

RESEARCH ARTICLE

Open Access



Regional-scale data assimilation with the Spatially Explicit Individual-based Dynamic Global Vegetation Model (SEIB-DGVM) over Siberia

Hazuki Arakida^{1,2*} , Shunji Kotsuki^{1,3,4}, Shigenori Otsuka^{1,5,6}, Yohei Sawada^{7,1} and Takemasa Miyoshi^{1,5,6,8,9*}

Abstract

This study examined the regional performance of a data assimilation (DA) system that couples the particle filter and the Spatially Explicit Individual-based Dynamic Global Vegetation Model (SEIB-DGVM). This DA system optimizes model parameters of defoliation and photosynthetic rate, which are sensitive to phenology in the SEIB-DGVM, by assimilating satellite-observed leaf area index (LAI). The experiments without DA overestimated LAIs over Siberia relative to the satellite-observed LAI, whereas the DA system successfully reduced the error. DA provided improved analyses for the LAI and other model variables consistently, with better match with satellite observed LAI and with previous studies for spatial distributions of the estimated overstory LAI, gross primary production (GPP), and aboveground biomass. However, three main issues still exist: (1) the estimated start date of defoliation for overstory was about 40 days earlier than the in situ observation, (2) the estimated LAI for understory was about half of the in situ observation, and (3) the estimated overstory LAI and the total GPP were overestimated compared to the previous studies. Further DA and modeling studies are needed to address these issues.

Keywords: Data assimilation, Particle filter, Individual-based DGVM, Overstory LAI, Phenology, Siberia

1 Introduction

Terrestrial ecosystem models (TEMs) have been developed as components of earth system models that simulate carbon, water, and energy cycles between the atmosphere and terrestrial ecosystems (Fisher et al. 2017; Peng 2000). These models are indispensable elements of procedures aimed at predicting (i) functional alterations in ecosystems under the changing climate, and (ii) the resulting changes in feedback processes. However, different studies used different parameterizations and climate forcing data and produced diverse simulation outputs. Such diversity indicates that the

projections of current TEMs have large uncertainties (Ahlström et al. 2012; Cheaib et al. 2012; Friend et al. 2014; Ito et al. 2017; Rogers et al. 2017).

Recent TEMs have been developed to incorporate data assimilation (DA) that mitigates such uncertainties by assimilating observations (Luo et al. 2011; Peng et al. 2011; Kaminski et al. 2013). At the flux measurement sites, we can use fine-timescale (e.g., 30-min collection interval) flux data and carbon stock data. Williams et al. (2005) estimated model parameters and carbon pools of a box-type TEM by assimilating daily averaged carbon flux data collected over 3 years and occasional carbon stock data. Likewise, in situ measurements have been assimilated into several TEMs to optimize model parameters which are sensitive to carbon flux, water flux, heat flux, and carbon

* Correspondence: arakidahk@hydrosoken.co.jp; takemasa.miyoshi@riken.jp

¹RIKEN Center for Computational Science, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

²Hydro Technology Institute Co., Ltd., Nakanoshima Daibiru 26F, 3-3-23, Nakanoshima, Kita-ku, Osaka 530-6126, Japan

Full list of author information is available at the end of the article

pools (Braswell et al. 2005; Knorr and Kattge 2005; Gao et al. 2011; Kato et al. 2013).

DA for optimizing phenology-related model parameters is crucial for reducing uncertainties in photosynthetic productivity estimates and water flux estimates. Satellite-based measurements are available for site scale to global scale DA for this purpose (Rayner et al. 2005; Demarty et al. 2007; Stöckli et al. 2011; MacBean et al. 2015; Kato et al. 2013; Yan et al. 2016; Arakida et al. 2017; Ise et al. 2018). For example, the fraction of photosynthetically active radiation (fPAR), the normalized difference vegetation index (NDVI), and the leaf area index (LAI) have been commonly used for satellite measurements.

LAI estimated from satellite-measured reflectance is a cumulative LAI value of overstory and understory. Namely, the satellite-derived LAI estimates are highly affected by understory (forest floor) reflectance (Eriksson et al. 2006). Earlier increment of LAI before the actual overstory foliation is thought to be understory LAI (e.g., Kobayashi et al. 2007, 2010). Therefore, satellite-observed LAI should be assimilated as the total LAI of overstory and understory. Here, forest structure is needed to be considered when phenology-related parameters for overstory are optimized with satellite-observed LAI. Individual-based dynamic global vegetation models (DGVMs) simulate vertical forest structure explicitly; understory is simulated at the forest floor separately from overstory. Arakida et al. (2017) developed a DA system with the Spatially Explicit Individual-based

DGVM (SEIB-DGVM; Sato et al. 2007) for the first time; it utilizes the satellite-observed LAI effectively and estimates overstory phenology separately from understory phenology.

In this study, we extended the previous experiment of Arakida et al. (2017) to examine the performance of the DA system at a large spatial domain across Siberia. Since this is the first regional-scale DA study with the SEIB-DGVM, we investigate the two research topics: (1) whether the DA system works well at a large spatial domain when frequent observations are available and (2) how DA estimates the regional distribution of unassimilated variables and model parameters for phenology. The DA system of Arakida et al. (2017) improved not only LAI but also other unassimilated variables such as carbon flux and biomass at a flux site in Siberia. This study also explores how assimilating LAI improves those unassimilated variables at a regional scale.

2 Methods

2.1 Study sites

We selected Siberia as the study area because a single overstory species larch is distributed over a large area, and the land use rarely changes for a long time. This area is suitable for the SEIB-DGVM that does not consider artificial changes in the land use. Siberia is also suitable for the first DA experiment at a regional scale, because the vegetation structure in Siberia is simple. We consider only two PFTs: deciduous needle-leaf tree and C3 grass.

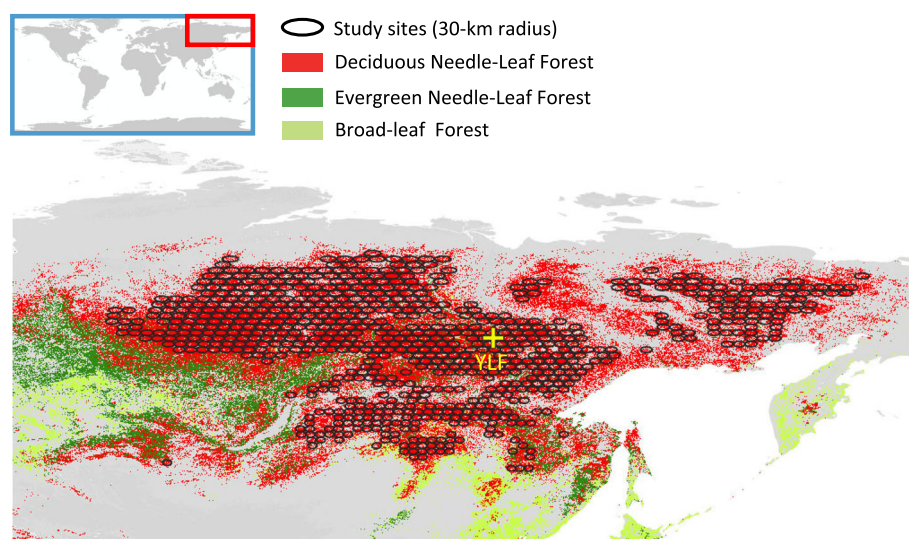


Fig. 1 Study sites (black circles). Yellow cross shows the Yakutsk larch forest, the study site of Arakida et al. (2017). The vegetation data shown by red, green, and light-green shades are from the global land cover dataset (GLC2000: European Commission, Joint Research Centre 2003). Gray areas indicate the planetary continental data provided by Esri, Global Mapping International, US Central Intelligence Agency (The World Factbook). The same continental data are used in subsequent maps

The non-overlapping circles in Fig. 1 illustrate the study sites, which represent the averaged vegetation state within circles. We first calculated the domain-averaged ratios of each PFT in each circle using the global land cover dataset (GLC2000: European Commission, Joint Research Centre 2003): deciduous needle-leaf forest (“tree cover, needle-leaved, deciduous”), evergreen needle-leaf forest (“tree cover, needle-leaved evergreen”), and broad-leaf forest (“tree cover, broad-leaved, deciduous, closed” + “tree cover, broad-leaved, deciduous, open” + “tree cover, broad-leaved, evergreen”). Next, we selected study sites covered mostly ($\geq 50\%$ cover) by deciduous needle leaf forest. Sites with $>10\%$ cover of evergreen needle-leaf forest or broad-leaf forest cover were excluded. We selected a total of 760 sites (black circles in Fig. 1).

2.2 The SEIB-DGVM

This study used a particle filter-based DA system with the SEIB-DGVM (Arakida et al. 2017). Refer to Arakida et al. (2017) for detailed descriptions. Here, we provide only a fundamental overview of Arakida et al. (2017) and additional configuration changes. The simulation started with bare ground. Forced by climate conditions, the model simulated the establishment, growth, and decay of individual trees within a virtual forest (Sato et al. 2007). Photosynthetically active radiation at the understory was attenuated by overstory; therefore, the amplitude of overstory LAI affected understory LAI. Carbon flux, water flux, heat flux, and vegetation structures (e.g., overstory LAI, biomasses of individual organs [leaf, root, trunk], and soil carbon) were also simulated through vegetation succession. For most processes, the model time step was a day. For mortality, establishment, and some of the adjustments of crown states, the model time step was a year. We used model version 2.71 (Sato and Ise 2012) with modifications described by Arakida et al. (2017). In addition, we corrected coding bugs and

modified some parameters for the experiment presented here (Table 1).

