Regional-Scale Data Assimilation with The Spatially Explicit Individual-Based Dynamic Global Vegetation Model (SEIB-DGVM) Over Siberia

Hazuki Arakida (✉ hazuki.arakida@riken.jp)

RIKEN Center for Computational Science <https://orcid.org/0000-0002-8570-7878>

Shunji Kotsuki

Chiba University

Shigenori Otsuka

RIKEN Center for Computational Science

Yohei Sawada

The university of Tokyo

Takemasa Miyoshi

RIKEN Center for Computational Science

Research article

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

Posted Date: October 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-88946/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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 dormancy 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 tree LAI, gross primary production (GPP), and above ground biomass. Most remarkably, the spatial distribution of tree LAI was estimated separately from undergrowth LAI because the SEIB-DGVM simulated the vertical structure of forest explicitly, and because satellite-observed LAI provided information on the onset and the end of the leaf season of tree and undergrowth, respectively. The DA system also provided the spatial distribution of the model parameters for tree separately from those of undergrowth. DA experiments started dormancy of trees more than a month earlier than the default phenology model and resulted in a decrease of the GPP.

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 pools (Braswell et al. 2005; Knorr and Kattage 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 tree and undergrowth. Namely, the satellite-derived LAI estimates are highly affected by undergrowth (forest floor) reflectance (Erikson et al. 2006). Earlier increment of LAI before the actual tree (overstory) foliation is thought to be undergrowth LAI (e.g. Kobayashi et al. 2007; 2010). Therefore, satellite-observed LAI should be assimilated as the total LAI of canopy trees and undergrowth. Here, forest structure is needed to be considered when phenology-related parameters for canopy trees are optimized with satellite-observed LAI. Individual-based dynamic global vegetation models (DGVMs) simulate vertical forest structure explicitly; undergrowth 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 undergrowth phenology.

The present study aims at optimizing model parameters for tree phenology at a large spatial domain. The DA system of Arakida et al. (2017) improved not only LAI but also other unobserved variables such as carbon flux and biomass at one of the flux sites in Siberia. This study also explores how assimilating LAI improves those unobserved variables at a regional scale. Therefore, in this study, we extended the previous experiment of Arakida et al. (2017) to examine the performance of the DA system in estimating tree LAI, plant functional type (PFT) specific model parameters, and other state variables across Siberia.

Methods

Study sites

We selected Siberia as the study area because a single tree 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 undergrowth.

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. No tree was simulated within light blue circles,

or some climate forcing data values were missing from the area within blue circles in Fig. 1. These sites were excluded from experiments. We selected total 750 sites (black circles in Fig. 1).

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 undergrowth was attenuated by tree; therefore, the amplitude of tree LAI affected undergrowth LAI. Carbon flux, water flux, heat flux, and vegetation structures (e.g. tree 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 (“tree”) and C3 grass (“undergrowth”) were selected for experiments. The major tree species in this area is larch (Ponomarev et al. 2016), and the forest undergrowth includes cowberry, grass, and shrubs (Ohta et al. 2014). These vegetations were included in the category C3 grass (in this study, simply “undergrowth”), following Sato et al. (2010).

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). Interannual differences were not considered in this study; a yearlong climate forcing data averaged between 2003 and 2014 was used to simulate the averaged vegetation structures and functions during this time period (Table 1).

First, we interpolated bilinearly each NCEP/NCAR and CRU data points into the center of each study site. We supplementary used the monthly CRU data so that the monthly mean air temperature and cloudiness were identical to those of the CRU. Likewise, precipitation and specific humidity were rescaled so that the monthly totals matched those of the CRU. In contrast, we used the NCEP/NCAR soil temperature and wind velocity data without scaling. Finally, the climate observations from 2003 through 2014 were averaged for each study site to provide daily climate forcing data.

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 8,000 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 dormancy start date (Dor; day of year [DOY]) for tree and undergrowth. The initial perturbation ranges were as follows: $P_{\max}$ for tree = [0, 60], $P_{\max}$ for undergrowth = [0, 15], Dor for tree = [200, 300], Dor for undergrowth = [200, 300]. Here, [a, b] denotes the uniform distribution for an interval between a and b. Next, we performed 8,000 parallel simulations with these parameter sets over 100 years, in which the averaged forcing climate data for the period of 2003–2014 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.

