**Supporting Information**

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

Hazuki Arakida1,2\*, Shunji Kotsuki1,3,4, Shigenori Otsuka1,5,6, Yohei Sawada7,1, and Takemasa Miyoshi1,5,6,8,9\*

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

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

3 Center for Environmental Remote Sensing (CEReS), Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba-shi, 263-8522, Japan

4 PRESTO, Japan Science and Technology Agency, 7, Gobancho, Chiyodaku, Tokyo,102-0076, Japan

5 RIKEN Interdisciplinary Theoretical and Mathematical Sciences Program, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo, 650-0047, Japan

6 RIKEN Cluster for Pioneering Research, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo, 650-0047, Japan

7 Institute of Engineering Innovation, The University of Tokyo, Yayoi 2-11-16, Bunkyo-ku, Tokyo 113-8656, Japan.

8 Department of Atmospheric and Oceanic Science, University of Maryland, 4254 Stadium Dr., College Park, MD 20742, USA.

9 Application Laboratory, Japan Agency for Marine-Earth Science and Technology, 3173-25, Showa-machi, Kanazawa-ku, Yokohama-city, Kanagawa, 236-0001, Japan.

Two Corresponding authors:

Hazuki Arakida1,2\*

Email: arakidahk@hydrosoken.co.jp

Takemasa Miyoshi1,5,6,8,9\*

Email: takemasa.miyoshi@riken.jp

Several existing studies were used for cross-comparisons in this study. The details for these studies are as follows. MODIS LAI (Knyazikhin et al. 1999) and the overstory LAI (Delbart et al. 2005; Kobayashi et al. 2010) have been estimated using a radiative transfer model and satellite-based data. Aboveground biomass has been estimated using the relationships between satellite-based vegetation optical depth and the aboveground biomass measured in the tropical regions (Liu et al. 2015). GPP of FLUXCOM is based on 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 optical-based vegetation indices. All datasets described above were generated from satellite observations. Liu et al. (2015) estimated aboveground biomass using microwave-based observations, but other studies, including this study, have been based on satellite observations with an optical sensor.

Figure S1 shows the results of this study for four years after spin-up at the same study site as Arakida et al. (2017). The left panels show the distribution of LAI and parameters without DA (“NODA”), and the right panels show the results with DA (“TEST”). In the course of spin-up, the perturbed parameters (Figures S1d–g, left) produced differences in the leaf-bearing season and the amplitude of LAI (Figures S1a–c, left). Other unassimilated variables were also perturbed (not shown, see Arakida et al. 2017). DA brought the distribution of LAI much closer to the observation standard deviation (Figure S1a, right). LAI for overstory (Figure S1b, right) and understory (Figure S1c, right) were estimated separately. The 1–99% quantile range for LAI and the model parameters (Figure S1) were greatly reduced for TEST compared to NODA. These results are consistent with those of Arakida et al. (2017), indicating that the DA system functioned well with the modifications in Table1.



Figure S1. Time series of leaf area index (LAI) in the Yakutsk larch forest (YLF): a) overstory + understory, b) overstory, c) understory, with (right) and without (left) data assimilation (DA) (Similar to Figure 7 in Arakida et al. (2017), with modifications listed in Table 1. The time series of parameters in YLF are also shown: d) maximum photosynthetic rate (Pmax) for overstory, e) Pmax for undergrowth, f) defoliation start date (DSD) for overstory, g) DSD for undergrowth, with (right) and without (left) DA (Similar to Figure 8 in Arakida et al. (2017) with modifications listed in Table 1. The red squares indicate the output in the fourth year after DA, which is used for the distribution maps.