First results

Marieke Wesselkamp

May 26, 2021

1 Approach

Intra- and inter-annual variability in CO_2 turnover can be attributed partly to climatic differences and partly to differences in ecosystem functions over time. When modeling the relationship between climatic factors and CO_2 turnover statistically, the parameters of the model account for the functional effects. These can also be understood as the biotic effects of the ecosystem on the CO_2 turnover.

To quantify the magnitude of these effects, in this work I will use an approach that follows Wu et al. [2012]. The gross primary production (GPP) of an ecosystem is modeled through two subsequent climatic time series $P1 = p_1, ..., p_t$ and $P2 = p_{t+1}, ..., p_n$. This results in two different models, $M1$ and $M2$. From each of the two models, a forecast for $P2$ is generated, that will be termed $\widehat{P2}_{m1}$ and $\widehat{P2}_{m1}$.

The difference between $\widehat{P2}_{m1}$ and $\widehat{P2}_{m2}$,

$$P_{diff} = \widehat{P2}_{m2} - \widehat{P2}_{m1} \tag{1}$$

will be due to differences in the parameter values of $M1$ and $M2$. When extrapolating with $M1$ to $P2$ with fixed parameters, this corresponds to assuming a constant ecosystem status [Wu et al., 2012]. The difference between P_{diff} we observe thus represents the magnitude of change over time due ecosystem functional differences. Accordingly, the difference in prediction error e_{diff} can be attributed to varying functional properties.

$$e_{diff} = e_{P2} - e_{P1} \tag{2}$$

When the time series for GPP are not observed but simulated from a physical model such as PRELES (PREdict Light-Use efficiency, Evapotranspiration and Soil water content)[Peltoniemi et al., 2015], there is no missing climatic information to model this relationship statistically. PRELES is purely deterministic and as the relationship between the climatic input and the GPP as output should be completely detectable, we expect e_{diff} to be smaller for simulated GPP than for observed GPP. That means, the difference between $\widehat{P2}_{m1}$ and $\widehat{P2}_{m1}$ represents the effect of the ecosystem functions that are represented in PRELES through 30 parameters.

2 Setup

The gross primary production (GPP) of a boreal forest stand was modeled using five climate variables while taking into account the annual cycle. For this purpose, 13 years (2000-2012) of Soro (DK) stand data from the Profound database were used [Reyer et al., 2020]. This database also includes preprocessed observations of GPP. The dataset containing these observations for GPP is referenced as $D1$ in the following. The semi-empirical process model PRELES was used to additionally simulate GPP for Soro with appropriate climate variables. This dataset is referenced in the following as $D2$. A selection of seven parameters from PREles was calibrated for these 13 years using Bayesian optimization [Hartig et al., 2019]. The remaining parameters were left at their default values. All experiments were performed with both data sets, unless otherwise noted.

The statistical model that I used was a feed-forward neural network (Mulitlayer Perceptron, MLP) with three hidden layers with 32, 32 and 16 hidden nodes, respectively. The architecture was selected using a random search in Thea cross-validation setting on the 2000 and 2001 Soro data. The network performance was quantified with the mean absolute error (MAE).

