1 **Supplementary Information**

2 *1. Data*

3 Since 1994, a large-scale ground inventory has been conducted on numerous permanent sample

- 4 plots established on the forested areas across the boreal region of Alaska. Inventory crew
- 5 collected size (DBH and percentage cover) and status of all the plant species including tree,
- 6 shrub, herbaceous vascular, and non-vascular species on an annual basis (1). The Alaska boreal
- 7 forest, the largest forest component of the Alaska landscape, is characteristic of some of the most
- 8 severe climatic conditions in the world with short growing seasons. Soil parent materials vary
- 9 greatly from alluvial, loess, glacial, and lacustrine deposits to various bedrock types, with
- 10 discontinuous permafrost present throughout the region. The sample area, stretching over
- 11 300,000 km² from Fairbanks, AK in the north to the Kenai Peninsula in the south and the
- 12 International Boundary in the east (**Fig. S1**), represents a wide range of species composition and
- 13 physiographic conditions. A set of permanent sample plots were used to capture the dynamics of
- 14 283 coexisting plant species in the region (**Table S1**)*.*
- 15 The fixed-size permanent sample plots, each square in shape and 0.04 ha in size, were re-
- 16 measured with a 5-year interval. The attributes measured include diameter at breast height,
- 17 height, tree status, the percentage cover of non-tree species, elevation, percentage slope,
- 18 thickness of organic horizons, and number of snags. The number of *Picea glauca* individuals in
- 19 comparative relationship to the total number of trees on a plot was calculated to represent the
- 20 sampling effect, as *Picea glauca* has the highest individual ANPP among all the tree species that
- 21 almost doubles the second highest species (*Betula neoalaskana*) (**Fig. S2**). In addition, two time
- 22 indicators (T_1, T_2) were introduced to represent from which inventory a particular observation
- 23 was made. From the 603 established permanent sample plots, 440 were selected for this study
- 24 because they have been measured at least twice and have not experienced any crown fire or
- 25 logging between two consecutive inventories.
- 26 For individual trees, we calculated the above-ground net primary productivity (ANPP) using the 27 following equation:

28
$$
ANPP_{ijk} = (V_{2,ijk} - V_{1,ijk})/(y_2 - y_1),
$$
 (M.1)

- 29 where the ANPP of the *k*th tree in species *i* on plot *j* is average annual increment in above-
- 30 ground biomass (AGB) of that tree from year y_1 to y_2 . V_1 and V_2 represent AGB of that tree at
- 31 year y_1 and y_2 , respectively, calculated from its measured DBH using the species-specific
- 32 biomass equations developed for the Alaska boreal forest(2). *ANPP* is zero for a tree that died
- 33 between year y_1 and y_2 . ANPP is the part of NPP that occurs aboveground and is composed of
- 34 wood and foliage productivity, whereas root growth represents belowground productivity, which
- 35 is extremely difficult to measure and contributes to ecosystem services, such as carbon storage,
- 36 but not to harvested fiber. ANPP of trees was selected as the response variable of our spatio-
- 1 temporal model, as it is more routinely measured and represents wood production, a surrogate for
- 2 economically important fiber production.
- 3 The total tree ANPP of a community (*TNPP*) was then calculated using the following equation:

4
$$
TNPP_j = \left(\sum_i \sum_k V_{2,j} - \sum_i \sum_k V_{1,j} + \sum_i \sum_k IG_j + \sum_i \sum_k MT_j\right) / (y_2 - y_1),
$$
 (M.2)

5 where $\Sigma \Sigma V_{2,i}$ and $\Sigma \Sigma V_{1,i}$ represent total AGB of trees on plot *j* that were alive at year *y*₂ and *y*₁,

- 6 and ΣΣ*IGj* and ΣΣ*MTj* are respectively total AGB of ingrowth and mortality on plot *j* between 7 year y_1 and y_2 .
- 8 The AGB of shrub, herbaceous vascular, and non-vascular species was calculated based on the
- 9 observed percentage cover and average height of that species using the species-specific biomass
- 10 equations calibrated for interior Alaska(3). On average, tree species accounts for most of the total
- 11 AGB, whereas shrub, herbaceous vascular, and non-vascular species accounts for less than 10
- 12 percent of the total AGB (**Fig. S3**).

13 In our empirical analysis, the authors relied upon two *measures of biodiversity*, which are widely 14 used in both ecological and economic research:

15 - Shannon index, *H*, which measures biodiversity with an entropic approach:

16
$$
H_{j} = -\sum_{i=1}^{n} \frac{V_{ij}}{V_{j}} \ln \left(\frac{V_{ij}}{V_{j}} \right),
$$
 (M.3)

17 where V_{ij} presents the live aboveground biomass of species *i* on plot *j*, and V_j the total live 18 aboveground biomass of Plot *j*.