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 observation error standard deviations were assigned to each pixel of MCD15A3 LAI at the original resolution of 1 km. 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 2003–2014 as the median for each 30-km radius (Table 1). We aggregated data within the 30-km radius. 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 prepared in the third step. 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.

Assessment of the DA results

To assess the impact of DA, an experiment without DA (“NODA” here after) was also performed with the parameter sets which were used for the initial perturbations. The experiments with DA (“TEST” here after) at the fourth year of DA was 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 was compared for total LAI (tree + undergrowth), tree LAI, GPP, and above ground 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 unobserved state variables.

Data used for cross-comparisons

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. 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), tree 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 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 2013 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 tree 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 2013 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 tree 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 correlation coefficient (Cor) to examine relationships in the data collected for the same area.

Results

Total LAI

Figures 2a-b display the spatial distributions of the annual maximum of total LAI (see the median in Figure S1a). The total LAI for NODA (Fig. 2a) was larger than the observations across the study sites. In contrast, DA reduced LAI for TEST (Fig. 2b) and made it closer to the observed LAI (Fig. 2c). The scatter plot (Fig. 2d) indicates that DA reduced LAI and raised the correlation coefficient from 0.39 to 0.99. These results indicate at least the necessary condition that the DA system worked properly.

Tree LAI

Figure 3a displays the spatial distributions of the annual maximum of tree LAI (see the median in Figure S1b, right). The spatial distribution of tree LAI for TEST (Fig. 3a) is similarly estimated to that of Kobayashi et al. (2010) (Fig. 3b), except for higher tree LAI for TEST at the middle to southern parts of Siberia. DA also reduced tree LAI for TEST (Fig. 3c), and the correlation coefficients between the tree LAI of the current study and that of Kobayashi et al. (2010) increased from 0.50 to 0.81. Because SEIB-DGVM calculates the vertical structure of vegetation, undergrowth LAI is also estimated separately from tree LAI (not shown).

Parameters

Model parameters uniformly distributed for NODA (see the parameter distributions in Figure S1d-g, left) throughout the study sites (not shown). The medians of tree Pmax and undergrowth Pmax for NODA were 30 and 7.5, respectively. Pmax for TEST varied spatially (Figs. 4a-b, left): 23.2 ± 4.6 (mean $\pm$ SD) for tree Pmax, and 7.1 ± 1.8 for undergrowth Pmax. Hence, Pmax must be optimized, especially for tree, to reproduce the observed LAI, and the optimum value varies spatially. The particle spread for Pmax was still large for TEST throughout the study area (Figs. 4a-b, right).

Tree Dor for TEST (Fig. 4c left) was about 60 days earlier than that of undergrowth (Fig. 4d, left), whereas the initial perturbation sizes for the tree and undergrowth Dor were identical. Namely, the DA system distinguished tree Dor from undergrowth Dor. The particle spread for Dor increased with latitude (Figs. 4c-d, right). Arakida et al. (2017) showed that the DA system distinguished tree and undergrowth Dor 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 annual maximum LAI was relatively large (larger than 2.0), and the particle spread for Dor at those sites was smaller than those at other sites.

Unobserved state variables

To explore how DA affected the unobserved state variables, we first compared the correlation coefficients with observed LAI. Table 2 shows that most of the 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 (tree LAI, biomass, and soil carbon) simultaneously. Tree height, tree density, DBH (diameter at breast height), NEE (net ecosystem exchange), evaporation, and SH (sensible heat) are not highly correlated with the observed LAI, indicating that these variables do not benefit from assimilating LAI in this system.

Among the variables in Table 2, GPP is directly related to LAI. Optimization of tree LAI should also affect the estimation of aboveground biomass, the other forest structural variable. For more detailed analysis, we compared the optimized GPP and aboveground biomass with other previous studies (Fig. 5). GPP and aboveground biomass for TEST were reduced by DA (Fig. 5) corresponding to the reduction in LAI (Fig. 2d).

