



Assessing model performance via the most limiting environmental driver in two differently stressed pine stands

DANIEL NADAL-SALA,^{1,6} RÜDIGER GROTE ,¹ BENJAMIN BIRAMI,¹ ANNA LINTUNEN,^{2,3} IVAN MAMMARELLA,² YAKIR PREISLER ,⁴ EYAL ROTENBERG,⁵ YANN SALMON,^{2,3} FEDOR TATARINOV,⁵ DAN YAKIR,⁵ AND NADINE K. RUEHR¹

¹Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research - Atmospheric Environmental Research (IMK-IFU), Garmisch-Partenkirchen 82467 Germany

²Faculty of Agriculture and Forestry, Institute for Atmospheric and Earth System Research/Forest Sciences, University of Helsinki, Latokartanonkaari 7, P.O. Box 27, Helsinki, 00014 Finland

³Institute for Atmospheric and Earth System Research/Physics, Faculty of Science, University of Helsinki, P.O. Box 68, Gustaf Hällströmin katu 2b, Helsinki, 00014 Finland

⁴Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138 USA

⁵Department of Environmental Sciences and Energy Research, Weizmann Institute of Science, Rehovot 76100 Israel

Citation: Nadal-Sala, D., R. Grote, B. Birami, A. Lintunen, I. Mammarella, Y. Preisler, E. Rotenberg, Y. Salmon, F. Tatarinov, D. Yakir, and N. K. Ruehr. 2021. Assessing model performance via the most limiting environmental driver in two differently stressed pine stands. *Ecological Applications* 31(4):e02312. 10.1002/eap.2312

Abstract. Climate change will impact forest productivity worldwide. Forecasting the magnitude of such impact, with multiple environmental stressors changing simultaneously, is only possible with the help of process-based models. In order to assess their performance, such models require careful evaluation against measurements. However, direct comparison of model outputs against observational data is often not reliable, as models may provide the right answers due to the wrong reasons. This would severely hinder forecasting abilities under unprecedented climate conditions. Here, we present a methodology for model assessment, which supplements the traditional output-to-observation model validation. It evaluates model performance through its ability to reproduce observed seasonal changes of the most limiting environmental driver (MLED) for a given process, here daily gross primary productivity (GPP). We analyzed seasonal changes of the MLED for GPP in two contrasting pine forests, the Mediterranean *Pinus halepensis* Mill. Yatir (Israel) and the boreal *Pinus sylvestris* L. Hyytiälä (Finland) from three years of eddy-covariance flux data. Then, we simulated the same period with a state-of-the-art process-based simulation model (LandscapeDNDC). Finally, we assessed if the model was able to reproduce both GPP observations and MLED seasonality. We found that the model reproduced the seasonality of GPP in both stands, but it was slightly overestimated without site-specific fine-tuning. Interestingly, although LandscapeDNDC properly captured the main MLED in Hyytiälä (temperature) and in Yatir (soil water availability), it failed to reproduce high-temperature and high-vapor pressure limitations of GPP in Yatir during spring and summer. We deduced that the most likely reason for this divergence is an incomplete description of stomatal behavior. In summary, this study validates the MLED approach as a model evaluation tool, and opens up new possibilities for model improvement.

Key words: Aleppo pine; gross primary productivity; model evaluation; most limiting environmental driver; productivity seasonality; random forest; Scots pine.

INTRODUCTION

Temperature, water availability, and irradiation are key drivers of forest productivity. In boreal regions, forest productivity is often limited by a combination of low temperatures and low radiation. In stark contrast, in semiarid regions forest productivity is mostly limited by water availability (Churkina and Running 1998, Nemani et al. 2003, Ciaia et al. 2005, Seneviratne et al. 2010,

Humphrey et al. 2018). Additionally, specific stressors occur that may directly affect tree functioning (e.g., ozone damage, herbivory grazing, pest occurrence), or swiftly increase water demand, as is the case of high atmospheric vapor pressure deficit (D). This leads to physiological adjustments, e.g., reduction of stomatal conductance, thus modifying the water transport in the soil–plant–atmosphere continuum and limiting leaf internal CO_2 supply for photosynthesis (e.g., Novick et al. 2016, Zhang et al. 2019, Grossiord et al. 2020).

During the next decades, global warming will modify the limitation strength of such environmental drivers differently around the globe. The consequences might

Manuscript received 15 July 2020; revised 6 November 2020; accepted 6 December 2020. Corresponding Editor: David S. Schimel.

⁶E-mail: daniel.sala@kit.edu

range from enhanced forest productivity in colder and more moist regions due to temperature increases (e.g., Hellmann et al. 2016, D'Orangeville et al. 2018, Park et al. 2019), to reductions of productivity in drier regions due to low water availability and higher atmospheric aridity (e.g., Ciais et al. 2005, Schlesinger et al. 2016, Brodrribb et al. 2020, Grossiord et al. 2020). Moreover, the slope of GPP responses to climate will most likely be modified by the parallel increase in atmospheric $[\text{CO}_2]$. Thus, climate change can be expected to affect not only the intensity but also the seasonality of the most limiting environmental drivers of forest productivity (e.g., Hellmann et al. 2016, Park et al. 2019).

The impact on forest productivity of multiple interacting environmental drivers changing simultaneously can only be assessed through process-based forest simulation models (e.g., Medlyn et al. 2015, Dietze et al. 2018). These models require a thorough evaluation against observations in order to assess their robustness. However, assessing model performance by direct comparison of model outputs with observations could be puzzling, as model outputs may match observations well, despite of wrong underlying model assumptions (e.g., Ruimy et al. 1999, Kirchner 2006, Medlyn et al. 2011a, 2015, Rollinson et al. 2017, Bugmann et al. 2019). Hence, an incomplete process understanding in a model may lead to unrealistic outputs. This needs particular attention when simulating unprecedented environmental conditions, as forecasting abilities may be hindered. In contrast to the traditional model validation *via* output-to-observation direct fit, validation of the underlying processes descriptions has been given considerably less attention.

To address this deficit, we developed the concept of the Most Limiting Environmental Driver (MLED). This method is based on the idea that suboptimal environmental conditions reduce potential forest gross primary productivity (GPP_{max}). It also assumes that we can identify the limitation that a given environmental driver exerts upon GPP_{max} in high temporal resolution, e.g., on a daily basis. The limitation is defined as the distance between the current state of this environmental driver and its state when $\text{GPP} = \text{GPP}_{\text{max}}$. By comparing the different (relative) limitations of different drivers, we can determine which one is the MLED, namely the one that is limiting GPP strongest within a given time-step (Fig. 1), even if environmental drivers may be additive in their limitation, e.g., vapor pressure deficit and soil water availability, Novick et al. (2016). Changes in the MLED should help us to understand how the environmental drivers are shifting in importance during the year, and therefore help us to determine the underlying ecophysiological processes that are most likely limiting forest productivity. Assessing model performance based on a model's ability to reproduce changes in the MLED could therefore become a new tool that goes beyond traditional model evaluation, and hence may facilitate to provide the right answers due to the right reasons.

