



ELSEVIER

Contents lists available at ScienceDirect

MethodsX

journal homepage: [www.elsevier.com/locate/mex](http://www.elsevier.com/locate/mex)

## Method Article

# Methodology to calculate forest stand level maximum potential productivity, potential achievable productivity and ecosystem fit



Angela M. Klock\*, Kristiina A. Vogt, Daniel J. Vogt, John C. Gordon

University of Washington, Yale University, USA

## A B S T R A C T

A modified Loomis-William model was originally developed to estimate the theoretical maximum yields of crops. That model was adapted in this paper to measure how much of the theoretical maximum potential productivity ( $tNpp_{tmax}$ ) is reached in any forest due to edaphic and climatic limits to growth, i.e., its “Ecosystem fit” (eFit). The procedure to calculate eFit has not been published except as a concept. Our goal is to describe the methodology in sufficient detail to facilitate its use by the scientific community and forest managers. To calculate eFit you need: 1) to convert all photosynthetically active radiation to a photosynthetic product for each forest plot or stand to calculate its  $tNpp_{tmax}$ , and 2) use field-collected data of total observed net primary productivity ( $tNpp_{obs}$ ). Theoretical maximum potential  $tNpp$  is calculated with a simple light-use efficiency model as the product of the efficiency at which forest canopies absorb solar radiation, the photosynthetic conversion efficiency into biomass, and remotely sensed solar radiation with temperature data extracted to the geographic coordinates for the site. Ecosystem fit represents a forest’s realized percent productive capacity and is the ratio of field-collected  $tNpp$  (i.e.,  $tNpp_{obs}$ ) to the theoretical maximum potential  $tNpp$  (i.e.,  $tNpp_{tmax}$ ).

- Available indices to assess forest productivity and adaptive capacity to land-use disturbance and climate change are sensitive at the small-to-meso spatio-ecophysiological scales.
- A more holistic index (such as eFit) will provide an informative picture of forest conditions where management practices are undertaken and the ecosystem’s capacity to adapt to environmental change.
- A comparison of eFit across similar forests within a climatic zone is an indication of the stressors or constraints that are being imposed locally and that limit  $tNpp_{obs}$ .

© 2022 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

DOI of original article: [10.1016/j.ecolind.2022.108973](https://doi.org/10.1016/j.ecolind.2022.108973)

\* Corresponding author.

E-mail address: [Angelk9@uw.edu](mailto:Angelk9@uw.edu) (A.M. Klock).

<https://doi.org/10.1016/j.mex.2022.101812>

2215-0161/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

## ARTICLE INFO

*Method name:* Calculating forest stand level adaptive capacity to produce biomass

*Keywords:* Random forest, Clustering, Binary trees, Reaction norm, Primary productivity, Ecosystem fit

*Article history:* Received 23 May 2022; Accepted 3 August 2022; Available online 6 August 2022

## Specifications table

Subject Area:	Environmental Science
More specific subject area:	Forest ecology and management
Method name:	Calculating forest stand level adaptive capacity to produce biomass
Name and reference of original method:	Loomis, R.S., Williams, W.A. 1963. Maximum crop productivity: An estimate. <i>Crop Sci.</i> , 3, 67-72. Gordon, J.C., Farnum, P., Timmis, R., 1983. Theoretical maximum phytomass yields as guides to yield improvement. In: B. Thielges, ed., <i>Proc. 7th N. Am. For. Bio. Work. Univ. Kentucky, Lexington, KY.</i> Gordon, J.C., Bormann, B.T., Kiestler, A.R., 1992. The physiology and genetics of ecosystems: A new target or "Forestry contemplates an entangled bank". <i>Proceedings of the 12th North American Forest Biology Workshop. Sault Ste. Marie, Ontario, Canada. Aug 17-20, 1992. Ontario Ministry of Natural Resources, Ontario Forest Research Institute and Forestry Canada, Ontario Region, pp. 1-14</i> N.A.
Resource availability:	N.A.