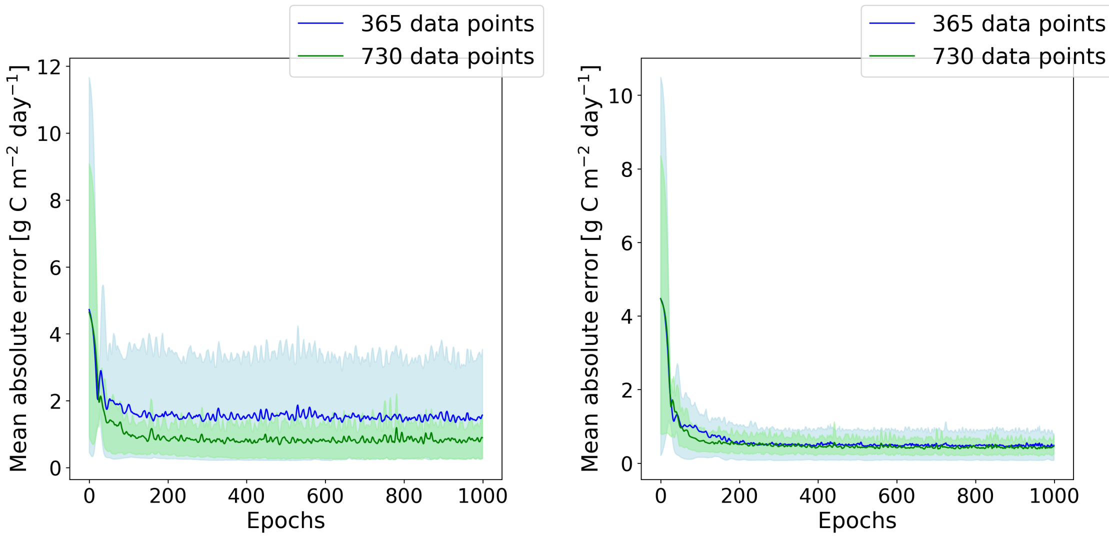
2.1 Splits

Data sets $D1$ and $D2$ were each split into two equally sized time series, D_P1 and D_P2. The length of the time series was determined based on the precision of the neural network predictions. A confidence interval was calculated from the results of the 5-fold cross-validation. While using one year of dataset D1 (i.e., 365 data points), the 95% confidence interval of the mean absolute error after convergence ranged between errors of 0.35 and 3.31, using two years (i.e., 730 data points) it was only 0.29 and 1.16 (see also figure 1. For data set $D2$, the overall precision was better, so that when using one year, the confidence interval was between 0.16 and 1.0, and when using two years, it was between 0.21 and 0.76. Based on these results, the experiment was first conducted with time series $P1$ and $P2$ each two years long, for a total of four years: 2000/2001 and 2002/2003, respectively (see figure 2).

3 Experiments

3.1 The effect of ecosystem functional changes

Following Wu et al. [2012], the model parameters for one data split, i.e. the years 2000 and 2001, were applied to predict the GPP for the years 2002 and 2003. The difference in modeled fluxes represents the variability in yearly GPP that is driven by the changing model parameters and can be interpreted as the ecosystems' functional changes. This was done for both data sets, D1 with observed GPP (see figure 3) and D2 with simulated GPP (see figure 4). The difference between predictions was quantified on the scale of the splits that is two years. Mean absolute prediction errors for the predicted time series for both models and their difference can be found in table 1.



(a) Training on D1.

(b) Training on D2.

Figure 1: Validation losses after training on time series of different size.