The objective of this paper is to evaluate the MLED approach as a tool for (1) analyzing the seasonality of environmental constraints on GPP and (2) model evaluation. We addressed these objectives by applying the MLED approach to two well-characterized pine forests growing under contrasting climate conditions. We focused on the daily limitation strength of four key environmental drivers on GPP, i.e., air temperature, incoming radiation, soil water availability, and vapor pressure deficit. The two forests are the Mediterranean Yatir forest, an Aleppo pine (*Pinus halepensis* Mill.) plantation in Israel, and the boreal Hyttiälä forest in Finland, which is dominated by Scots pine (*Pinus sylvestris* L.). For both stands, eddy-covariance flux measurements as well as accompanying information about climate, forest management, soil and stand properties, are available (Rannik et al. 2004, Tatarinov et al. 2016, Reyer et al. 2020). In parallel, the same approach has also been applied to daily GPP simulations derived by a state-of-the-art physiologically oriented model, in order to compare observed and simulated MLED dynamics and to assess model performance.

METHODS

Study sites

For this study, we selected two pine stands from the FLUXNET network, i.e., the Israeli Yatir *P. halepensis* Mill. forest and the Finnish Hyttiälä site, where *P. sylvestris* L. is the dominant species, in which long-term eddy-covariance flux measurements are available. Site characteristics for both forests are described in Table 1. Flux data for Hyttiälä was obtained from the PROFOUND database (Reyer et al. 2020). Flux data from Yatir stand was measured in situ following the Euroflux methodology (Aubinet et al. 1999, Grünzweig et al. 2003, Qubaja et al. 2019). We analyzed a period of 3 yr for both stands, in order to ascertain that, on the one hand, we had a large enough pool of observations to train and evaluate the machine learning algorithms, but on the other hand, the period was short enough to neglect the impact of increasing atmospheric CO_2 concentration on GPP (e.g., Norby et al. 2005).

Carbon flux measurements for evaluation

For both stands, GPP_t was calculated half-hourly as $\text{GPP}_t = \text{NEP}_t + \text{ER}_t$ (gross primary productivity equals net ecosystem productivity plus ecosystem respiration), and then integrated into daily GPP. ER_t was inferred from nighttime ecosystem respiration (Reichstein et al. 2005) following a site-specific algorithm both for Yatir (Tatarinov et al. 2016, Qubaja et al. 2019), and for Hyttiälä (Kolari et al. 2009, Peltoniemi et al. 2015). A conservative criterion of including only good and very good data quality was applied by restricting to use NEP_t half-hourly observations of category 0 and 1 (Fluxnet

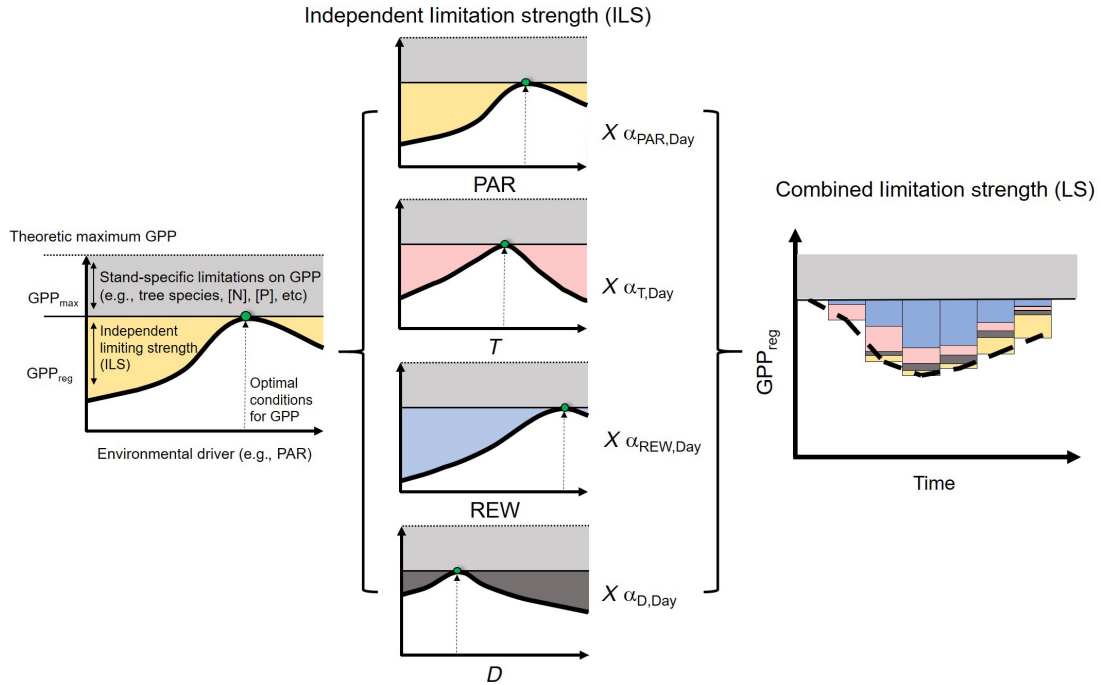


FIG. 1. Conceptual framework of the Most Limiting Environmental Driver (MLED) approach. First, a stand-specific maximum gross primary production (GPP_{max}) under optimal climate conditions is defined from observations. In a second step, the independent limitation strength of each environmental driver (ILS_i) over GPP_{max} is calculated. Here we accounted for incoming photosynthetic active radiation (PAR), air temperature (T), relative extractable water (REW), and vapor pressure deficit (D). Third, the limitation strength for each i driver (LS_i) is calculated daily by combining its independent limitation strength (ILS_i) with the daily α_i coefficient. This coefficient ranges between 0 and 1. It is obtained daily via optimization as described in *Methods: LandscapeDNDC*. It accounts for the interaction between the ILS_i of the different environmental drivers at any given day. Finally, the combined LS is calculated by adding the daily LS_i of each environmental driver. Then, the daily MLED is assumed to be the driver that accounts for most of the LS.

2017), all other data were considered as missing values. For days with two or less missing half-hourly values, missing data was replaced by daily-averaged daytime GPP_i if the gap was during the day, and equal to daily average nighttime GPP_i if the gap was during the night. Days with more than two missing GPP data points were excluded from the analysis (Yatir ~30% days excluded, and Hyytiälä ~1% days excluded).