## Ecosystem Fit rationale to measure adaptive capacity of forests to grow biomass

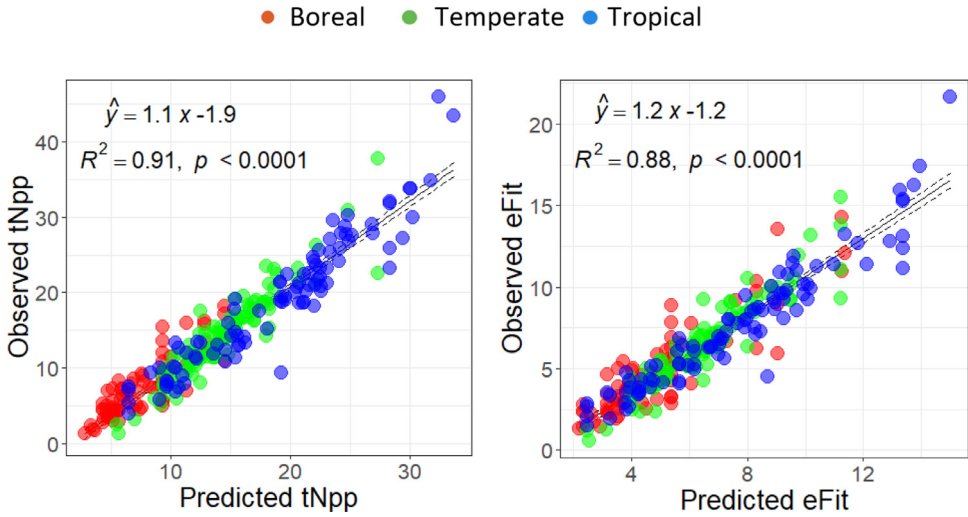
The "Ecosystem Fit" (eFit) of a forest [8] is the ratio of  $tNpp_{obs}/tNpp_{tmax}$  (where  $tNpp_{obs}$  = total net primary productivity measured at each site and  $tNpp_{tmax}$  = theoretical maximum potential total net primary productivity possible when there are no limits to growth), each in units of dry biomass ( $Mg\ ha^{-1}\ yr^{-1}$ ) in this paper. It expresses how well a forest is adapted to its site and its potential for growth, i.e., is it possible to grow more biomass than what is measured in the field. Ecosystem Fit [7] directly compares the measured  $tNpp$  (or  $tNpp_{obs}$  as used in this paper) to growth that is theoretically possible for that site without site-level constraints (i.e.,  $tNpp_{tmax}$  as used in this paper).

Current measures of  $tNpp$  do not allow an assessment of how vulnerable a forest is to climate change since this metric does not identify ranked thresholds of low, medium, or high  $tNpp$  by site (*sensu* [9,21]). In addition,  $tNpp$  does not inform how well forests perform in relation to site limiting factors. Thus, comparison of sites requires a productivity metric that defines the maximum potential productivity for each site. For example, forest stand-level comparisons for management strategies could be misleading as an internal reference of gross carbon flux since it could differ unpredictably between sites due to the variability of stand-specific factors affecting productive capacity. Whereas calculating eFit of an ecosystem relative to its current environment can be obtained with the calculation of  $tNpp_{tmax}$  as a reference condition to assess the potential of any site to grow biomass [7,8].

Since photosynthesis is the ecophysiological framework where carbon assimilation occurs, it is an indicator to quantify the environment's functional limitations on production [8,14]. Trees convert solar energy into biomass within their genetic and environmental constraints. When all the structural and functional attributes of trees operate at their optimal levels, and the environmental conditions are within the normal range of function, productivity will be maximum for that ecosystem under the conditions present [7].

Observed  $tNpp$ 

The dataset used in this paper consisted of 267 natural forest field sites and 40 forest plantation sites (for a total of  $n = 307$ ) to develop the methodology to calculate plot level maximum potential productivity ( $tNpp_{tmax}$ ), observed or actual productivity ( $tNpp_{obs}$ ), and Ecosystem Fit [9]. The data and variables included in the dataset were comprised of site-level field measures that included above- and below-ground productivity collected from mature boreal, temperate and tropical



**Fig. 1.** Validation of full Random Forest regression model of field measures of total productivity ( $tNpp_{obs}$ ) and calculated Ecosystem Fit (eFit) for natural forests ( $n = 267$ ). Colors indicate the distribution of the different climatic zones and  $R^2$  values are for the full dataset.

natural and plantation forests (from [9], Table 1 – variables used in the study database; Fig. 1 – Geographic distribution of sites; and references to individual sites included in our database are in the Supplementary data file - <https://doi.org/10.1016/j.ecolind.2022.108973>).

### Method details for calculating $tNpp_{tmax}$ and eFit

#### Calculating $tNpp_{tmax}$ and eFit

A basic light-use efficiency model was first used to calculate the site-specific theoretical maximum potential net primary productivity without site-level constraints limiting growth. This provided a site-specific maximum  $tNpp$  that was then used to calculate Ecosystem Fit. Theoretical potential maximum  $tNpp$  (or  $tNpp_{tmax}$ ) is the product of photosynthetically active radiation (PAR) and plant physiological parameters to calculate biomass (*sensu* [4]). This simple efficiency model calculates growth as the product of mean solar radiation available during the growing season [5], the interception efficiency ( $E_i$ ) at which forest canopies absorb solar radiation, and the conversion efficiency ( $E_c$ ) or rate at which absorbed solar radiation is converted into tree biomass. Mathematical equations and model parameter specifications were from Bernacchi et al. [1,2] and the Global Change & Energy Project [6].