As for GPP, DA increased the correlation between this study and those of FLUXCOM from 0.44 to 0.88 (artificial neural network: ANN), from 0.47 to 0.82 (multivariate regression splines: MARS), and from 0.56 to 0.92 (random forest: 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). As for aboveground biomass, the correlation coefficient (Fig. 5d) increased from 0.12 (NODA) to 0.40 (TEST).

Discussion

Performance of the DA system at regional scales

We applied the DA system with an individual-based 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 tree LAI separately from that of undergrowth LAI. Kobayashi et al. (2010) pioneered to estimate seasonal change of tree (overstory) LAI separately from that of undergrowth LAI at a regional scale with a three-dimensional radiative transfer model and satellite-observed data. The correlation coefficients between the tree LAI of the current study and that of Kobayashi et al. (2010) increased from 0.50 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 higher tree LAI (> 2.0) estimates in this study tended to exceed those of Kobayashi et al. (2010) (Figure. 3c). This 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. The maximum tree LAI in the Yakutsk larch forest (YLF) was about 1.6 in this study, close to the estimate of 1.56 of Ohta et al. (2008). Additional LAI observation data are also needed for further validation.

Parameter estimation

The DA system estimated the spatial distribution of model parameters for tree separately from the undergrowth. Although tree and undergrowth Pmax values in the original model were fixed at 13 and 8, respectively, the estimated Pmax in the current study varied spatially. The estimated Pmax may reflect ecological characteristics, but further studies are needed for further discussions. The reduction of the particle spread for tree 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). Furthermore, the particle spread for tree Pmax (Fig. 4a, right) increased along with increases in the observed LAI (Fig. 2c) and its standard deviation (not shown). Two possible reasons are 1) uncertainties in the observation data may affect the parameter estimation, or 2) other parameters may affect the amplitude of LAI at the regions with higher LAI. 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.

Dor for tree was the most well-estimated parameter with a smaller particle spread. Figure 6 shows the two-dimensional kernel density of the scatter plot of estimated tree Dor in this study (TEST) against that in the original SEIB-DGVM. SEIB-DGVM identifies the start date of dormancy when the 10-day running average of daily mean air temperature was lower than 7°C based on a single flux site in Siberia (Sato et al. 2010). In general, the estimated tree Dor in the current study was about 40 days earlier than that in SEIB-DGVM (Fig. 6). Dor of the current SEIB-DGVM is based on a single flux site which is included in our study sites. For the regional simulation, the phenology model must be constructed using regional data (Ise et al. 2018). Therefore, a more sophisticated model equation for tree Dor at this region is necessary, and for such a development, the estimated parameter by DA might be useful.

DA impact on unobserved state variables

DA reduced LAI estimates throughout the study area, which made the estimated LAI very close to the observed LAI (Fig. 2d). In addition to the reduction in LAI, DA reduced values of most of the unobserved variables (e.g. Figure 3c, Fig. 5). The resulting spatial correlations between the estimated variables and the observed LAI were generally high (Table 2). Hence, optimization of LAI markedly affected the spatial distributions of fluxes and vegetation structure at the regional scale.

The estimates obtained in this study were highly correlated with those of previous investigations using satellite observations with an optical sensor for FLUXCOM (Figs. 5b: Tramontana et al. 2016; Jung et al. 2017) and tree LAI (Fig. 3c: 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 by DA are a natural consequence as long as the MODIS LAI is assimilated correctly. However, improvements in the aboveground biomass correlations were less than those for carbon flux and tree LAI (Fig. 5c, Liu et al. 2015). Liu et al. (2015) used microwave-based observations, and the principles of the observation are fundamentally different from this study. Hence, both observational (optical vs. microwave) and estimation procedures influence model outputs. 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.

Limitations

This study demonstrated that the DA system was able to estimate the structure of the virtual forests (i.e., tree and undergrowth) at the regional scale. However, tree density was not highly correlated with observational data (Table 2). This suggests that the number of trees be unidentifiable by LAI.