Among 14 prescribed PFTs, deciduous needle-leaf trees (“overstory”) and C3 grass (“understory”) were selected for experiments. The major tree species in this area is larch (Ponomarev et al. 2016), and the forest understory includes cowberry, grass, shrubs, mosses, and lichens (Ohta et al. 2014; Suzuki et al. 2007). The SEIB-DGVM simulates understory as deciduous grass, and the setting has been used in the previous studies in Siberia (Sato et al. 2010, 2016). In this study, we also used the same setting for the DA experiment.

2.3 Climate forcing data

Daily climate forcing data were generated using the monthly Climate Research Unit observation-based dataset (CRU-TS3.23 0.5° monthly climate time series: University of East Anglia Climatic Research Unit et al. 2015) and the daily data at the spatial resolution of T62 with a Gaussian grid (about 1.9°) from the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis (Kalnay et al. 1996). We chose a 10-year period from 2003 to 2012 in which the MODIS LAI observation data for the DA experiment and the data of the previous studies for inter-comparison were available. As Sato et al. (2007), interannual differences were not considered in this study. A yearlong climate forcing data averaged from 2003 and 2012 were used (Table 1) corresponding to the observation data described in the section “Particle filter-based data assimilation (DA) and observational data.”

First, we interpolated bilinearly each NCEP/NCAR and CRU data points into the center of each study site. Air temperature and cloudiness of NCEP/NCAR reanalysis data were corrected by CRU data so that the monthly mean was identical to those of the CRU. Likewise, precipitation and specific humidity were rescaled so that the monthly totals matched those of the CRU. In

Table 1 Modifications to the data assimilation (DA) system of Arakida et al. (2017)

Modifications		This study
Radius of the study site	10km	30km
Bugfix	Bugfix in Appendix A of Arakida et al. (2017)	Additional bug fixes, which were found unfixed between ver. 2.71 and ver. 2.82
Parameters	Modifications in Appendix A of Arakida et al. (2017)	Additional modifications, which were found in ver. 3.00 (Sato et al. 2016) TO_f (year ⁻¹) for overstory: turnover rate for foliage = 3.19 $Lue0$ (mol CO ₂ mol photon ⁻¹) for overstory: control value of light-use efficiency = 0.041 $M2$ (dimensionless) for overstory: a parameter for mortality = 1.5
Climate forcing data	Interannual data from 2004 to 2007	Averaged data from 2003 to 2012
Observation data	Interannual data from 2004 to 2007	Median of the data from 2003 to 2012 LAI which was calculated from insufficient numbers of the observation data was excluded. The threshold of standard deviation was set to 0.5.

contrast, we used the NCEP/NCAR soil temperature and wind velocity data without scaling. Finally, the climate observations from 2003 through 2012 were averaged for each study site to provide daily climate forcing data.

2.4 Particle filter-based data assimilation (DA) and observational data

Arakida et al. (2017) demonstrated that a well-known particle filter procedure, sequential importance resampling (SIR: Gordon et al. 1993), performed well as a DA method for the SEIB-DGVM. Particle filtering is a Bayesian process that remains robust when an ecological process such as phenology exhibits non-linearity; e.g., the state changes suddenly at the beginning and end of the leaf-bearing season (Arakida et al. 2017; Ise et al. 2018). In addition, particle filters can handle phase space variability in an individual-based model, such as occasional establishment or death of a tree. As proposed by Kitagawa (1998), the probability densities of the state variables and the model parameters were denoted by parallel simulations (particles), and the distributions were updated sequentially when the observational data were assimilated. The experiments in this study comprised three steps: (i) initial perturbation, (ii) spin-up, and (iii) DA. We used the particle, initial perturbation, and re-sampling perturbation sizes described by Arakida

et al. (2017). Although the radius of the study sites was 10 km in Arakida et al. (2017), it was increased to 30 km (Table 1) to reduce the number of sites and therefore, the computational cost.

First, we created 8000 random combinations of parameters from the initial perturbation ranges provided by Arakida et al. (2017) for the maximum photosynthetic rate (P_{\max} ; $\mu\text{mol CO}_2 \text{ m}^{-2} \text{ s}^{-1}$) and the defoliation start date (DSD; day of year [DOY]) for overstory and understory. Figure 2a, b shows the conceptual diagrams for the parameters. P_{\max} affects the LAI amplitude, and DSD affects the start of LAI decay. A larger P_{\max} produces a larger LAI, and a later DSD produces a longer leaf season, and vice versa. The SEIB-DGVM simulates understory as deciduous grass, but the “defoliation” is not suitable for non-deciduous vegetation, such as cowberry trees and moss. The DA system did not work well when the understory defoliated earlier than the overstory. Therefore, DSD for understory was optimized with overstory to stabilize the DA system.

The initial perturbation ranges were as follows: P_{\max} for overstory = [0, 60], P_{\max} for understory = [0, 15], DSD for overstory = [200, 300], and DSD for understory = [200, 300]. Here, [a, b] denotes the uniform distribution for an interval between a and b. Next, we performed 8000 parallel simulations with these parameter sets over

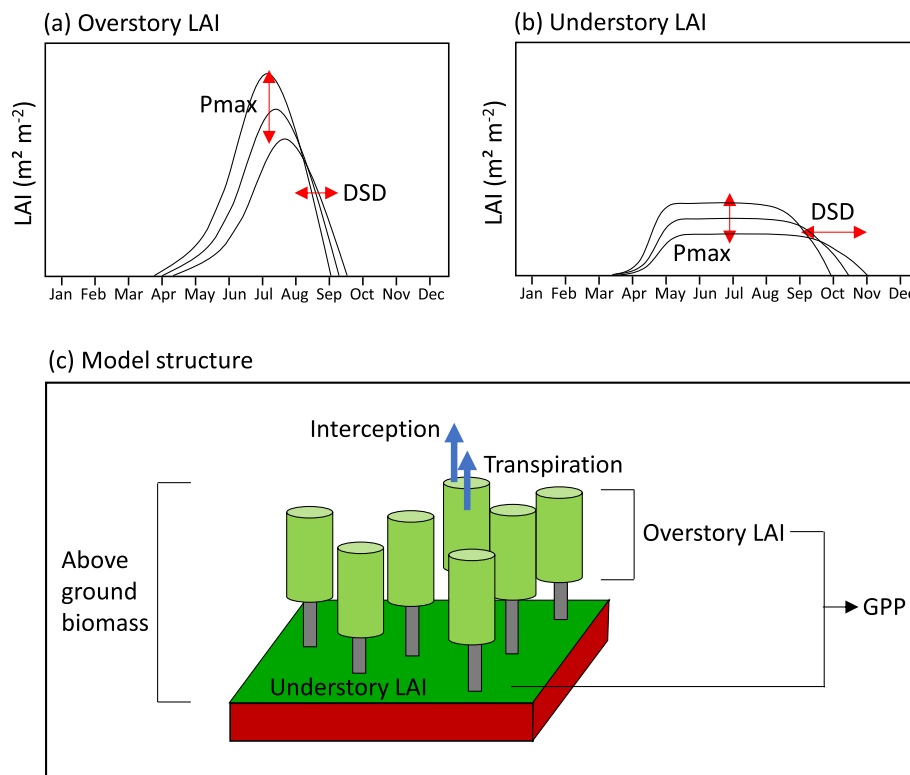


Fig. 2 The schematic diagram for **a** P_{\max} , **b** DSD, and **c** model structure

100 years, in which the averaged forcing climate data for the period of 2003–2012 were used repeatedly for spin up. During the spin-up period, the perturbed parameters led to different leaf season, the amplitude of LAI, and forest states for each particle. The detailed description for the spin-up period is shown in the “Discussion” section.

A yearlong time series of satellite-observed LAI is prepared for each site; we selected the MODIS LAI product of MCD15A3 with a 4-day interval (Knyazikhin et al. 1999) using the quality-control procedure described by Arakida et al. (2017). The strict quality-control makes observation sparse. One of our research goals is to investigate whether the DA system works well at a large spatial domain when frequent observations are available. Here, we aggregated the observation data spatially and temporally. The observation error standard deviations were assigned to each pixel of MCD15A3 LAI at the original resolution of 1 km. We used the median of this error as the observation error. The 4-day-interval LAI data and its error standard deviations for each study site were calculated using data for the same DOY in the period of 2003–2012 as the median for each 30-km radius (Table 1). The aggregation helped complement the lack of data due to the strict quality control; however, aggregated observations with insufficient data tended to produce erroneous data. To exclude such erroneous data, we used only observations calculated from one-eighth or more of the quality-controlled set in each circle at each time step (Table 1). In addition, LAI data lower than 0.5 were not assimilated, following the procedure of Arakida et al. (2017). Observations with small error standard deviations made the DA system unstable; therefore, standard deviations lower than 0.5 were fixed at 0.5 (Table 1).

Finally, DA was performed for four years repeatedly using the yearlong time series of the satellite-observed LAI. When an observation was assimilated, the likelihood was calculated for each particle, and the posterior

distribution of the particles was updated with resampling using likelihood as the weighting factor. We used the resampling perturbation size as described by Arakida et al. (2017) to avoid particle degeneracy.

2.5 Assessment of the DA results

To assess the impact of DA, an experiment without DA (“NODA” hereafter) was also performed with the parameter sets which were used for the initial perturbations. The experiments with DA (“TEST” hereafter) at the fourth year of DA were compared with NODA at the same period (i.e., the 104th year of the simulation). TEST for model parameters was compared with the medians and ranges of the initial perturbations. NODA and TEST were compared for total LAI (overstory + understory), overstory LAI, GPP, and aboveground biomass. These results were also compared with the existing studies. Other variables were only compared with the observed LAI to explore the extent to which DA affected the unassimilated state variables. Figure 2c shows the schematic diagram for the relations between LAI and these variables.