The calculation of  $E_c$  or the conversion efficiency of forest canopies used the average annual atmospheric  $CO_2$  concentration the year of data collection (1981–2011) for each site as measured by the Mauna Loa Observatory (<https://www.sealevel.info/co2.html>). Model assumptions were that photorespiration was a constant fraction of photosynthesis, PAR was 45% of total solar radiation, and forest canopies absorbed 90% of incoming PAR during active growth [1]. In addition, we assumed no nutrient or water limitations to growth. Finally, the growing season was calculated as total month-days when mean monthly temperature exceeded zero.

Ecosystem Fit (eFit) is the ratio of measured  $tNpp$  (or  $tNpp_{obs}$ ) to the estimated theoretical maximum potential  $tNpp$  (or  $tNpp_{tmax}$ ), which is percent eFit ( $tNpp_{obs}/tNpp_{tmax}$ ) as shown in Eq. 1:

$$eFit = \frac{tNpp_{obs}}{tNpp_{tmax}} \times 100 \quad (1)$$

where eFit = Ecosystem Fit,  $tNpp_{obs}$  = total above- and belowground productivity from field collected  $tNpp$  data ( $Mg\ ha^{-1}\ yr^{-1}$ ), and  $tNpp_{tmax}$  = theoretical maximum potential  $tNpp$  ( $Mg$

$\text{ha}^{-1} \text{yr}^{-1}$ ). Determination of the environmental variables driving eFit percentages were determined independently at each scale by machine learning algorithms with the R packages “randomForest” [3,13], “randomForestSRC,” [18] “caret” [11], and “pdp” [19].

#### Required data acquisition and processing

The methods involved in the development of  $\text{tNpp}_{\text{tmax}}$ , eFit, and metrics of forest ecosystem resilience require the following: (1) direct field measures of primary productivity and site-level environmental conditions, (2) global proximal-sensed solar radiation and temperature data, (3) identification of productivity groups using unsupervised clustering, (4) determination of response thresholds, correlations of productivity for each explanatory variable, and the most important environmental variables using binary regression tree analysis, and (5) visualization of response dependence to environmental gradients and low-order interactions that are often critical to understanding ecological processes.

The Random Forest algorithm is not affected by violations of the typical model assumptions of linearity and normality with little/no multicollinearity, and homoscedastic variance. This method has several desirable characteristics: (i) the ability to handle multiple types of data; (ii) the ability to assess interactions and nonlinearity; (iii) the ability to calculate variable importance, and iv) the ability to handle unbalanced data sets [12].

#### Required data

- 1) Field measures of climatic, environmental, edaphic, and biotic metrics.
- 2) Total net primary productivity ( $\text{tNpp}$ ) calculated as the sum of field measures of aboveground net primary productivity ( $\text{aNpp}$ ) and belowground net primary productivity ( $\text{bNpp}$ ), and therefore is equivalent to  $\text{tNpp}_{\text{obs}}$  as used in this paper.
- 3) High spatial resolution average monthly/daily temperature and solar radiation data extracted for the geographic coordinates of each site (buffer zones can be used to calculate a mean for a given area in lieu of point data).

#### Required derived data

- 1) Determine the length of the growing season, i.e., the number of grow days when the temperature exceeds zero ( $G_d$ ).
- 2) Calculate mean temperature for the growing season ( $G_t$ ).
- 3) Calculate mean solar radiation for the growing season ( $G_{\text{srad}}$ ).

#### Model assumptions and parameterization

$R$  = molar gas constant ( $8.314 \text{ J mol}^{-1} \text{ K}^{-1}$ )

$K$  = specific energy of plant dry biomass ( $18.2 \text{ MJ kg}^{-1}$ ) [6,16]

$c_a$  = standard or site-/time-level atmospheric concentration of  $[\text{CO}_2 \text{ ppm}]$

$c_i$  =  $c_a \times 0.7$  fraction of intercellular leaf  $[\text{CO}_2 \mu\text{mol mol}^{-1}]$  [1]

$\varepsilon_i$  = interception efficiency at which tree canopies absorb solar radiation (0.90)