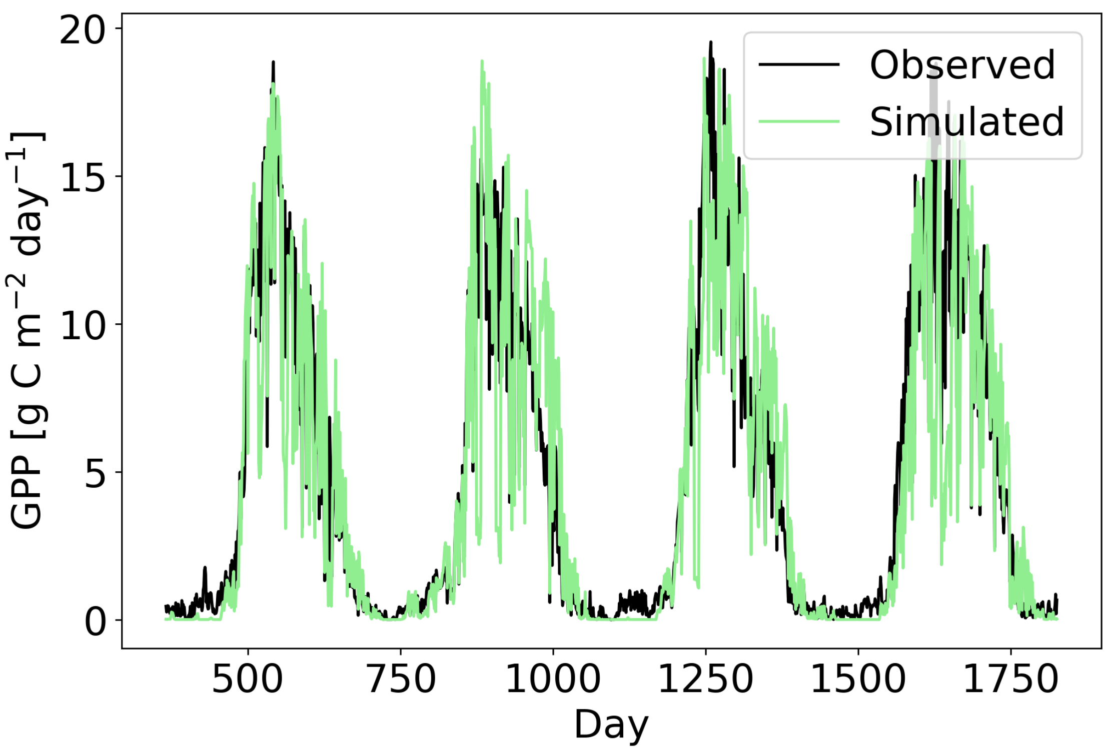


Figure 2: The observed and simulated gross primary production at the Soro site (DK) for the years from 2000 to 2003. The years 2000/2001 correspond to split P1, the years 2002/2003 to split P2.

Table 1: Caption

	$MAE_{\widehat{P2}_{m1}}$	$MAE_{\widehat{P2}_{m2}}$	$MAE_{\widehat{P2}_{diff}}$
D1	0.93	0.68	0.25
D2	0.30	0.21	0.09

The same experiment was subsequently repeated with focus on different data splits. Doing so, first the length of the time series in both splits, $P1$ and $P2$, was step wise increased from 250 up to 2000 days. The step size was ten days. Results are displayed in figure 6.

3.2 Bivariate Correlation Analysis

To come...