19 – Simpson index, *D*. The same index known as the Herfindahl-Hirschman index (HHI) is a 20 measure of diversity widely used in competition law, antitrust, and management:

21
$$
D_j = 1 - \sum_{i=1}^{n} \left(\frac{V_{ij}}{V_j} \right)^2
$$
. (M.4)

22 *2. The Spatio-Temporal Model*

23 In investigating the empirical relationship between biodiversity and ANPP, the authors elected to

24 develop a regression-based spatio-temporal model to account for spatial autocorrelation which

- 25 has been found prevalent in attributes of the Alaska boreal forest, such as stand basal area, tree
- 26 diversity, and forest productivity (4). Failing to account for such correlation in space will render 27 model outputs less reliable (5). Although an alternate statistical approach, namely structural
- 28 equation modeling (SEM), has been employed in ecology for the testing of causal relationship in
- 29 complex datasets, spatial autocorrelation violates basic SEM assumptions and remains difficult
- 30 to model in a SEM framework (6). Furthermore, computational constraints of existing SEM
- 1 software have limited applicability of SEM on this study because our dataset consisted of tree
- 2 records from 40,526 spatially explicit locations.
- 3 Autocorrelation in both space and time were accounted for in our model to avoid biased or
- 4 inconsistent coefficient estimates and significance levels. Spatial autocorrelation in above-
- 5 ground net primary productivity was estimated using a nonparametric spatial covariance function.
- 6 This method incorporates a smoothing spline to measure the correlation between values (in our
- 7 case ANPP) from pairs of samples over a continuous function of the distance separating samples,
- 8 without assuming any functional form *a priori*. However, because of the large size (over 40,000
- 9 entries) of our spatially explicit dataset, it was not feasible to estimate spatial autocorrelation
- 10 based upon the entire dataset due to computational constraints. To this end, a bootstrap
- 11 approach(7) was used to quantify the spatial covariance function. We iteratively and randomly
- 12 selected approximately 10% of the data, and then repeated this procedure 500 times. From this
- 13 distribution of 500 replicated estimates of the spatial covariance function, the mean was
- 14 determined and 95% confidence intervals were estimated as the 0.025% and 0.975% quantiles of
- 15 the distribution(7).
- 16 Across the 500 replications, the mean number of sampling locations selected (from the 40,526
- 17 locations) per iteration was 4056, and ranged from 3838 to 4216. The authors detected spatial
- 18 correlation in ANPP (**Fig. S4**), and estimated a mean local spatial autocorrelation (the correlation
- 19 as the spatial distance between pairs of sample ANPP values approaches 0) of 0.09 and a range
- 20 of spatial dependence (the lag distance at which spatial autocorrelation approaches 0) of \sim 10 km.
- 21 Moreover, the functional form of the spatial covariance function over lag distance suggested that
- 22 the use of spherical model would be the most appropriate one to use as an error term when
- 23 assessing the significance of potential explanatory variables on ANPP.
- 24 To avoid compromised type-I error rates and severe artifacts associated with common model
- 25 selection procedures, hierarchical partitioning (HP) was used here to evaluate the goodness-of-fit
- 26 of all the possible combinations of explanatory variables in determining the average independent
- 27 contribution of each variable(8). The threshold value of independent contribution was set at 8
- 28 percent, the average value with 12 explanatory variables (**Table S2**). Diversity variables and
- 29 time indicators were always present in the model, regardless of their independent contribution.
- 30 HP was conducted with the hier.part package of the R program. Total aboveground biomass, *V*,
- 31 the number of snags, *S*, and the frequency of *Picea glauca*, *P*, were selected as control variables
- 32 as each of them has an independent contribution to the individual ANPP no less than 8 percent
- 33 (**Fig. S5**). Biodiversity *ω*(**s**) was represented in this model by the full quadratic forms of Shannon
- 34 index, *H*, and Simpson index, *D*, both of which provide a continuous measure of biodiversity.
- 35 The following spatio-temporal model was developed to examine the effect of biodiversity on
- 36 individual plant productivity (ANPP model):

$$
37 \quad \text{ANPP}_{ijk}(\mathbf{s},\omega) = \alpha_0 + \alpha_1 \cdot \omega(\mathbf{s}) + \alpha_2 \cdot T_1(\mathbf{s}) + \alpha_3 \cdot T_2(\mathbf{s}) + \beta \cdot \mathbf{z}(\mathbf{s}) + \varepsilon_i(\mathbf{s})\,,\tag{M.5}
$$