$\beta$  = fraction of absorbed quanta reaching Photosystem II (PSII) (0.45) [1]

$\Gamma^*$  =  $\text{CO}_2$  compensation point in the absence of dark respiration ( $\mu\text{mol mol}^{-1}$ )

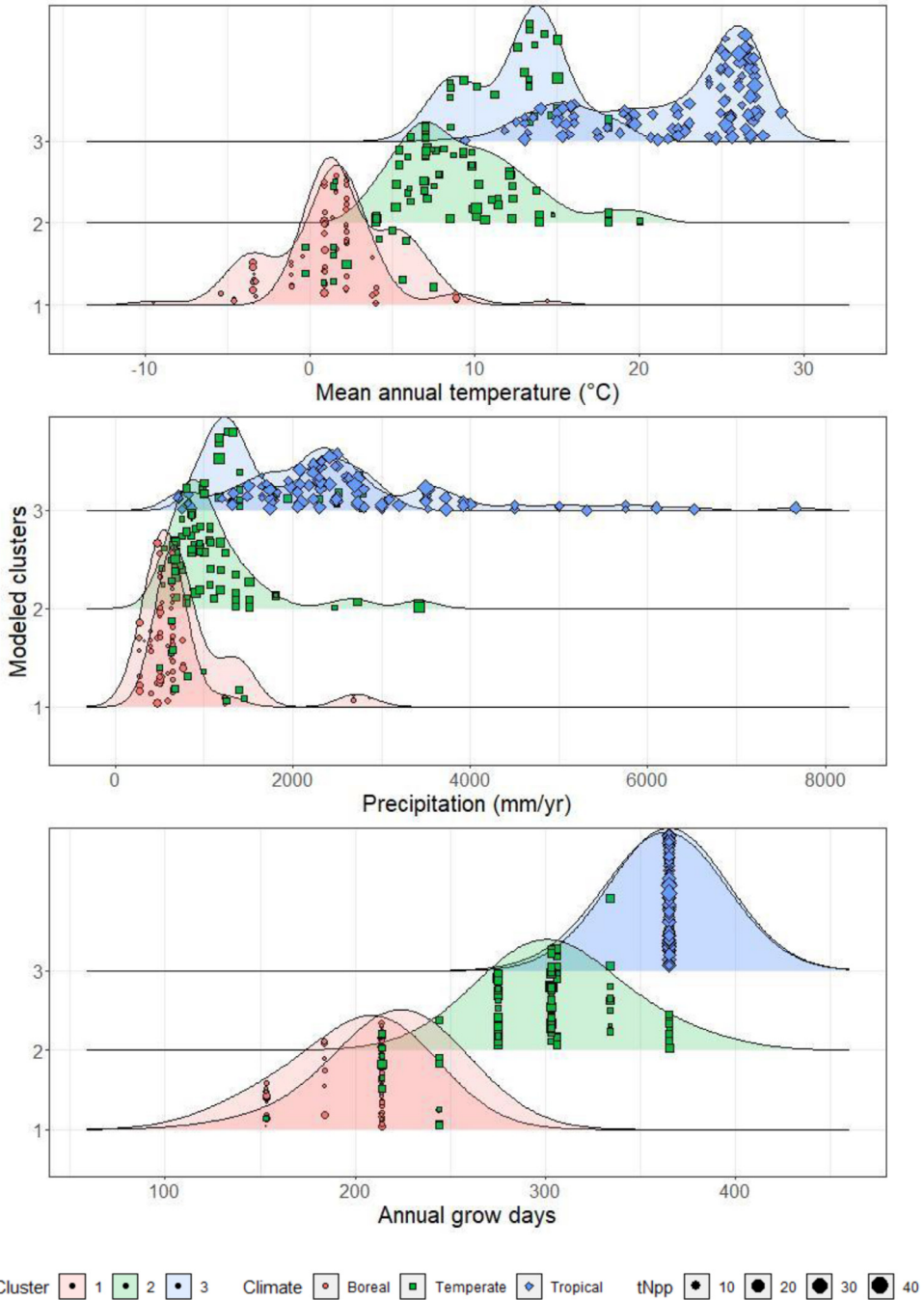
$$\Gamma^* = \exp \left( \frac{c - \Delta H_a}{R \times (^\circ\text{C} + 273.15)} \right) \quad (2)$$

where  $c$  (kJ/mol) represents a scaling constant and  $\Delta H_a$  (kJ/mol) represents activation energy,  $R$  is the molar gas constant, and the temperature  $T$  ( $^\circ \text{K}$ ) measured at each site.