The DA system was unstable in the study area at latitudes lower than 60°N, and the estimated LAI for undergrowth exceeded that for trees. In these unstable cases, apparently the particles converged to an unrealistic vegetation structure. In these areas, tree Dor was later than undergrowth Dor (Figs. 4c-d, left), and tree LAI was low (Fig. 3a). To resolve these contradictions, the DA system must be further improved for the lower latitudes.

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 tree LAI at a regional scale; this leads to the estimation of spatial distributions of model parameters for tree and undergrowth 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:normalised 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; Dor:Dormancy start date; DOY:Day Of Year; ANN:Artificial Neural Network; MARS:Multivariate Regression Splines; RF:Random Forest; Cor:(correlation) Coefficient; NODA:results without DA; TEST:results with DA; SD:Standard Deviation; RE:Ecosystem Respiration; NEE:Net Ecosystem Exchange; LH:Latent Heat; SH:Sensible Heat; DBH:Diameter at Breast Height; SOM:Soil Organic Matter; YLF:Yakutsk Larch Forest

Declarations

Availability of data and material

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/>).

Competing interests

No competing interest.

Funding

None

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, YS collaborated with the corresponding authors (HA and TM) in the construction of the manuscript. All authors read and approved the final manuscript.

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 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 tree LAI data based on Delbart et al. (2005) and Kobayashi et al. (2010) was retrieved from the author's website (<http://flies.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>).

References

1. 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* 7:044008. doi:10.1088/1748-9326/7/4/044008.
2. 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 Proc Geoph* 24:553-567. doi:10.5194/npg-24-553-2017.
3. 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. *Global Change Biology* 11:335-355. doi:10.1111/j.1365-2486.2005.00897.x.
4. Cheaib A, Badeau V, Boe J, Chuine I, Delire C, Dufrêne 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* 15:533–544. doi:10.1111/j.1461-0248.2012.01764.x.
5. 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* 97:26-38. doi:10.1016/j.rse.2005.03.011.
6. 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: L15402. doi:10.1029/2007GL030014.
7. 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* 103:408-418. doi: 10.1016/j.rse.2006.04.005.
8. European Commission, Joint Research Centre (2003) The global land cover map for the year 2000, GLC2000 database. European commission joint research centre. <http://www-gem.jrc.it/glc2000>. Accessed 6 Jan 2017.