Environmental drivers

In both stands, meteorological time series were obtained from the meteorological stations accompanying the eddy-covariance measurements. Precipitation (P , mm/h), photosynthetic active radiation (PAR, $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$), air relative humidity (RH, %), wind speed (WS, m/s) and air temperature (T , °C) were measured above the forest canopy both in Yatir (Tatarinov et al. 2016), and in Hyytiälä (Markkanen et al. 2001) at a half-hourly basis. Soil water content (SWC, %) was also measured half-hourly using three (Yatir) and five (Hyytiälä) time domain reflectometer (TDR) sensors (TRIME, IMKO, Ettlingen, Germany for Yatir, and Campbell TDR100 Time-Domain Reflectometer for

Hyytiälä) measuring at 25 cm depth in Yatir (Klein et al. 2014), and at 22.5 cm depth in Hyytiälä (Peltoniemi et al. 2015, Reyer et al. 2020), which are supposed to be representative for the rooted soil depth (see simulation setup). We calculated half-hourly vapor pressure deficit based on T and RH. Then we averaged daytime daily vapor pressure deficit (D , kPa). Relative extractable water (REW, %) was determined for each location from daily average values of SWC (SWC_i) relative to maximum and minimum measured daily average SWC (SWC_{max} and SWC_{min} , respectively) during the observed period according to Granier et al. (2000)

$$REW = \frac{SWC_i - SWC_{min}}{SWC_{max} - SWC_{min}}. \quad (1)$$

It should be noted that the minimum SWC value in Hyytiälä might not be representative for the wilting point, i.e., the SWC at which trees cannot take up any more water, as this threshold may have not been reached within our investigation period. Another uncertainty derives from inhomogeneous soil and stand properties within the footprint area of the flux tower, particularly for Hyytiälä. Moreover, the high SWC that has been measured in

TABLE 1. Main characteristics of the two forest stands including location and dominant tree species as well as soil type and maximum soil depth.

Characteristic	Yatir	Hyytiälä
Country	Israel	Finland
Latitude	31°20' N	61°51' N
Longitude	35°03' E	24°17' E
Elevation (m above sea level)	650	185
Dominant species	<i>Pinus halepensis</i> Mill.	<i>Pinus sylvestris</i> L.
Age (yr)	55	50
Tree density (trees/ha)	300	684
DBH (cm)	18.5	20.6
Tree height (m)	9.3	17.8
Soil type	Rendzic Leptosol	Haplic Podzol
Maximum soil depth	100 cm	90 cm
Mean <i>T</i> (°C)	18.9	4.4
Mean annual P (mm year ⁻¹)	290	604
Period	2013-2015	2012-2014
Main references	Grünzweig et al. (2007), Rotenberg and Yakir (2010), Tatarinov et al. (2016)	Haataja and Vesala (1997), Rannik et al. (2004)

Notes: The age of the plantation, tree density, diameter at breast height (DBH), and average tree height at the beginning of the simulations are given. The average annual air temperature (°C) and average annual precipitation and the period of ecosystem flux data considered in this study are also given.

Hyytiälä during winter may not be available for plant uptake since the soil water is likely to be frozen. Thus, REW values in Hyytiälä must be considered with particular care (see Appendix S1: Fig. S1 for further information on monthly climate conditions for both stands).

LandscapeDNDC

LandscapeDNDC (LDNDC) is a forest simulation framework that uses the PSIM vegetation model (Grote 2007, Grote et al. 2009, Haas et al. 2013) for representation of structured forests based on physiological processes. It calculates carbon, nitrogen, and water fluxes as well as stocks within a forest stand at daily or hourly time steps (Grote et al. 2011a, b). The stand can include one or more cohorts, which are defined by species, average tree dimension and tree density. Canopy microclimate, nutrient, and soil water availability are calculated for a user-defined number of canopy and soil layers (Grote et al. 2011b). In PSIM, photosynthesis is implemented according to the Farquhar, Von Caemmerer, and Berry model (Farquhar et al. 1980) and stomatal conductance is calculated according to Leuning (1995). Reduced soil water availability restricts stomatal conductance following Knauer et al. (2015) and nitrogen

availability might affect maximum carboxylation velocity (Grote et al. 2009). Phenology is defined by bud burst, depending on temperature and cumulated growing degree-days. Foliage turnover time is derived from species-specific senescence parameters (Grote 2007). Respiration is composed of a fixed fraction of growth respiration from GPP, nitrogen transport costs, and a nitrogen and temperature dependent component of maintenance respiration (Thornley and Cannell 2000). At the ecosystem level, PSIM is complemented with the DeNitrification-DeComposition (DNDC) model that describes microbial release of CO₂ explicitly for decomposition, nitrification and denitrification processes (Li et al. 1992). The PSIM module within the LDNDC or one of its predecessor frameworks has been applied Europe-wide, being tested against tree growth, biogenic volatile compound emissions, and carbon and water fluxes in coniferous, evergreen, and deciduous forest stands (e.g., Grote et al. 2009, 2011a, b, Holst et al. 2010, Schweier et al. 2017, Dirnböck et al. 2020).

Simulation setup

The model has been initialized for the two sites based on available measurements as follows. Soil properties such as bulk density, grain structure, carbon and nitrogen content have been taken directly from site measurements to a depth of 100 cm for Yatir and 60 cm for Hyytiälä (for references, see Table 1). Forests at both sites have been treated as one cohort assuming homogeneous stand conditions, with average stem diameter at breast height, stem height, and stem density (number per hectare) set from recent inventories as depicted in Table 1. The parameterizations for the two species *P. halepensis* and *P. sylvestris* have been done based on available data from the literature. For *P. sylvestris*, these have been published in Grote et al. (2011b). For *P. halepensis*, gas exchange parameters as well as phenological developments such as leaf flushing and maximum leaf area were available from site measurements (Maseyk et al. 2008), supplemented with activation energy terms published in Simioni et al. (2016). The model is run without any spin-up time because this affects mostly the development of various soil carbon and nitrogen pools from initialized total carbon and nitrogen distribution, which is assumed to have no significant effect on GPP for the short-term simulations carried out in this study. During the simulation, the model uses daily input values for minimum, maximum, and average temperature, global radiation, relative humidity, and precipitation that have been measured at the sites. As we ran the model at hourly basis, the meteorological variables were down-scaled to hourly values. CO₂ concentration was assumed to change yearly following the average annual CO₂ concentrations from Mauna Loa (data available online).⁷

⁷<https://www.co2.earth/>

Finally, we integrated hourly simulated GPP into daily cumulated GPP_{sim} values.

Sensitivity of GPP to environmental drivers and most limiting environmental driver

We assessed GPP sensitivity to each environmental driver using a random forest (RF) algorithm. RF allows us to relate a single response variable, in our case daily GPP, to any set of given explanatory variables, here T , D , REW, and PAR, in a nonadditive and nonlinear way. RF works iteratively, growing a population of small regression trees from the data. Briefly, from the training data set, each iteration (tree) first randomly selects 10% of the training data. Then, a regression tree is built by splitting the observations into increasingly homogeneous groups, or “leaves,” based on a random subset of explanatory variables. The number of candidate explanatory variables for each split was set to 2 (following Liaw and Wiener 2002), until a given minimum of observations per leaf is reached (here $n = 7$ observations). This process is repeated for a predefined number of times (here, $n = 1,000$ “trees”). Final RF projections are the average from all repetitions (De’ath and Fabricius 2000, Liaw and Wiener 2002). RF analysis is able to identify subtle nonlinear responses within the data, even when there may be high correlation between the explanatory variables (see Appendix S1: Table S1), and it is suited for time-series analysis as it breaks temporal autocorrelation due to the random sampling of the training data set (Breiman 2001, e.g., Zhang et al. 2017, Xu et al. 2020).