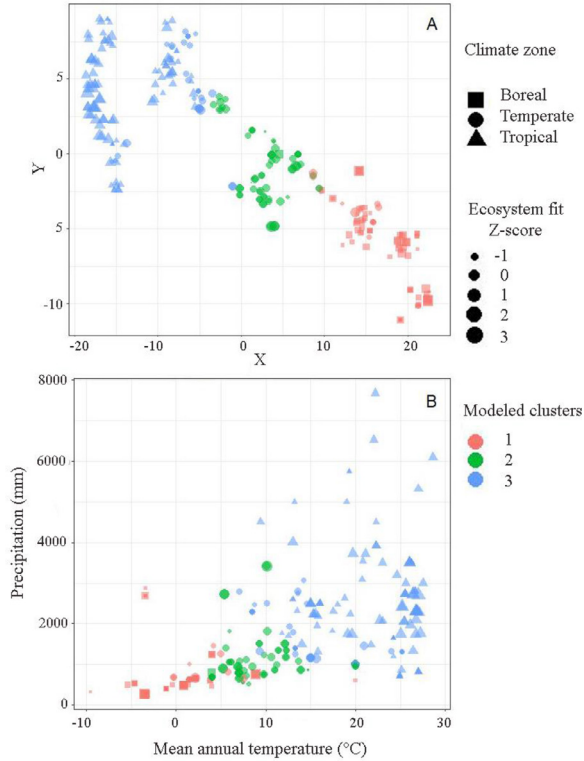
$A$  = net rate of  $\text{CO}_2$  uptake after accounting for carboxylation and oxygenation

$$A_i = \left( 1 - \frac{\Gamma^*}{c_i} \right) \times W_i \quad (3)$$

where  $W_i$  is the rubisco-limited rate of carboxylation ( $\mu\text{mol m}^{-2} \text{ s}^{-1}$ )



**Fig. 2.** Prediction of climatic classification and density distributions of  $tNpp_{obs}$  to the environmental gradient of the three most important variables identified by cluster analysis (represented by the foreground density fill) to the climatic zone designations assigned to the data (represented by the background density fill). Symbols represent the original climatic classification. Color indicates modeled cluster groups and climatic zone. Symbol size represents the decile breaks of  $tNpp_{obs}$  for each natural forest ecosystem.



**Fig. 3.** Comparison of model-generated clusters (color) identified by t-SNE and climatic classifications of the data (shapes) of natural forests in (A) semantic similarities in a reduced dimensional space where proximity indicates similarity among forest stands (X and Y axes units have no intrinsic meaning), and (B) distribution in relation to total annual precipitation and mean annual temperature. Symbol size represents the Z-score (standard deviation) of Ecosystem fit.

$W_i = c_i / (4.5 \times c_i + 10.5 \times \Gamma^*)$  [1,2]  $\varepsilon_c$  = photosynthetic conversion efficiency (rate at which C3 plants convert absorbed solar energy into tissue)

$$\varepsilon_c = A \times \beta \times \varepsilon_i \quad (4)$$

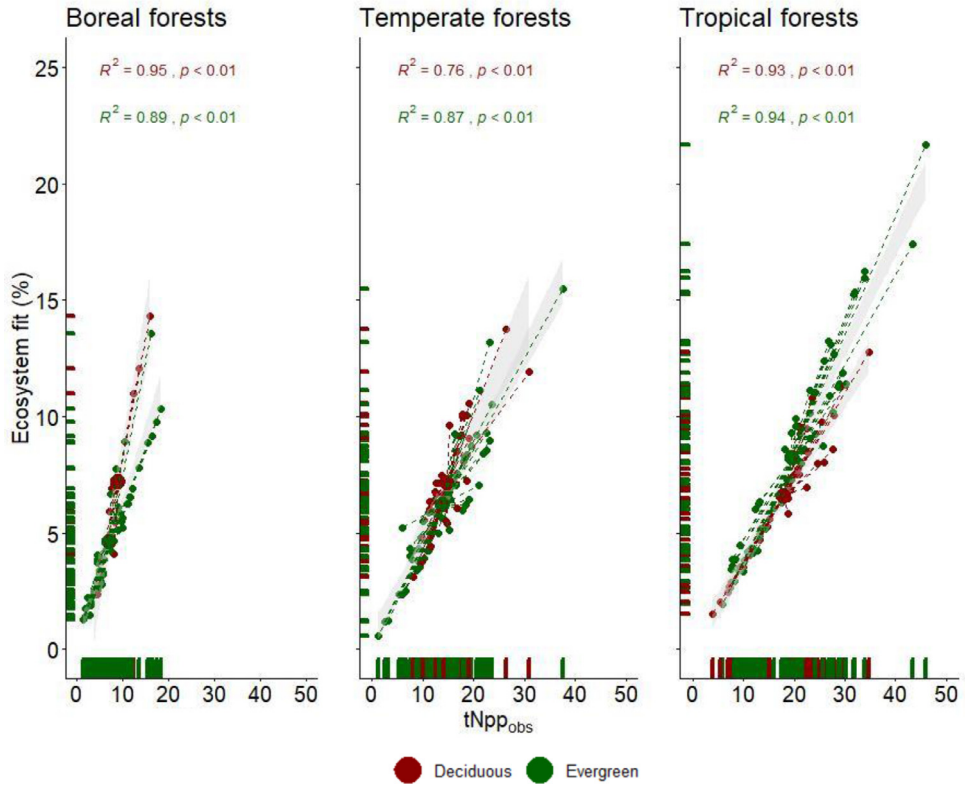
Derive theoretical maximum primary productivity ( $tNpp_{tmax}$ ) and Ecosystem Fit (eFit)