In situ observation was not widely available in Siberia. Hence, we compared the results with those of existing studies to investigate the characteristics of the DA system (Table 2). The estimated LAI was compared with assimilated observation data (MODIS LAI: Knyazikhin et al. 1999) to confirm that the DA system worked properly. Other estimated variables were compared with those of existing studies: gross primary production (GPP) of FLUXCOM (Tramontana et al. 2016; Jung et al. 2017), overstory LAI (Delbart et al. 2005; Kobayashi et al. 2010), and aboveground biomass (Liu et al. 2015). Supporting information includes more details about the existing studies.

To compare the results of this study with those of the earlier works, we consider differences of the spatiotemporal resolution. As for the LAI, we used the annual maximum LAI for cross-comparisons. The spatial

Table 2 Data used for cross-comparisons

Data	Period	Description	References
LAI (MODIS)	2003–2012	LAI data assimilated in this study. Estimation using a radiative transfer model and satellite-observed data.	Knyazikhin et al. (1999)
Overstory LAI	2003–2012	Estimation using a radiative transfer model and satellite-observed data (SPOT-VEGETATION).	Delbart et al. (2005); Kobayashi et al. (2010)
GPP (FLUXCOM)	2003–2012	Estimation with three machine learning methods: the artificial neural network (ANN), multivariate regression splines (MARS), and random forest (RF) using the flux partitioning method of Reichstein et al. (2005). These machine learning methods make use of in-situ flux measurements and diverse explanatory variables, such as meteorological data and satellite-based vegetation indices.	Tramontana et al. (2016); Jung et al. (2017)
Aboveground biomass	2003–2012	Estimation using the relationships between satellite-observed vegetation optical depth and the aboveground biomass.	Liu et al. (2015)

resolution of 30 km was identical for the observed LAI used in the DA experiment and the estimated LAI from the DA experiment, but the temporal resolution of the observation was different (i.e., 4 days). Both observed and estimated LAI reached the maximum value and did not greatly change in mid-summer. If the observation was successfully assimilated, the amplitude of LAI for the DA should have been close to that of the observed amplitude.

As for GPP, the spatial resolution for GPP of FLUXCOM was 0.5°, and the temporal resolution was a year. We first averaged the FLUXCOM data from 2003 to 2012 and multiplied it by 365 to calculate the annual total. Next, FLUXCOM data were interpolated bilinearly to the center of each study site. The interpolated GPP of FLUXCOM were compared to the annual total GPP estimated in this study.

The aboveground biomass of Liu et al. (2015), 0.25° annual mean data, was averaged from 2003 to 2012. Next, it was spatially interpolated by the same procedure used for interpolating FLUXCOM parameters, and cross-compared with the annual mean aboveground biomass estimated in this study.

The spatial resolution for overstory LAI of Kobayashi et al. (2010) was 1/112 of a degree, and the temporal resolution was 10 days. Kobayashi et al. (2010) had a higher spatial resolution than the one in this study. We first calculated the average LAI from 2003 to 2012 for each grid/DOY combination and subsequently calculated the annual maximum for each grid. Finally, we calculated the median of the maximum at each study site (i.e., within a 30-km radius) for comparison with the annual maximum overstory LAI that we estimated.

The results for the lower latitudes south of 60° N were not stable: this will be discussed as “limitations of this study” in the discussion section. We therefore used results for latitudes higher than 60° N to construct a scatter plot of the relationships between estimates in this study and those of previous reports. We used Pearson’s product-moment correlation coefficient (r) and RMSE to examine relationships in the data collected for the same area. P values for r are calculated for reference without considering spatial autocorrelation. The null hypothesis is $r = 0$. Therefore, the results should be interpreted with caution.

3 Results

3.1 Total LAI

Figure 3a, b displays the spatial distributions of the annual maximum of total LAI (see the median in Figure S1a). The total LAI for NODA (Fig. 3a) was larger than the observations across the study sites. In contrast, DA reduced LAI for TEST (Fig. 3b) and made it closer to the observed LAI (Fig. 3c). The scatter plot (Fig. 3d)

indicates that DA reduced LAI and raised the correlation coefficient from 0.46 to 0.99 and reduced RMSE from 1.94 to 0.17. These results indicate at least the necessary condition that the DA system worked properly.

3.2 Unassimilated state variables

To explore how DA affected unobserved state variables, we first compared the correlation coefficients with observed LAI. Table 3 shows that the unassimilated variables estimated by DA are highly correlated with the observed LAI. Namely, optimizing states and parameters with respect to LAI impacts the spatial distributions of carbon flux, water flux, and vegetation structures (overstory LAI and biomass) simultaneously. For a more detailed assessment of the DA results, we compared NODA and TEST for overstory LAI, GPP, and aboveground biomass with the previous studies.

Figure 4a displays the spatial distributions of the annual maximum of overstory LAI (see the median in Figure S1b, right). The spatial distribution of overstory LAI for TEST (Fig. 4a) is similarly estimated to that of Kobayashi et al. (2010) (Fig. 4b), except for higher overstory LAI for TEST at the middle to southern parts of Siberia. DA also reduced overstory LAI for TEST (Fig. 4c), and the correlation coefficients between the overstory LAI of the current study and that of Kobayashi et al. (2010) increased from 0.53 to 0.81 and RMSE decreased from 1.99 to 0.54. Because SEIB-DGVM calculates the vertical structure of vegetation, understory LAI is also estimated separately from overstory LAI (not shown).

GPP and aboveground biomass for TEST were reduced by DA (Fig. 5) corresponding to the reduction in LAI (Fig. 3d). As for GPP, DA increased the correlation between this study and those of FLUXCOM from 0.45 to 0.88 (artificial neural network: ANN), from 0.51 to 0.82 (multivariate regression splines: MARS), and from 0.58 to 0.92 (random forest: RF). RMSE was also decreased from 1109 to 434 (ANN), from 932 to 300 (MARS), and from 959 to 289 (RF). However, TEST in this study was 2–3 times higher than that of FLUXCOM at higher GPP. FLUXCOM used only a limited number of in situ observations and is not used as the verification truth in this study. Nevertheless, at least notable outliers were not found in the scatter plot (Fig. 5b). This result indicates that the DA system estimates GPP similarly to regional trends in the previous study. As for aboveground biomass, the correlation coefficient (Fig. 5c, d) increased from 0.36 (NODA) to 0.72 (TEST) and RMSE decreased from 17.4 (NODA) to 15.0 (TEST).

3.3 Parameters

Model parameters uniformly distributed for NODA (see the parameter distributions in Figure S1d–g, left)

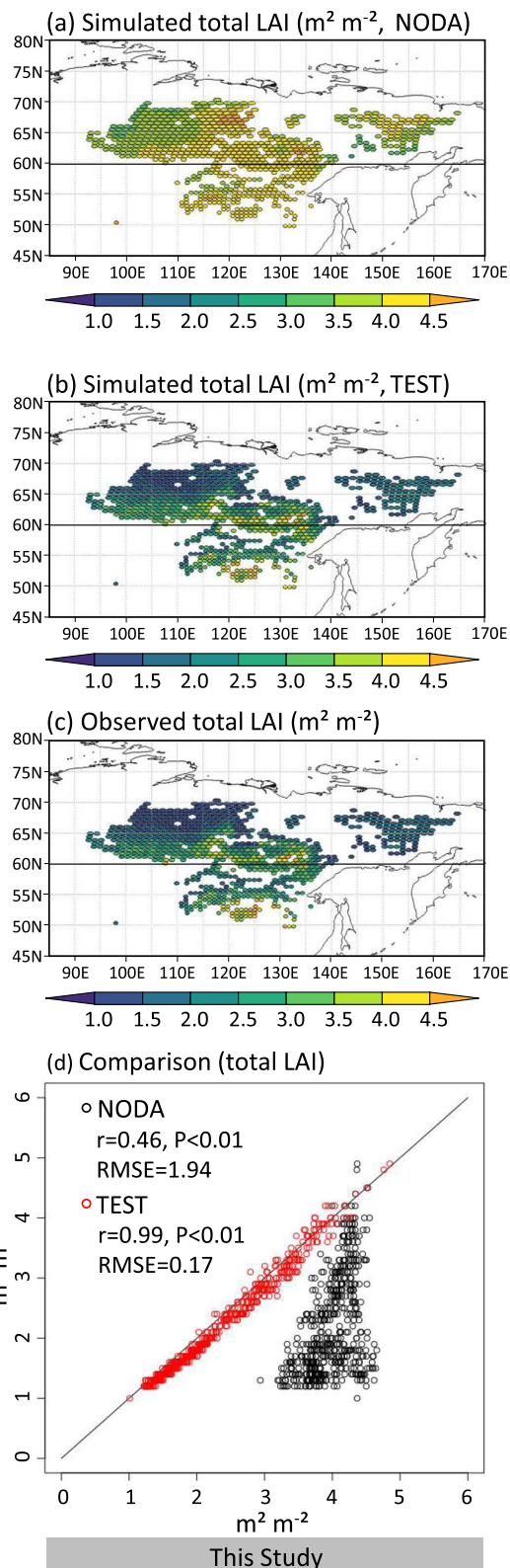


Fig. 3 Total LAI (annual maximum of median). **a** NODA, **b** TEST, **c** observation, and **d** scatter plot of simulation results (NODA and TEST) against observational data. The estimates for sites south of 60° N were not stable. Only results from sites north of 60° N (latitude line shown in each map) were used in subsequent scatter plots

throughout the study sites (not shown). The medians of overstory Pmax and understory Pmax for NODA were 30 and 7.5, respectively. The median of Pmax for TEST varied spatially (Fig. 6a,b, left): 23.2 ± 4.3 (mean \pm SD) for overstory Pmax, and 7.1 ± 1.8 for understory Pmax. Hence, Pmax must be optimized, especially for overstory, to reproduce the observed LAI, and the optimum value varies spatially. The 1–99% quantile range for Pmax was still large for TEST throughout the study area (Fig. 6a, b, right).