For all RF analyses, we first randomly split the observations as well as the simulation output into training (75%) and validation (25%) data. We trained the RF algorithm with the first, and assessed model performance with the second data set, accounting for temporal autocorrelation through bootstrapping. All RF analyses were performed in R (version 3.6.1; R Core Team 2020) using the RandomForest package (Liaw and Wiener 2002; version 4.6-14). The importance of each environmental driver for each RF was defined as the percent of increase in Root Mean Square Error if the values for such driver were randomized when performing the RF analysis (Hastie et al. 2009). In order to assess if the ranking of different environmental drivers was robust, we repeated the RF procedure 100 times, each time with a random data sample of 200 observations from each of the four data sets used for training and validation of the two sites. From this analysis, we report the mean 95% confidence interval of the resulting importance distribution (see Appendix S1: Fig. S2 and Appendix S1: Table S2). Based on the RF results, we calculated daily GPP as if only one driver was limiting ($GPP_{pred,i}$) by setting all other drivers to optimum values. From this procedure we obtain the independent limitation strength (ILS_i) of GPP for each driver (i) at a daily basis according to

$$ILS_i = 1 - \frac{GPP_{pred,i}}{GPP_{max}} \quad (2)$$

where ILS_i is the decrease of GPP due to the i th environmental driver, GPP_{max} is the maximum GPP measured for a given stand, defined as the average of the 5% highest daily observed flux measurements within the whole time-series, and $GPP_{pred,i}$ is the daily predicted GPP if only the i th environmental driver is limiting GPP_{max} . All GPP values are provided as $kg \cdot ha^{-1} \cdot day^{-1}$.

After calculating ILS_i , we considered the different daily ILS_i to be additive in their limitation of GPP. However, simply adding the different ILS_i might result in a limitation higher than 1 during less productive periods, which would lead to negative GPP values. Therefore, the α_i coefficient was introduced

$$GPP_{reg} = GPP_{max} \left(1 - \sum_{i=1}^n ILS_i \times \alpha_i \right) \quad (3)$$

where GPP_{reg} is the GPP for a given day obtained either from direct observation or LDNDC simulations. The α_i coefficients range from 0 to 1 and are estimated for each data set (training and validation for both sites) via Bayesian inversion (see Hartig et al. [2012] for a review) using the DEOptim package (Mullen et al. 2011; Version 2.2-4). More specifically, we considered daily GPP_{reg}/GPP_{max} as observations and ILS_i as constants, assumed a Gaussian distribution for the likelihood function, and a flat uniform prior for each α_i between 0 and 1. After the optimization, the daily limitation strength (LS_i) for a given environmental driver was calculated as $ILS_i \times \alpha_i$. After computing the LS_i for each environmental driver, the combined limitation strength (LS) has been defined as the sum of the four individual LS_i . The MLED of GPP for a given day was therefore the environmental driver with the highest LS_i at this particular day (Fig. 1).

Statistical analysis

All statistical analyses were performed in R (Version 3.6.1, R Core Team 2020). In order to additionally evaluate modeled GPP values against observations as is routinely done by most studies, we performed a least square approach for both stands. We assessed model performance, i.e., homoscedasticity and normality of the residuals, via visual inspection, and we applied a log-log transformation to linearize the relationship in case that such criteria were not met.

RESULTS

Importance of environmental drivers in limiting GPP observed

In both stands observed GPP (GPP_{obs}) showed a high seasonality, although with contrasting patterns (Figs. 2

and 3). At Yatir, GPP_{obs} peaked during winter, while in Hyytiälä, GPP_{obs} peaked during summer, the growing season for high-latitudes in the Northern Hemisphere. The RF analysis properly captured this GPP_{obs} seasonality based on the combination of the four environmental drivers (Figs. 2a, 3a, Table 2). In Yatir, the environmental driver that best explained seasonal dynamics in GPP was relative extractable soil water (REW, 112% increase in RMSE if it was randomized, Table 2). In stark contrast, in Hyytiälä the driver that best explained GPP variability was daily-averaged air temperature (T , 78% increase in RMSE, Table 2). Stability analyses showed that the order of importance for all of the variables remained the same when reducing sample size and randomizing the training data set (see Appendix S1: Fig. S3).

Evaluation of LDNDC simulations

LDNDC overestimated GPP in both stands (Fig. 4), with projected daily GPP (GPP_{sim}) being on average

10% higher than the observations in Hyytiälä and 22% higher in Yatir. LDNDC tended to agree better with the observations when GPP_{obs} was low, but to overestimate GPP_{obs} during periods of high productivity in both stands. In agreement with the observations, RF analysis identified REW in Yatir, and air temperature in Hyytiälä as the most important environmental drivers of GPP_{sim} (Table 2). As LDNDC is a mathematical construct and its output is generated in a deterministic environment, GPP_{sim} was better reproduced by the RF algorithm than GPP_{obs} (Figs. 2c and 3c, higher R^2 and lower RMSE).

Sensitivity to environmental drivers

GPP_{obs} in the Yatir stand was strongly sensitive to changes in REW (Fig. 5), with GPP_{obs} declining as the soil dried. GPP_{obs} also increased with daylight average PAR, reaching an optimum at approximately $1,300 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$. Observations suggested that GPP_{obs} decreased as daily averaged air temperature