Calculate  $tNpp_{tmax}$  as the product of solar radiation during the growing season and plant physiological parameters.

$$tNpp_{tmax} = G_{srad} \times \varepsilon_c \times \varepsilon_i \times K \quad (5)$$

where  $G_{srad}$  is mean solar radiation available during the growing season and  $K$  is the inverse specific energy of plant biomass ( $18.4 \text{ MJ kg}^{-1}$ ).

Photosynthesis is the eco-physiological framework where carbon assimilation occurs. Thus, it is an indicator to quantify the environment's functional limitations on primary productivity [8]. Trees convert solar energy into biomass within their genetic and environmental constraints. If we suppose all structural and functional attributes of trees are operating at their optimal levels, and the environmental conditions are within the tree's normal range of operation, productivity will be maximized for that ecosystem under the conditions present [7]. Based on these constructs, eFit is described as a ratio of measured  $tNpp_{obs}$  to the estimated  $tNpp_{tmax}$  ( $tNpp_{obs}/tNpp_{tmax}$ ) or percent eFit (see Eq. 1).



**Fig. 4.** Comparison of the relationship between eFit and tNpp<sub>obs</sub> in each climatic zone (left to right) and leaf phenology (red = deciduous, green = evergreen). Colored plot rug along the x-axis indicates observations by phenology. The large point is the centroid which represents the mean of all observations and the grey background represents the 95% confidence interval. The statistic is the squared Pearson product-moment correlation coefficient for the corresponding data pairs for deciduous and evergreen natural forests ( $n = 267$ ).

### Model assumptions and validation of field collected data

This study performed statistical functions in R version 4.0.3 [17], and all analyses were subject to the Shapiro-Wilk tests of normality [22]. Correlations were considered significantly different at  $p < 0.05$  and linear regressions were evaluated by behavior of residuals and variance explained. To demonstrate the independence of tNpp and eFit to phenological structure at the global scale and their reliance on stand-level site conditions, comparisons were made with Wilcoxon signed-rank tests for non-parametric and unbalanced data for climatic classifications. Random forest model performance was evaluated by variable importance and which variables (or combinations of variables) and their interactions had the most predictive power. We validated the random forest regression models with visualization of the discrepancies between field observations and predictions of tNpp<sub>obs</sub> and eFit and calculation of Pearson product-moment correlation coefficient for each (Fig. 1).

To validate the random forest classification model of the climatic zones we first evaluated the underlying structure of the data by first removing all phenological and climatic membership identifiers. Dissimilarity matrices (distances between all pairs of data points) were calculated for the three most important variables identified by the random forest regression, mean annual temperature, precipitation, and annual grow season. A multivariate random forest model was used to predict the clusters. The data were then aggregated using the partitioning around medoids (PAM) method from

the R package “cluster” [20], which repeatedly iterates until the medoid position is stabilized. The medoid of the cluster represents the median of all the attributes included. The optimal number of clusters was selected by evaluating the average silhouette width and the variance explained by the first two principal components. The clusters were visualized with nonlinear dimensional reduction with the R package “Rtsne” [10] utilizing the t-distributed stochastic neighbor embedding technique (t-SNE) [15]. The algorithm first creates probability distributions of the distance relationships between points in their high-dimensional space and then projects them into a reduced 2-dimensional semantic space that retains the proximities of the relationships. This allows identification of misclassifications and density distributions for classes to each environmental variable (Fig. 2).