The medians of overstory DSD for NODA were 250 both for overstory and understory. The median of DSD for TEST also varied spatially (Fig. 6c, d, left): 214.0 ± 6.7 (mean \pm SD) for overstory DSD, and 270.3 ± 14.7 for understory DSD. Overstory DSD for TEST (Fig. 6c, left) was about 60 days earlier than that of understory (Fig. 6d, left), whereas the initial perturbation sizes for the overstory and understory DSD were identical. Namely, the DA system distinguished overstory DSD from understory DSD. The particle spread for DSD increased with latitude (Fig. 6c, d, right). Arakida et al. (2017) showed that the DA system distinguished overstory and understory DSD when consecutive low LAI observations were available near the observation threshold (LAI = 0.5) at the start of the growing season and at the end of the leaf-bearing season. We found that consecutive low LAIs near the threshold occurred only at the sites where the annual maximum LAI was relatively large (larger than 2.0), and the particle spread for DSD at those sites was smaller than those at other sites.

Table 3 Correlation coefficients for the spatial relationships between the variables and observed leaf area indices (LAI; annual maximum) for NODA and TEST. Asterisk * shows *P* values less than 0.01

State variable	NODA	TEST
Total LAI (annual maximum)	0.46 *	0.99 *
Overstory LAI (annual maximum)	0.53 *	0.98 *
Aboveground biomass (annual mean)	0.38 *	0.93 *
GPP (annual total)	0.52 *	0.99 *
Interception (annual total)	0.68 *	0.89 *
Transpiration (annual total)	0.53 *	0.96 *
Bowen ratio (annual minimum)	– 0.41 *	– 0.83 *

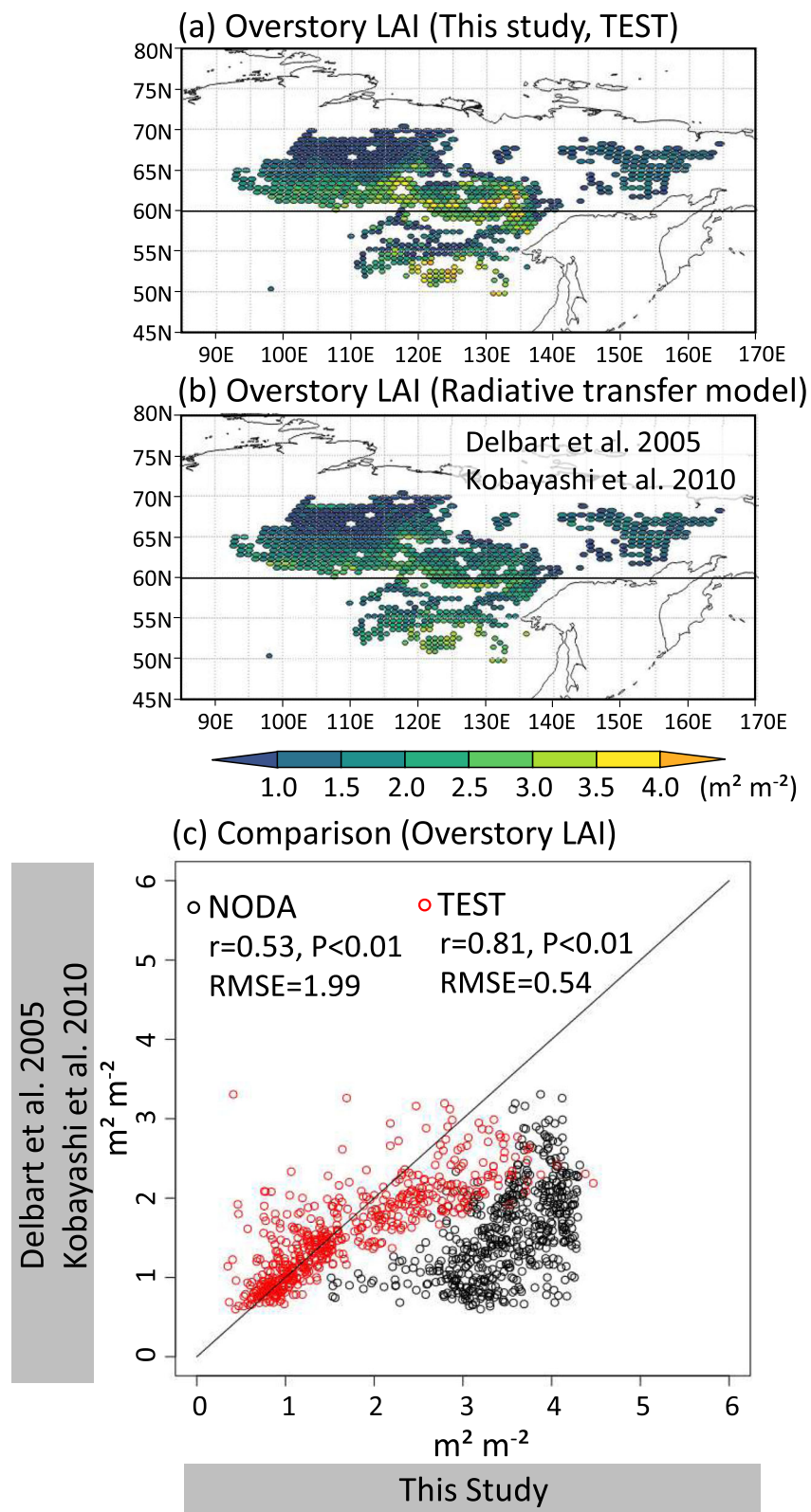


Fig. 4 Overstory LAI (annual maximum of median). **a** This study (TEST), **b** estimation with three-dimensional radiative transfer model (Delbart et al. 2005; Kobayashi et al. 2010), and **c** scatter plot of this study against the estimates from Delbart et al. (2005) and Kobayashi et al. (2010)

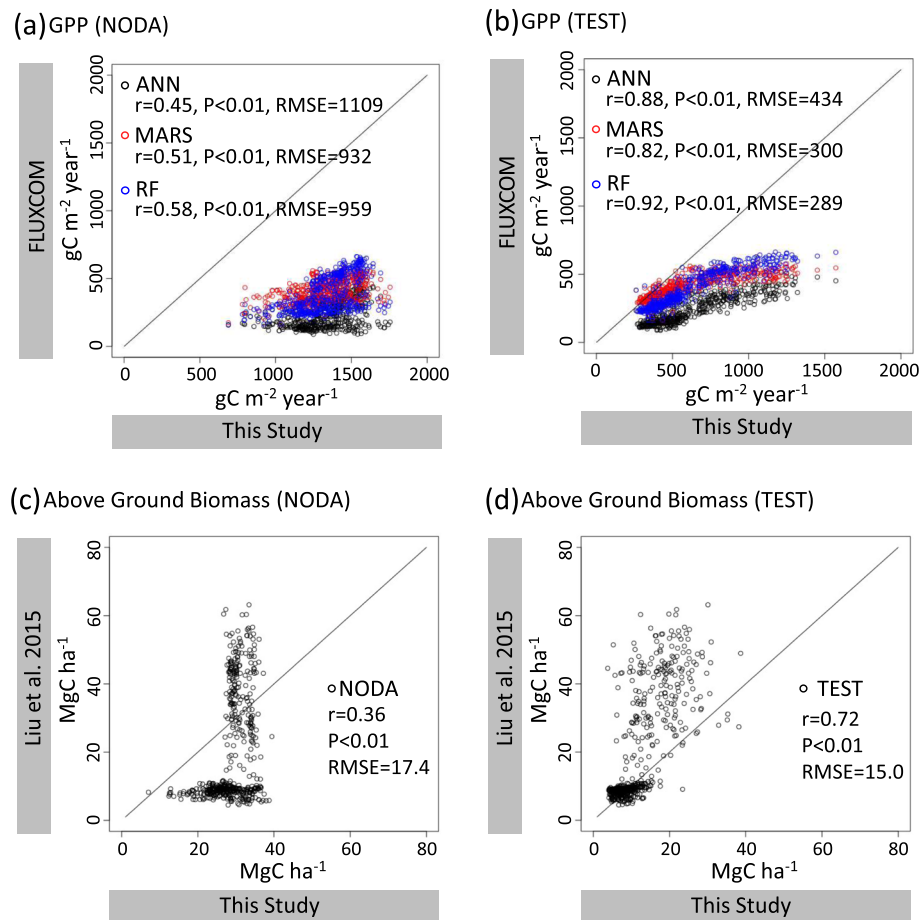


Fig. 5 Spatial correlation coefficients between this study and the previous studies. The upper panels display the scatter plots of GPP of this study against that of FLUXCOM for NODA (a) and TEST (b): the artificial neural network (ANN), multivariate regression splines (MARS), and random forest (RF) (Tramontana et al. 2016; Jung et al. 2017), with the observation threshold of Reichstein et al. (2005). The lower panels display the scatter plots of aboveground biomass of this study against that of Liu et al. (2015) for NODA (c) and TEST (d)

4 Discussion

4.1 Performance of the DA system at regional scales

We applied the DA system with an individual-based vegetation model and satellite observations to the regional scale for the first time. The results demonstrated that the satellite-observed LAI was successfully assimilated into the SEIB-DGVM. In addition, the DA system estimated the spatiotemporal distribution of overstory LAI separately from that of understory LAI. Kobayashi et al. (2010) pioneered to estimate the seasonal change of overstory LAI separately from that of understory LAI at a regional scale with a three-dimensional radiative transfer model and satellite-observed data. The correlation coefficients between the overstory LAI of the current study and that of Kobayashi et al. (2010) increased from 0.53 to 0.81 by DA. This indicated that the DA system with an individual-based DGVM also functioned well in estimating forest structure at a large spatial domain. In addition, the DA system affected the

regional estimations of carbon and water fluxes, which were not simulated by the three-dimensional radiative transfer model in Kobayashi et al. (2010).