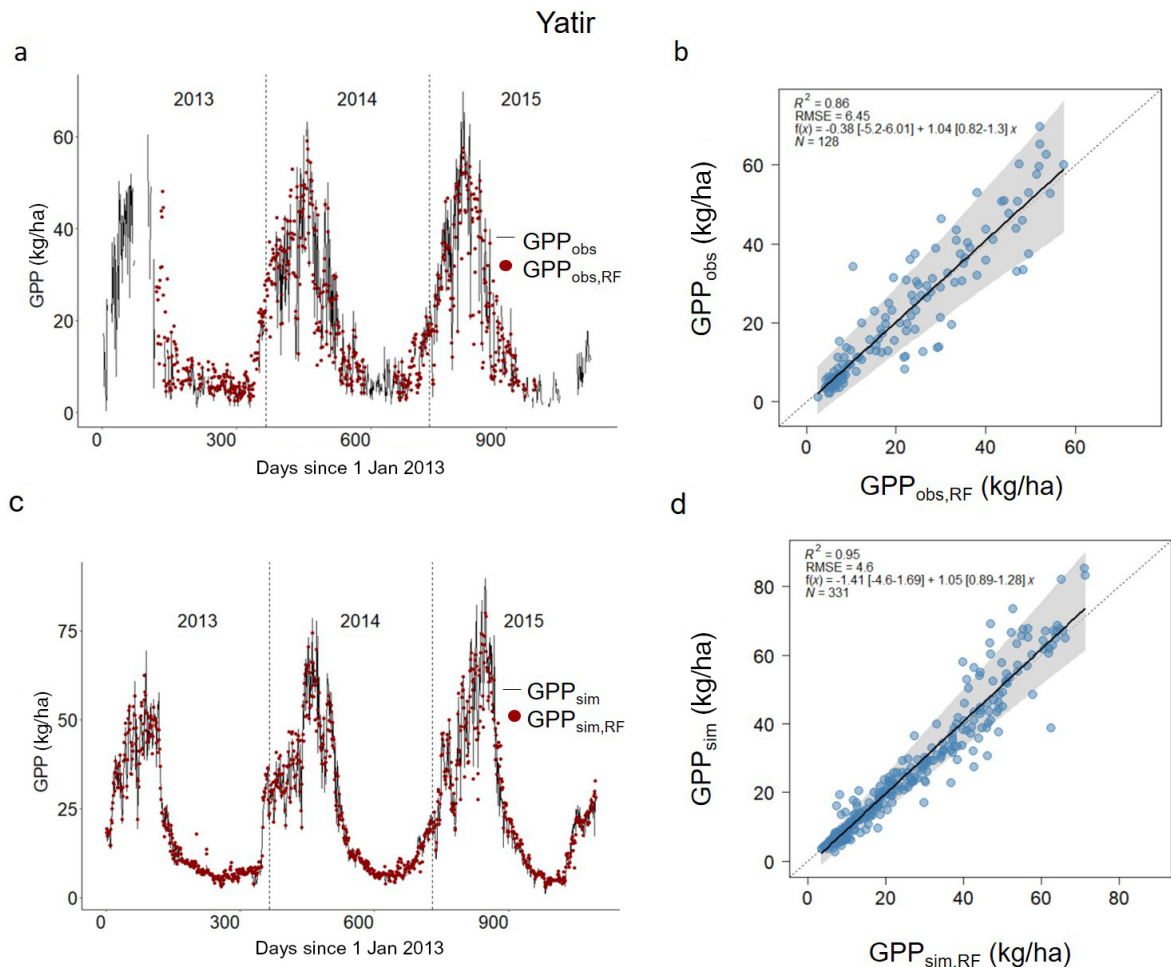


FIG. 2. Evaluation of the GPP derived from the Random Forest analyses (GPP_{RF}) for the Yatir *Pinus halepensis* stand from 2013 to 2015 is shown. For each analysis, daily observed GPP (GPP_{obs} , a and b), and daily LDNDC-simulated GPP (GPP_{sim} c and d) are related to four environmental drivers, which are average relative extractable water (REW), average air temperature (T), daily daylight average vapor pressure deficit (D), and daily daylight average photosynthetic active radiation (PAR).

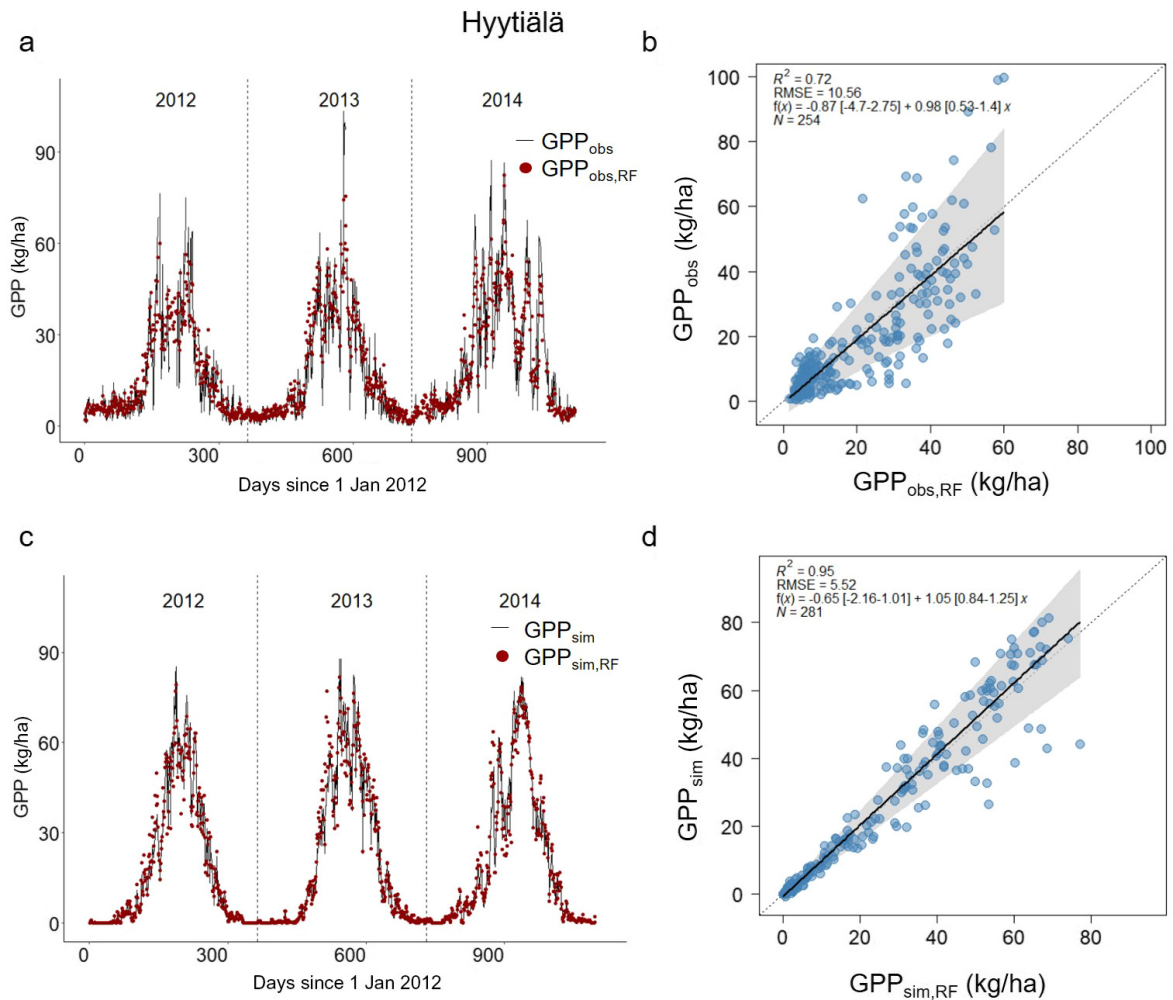


FIG. 3. Evaluation of the GPP derived from the Random Forest analyses (GPP_{RF}) Hyttiälä *Pinus sylvestris* stand from 2012 to 2014 is shown. For each analysis, daily observed GPP (GPP_{obs} , a and b), and daily LDNDC-simulated GPP (GPP_{sim} , c and d) are related to four environmental drivers, which are average relative extractable water (REW), average air temperature (T), daily daylight average vapor pressure deficit (D), and daily daylight average photosynthetic active radiation (PAR).

raised above 12°C , and declined at $D > 1$ kPa. LDNDC captured the observed sensitivity of GPP_{obs} to REW and PAR. Interestingly, model outputs deviated from observations when responses to T and D were considered. GPP_{sim} responded positively to increasing T and D , with optimal GPP_{sim} occurring at a daytime-averaged D of ~ 3 kPa.