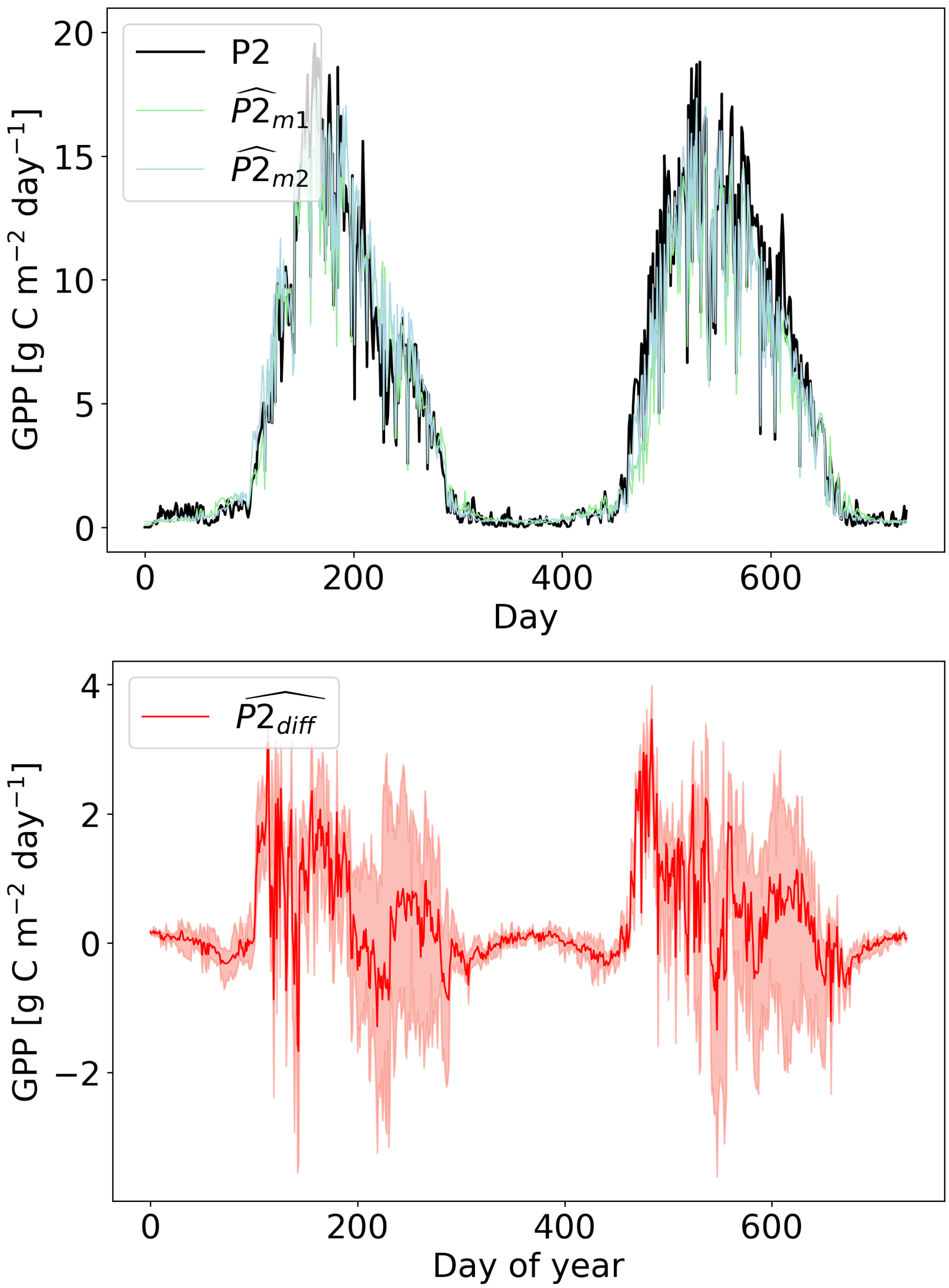


Figure 3: Top: Predictions of two neural networks (M1 and M2) for two years (2002,2003) of measured gross primary productivity (D1) at the Soro site (DK). Bottom: Difference in prediction.