The maximum overstory LAI in the Yakutsk larch forest (YLF) was about 1.6 in this study, close to the estimate from 1.6 to 2.0 of in situ observations (Ohta et al. 2008; Iida et al. 2009). On the other hand, the maximum understory LAI in YLF was about 1.0 in this study, about half of 2.1 of Iida et al. (2009). The total LAI in this study was close to the MODIS LAI (Figure S1a), therefore, the underestimate of the understory LAI would be caused by the observation bias. The higher overstory LAI (> 2.0) estimates in this study tended to exceed those of Kobayashi et al. (2010) (Fig. 4c). This also may be related to the observation bias. Further DA studies with LAI products from other satellite observations (e.g., Kobayashi et al. 2010) may improve the understanding of the bias in the observation data.

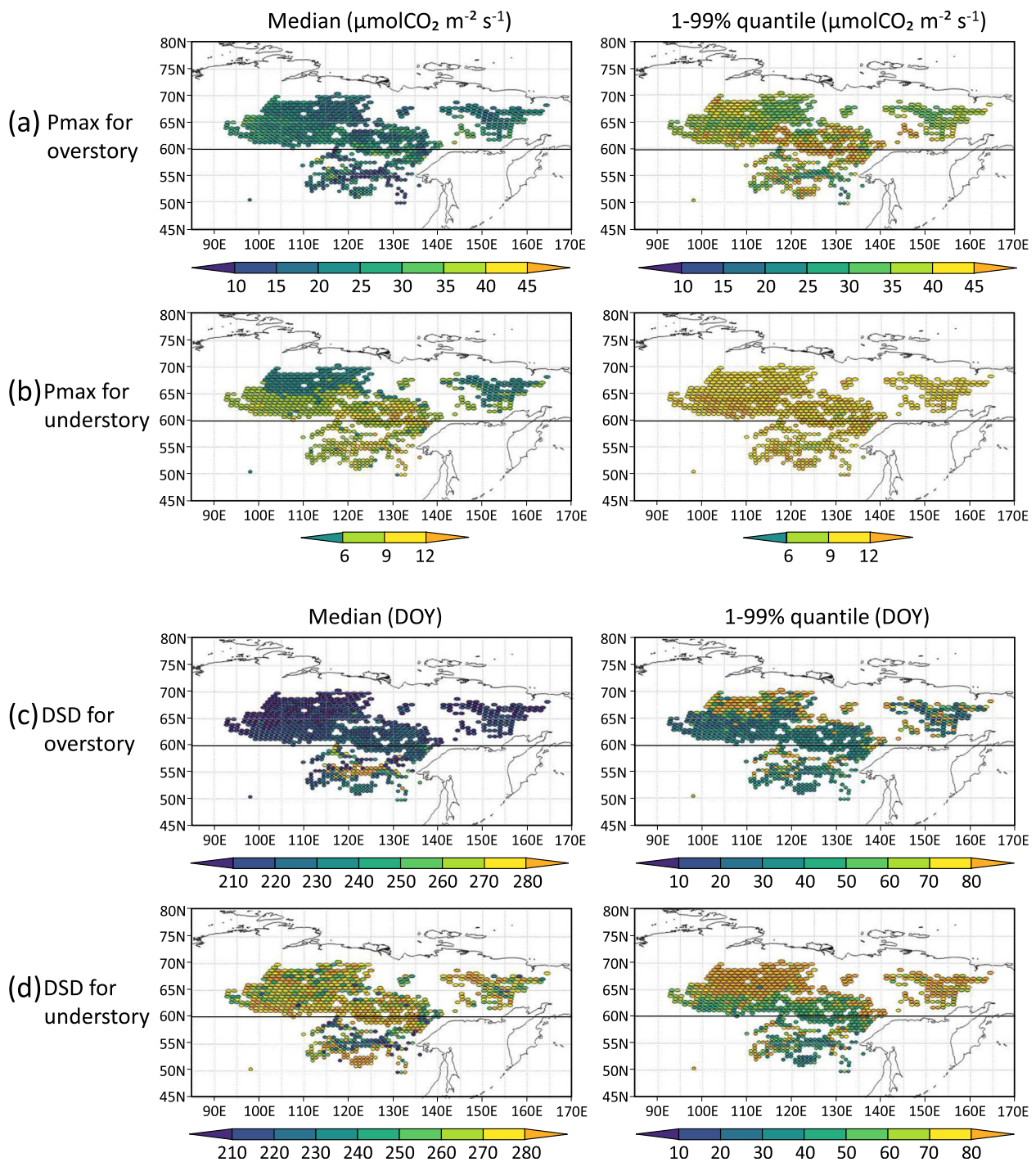


Fig. 6 Estimated parameters. Annual means of the median (left) and 1–99% quantile range (right) of the parameters for TEST: **a** maximum photosynthetic rate (Pmax) for overstory, **b** Pmax for understory, **c** defoliation start date (DSD) for overstory, and **d** DSD for understory

DA reduced LAI estimates throughout the study area, which made the estimated LAI very close to the observed LAI (Fig. 3d). In addition to the reduction in LAI, DA reduced values of the unassimilated variables (Figs. 4c and 5b, d). The resulting spatial

correlations between the estimated variables and the observed LAI were generally high (Table 3). Hence, optimization of LAI markedly affected the spatial distributions of fluxes and vegetation structure at the regional scale.

The relations between LAI and unassimilated variables are shown in Fig. 2c. According to the model description paper of the SEIB-DGVM (Sato et al. 2007), the relationships among Pmax, DSD, GPP, and LAI are explained as follows. The single-leaf photosynthetic rate under light saturation is calculated by multiplying Pmax by the coefficients of temperature, CO₂ level, and soil water effects. GPP is calculated using the equation including the light-saturated photosynthesis rate and LAI. On the other hand, the leaf mass increment is determined by the distribution from GPP. Therefore, there is a mutual relationship between LAI and GPP. In the SEIB-DGVM, LAI is modeled to decrease linearly from the DSD, and photosynthesis stops suddenly at the DSD. By assimilating satellite-observed LAI, DA made overstory DSD earlier and overstory Pmax smaller than those of the median of the NODA experiment. Therefore, DA shortened the period for photosynthesis and reduced the photosynthesis rate, which led to the reduction of annual total GPP. As for water flux, interception and transpiration are also calculated using LAI in the SEIB-DGVM. Bowen ratio is affected by those water fluxes. Therefore, the high correlation between these variables and the LAI is a natural result if the LAI is properly assimilated. In this way, our DA system has a significant impact on carbon, water, and energy fluxes.

Another interesting result is the high correlation between the aboveground biomass and the observed LAI (Table 3) and the increase of the correlation with the previous study (Fig. 5d). Although the aboveground biomass is not directly calculated from the LAI, the biomass allometry to each organ (i.e., leaves, roots, trunk, and storage resources) is parameterized in the model (Sato et al. 2007). Since the LAI of leaves corresponds to the leaf biomass, there is a possibility that the total biomass could be estimated inversely by assimilating LAI. Further study for this issue is needed in future studies.

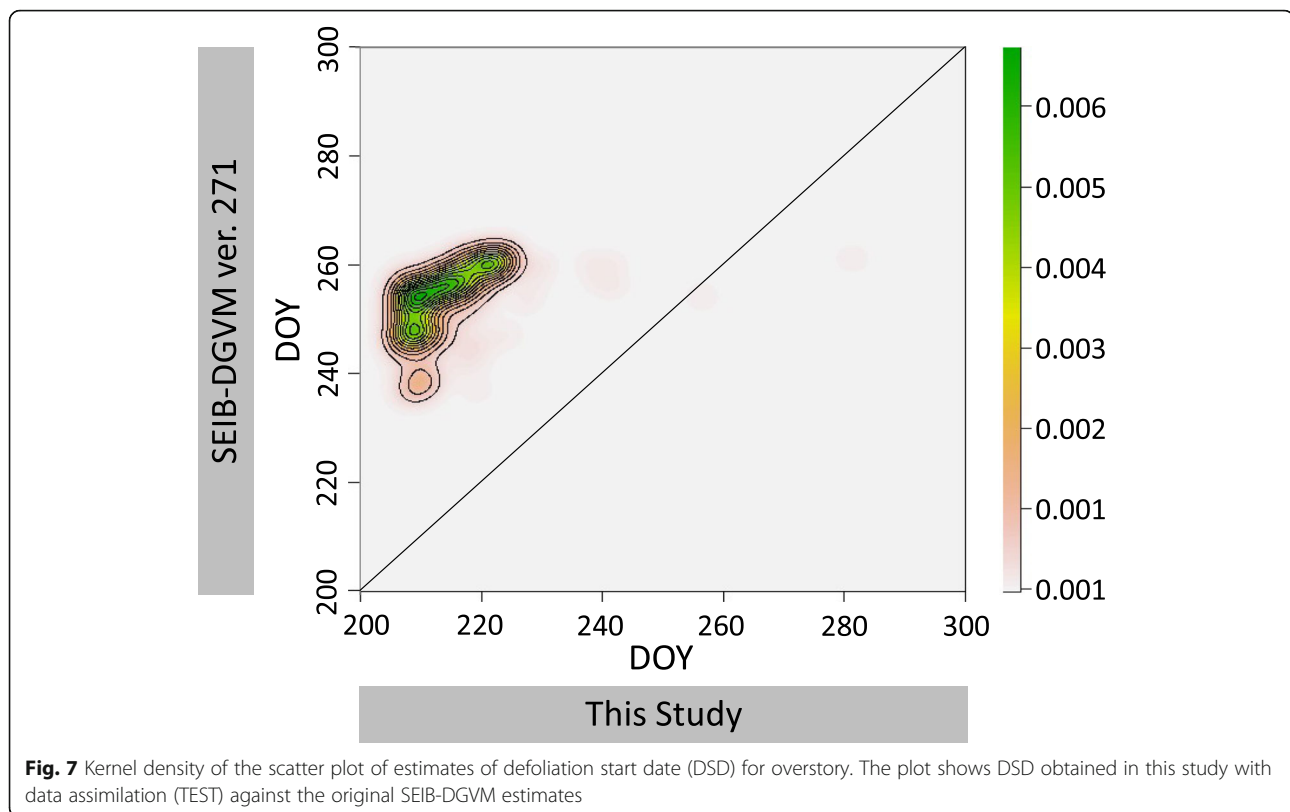
The estimates obtained in this study were highly correlated with those of previous investigations using satellite observations with an optical sensor for FLUXCOM (Fig. 5b: Tramontana et al. 2016; Jung et al. 2017) and overstory LAI (Fig. 4c: Delbart et al. 2005; Kobayashi et al. 2010). The principles of the observations used in these works were identical to those of this study. Therefore, the improved correlations and RMSE by DA are a natural consequence as long as the MODIS LAI is assimilated correctly. The correlation for aboveground biomass was also improved (Fig. 5d, Liu et al. 2015) nevertheless they used microwave-based observations, the principles of the observation are fundamentally different from this study. This improvement suggests that the DA system worked to optimize not only LAI but also biomass, as mentioned earlier. The estimated NEE was