GPP_{obs} in the Hyttiälä forest was most sensitive to changes in air temperature (Fig. 6). The particular pattern of soil water availability in Hyttiälä, with maximum REW values during winter and minimum REW values during summer, resulted apparently in high REW limiting GPP_{obs} . This unrealistic behavior predicted by the RF model could be due to two reasons. First, REW depends on maximum and minimum recorded SWC, but SWC values in Hyttiälä were relatively high even during summer and therefore were not likely limiting GPP (see

Appendix S1: Fig. S1). Therefore, REW values were steeply declining during summer in a few occasions without necessarily indicating low water availability (Fig. 6). Second, the absence of GPP_{obs} data during combined high REW and high T conditions may have biased RF outputs due to the high sensitivity of GPP_{obs} to T . In Hyttiälä, the sensitivity of GPP_{obs} to increases both in PAR and in D was relatively low, although increases in both drivers generally resulted in larger GPP_{obs} . It is worth noting that average daylight D values in Hyttiälä were ~ 0.3 kPa, which is five times lower than the average D values in Yatir (average ~ 1.5 kPa). This may explain the divergent responses to D between a forest that is water-supply limited (Yatir) and a forest that is water-demand limited (Hyttiälä). Although LDNDC simulations for Hyttiälä properly captured the observed impact of the four environmental drivers considered,

TABLE 2. Analysis of the importance for the environmental drivers included in the Random Forest analysis (RF), both for observed GPP and simulated GPP with LandscapeDNDC.

Characteristic/ statistic	Yatir		Hyytiälä	
	Observed	LDNDC	Observed	LDNDC
REW (%)	112	172	44.5	26.3
Mean T (°C)	51.3	33.4	78.2	90.4
Mean D (kPa)	34.7	21.17	25.5	18.1
PAR ($\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$)	62.6	76.8	28.7	33.2
R^2	0.86	0.95	0.72	0.95
RMSE	6.45	4.6	10.6	5.5
Slope	1.04†	1.05†	0.98†	1.05†

Notes: The importance of each environmental driver is provided as the percent increase in root mean square error (RMSE) if the environmental driver was randomized in the input training data set. REW, relative extractable water; T , air temperature; D , daily vapor pressure deficit; PAR, photosynthetic active radiation. The environmental driver with the highest importance is highlighted in bold letters. Further, the validation of the RF analysis outputs against the validation data set is provided. R^2 , RMSE and slope were calculated accounting for temporal autocorrelation within the data set.

†Indicates that 95% CI include slope = 1.

modeled sensitivity to T was slightly stronger, with a T optimum for GPP_{sim} at 23°C compared to GPP_{obs} at 20°C.

Seasonality of the most limiting environmental driver for GPP

The MLED of GPP_{obs} at Yatir was mainly REW (56.1% of the days of the year), followed by PAR, T , and

D (24.5%, 11.8%, and 7.6%, respectively). There was a strong seasonal component, with low water availability during late spring, summer, and autumn being most important (Fig. 7a). However, during winter, when productivity reached its maximum due to high soil water availability and mild temperatures, the MLED was mostly PAR. During spring and early summer, the MLED of GPP_{obs} shifted between T , D , and REW, indicating a combination of environmental limitations for GPP during those months, with a noteworthy fraction of days in which either high T or high D were the MLED of GPP_{obs} . The transformation of the independent limitation strength (ILS_i) to the combined limitation strength (LS_i) had only little effect on the MLED (Appendix S1: Table S3), with minor differences regarding T and D at Yatir during spring, and regarding REW and T at Hyytiälä during summer (data not shown).

Considering the simulations, we found REW to be the MLED of GPP_{sim} during most of the year in Yatir (67% of days), while PAR was again the dominant MLED during winter (16.5% of days), accompanied by a considerable fraction of T being the most important driver (7.2% of days). Interestingly, LDNDC simulations indicated a much larger number of days where T was the MLED of GPP_{sim} during winter compared to the observations (Fig. 7a, lower panel). The overall percentage of D being the MLED during the course of the year was similar between GPP_{sim} and GPP_{obs} . However, GPP_{obs} was limited predominantly by high D while GPP_{sim} was mostly limited by low D (see Fig 5 and see Appendix S1: Fig. S3).

At Hyytiälä, air temperature was the MLED of GPP_{obs} during 83% of the days per year. This trend was matched well by the model (Fig. 7b). Only for a few days

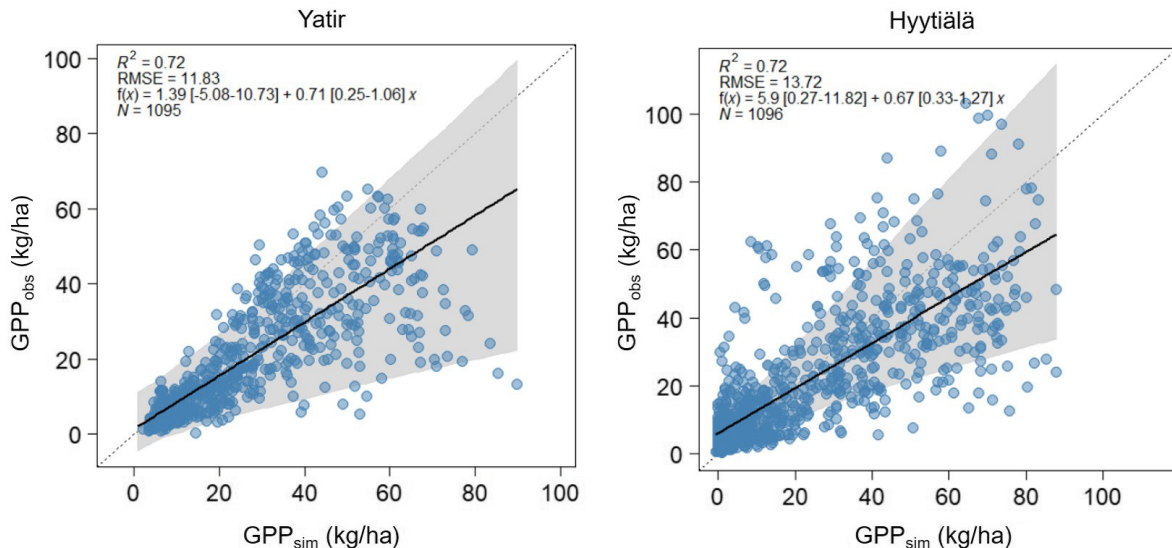


FIG. 4. Comparison between observed GPP (GPP_{obs}) and LDNDC-simulated GPP (GPP_{sim}) in the Yatir *Pinus halepensis* stand and the Hyytiälä *Pinus sylvestris* stand. The uncertainty associated with temporal autocorrelation is given as 95% CI (shaded area). The statistical results of a least squares approach accounting for temporal autocorrelation are also notated within the plot.