- 1 where $\mathbf{s} = (\lambda, \varphi) \in \mathcal{D} \subset \mathbb{R}^2$ represents the geographic location of a plot *j* within the Alaska
- 2 boreal region \mathcal{D} . In this specification, individual ANPP at geographic location **s** is a linear
- 3 function of the biodiversity of the same plot, the temporal indicators T_1 and T_2 , and the vector of
- 4 control variables **z**. *εi*(**s**) is a spatially autocorrelated residual term best described as a spherical
- 5 variogram 2*ς* model with intrinsic stationarity and isotropic covariance. *α*0, *α*1, *α*2, *α*3, **β**, and *ς* are
- 6 parameters to be estimated with generalized least squares (GLS). The two time indicators (T_1, T_2)
- 7 were introduced to the model to account for potential factors of ANPP that are not controlled for
- 8 by the explanatory variables in this model, such as the temporal difference in sample plot
- 9 locations.
- 10 Due to computational constraints, a bootstrap approach was used to obtain model coefficients
- 11 and predicts. The authors iteratively and randomly selected approximately 1% of the data from
- 12 the entire dataset, obtained GLS estimates, log-likelihood, and predictions across the full
- 13 observed range of each biodiversity measure for the sensitivity analysis, and then repeated this
- 14 procedure 1,000 times. From the distribution of 1,000 replicated estimates and predicts of the
- 15 model, the mean and 95% confidence intervals were determined.
- 16 Estimates and goodness-of-fit of the spatio-temporal model are summarized in **Table S3**. All the
- 17 coefficients are significant at the 5 percent level. The signs of total aboveground biomass, *V*, and
- 18 the frequency of *Picea glauca*, *P*, are consistently positive, indicating an expected positive effect
- 19 of total biomass and sampling effect on individual tree ANPP. The negative signs of the number
- 20 of snags, *S*, and the two temporal indicators T_1 and T_2 are also consistent. The two goodness-of-
- 21 fit measures Akaike information criterion (AIC) and Bayesian information criterion (BIC) were
- 22 calculated based on the average log-likelihood values of all the bootstrap iterations, the number
- 23 of coefficients, and the total number of observations. The present spatio-temporal models were
- 24 able to capture an interaction in space and time, as the negative coefficients of T_1 and T_2 indicate
- 25 a declining trend of ANPP over time, which is likely caused by the fact that most plots in the
- 26 third and fourth inventories are centered around Fairbanks, AK, where the sites are in general
- 27 less productive.
- 28

29 **SI REFERNCES**

- 30 1. Malone T, Liang J, & Packee. EC (2009) Cooperative Alaska Forest Inventory. (Gen. 31 Tech. Rep. PNW-GTR-785, USDA Forest Service, Pacific Northwest Research Station, 32 Portland, OR), p 42.
- 33 2. Yarie J, Kane ES, & Mack MC (2007) Aboveground biomass equations for trees of 34 Interior Alaska. (Agricultural and Forestry Experiment Station, University of Alaska 35 Fairbanks, Fairbanks, AK), p 15.
- 36 3. Mack M*, et al.* (2008) Recovery of Aboveground Plant Biomass and Productivity After 37 Fire in Mesic and Dry Black Spruce Forests of Interior Alaska. *Ecosystems* 11(2):209- 38 225.
- 1 4. Liang J (2012) Mapping large-scale forest dynamics: a geospatial approach. *Landscape*
- 2 *Ecol.* 27(8):1091-1108.
3 5. Legendre P (1993) Spat: 3 5. Legendre P (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology* 4 74(6):1659-1673.
- 5 6. Lamb EG, Mengersen KL, Stewart KJ, Attanayake U, & Siciliano SD (2014) Spatially explicit structural equation modeling. *Ecology* 95(9):2434-2442. 6 explicit structural equation modeling. *Ecology* 95(9):2434-2442.
- 7 7. Efron B & Tibshirani RJ (1993) *An Introduction to the Bootstrap* (Chapman & Hall, New 8 York) p 436.
9 8. Chevan A &
- 9 8. Chevan A & Sutherland M (1991) Hierarchical partitioning. *Amer. Statistician* 45:90-96.
- 10
- 11
- 12

- 2 **Fig. S1 Geographic distribution of the 440 permanent sample plots and their relative**
- 3 **location within the State of Alaska (inset)**. All the plots have been measured at least twice with
- 4 an interval of 5 years, 280 have been measured in 3 inventories, and 65 in 4 inventories, making a total of 785 plot observations.
- a total of 785 plot observations.

- **Fig. S2 Total net primary productivity (TNPP) (primary axis) and average individual**
- **ANPP (secondary axis) by tree species with standard error bars.**

Fig. S3 Total above-ground biomass (AGB) by species and plant type with standard error bars.

- 2 **Fig. S4 Mean (solid line) and 95% confidence intervals (dashed lines) of the nonparametric**
- 3 **spatial covariance function**, showing local spatial autocorrelation in values of above-ground net
- 4 primary productivity, and a range of spatial dependence of ~10 km.

- **Fig. S5 Independent contribution of biodiversity measures (blue wedges) and control**
- **variables (red wedges) to the goodness-of-fit of individual ANPP.**

Table S1 | List of plant species. A total of 283 coexisting plant species studied in this paper are

2 categorized into nine general functional groups—tree, shrub, forb, grass and rush, sedge, fern, 3 lichen, moss, and liverwort.

1 **Table S2 | Definition, unit, and summary statistics of the variables used in this study.**

1 **Table S3 | Estimates and goodness-of-fit of the spatio-temporal model supporting N–E.**

2 Initial explanatory variables were selected with HP, and parameters were estimated using GLS.
3 Biodiversity variables and their significance levels were in shade.

Biodiversity variables and their significance levels were in shade.

4 *const*: constant
5 AIC: Akaike in 5 AIC: Akaike information criterion
6 BIC: Bayesian information criterion

6 BIC: Bayesian information criterion
 $\overline{7}$ \overline{n} : number of observations that is u

 n : number of observations that is used in the bootstrap simulation

8

9