within a reasonable range at the two flux sites in Siberia (not shown), i.e., the Tura (Nakai et al. 2008) and Yakutsk (Ohta et al. 2001, 2008, 2014) larch forests. Additional in situ observations are needed for further validation.

4.2 Parameter estimation

According to the observations at YLF Asiaflux mixed forest site in 1997–2000 (Suzuki et al. 2001), the leaf senescence of birch started at the end of August (DOY 237–242), and that of larch started in mid-September (DOY around 258). The default SEIB-DGVM identifies the start date of defoliation when the 10-day running average of daily mean air temperature is lower than 7 °C, tuned to coincide with the onset of the leaf senescence of larch at YLF (Sato et al. 2010). On the other hand, DSD for TEST in this study was DOY 218, about 40 days earlier than the observation for larch by Suzuki et al. 2001. Figure 7 shows the two-dimensional kernel density of the scatter plot of estimated overstory DSD in this study (TEST) against that in the original SEIB-DGVM. In general, the estimated overstory DSD in the present study was about 40 days earlier than that in SEIB-DGVM. As shown in Figure S1a, the phenology at the YLF Asiaflux site was well optimized by DA for MODIS LAI. Overstory DSD was also estimated with a smaller particle spread at the regional scale (Fig. 6c). Therefore, such an early DSD was likely caused by the bias of MODIS itself. Wang et al. (2005) validated seasonal patterns of MODIS LAI at two deciduous forests in Europe, and they also showed that the start of defoliation in MODIS LAI was up to 18 days earlier than the local observation. This suggests that to better estimate the defoliation start date, it would be necessary to use observations that properly reflect phenology.

Since the correlation between the observed and TEST LAIs was very high in this DA experiment (Fig. 3d), we can assume that the optimization of Pmax at least contributed to the optimization of the amplitude of LAI. However, several issues remain: (1) averaged forcing climate data was used for the DA experiment which affects photosynthesis rate, (2) the amplitude of LAI was optimized only by Pmax, and (3) the 1–99% quantile range for Pmax was still large for TEST throughout the study area. Therefore, a more appropriate DA system is needed for the optimization of the amplitude of LAI. The reduction of the particle spread for overstory Pmax from the initial perturbation was limited, and it did not greatly improve across the study area, likely due to the large resampling perturbation size used to avoid particle degeneracy (Arakida et al. 2017). Our future research will improve the DA system with a smaller resampling perturbation size and investigate Pmax more carefully.



The sensitivity of other parameters to the amplitude of LAI should also be investigated.

4.3 Limitations

The SEIB-DGVM simulates the understory as deciduous grass, and this setting was not modified for the DA experiment in this study. For a more realistic simulation in Siberia, it is necessary to develop a DA system that takes into account the evergreen understory, because understory also highly contributes to GPP (Kotani et al. 2019) and evapotranspiration (Iida et al. 2009) in Siberia. Further study for the underestimation of LAI for the understory is also needed. The DA system was unstable in the study area at latitudes lower than 60° N, and the estimated LAI for understory exceeded that for overstory. In these unstable cases, apparently, the particles converged to an unrealistic vegetation structure. In these areas, overstory DSD was later than understory DSD (Fig. 6c, d, left), and overstory LAI was low (Fig. 4a). To resolve these contradictions, the DA system must be further improved for the lower latitudes. The Larch canopy in this region is scarce and the understory is dominated by mosses and lichens (Suzuki et al. 2007). Classification of understory may be a key to improve the DA system, especially in this region.

The correlations between TEST and FLUXCOM for GPP were improved by DA (Fig. 5b), but the estimated

values were overestimated for TEST. We performed a simple regression analysis with the MODIS LAI (annual maximum) and GPP (annual total) for further understanding. The slope for TEST was 328 and that for FLUXCOM was 109 for ANN, 82 for MARS, and 149 for RF, which indicates the slope for TEST was from 2.2 to 4 times higher than that of FLUXCOM. This may be related to the overestimation of the overstory LAI and the DA setting which optimizes the amplitude of LAI only by Pmax. Further study is also needed on this issue.

In the previous studies which have been used the SEIB-DGVM, spin-up was performed over some thousands of years to simulate soil carbon accumulations (e.g., Sato et al. 2010). In this study, we used the particle filter method with 8000 parallel simulations at each location. The calculation cost is huge for such a long spin-up period. Therefore, we only validated aboveground forest states such as LAI, aboveground biomass, and GPP with previous studies. We have performed a preliminarily NODA experiment with 100 particles and confirmed that a spin-up period of 100 years was enough for saturation of LAI, aboveground biomass, and GPP for all of the study sites. Only above ground biomass was decreased after 100 years because of the death of the trees. Sensitivity to the spin-up period from this aspect may be needed in future studies.

We used averaged forcing climate data corresponding to the temporal aggregation of the observation data. We prioritized the stability of the DA system over the realism of the climate forcing data. This averaging may produce unrealistic conditions, especially for water status. If more good quality observations are available, it is better to use year-to-year climate forcing data for spin-up and DA. In addition, the forcing data with higher spatial resolution such as ECMWF Reanalysis v5 (ERA5) and Global Soil Wetness Project Phase 3 (GSWP3) may improve the DA results.

5 Conclusion

We tested the performance of the SEIB-DGVM-based DA system over Siberia and found that it generally performed well at a large spatial domain. The study revealed that the DA system with an individual-based DGVM estimates overstory LAI at a regional scale; this leads to the estimation of spatial distributions of model parameters for overstory and understory separately. In addition, this study corroborated the previous DA studies with vegetation models and showed that LAI was crucial for the estimation of carbon flux at a regional scale.

Remaining issues for future studies include that the DA system fails in the southern sectors of Siberia. Improvements are required for application to lower latitudes. Second, we assimilated only the satellite-observed MODIS LAI. Assimilation of other observations along with LAI would be beneficial such as using microwave-based aboveground biomass and GPP estimated from solar-induced chlorophyll fluorescence (e.g., Frankenberg et al. 2011). We may increase the number of parameters to estimate by assimilating these additional observations. LAI products from other satellite data may also help understand the influence of observation bias on the DA system.

Abbreviations

DA: Data assimilation; SEIB-DGVM: Spatially Explicit Individual-Based Dynamic Global Vegetation Model; MODIS: MODerate resolution Imaging Spectroradiometer; LAI: Leaf area index; GPP: Gross primary production; TEMs: Terrestrial ecosystem models; PFT: Plant functional type; NDWI: Normalized difference water index; OSSE: Observing System Simulation Experiment; GLC 2000: Global Land Cover dataset 2000; CRU: Climate Research Unit; NCEP/NCAR: National Centers for Environmental Prediction/National Center for Atmospheric Research; SIR: Sequential importance resampling; Pmax: Maximum Photosynthetic rate; DSD: Defoliation start date; DOY: Day of year; ANN: Artificial neural network; MARS: Multivariate regression splines; RF: Random forest; NODA: Results without DA; TEST: Results with DA; SD: Standard deviation; YLF: Yakutsk Larch Forest; ERA5: ECMWF Reanalysis v5; GSWP3: Global Soil Wetness Project Phase 3

6 Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40645-021-00443-6>.

Additional file 1. Supporting information.

Acknowledgements

The authors thank Hisashi Sato, the main developer of SEIB-DGVM, for useful discussions and for providing the processing code for the climate forcing data. The authors also thank the two reviewers for their useful comments to improve the manuscript. The GLC 2000 data was retrieved from the Global Land Cover 2000 database (<http://forobs.jrc.ec.europa.eu/products/glc2000/glc2000.php>). The source code of the SEIB-DGVM was retrieved from the developer's website (<http://seib-dgvm.com/>). The CRU-TS3.23 data was retrieved from the database of NCAS British atmospheric data centre (<http://catalogue.ceda.ac.uk/uuid/3f8944800cc48e1cbc29a5ee12d8542d>). The NCEP/NCAR Reanalysis 1 data was retrieved from the database of NOAA/OAR/ESRL PSD, Boulder, Colorado, USA (<https://www.esrl.noaa.gov/psd/>). MODIS LAI product of MCD15A3 was retrieved from the online data pool, courtesy of the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota (https://lpdaac.usgs.gov/data_access/data_pool). The overstory LAI data based on Delbart et al. (2005) and Kobayashi et al. (2010) was retrieved from the author's website (<http://files.sakura.ne.jp/RSdata/SiberiaLAI/README.html>). Aboveground biomass data of Liu et al. (2015) was retrieved at the author's website (Version 1.0: <http://www.wenfo.org/wald/global-biomass/>). Carbon flux estimations of FLUXCOM (Tramontana et al. 2016; Jung et al. 2017) were retrieved at the data portal of the Max Planck institute for biogeochemistry (<https://www.bgc-jena.mpg.de/geodb/projects/Home.php>).