9. 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 Change Biol* 24:35-54. doi: 10.1111/gcb.13910.
10. 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. doi:10.1029/2011GL048738.
11. 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* 111:3280-3285. doi:10.1073/pnas.1222477110.
12. Gao C, Wang H, Weng E, Lakshmivarahan 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* 21:1461-1473. doi: 10.1890/09-1234.1.
13. Gordon NJ, Salmond DJ, Smith AFM (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proc F* 140:107–113. doi:10.1049/ip-f-2.1993.0015.
14. 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. doi: 10.3389/fenvs.2018.00095.
15. 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* 12:085001. doi: 10.1088/1748-9326/aa7a19.
16. 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* 541:516-520. doi:10.1038/nature20780.
17. 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. doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.
18. 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-Biogeosci* 118:1414–1426. doi: 10.1002/jgrg.20118.
19. 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* 10:789–802. doi:10.5194/bg-10-789-2013.
 20. Kitagawa G (1998) A self-organizing state-space model. *J Am Stat Assoc* 93:1203-1215. doi: 10.2307/2669862.
 21. Knorr W, Kattge J (2005) Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. *Glob Change Biol* 11:1333-1351. doi: 10.1111/j.1365-2486.2005.00977.x.
 22. Knyazikhin Y, Glassy J, Privette JL, Tian Y, Lotsch 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, MD 20771, USA.
 23. 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* 106:238-252. doi: 10.1016/j.rse.2006.08.011.
 24. 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:G01002. doi: 10.1029/2009JG000939.
 25. 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 Change* 5:470-474. doi:10.1038/nclimate2581.
 26. 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* 21:1429–1442. doi: 10.1890/09-1275.1.
 27. 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* 12:7185-7208. doi:10.5194/bg-12-7185-2015.
 28. 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* 93:133–147. doi:10.1007/s00704-007-0337-x.
 29. Ohta T, Hiyama T, Tanaka H, Kuwada T, Maximov TC, Ohata T, Fukushima Y (2001) Seasonal variation in the energy and water exchanges above and below a larch forest in eastern Siberia. *Hydrol Process* 15:1459–1476. doi: 10.1002/hyp.219.
 30. 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). *Agr. Forest Meteorol* 148:1941–1953. doi:10.1016/j.agrformet.2008.04.012.
31. 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. *Agricultural and forest meteorology* 188:64–75. doi:10.1016/j.agrformet.2013.12.012.
 32. Peng C (2000) From static biogeographical model to dynamic global vegetation model: a global perspective on modelling vegetation dynamics. *Ecol Model* 135:33–54. doi:10.1016/S0304-3800(00)00348-3.
 33. 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* 14:522–536. doi: 10.1111/j.1461-0248.2011.01603.x.
 34. Ponomarev EI, Kharuk VI, Ranson KJ (2016) Wildfires dynamics in Siberian larch forests. *Forests* 7:125. doi:10.3390/f7060125.
 35. 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:GB2026. doi: 10.1029/2004GB002254.
 36. Reichstein M, Falge E, Baldocchi D, Papale D, Aubinet M, Berbigier P, Bernhofer C, Buchmann N, Gilmanov T, Granier A, Grünwald T, Havránková 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 Change Biol* 11:1424–1439. doi:10.1111/j.1365-2486.2005.001002.x.
 37. 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* 213:22–42. doi: 10.1111/nph.14283.
 38. Sato H, Itoh A, Kohyama T (2007) SEIB–DGVM: A new Dynamic Global Vegetation Model using a spatially explicit individualbased approach. *Ecol Model* 200:279–307. doi: 10.1016/j.ecolmodel.2006.09.006.
 39. 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* 259:301–311. doi:10.1016/j.foreco.2009.10.019.
 40. 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:G03017. doi:10.1029/2012JG002056.
 41. Sato H, Kobayashi H, Iwahana G, Ohta T (2016) Endurance of larch forest ecosystems in eastern Siberia under warming trends. *Ecol Evol* 6:5690–5704. doi:10.1002/ece3.2285.

42. Stöckli R, Rutishauser T, Baker I, Liniger MA, Denning AS (2011) A global reanalysis of vegetation phenology. *J Geophys Res* 116:G03020. doi:10.1029/2010JG001545.
43. 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* 13:4291-4313. doi:10.5194/bg-13-4291-2016.
44. 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, 09 November 2015. doi:10.5285/4c7fdfa6-f176-4c58-acee-683d5e9d2ed5.
45. Williams M, Schwarz PA, Law BE, Irvine J, Kurpius MR (2005) An improved analysis of forest carbon dynamics using DA. *Glob Change Biol* 11:89–105. doi: 10.1111/j.1365-2486.2004.00891.x.
46. 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* 8:567. doi: 10.3390/rs8070567.

Tables

Due to technical limitations, table 1,2 is only available as a download in the Supplemental Files section.

Figures

Figure 1

Study sites (black circles). Light-blue and blue circles were excluded because no trees were simulated in light-blue circles, and because the forcing data included missing values in blue 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, U.S. Central Intelligence Agency (The World Factbook). The same continental data are used in subsequent maps.

Figure 2

Total LAI (annual maximum of median). (a) NODA, (b) TEST, (c) observation, (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.

Figure 3

Tree 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), (c) scatter plot of this study against the estimates from Delbart et al. (2005) and Kobayashi et al. (2010).

Figure 4

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 trees, (b) Pmax for undergrowth, (c) dormancy start date (Dor) for trees, (d) Dor for undergrowth.

Figure 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). Data were missing from three study sites examined by Liu et al. (2015), so that these sites are not included in the figures.

Figure 6

Kernel density of the scatter plot of estimates of tree dormancy start date (Dor). The plot shows Dor obtained in this study with data assimilation (TEST) against the original SEIB-DGVM estimates.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Graphicalabstractimage.png](#)
- [Table1.JPG](#)
- [Table2.JPG](#)

- [20200925arakidaSupportingInformation.docx](#)