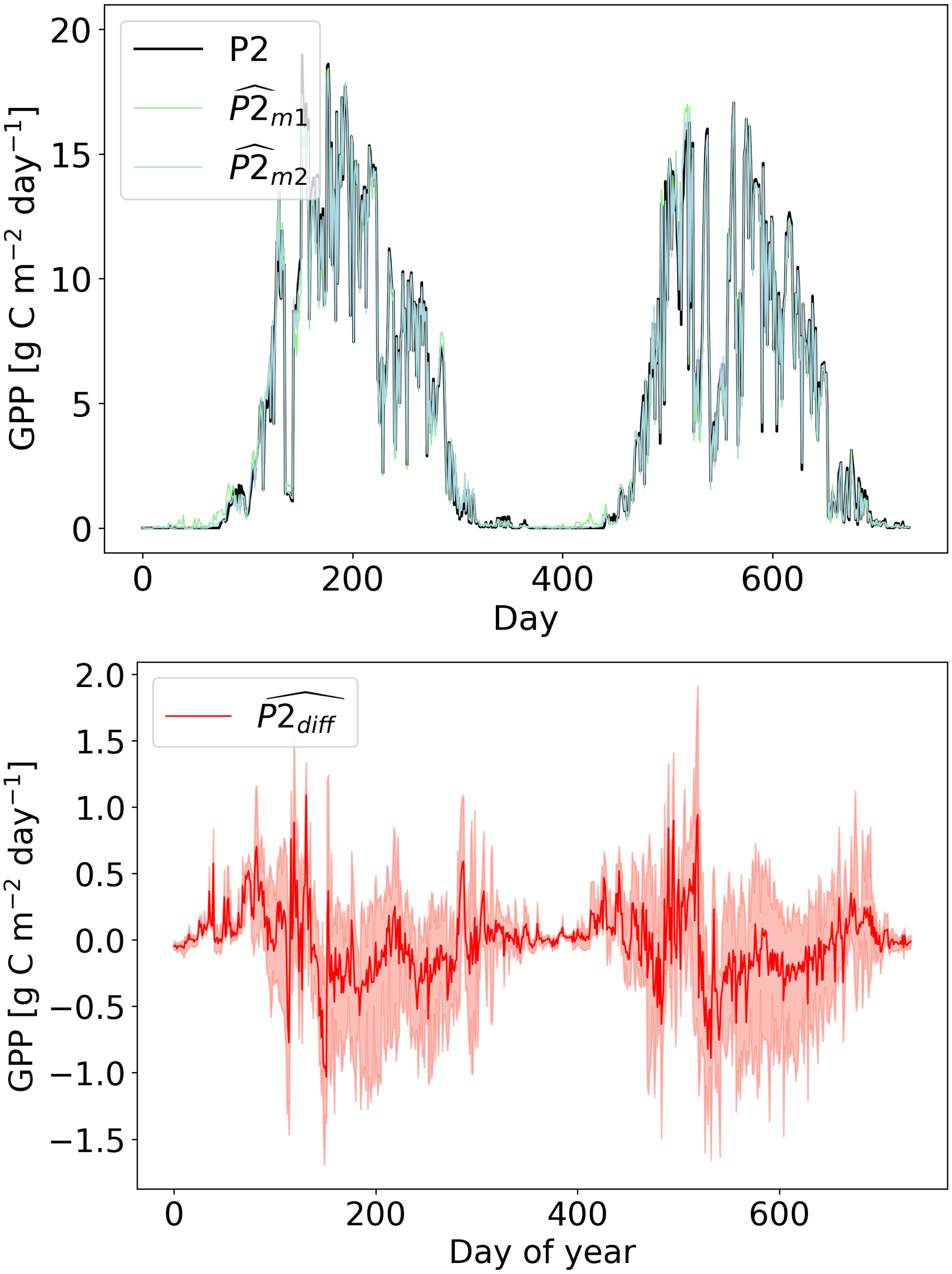


Figure 4: Top: Predictions of two neural networks (M1 and M2) for two years (2002,2003) of simulated gross primary productivity (D2) at the Soro site (DK). Bottom: Difference in prediction.

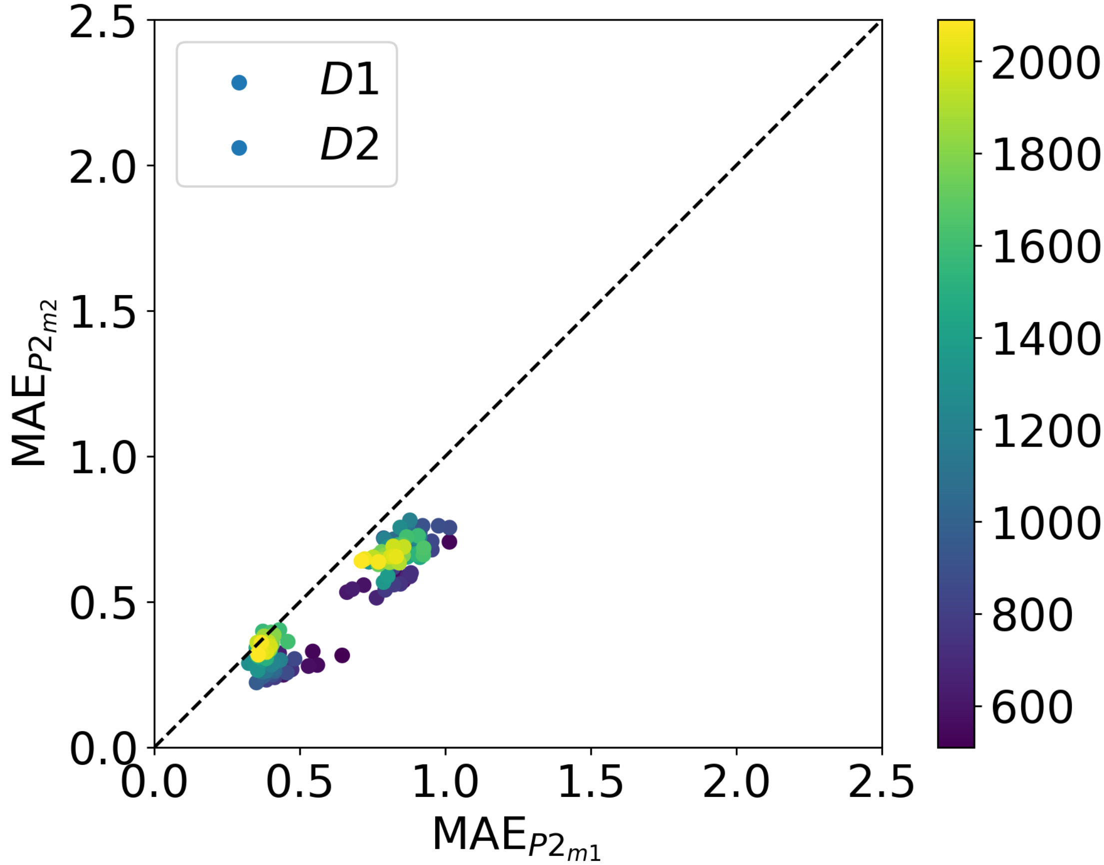


Figure 5: Mean absolute errors over the two years for predictions of two neural networks (M1 and M2) for time series of increasing length of measured (D1, upper blob) and simulated (D1, lower blob) gross primary productivity at the Soro site (DK)

