flags of the daily ET measurements were lower than 0.7. In the left up corner of the plotting area a table with the R^2 and slopes are reported for M-S and S-S.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

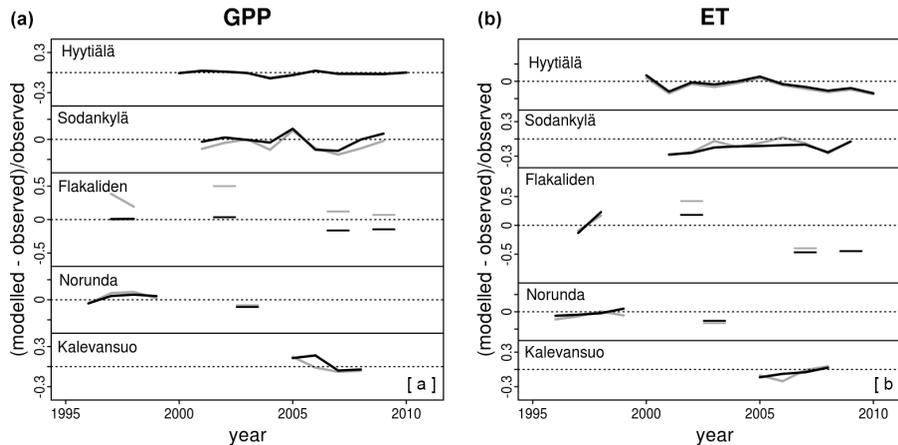


Figure 9. (a) Relative differences (i.e., $y = (\text{observed} - \text{modelled}) / \text{observed}$) between the modelled and observed annual sums of gross primary production; note that only days with flag lower than 0.7 were considered for annual sum calculations. The grey line refers to the predictions generated with the multi-site version, while the black line refers to the site-specific calibration. **(b)** Relative differences (i.e., $y = (\text{observed} - \text{modelled}) / \text{observed}$) between the modelled and observed annual sums of evapotranspiration; note that only days with flag lower than 0.7 were considered for annual sum calculations. The grey line refers to the predictions generated with the multi-site version, while the green line refers to the site-specific calibration.

Is a generic calibration sufficient?

F. Minunno et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

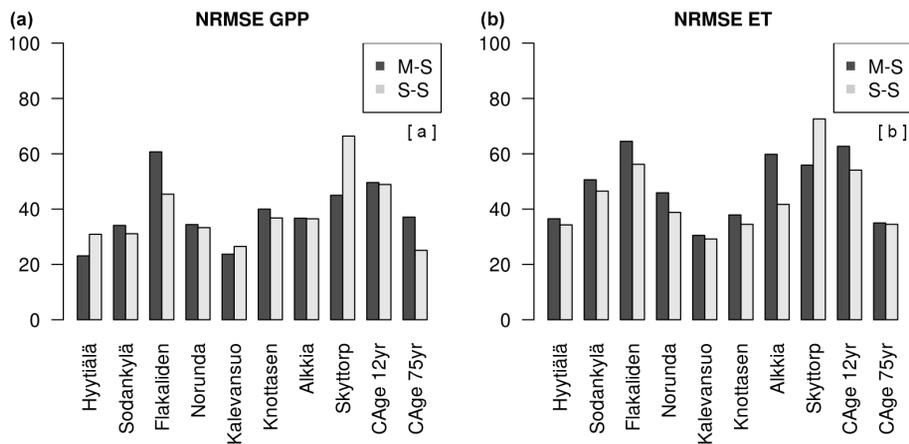


Figure 10. (a) Normalized root mean squared errors, for GPP. MSEs were normalized using the standard deviations of the observations. Sites are ordered from left to right according to the number of data points available for model calibration and evaluation. M-S and S-S refer to the multi-site and the site-specific calibration, respectively. (b) Normalized root mean squared errors, for ET. MSEs were normalized using the standard deviations of the observations. Sites are ordered from left to right according to the number of data points available for model calibration and evaluation. M-S and S-S refer to the multi-site and the site-specific calibration, respectively.

Is a generic calibration sufficient?

F. Minunno et al.

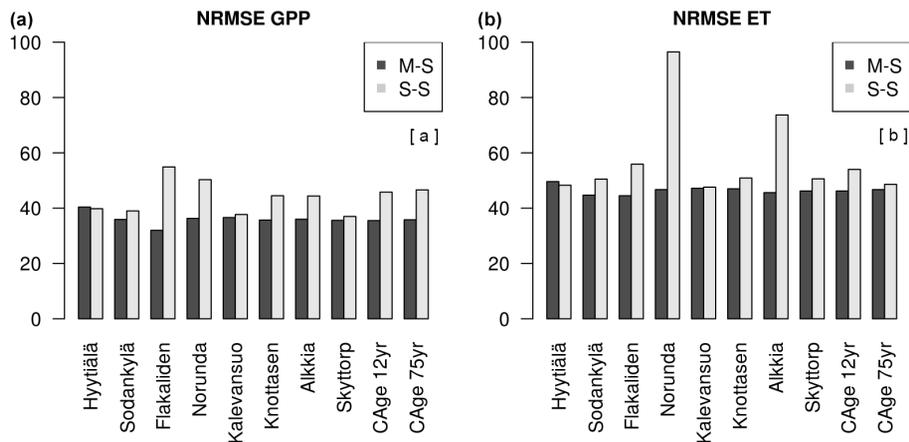


Figure 11. (a) Normalized root mean squared errors, for GPP. MSEs were normalized using the standard deviations of the observations. Sites are ordered from left to right according to the number of data points available for model calibration and evaluation. (b) Normalized root mean squared errors, for ET. MSEs were normalized using the standard deviations of the observations. Sites are ordered from left to right according to the number of data points available for model calibration and evaluation.

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)

[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)