Authors' contributions

HA designed and carried out the study, and TM directed the research. SK prepared the satellite-observed LAI. SO contributed to the computational parallelization. YS helped in preparing the previous studies which were compared with this study. SK, SO, and YS collaborated with the corresponding authors (HA and TM) in the construction of the manuscript. All authors read and approved the final manuscript.

Funding

None

Availability of data and materials

The data assimilation code using in this study and the datasets supporting the conclusions of this article are available in the internet in the RIKEN domain (<https://data-assimilation.riken.jp/~opendata/SEIB-PF02/>).

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹RIKEN Center for Computational Science, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. ²Hydro Technology Institute Co., Ltd., Nakanoshima Daibiru 26F, 3-3-23, Nakanoshima, Kita-ku, Osaka 530-6126, Japan. ³Center for Environmental Remote Sensing (CEReS), Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba-shi 263-8522, Japan. ⁴PRESTO, Japan Science and Technology Agency, 7, Gobancho, Chiyodaku, Tokyo 102-0076, Japan. ⁵RIKEN Interdisciplinary Theoretical and Mathematical Sciences Program, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. ⁶RIKEN Cluster for Pioneering Research, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan. ⁷Institute of Engineering Innovation, The University of Tokyo, Yayoi 2-11-16, Bunkyo-ku, Tokyo 113-8656, Japan. ⁸Department of Atmospheric and Oceanic Science, University of Maryland, 4254 Stadium Dr, College Park, MD 20742, USA. ⁹Application Laboratory, Japan Agency for Marine-Earth Science and Technology, 3173-25, Showa-machi, Kanazawa-ku, Yokohama City, Kanagawa 236-0001, Japan.

Received: 2 October 2020 Accepted: 14 August 2021

Published online: 14 September 2021

References

- Ahlström A, Schurgers G, Arneth A, Smith B (2012) Robustness and uncertainty in terrestrial ecosystem carbon response to CMIP5 climate change projections. *Environ Res Lett* 7(4):044008. <https://doi.org/10.1088/1748-9326/7/4/044008>
- Arakida H, Miyoshi T, Ise T, Shima S, Kotsuki S (2017) Non-Gaussian DA of satellite-based leaf area index observations with an individual-based dynamic

- global vegetation model. *Nonlinear Process Geophys* 24(3):553–567. <https://doi.org/10.5194/npg-24-553-2017>
- Braswell BH, Sacks WJ, Linder E, Schimel DS (2005) Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Glob Chang Biol* 11(2): 335–355. <https://doi.org/10.1111/j.1365-2486.2005.00897.x>
- Cheait A, Badeau V, Boe J, Chuine I, Delire C, Dufrene E, François C, Gritti ES, Legay M, Pagé C, Thuiller W, Viovy N, Leadley P (2012) Climate change impacts on tree ranges: model intercomparison facilitates understanding and quantification of uncertainty. *Ecol Lett* 15(6):533–544. <https://doi.org/10.1111/j.1461-0248.2012.01764.x>
- Delbart N, Kergoat L, Toan TL, Lhermitte J, Picard G (2005) Determination of phenological dates in boreal regions using normalized difference water index. *Remote Sens Environ* 97(1):26–38. <https://doi.org/10.1016/j.rse.2005.03.011>
- Demarty J, Chevallier F, Friend AD, Viovy N, Piao S, Ciais P (2007) Assimilation of global MODIS leaf area index retrievals within a terrestrial biosphere model. *Geophys Res Lett* 34(15):L15402. <https://doi.org/10.1029/2007GL030014>
- Eriksson HM, Eklundh L, Kuusk A, Nilson T (2006) Impact of understory vegetation on forest canopy reflectance and remotely sensed LAI estimates. *Remote Sens Environ* 103(4):408–418. <https://doi.org/10.1016/j.rse.2006.04.005>
- European Commission, Joint Research Centre (2003) The global land cover map for the year 2000, GLC2000 database. European commission joint research centre <https://forobs.jrc.ec.europa.eu/products/glc2000/glc2000.php>. Accessed 6 Jan 2017
- Fisher RA, Koven CD, Anderegg WRL, Christoffersen BO, Dietze MC, Farrior CE, Holm JA, Hurtt GC, Knox RG, Lawrence PJ, Lichstein JW, Longo M, Matheny AM, Medvigy D, Muller-Landau HC, Powell TL, Serbin SP, Sato H, Shuman JK, Smith B, Trugman AT, Viskari T, Verbeeck H, Weng E, Xu C, Xu X, Zhang T, Moorcroft PR (2017) Vegetation demographics in earth system models: A review of progress and priorities. *Glob Chang Biol* 24(1):35–54. <https://doi.org/10.1111/gcb.13910>
- Frankenberg C, Fisher JB, Worden J, Badgley G, Saatchi SS, Lee JE, Toon GC, Butz A, Jung M, Kuze A, Yokota T (2011) New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophys Res Lett* 38(17):17. <https://doi.org/10.1029/2011GL048738>
- Friend AD, Lucht W, Rademacher TT, Keribin R, Betts R, Cadule P, Ciais P, Clark DB, Dankers R, Falloon PD, Ito A, Kahana R, Kleidon A, Lomas MR, Nishina K, Ostberg S, Pavlick R, Peylin P, Schaphoff S, Vuichard N, Warszawski L, Wiltshire A, Woodward FI (2014) Carbon residence time dominates uncertainty in terrestrial vegetation responses to future climate and atmospheric CO₂. *Proc Natl Acad Sci* 111(9):3280–3285. <https://doi.org/10.1073/pnas.1222477110>
- Gao C, Wang H, Weng E, Lakshminarayanan S, Zhang Y, Luo Y (2011) Assimilation of multiple data sets with the ensemble Kalman filter to improve forecasts of forest carbon dynamics. *Ecol Appl* 21(5):1461–1473. <https://doi.org/10.1890/09-1234.1>
- Gordon NJ, Salmond DJ, Smith AFM (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc F* 140(2):107–113. <https://doi.org/10.1049/ip-f-2.1993.0015>
- Iida S, Ohta T, Matsumoto K, Nakai T, Kuwada T, Kononov AV, Maximov TC, van der Molen MK, Dolman H, Tanaka H, Yabuki H (2009) Evapotranspiration from understory vegetation in an eastern Siberian boreal larch forest. *Agric Forest Met* 149(6–7):1129–1139. <https://doi.org/10.1016/j.agrformet.2009.02.003>
- Ise T, Ikeda S, Watanabe S, Ichii K (2018) Regional-scale data assimilation of a terrestrial ecosystem model: leaf phenology parameters are dependent on local climatic conditions. *Front Environ Sci* 6:95. <https://doi.org/10.3389/fenvs.2018.00095>
- Ito A, Nishina K, Reyer CPO, François L, Henrot AJ, Munhoven G, Jacquemin I, Tian H, Yang J, Pan S, Morfopoulos C, Betts R, Hickler T, Steinkamp J, Ostberg S, Schaphoff S, Ciais P, Chang J, Rafique R, Zeng N, Zhao F (2017) Photosynthetic productivity and its efficiencies in ISIMIP2a biome models: benchmarking for impact assessment studies. *Environ Res Lett* 12(8):085001. <https://doi.org/10.1088/1748-9326/aa7a19>
- Jung M, Reichstein M, Schwalm CR, Huntingford C, Sitch S, Ahlström A, Arneth A, Camps-Valls G, Ciais P, Friedlingstein P, Gans F, Ichii K, Jain AK, Kato E, Papale D, Poulter B, Raduly B, Rödenbeck C, Tramontana G, Viovy N, Wang YP, Weber U, Zaehle S, Zeng N (2017) Compensatory water effects link yearly global land CO₂ sink changes to temperature. *Nature* 541(7638):516–520. <https://doi.org/10.1038/nature20780>
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. *B Am Meteorol Soc* 77: 437–471. <http://www.forobs.jrc.ec.europa.eu/products/glc2000/glc2000.php>
- Kaminski T, Knorr W, Schürmann G, Scholze M, Rayner PJ, Zaehle S, Blessing S, Dorigo W, Gayler V, Giering R, Gobron N, Grant JP, Heimann M, Hooker-Stroud A, Houweling S, Kato T, Kattge J, Kelley D, Kemp S, Koffi EN, Köstler C, Mathieu P-P, Pinty B, Reick CH, Rödenbeck C, Schnur R, Scipal K, Sebald C, Stacke T, Terwisscha van Scheltinga A, Vossbeck M, Widmann H, Ziehn T (2013) The BETHY/JSBACH Carbon Cycle DA System: experiences and challenges. *J Geophys Res Biogeo* 118(4):1414–1426. <https://doi.org/10.1002/jgrg.20118>
- Kato T, Knorr W, Scholze M, Veenendaal E, Kaminski T, Kattge J, Gobron N (2013) Simultaneous assimilation of satellite and eddy covariance data for improving terrestrial water and carbon simulations at a semi-arid woodland site in Botswana. *Biogeosciences* 10(2):789–802. <https://doi.org/10.5194/bg-10-789-2013>
- Kitagawa G (1998) A self-organizing state-space model. *J Am Stat Assoc* 93(443): 1203–1215. <https://doi.org/10.2307/2669862>
- Knorr W, Kattge J (2005) Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. *Glob Chang Biol* 11(8):1333–1351. <https://doi.org/10.1111/j.1365-2486.2005.00977.x>
- Knyazikhin Y, Glassy J, Privette JL, Tian Y, Lötters A, Zhang Y, Wang Y, Morisette JT, Votava P, Myneni RB, Nemani RR, Running SW (1999) MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) product (MOD15) Algorithm. Theoretical Basis Document. NASA Goddard Space Flight Center, Greenbelt
- Kobayashi H, Delbart N, Suzuki R, Kushida K (2010) A satellite-based method for monitoring seasonality in the overstory leaf area index of Siberian larch forest. *J Geophys Res* 115(G1):G01002. <https://doi.org/10.1029/2009JG000939>
- Kobayashi H, Suzuki R, Kobayashi S (2007) Reflectance seasonality and its relation to the canopy leaf area index in an eastern Siberian larch forest: Multi-satellite data and radiative transfer analyses. *Remote Sens Environ* 106(2): 238–252. <https://doi.org/10.1016/j.rse.2006.08.011>
- Kotani A, Saito A, Kononov AV, Petrov RE, Maximov TC, Iijima Y, Ohta T (2019) Impact of unusually wet permafrost soil on understory vegetation and CO₂ exchange in a larch forest in eastern Siberia. *Agric Forest Met* 265:295–309. <https://doi.org/10.1016/j.agrformet.2018.11.025>
- Liu YY, van Dijk AIJM, de Jeu RAM, Canadell JG, McCabe MF, Evans JP, Wang G (2015) Recent reversal in loss of global terrestrial biomass. *Nat Clim Chang* 5(5):470–474. <https://doi.org/10.1038/nclimate2581>
- Luo Y, Ogle K, Tucker C, Fei S, Gao C, LaDeau S, Clark JS, Schimel DS (2011) Ecological forecasting and DA in a data-rich era. *Ecol Appl* 21(5):1429–1442. <https://doi.org/10.1890/09-1275.1>
- MacBean N, Maignan F, Peylin P, Bacour C, Bréon FM, Ciais P (2015) Using satellite data to improve the leaf phenology of a global terrestrial biosphere model. *Biogeosciences* 12(23):7185–7208. <https://doi.org/10.5194/bg-12-7185-2015>
- Nakai Y, Matsuura T, Kajimoto T, Abaimov AP, Yamamoto S, Zyryanova OA (2008) Eddy covariance CO₂ flux above a Gmelin larch forest on continuous permafrost in central Siberia during a growing season. *Theor Appl Climatol* 93(3–4):133–147. <https://doi.org/10.1007/s00704-007-0337-x>
- Ohta T, Hiyama T, Tanaka H, Kuwada T, Maximov TC, Ohta T, Fukushima Y (2001) Seasonal variation in the energy and water exchanges above and below a larch forest in eastern Siberia. *Hydrol Process* 15(8):1459–1476. <https://doi.org/10.1002/hyp.219>
- Ohta T, Kotani A, Iijima Y, Maximov TC, Ito S, Hanamura M, Kononov AV, Maximov AP (2014) Effects of waterlogging on water and carbon dioxide fluxes and environmental variables in a Siberian larch forest, 1998–2011. *Agric For Meteorol* 188:64–75. <https://doi.org/10.1016/j.agrformet.2013.12.012>
- Ohta T, Maximov TC, Dolman AJ, Nakai T, van der Molen MK, Kononov AV, Maximov AP, Hiyama T, Iijima Y, Moors EJ, Tanaka H, Toba T, Yabuki H (2008) Interannual variation of water balance and summer evapotranspiration in an eastern Siberian larch forest over a 7-year period (1998–2006). *Agric For Meteorol* 148(12):1941–1953. <https://doi.org/10.1016/j.agrformet.2008.04.012>
- Peng C (2000) From static biogeographical model to dynamic global vegetation model: a global perspective on modelling vegetation dynamics. *Ecol Model* 135(1):33–54. [https://doi.org/10.1016/S0304-3800\(00\)00348-3](https://doi.org/10.1016/S0304-3800(00)00348-3)