All class designations were validated by comparing the climatic assignment to model-generated clusters (Fig 3. A) and to the environmental gradients of annual precipitation and mean annual temperature (Fig. 3 B). The Z-score for eFit was added as proportional-sized shapes.

The relationships between tNpp and eFit vary by biome and phenological behavior (Fig. 4). The amount of variance of eFit explained by tNpp was lowest in the temperate forest biome, and especially for deciduous forests where 76% of the variance of eFit was explained by tNpp compared to the evergreen forests (87%). This may be due to the high seasonal variability of the temperate zone compared to the stability of the tropics and strong seasonal forcing and limits of the boreal climate. Thus, if a researcher wants to explain tNpp and eFit in deciduous forests in the temperate climatic zone, other variables need to be explored to improve the strength of the relationship between eFit and tNpp.

#### Sample R script:

```
##-----SETUP: LOAD REQUIRED PACKAGES-----
require(raster); require(rockchalk); require(dplyr)
##-----LOAD DATA-----
data<-read.csv()
samples <- data.frame(X, Y)
##...create raster stack of mean temperature;
mean.temp <- list.files("", ".tif", full.names = TRUE)
meantemp<-stack(mean.temp)
##...rename layers in the raster stack
month <- c("Jan", "Feb", ...)
names(meantemp)<-month
##...extract data from raster layer
meantemp.data <- extract(meantemp, samples)
x<-as.data.frame(meantemp.data)
##...set all negative values to NA
x[x<0]<-NA
##...create raster stack objects for solar radiation;
srad.files <- list.files("", ".tif", full.names=TRUE)
srad <- stack(srad.files)
##...unique names for layers in the raster stack
month <- c("Jan1", "Feb1", ...)
names(srad)<-month
##...extract data from raster layer
srad.data <- extract(srad, samples);
srad.data<-as.data.frame(srad.data)
##...select solar months based on temperature months
tempsol<-data.frame(x, srad.data)
##...change all NA temps to NA solars
tempsol$Jan1 <- ifelse(tempsol$Jan == "NA", tempsol$Jan1 <-NA,
tempsol$Jan1)
...
##...calculate the row means for temps => 0 (growing season)
tempsol$GrowTemp<-rowMeans(tempsol[1:12], na.rm=T)
```



```

##...Calculate the row means for solar radiation ( $G_{srad}$ ) ###
tempsol$GrowSolar <- rowMeans(tempsol[13:24], na.rm=T)
tempsol$GrowDaysJan<-ifelse(tempsol$Jan1 == "NA", tempsol$GrowDaysJan
<- 0, tempsol$GrowDaysJan <- 31)
tempsol$GrowDaysFeb<-ifelse(tempsol$Feb1 == "NA", tempsol$GrowDaysFeb
<- 0, tempsol$GrowDaysFeb <- 28)
...
##...column bind variables to new data
dataSoRad<-cbind(data, tempsol[c(25:26,39)])
##...model parameters up front
R <-0.008314 #(kJ mol-1 K-1)
Ca <-400 #CO2 concentration (standard or observed)
Ci <-ca*0.7 #280
Beta <-0.45
Alpha <-0.9
Ei <-0.9
K <-18.2 #Inverse specific energy of plant biomass
##...calculate gamma (photorespiration compensation point)
gamma <- exp(19.02 - 37.83 / (R*(GrowTemp + 273.15)))
Wj <- ci / (4.5 * ci + 10.5 * gamma)
##...calculate alpha
A <-(1 -- gamma / ci) * Wj
##...calculate photosynthetic efficiency (Ec)
Ec <- A * beta * alpha
##...calculate theoretical maximum NPP (tNpptmax) w/attention to unit
conversion
tNpptmax <- GrowSolar * GrowDays * 0.45 * Ei * Ec * K / units
##...Calculate ecosystem fit
eFit <- tNppobs / tNpptmax # ratio or percent

```

## Data availability

See supplemental materials [9].

## Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] C.J. Bernacchi, C. Pimentel, S.P. Long, In vivo temperature response functions of parameters required to model RuBP-limited photosynthesis, *Plant Cell Environ.* 26 (9) (2003) 1419–1430, doi:[10.1046/j.0016-8025.2003.01050.x](https://doi.org/10.1046/j.0016-8025.2003.01050.x).
- [2] C.J. Bernacchi, E.L. Singaas, C. Pimentel, A.R. Portis Jr, S.P. Long, Improved temperature response functions for models of Rubisco-limited photosynthesis, *Plant* 24 (2001) 253–259.
- [3] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32, doi:[10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [4] E.H. DeLucia, N. Gomez-Casanovas, J.A. Greenberg, T.W. Hudiburg, I.B. Kantola, S.P. Long, A.D. Miller, D.R. Ort, W.J. Parton, The theoretical limit to plant productivity, *Environ. Sci. Technol.* 48 (16) (2014) 9471–9477, doi:[10.1021/es502348e](https://doi.org/10.1021/es502348e).
- [5] S.E. Fick, R.J. Hijmans, WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas, *Int. J. Climatol.* 37 (12) (2017) 4302–4315, doi:[10.1002/joc.5086](https://doi.org/10.1002/joc.5086).
- [6] Global Change & Energy Project, An Assessment of Carbon Capture Technology and Research Opportunities, Assessment (2005) 1–20 [http://gcep.stanford.edu/pdfs/assessments/carbon\\_capture\\_assessment.pdf](http://gcep.stanford.edu/pdfs/assessments/carbon_capture_assessment.pdf).
- [7] J.C. Gordon, P. Farnum, R. Timmis, Theoretical maximum phytomass yields as guides to yield improvement, in: B. Thielges (Ed.), *Proc. 7th N. Am. For. Bio. Work. Univ. Kentucky, Lexington, KY, 1983*.
- [8] J.C. Gordon, B.T. Bormann, A.R. Kiester, The physiology and genetics of ecosystems: a new target or “Forestry contemplates an entangled bank”, in: *Proceedings of the 12th North American Forest Biology Workshop*. Sault Ste. Marie, Ontario, Canada. Aug 17-20, 1992, Ontario Ministry of Natural Resources, Ontario Forest Research Institute and Forestry Canada, Ontario Region, 1992, pp. 1–14.

- [9] A.M. Klock, K.A. Vogt, D.J. Vogt, J.C. Gordon, J.J. Scullion, A.S. Suntana, K.K. Mafune, A.Y. Polyakov, S.J. Gmur, C. Gómez de la Rosa, See the forest not the trees! Ecosystem-based assessment of response, resilience, and scope for growth of global forests, *Ecol. Indicators* 140 (2022), doi:[10.1016/j.ecolind.2022.108973](https://doi.org/10.1016/j.ecolind.2022.108973).
- [10] J.H. Krijthe, Rtsne: T-Distributed Stochastic Neighbor Embedding using a Barnes-Hut Implementation, (2015) URL: <https://github.com/jkrijthe/Rtsne>.
- [11] M. Kuhn, caret: Classification and Regression Training. R package version 6.0-92. (2022) <https://CRAN.R-project.org/package=caret>.
- [12] S.C. Lemon, J. Roy, M.A. Clark, P.D. Friedmann, W. Rakowski, Classification and regression tree analysis in public health: methodological review and comparison with logistic regression, *Soc. Behav. Med.* (2003) 172–181.
- [13] A. Liaw, M. Wiener, Classification and Regression by randomForest, *R News* 2 (3) (2002) 18–22.
- [14] R.S. Loomis, W.A. Williams, Maximum crop productivity: an estimate, *Crop Sci.* 3 (1963) 67–72.
- [15] L. Van Der Maaten, G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* 9 (2008) 2579–2605.
- [16] J.L. Monteith, Climate and efficiency of crop production in Britain, *Philos. Trans. R. Soc. Lond.* 281 (1977) 277–295 <https://royalsocietypublishing.org/doi/10.1098/rstb.1977.0140>.
- [17] R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2020 <https://www.r-project.org/>.
- [18] H. Ishwaran, U.B. Kogalur, Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC), R package version 3.0.0 (2022).
- [19] B.M. Greenwell, pdp: An R Package for Constructing Partial Dependence Plots, *The R Journal* 9 (1) (2017) 421–436 <https://journal.r-project.org/archive/2017/RJ-2017-016/index.html>.
- [20] M. Maechler, P. Rousseeuw, A. Struyf, M. Hubert, K. Hornik, cluster: Cluster Analysis Basics and Extensions. R package version 2.1.2. — See the “Changelog” file (in the package source). /package=cluster. (2019) <https://CRAN.R-project.org>.
- [21] D.J. Vogt, K.A. Vogt, S.J. Gmur, J.J. Scullion, A.S. Suntana, S. Daryanto, R. Sigurdardotti, Vulnerability of tropical forest ecosystems and forest dependent communities to drought, *Environ. Res.* 144 (2016) 27–38, doi:[10.1016/j.envres.2015.10.022](https://doi.org/10.1016/j.envres.2015.10.022).
- [22] J. Peat, B. Barton, *Medical Statistics: A guide to data analysis and critical appraisal*, Blackwell Publishing, 2005.