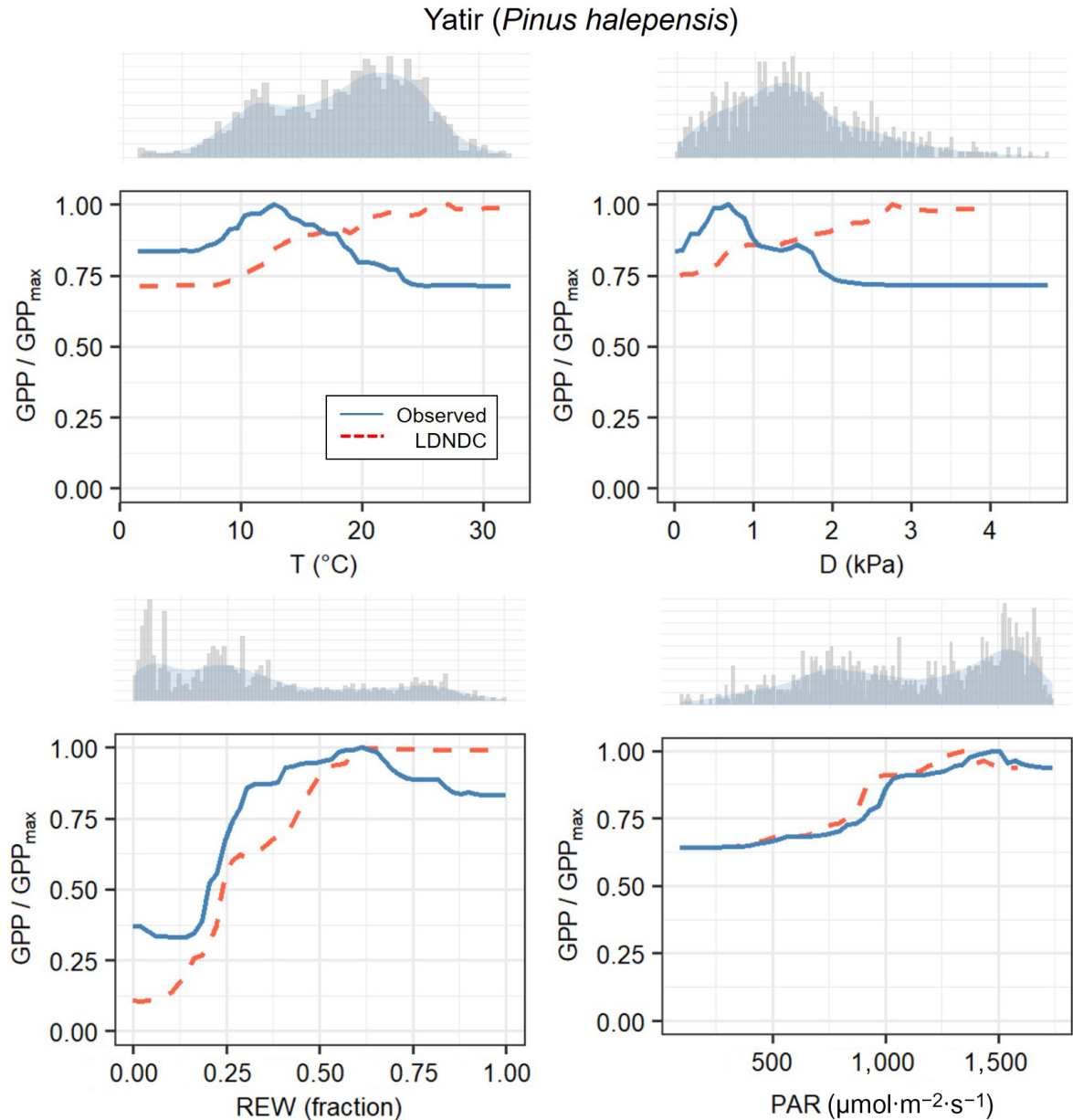


FIG. 5. Sensitivity of observed and simulated GPP in the Yatir stand to the four environmental drivers considered: i.e., average air temperature (T), average daylight vapor pressure deficit (D), average relative extractable water (REW) and average photosynthetic active radiation (PAR). The sensitivity is expressed as the GPP in relation to GPP_{\max} when all environmental drivers but the analyzed one are at their optimal values, for both observations (blue solid line) and LDNDC outputs (red dashed line). Shaded area in the panels above each figure show the observed distribution of measured daily values for each environmental driver.

during late spring and summer months, when temperatures were relatively high, other environmental drivers such as PAR were the MLED of GPP_{obs} . As mentioned before, the occasional occurrence of REW as MLED of both GPP_{obs} and GPP_{sim} during summer is likely caused by a methodological bias. In general, the agreement of the MLED between the LDNDC model and the observations was larger for Hyttiälä (74.2%) than for Yatir (61.7%), likely because of a concomitant T -limitation on

GPP at Hyttiälä during all seasons, whereas the drivers that mostly limited GPP at Yatir changed throughout the seasons.

DISCUSSION

We identified soil water availability as the MLED of GPP in a semiarid Mediterranean pine forest, while irradiance being an important factor during the relatively

Hyytiälä (*Pinus sylvestris*)