- Peng C, Guiot J, Wu H, Jiang H, Luo Y (2011) Integrating models with data in ecology and palaeoecology: advances towards a model–data fusion approach. *Ecol Lett* 14(5):522–536. <https://doi.org/10.1111/j.1461-0248.2011.01603.x>
- Ponomarev EI, Kharuk VI, Ranson KJ (2016) Wildfires dynamics in Siberian larch forests. *Forests* 7(12):125. <https://doi.org/10.3390/f7060125>
- Rayner PJ, Scholze M, Knorr W, Kaminski T, Giering R, Widmann H (2005) Two decades of terrestrial carbon fluxes from a carbon cycle DA system (CCDAS). *Global Biogeochem Cy* 19(2):GB2026. <https://doi.org/10.1029/2004GB002254>
- Reichstein M, Falge E, Baldocchi D, Papale D, Aubinet M, Berbigier P, Bernhofer C, Buchmann N, Gilmanov T, Granier A, Grünwald T, Havráňková K, Ilvesniemi H, Janous D, Knohl A, Laurila T, Lohila A, Loustau D, Matteucci G, Meyers T, Miglietta F, Ourcival JM, Pumpanen J, Rambal S, Rotenberg E, Sanz M, Tenhunen J, Seufert G, Vaccari F, Vesala T, Yakir D, Valentini R (2005) On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Glob Chang Biol* 11(9):1424–1439. <https://doi.org/10.1111/j.1365-2486.2005.001002.x>
- Rogers A, Medlyn BE, Dukes JS, Bonan G, von Caemmerer S, Dietze MC, Kattge J, Leakey ADB, Mercado LM, Niinemets Ü, Prentice IC, Serbin SP, Sitch S, Way DA, Zaehle S (2017) A roadmap for improving the representation of photosynthesis in Earth system models. *New Phytol* 213(1):22–42. <https://doi.org/10.1111/nph.14283>
- Sato H, Ise T (2012) Effect of plant dynamic processes on African vegetation responses to climate change: Analysis using the spatially explicit individual-based dynamic global vegetation model (SEIB-DGVM). *J Geophys Res* 117(G3):G03017. <https://doi.org/10.1029/2012JG002056>
- Sato H, Itoh A, Kohyama T (2007) SEIB–DGVM: A new Dynamic Global Vegetation Model using a spatially explicit individualbased approach. *Ecol Model* 200(3–4):279–307. <https://doi.org/10.1016/j.ecolmodel.2006.09.006>
- Sato H, Kobayashi H, Delbart N (2010) Simulation study of the vegetation structure and function in eastern Siberian larch forests using the individual-based vegetation model SEIB-DGVM. *Forest Ecol Manag* 259(3):301–311. <https://doi.org/10.1016/j.foreco.2009.10.019>
- Sato H, Kobayashi H, Iwahana G, Ohta T (2016) Endurance of larch forest ecosystems in eastern Siberia under warming trends. *Ecol Evol* 6(16):5690–5704. <https://doi.org/10.1002/ece3.2285>
- Stöckli R, Rutishauser T, Baker I, Liniger MA, Denning AS (2011) A global reanalysis of vegetation phenology. *J Geophys Res* 116(G3):G03020. <https://doi.org/10.1029/2010JG001545>
- Suzuki K, Kubota J, Yabuki H, Ohata T, Vuglinsky V (2007) Moss beneath a leafless larch canopy: influence on water and energy balances in the southern mountainous taiga of eastern Siberia. *Hydrol Process* 21(15):1982–1991. <https://doi.org/10.1002/hyp.6709>
- Suzuki R, Yoshikawa K, Maximov TC (2001) Phenological photographs of Siberian larch forest from 1997 to 2000 at Spasskaya Pad, Republic of Sakha, Russia. ACDAP, JAMSTEC, Digital Media, Yokosuka
- Tramontana G, Jung M, Schwalm CR, Ichii K, Camps-Valls G, Ráduly B, Reichstein M, Arain MA, Cescatti A, Kiely G, Merbold L, Serrano-Ortiz P, Sickert S, Wolf S, Papale D (2016) Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms. *Biogeosciences* 13(14):4291–4313. <https://doi.org/10.5194/bg-13-4291-2016>
- University of East Anglia Climatic Research Unit, Harris IC, Jones PD (2015) CRU TS3.23: Climatic Research Unit (CRU) Time-Series (TS) version 3.23 of high resolution gridded data of month-by-month variation in climate (Jan. 1901–Dec. 2014). Centre for Environmental Data Analysis. <https://doi.org/10.5285/4c7fdfa6-f176-4c58-acee-683d5e9d2ed5>
- Wang Q, Tenhunen J, Dinh NQ, Reichstein M, Otieno D, Granier A, Pilegarrd K (2005) Evaluation of seasonal variation of MODIS derived leaf area index at two European deciduous broadleaf forest sites. *Remote Sens Environ* 96(3–4):475–484. <https://doi.org/10.1016/j.rse.2005.04.003>
- Williams M, Schwarz PA, Law BE, Irvine J, Kurpius MR (2005) An improved analysis of forest carbon dynamics using DA. *Glob Chang Biol* 11(1):89–105. <https://doi.org/10.1111/j.1365-2486.2004.00891.x>
- Yan M, Tian X, Li Z, Chen E, Wang X, Han Z, Sun H (2016) Simulation of forest carbon fluxes using model incorporation and DA. *Remote Sens* 8(7):567. <https://doi.org/10.3390/rs8070567>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)