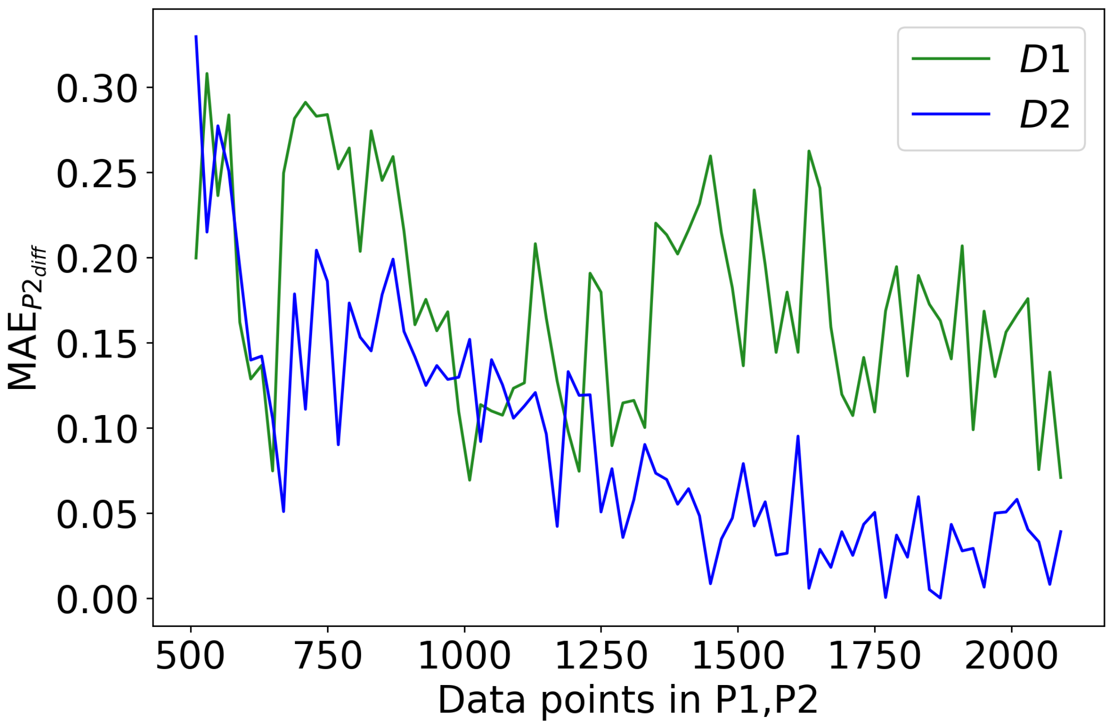


Figure 6: Difference in mean absolute prediction error of the observed (D1) and simulated (D2) gross primary productivity for time series of increasing length at the Soro (DK) site.

References

J. Wu, L. van der Linden, G. Lasslop, N. Carvalhais, K. Pilegaard, C. Beier, and A. Ibrom. Effects of climate variability and functional changes on the interannual variation of the carbon balance in a temperate deciduous forest. *Biogeosciences*, 9(1):13–28, 2012. doi: 10.5194/bg-9-13-2012. URL `https://bg.copernicus.org/articles/9/13/2012/`.

Mikko Peltoniemi, Minna Pulkkinen, Mika Aurela, Jukka Pumpanen, Pasi Kolari, and Annikki Mäkelä. A semi-empirical model of boreal-forest gross primary production, evapotranspiration, and soil water-calibration and sensitivity analysis. 2015.

Christopher PO Reyer, Ramiro Silveyra Gonzalez, Klara Dolos, Florian Hartig, Ylva Hauf, Matthias Noack, Petra Lasch-Born, Thomas Rötzer, Hans Pretzsch, Henning Meesenburg, et al. The profound database for evaluating vegetation models and simulating climate impacts on european forests. *Earth system science data*, 12(2):1295–1320, 2020.

Florian Hartig, Francesco Minunno, and Stefan Paul. *BayesianTools: General-Purpose MCMC and SMC Samplers and Tools for Bayesian Statistics*, 2019. URL `https://CRAN.R-project.org/package=BayesianTools`. R package version 0.1.7.