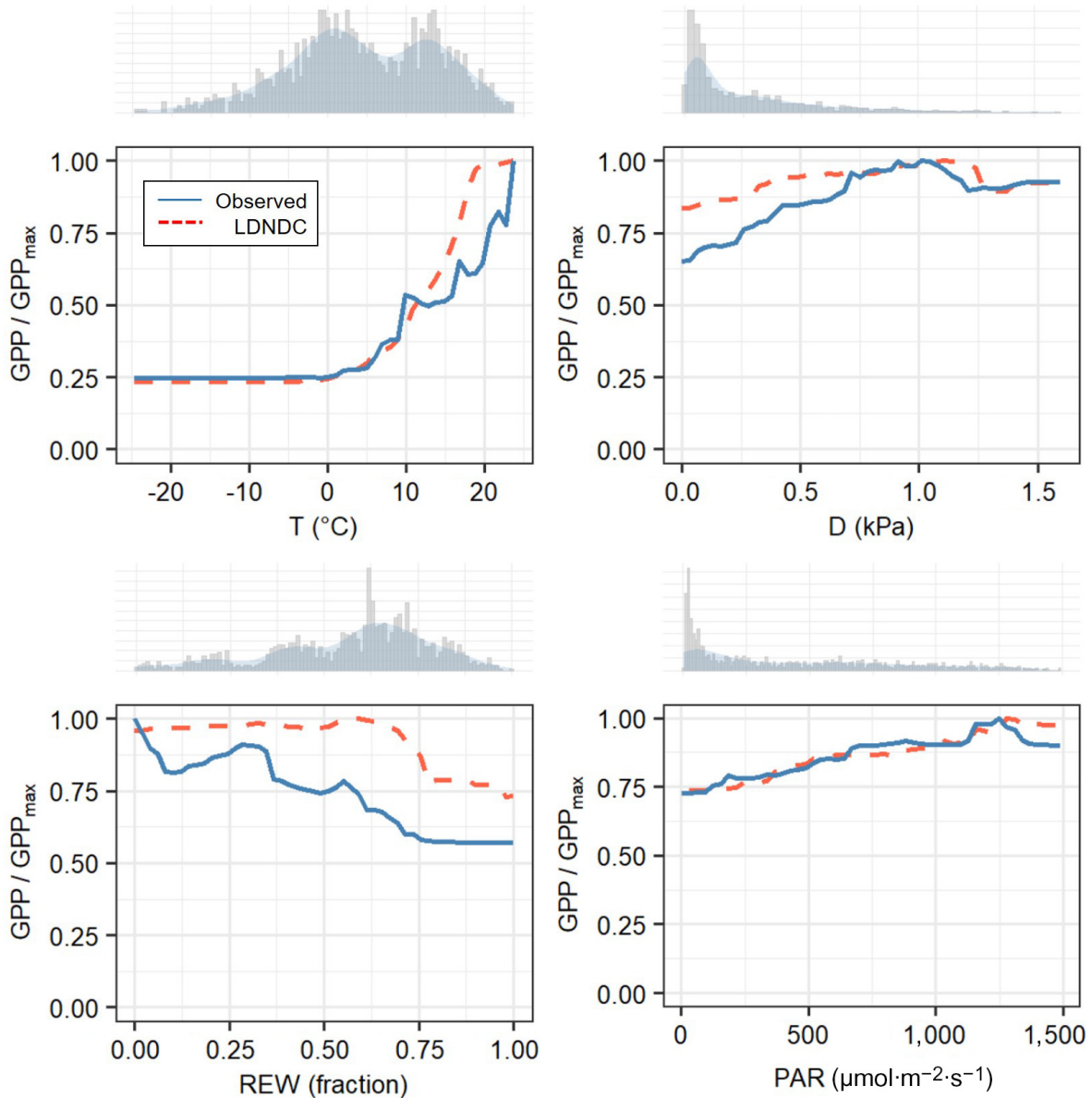


FIG. 6. Sensitivity of observed and simulated GPP in the Yatir stand to the four environmental drivers considered: i.e., average air temperature (T), average daylight vapor pressure deficit (D), average relative extractable water (REW) and average daylight photosynthetic active radiation (PAR). The sensitivity is expressed as the GPP in relation to GPP_{max} when all environmental drivers but the analyzed one are at their optimal values, for both observations (blue solid line) and LDNDC outputs (red dashed line). Shaded area in the panels above each figure show the observed distribution of measured daily values for each environmental driver.

wet and productive winter months. In addition, high temperature and high vapor pressure deficit were the MLED for a number of days during spring and early summer. Contrastingly, low temperature was the dominating MLED of GPP in a boreal pine forest, showing little seasonality. The LDNDC model reproduced the seasonal MLED dynamics in both stands reasonably

well. However, we found inconsistencies between the limitation of high- D and high- T conditions regarding GPP_{mod} and GPP_{obs} in the Mediterranean forest, particularly during spring and early summer. Such discrepancies likely originate from an incomplete model description regarding the response of stomatal conductance to rising D , which results in an overestimation of

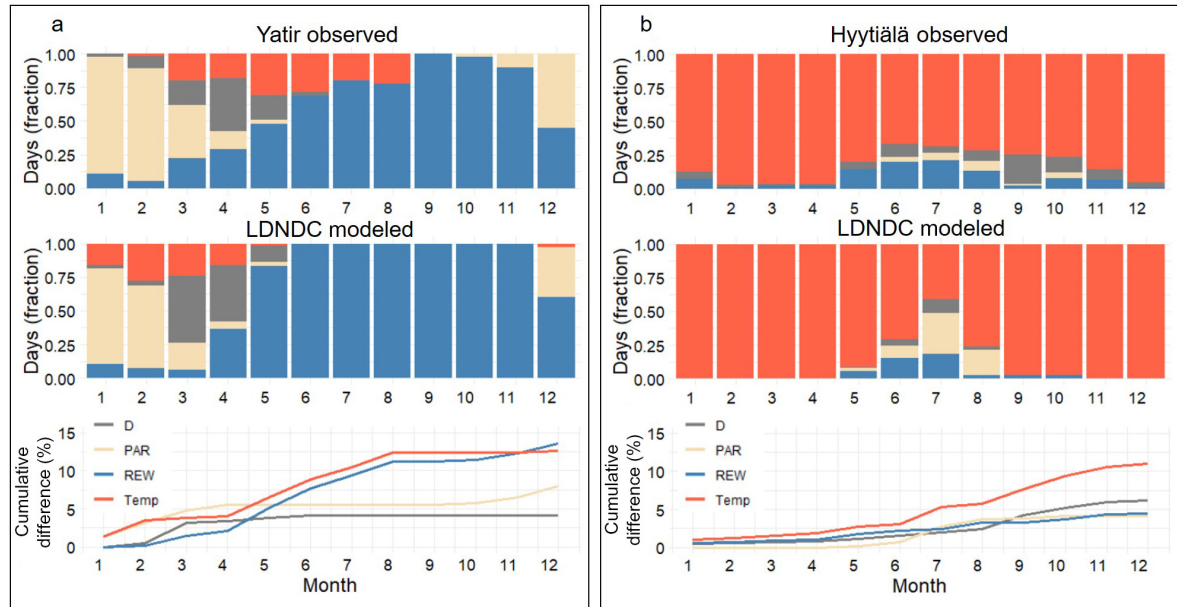


Fig. 7. Monthly distribution of the most limiting environmental driver (MLED) of GPP, in the Yatir *Pinus halepensis* Mill. stand (a), and the Hyttiälä *Pinus sylvestris* L. stand (b). The fraction of days per month that a given environmental driver is the MLED of GPP_{obs} (upper panels) and GPP_{sim} (middle panels) are shown. The monthly cumulated percent difference between observed and LDNDC-simulated MLED of GPP for each environmental driver is given. Vapor pressure deficit (D) is shown in gray, photosynthetic active radiation (PAR) in yellow, relative extractable water (REW) in blue, and temperature (Temp) in red.

C gain under high D conditions. Under such circumstances, in LDNDC the enhancement of photosynthesis kinetics due to high temperature is not offset by a lower C supply due to stomatal closure at high D , and hence modeled GPP_{mod} diverged from the observations.

Potential applications and limitations of the MLED approach

Model assessment is routinely performed through direct comparison of model outputs against observations. Recently, new methods are being proposed, such as model output adjustment through a posteriori recalibration to improve their fit to observations (Dormann 2020) or reanalysis of model outputs to assess their sensitivity to the different environmental drivers (e.g., Rollinson et al. 2017), but such methods are not routinely implemented. Furthermore, most output-to-observation evaluations are typically performed (1) without accounting for temporal autocorrelation in the resulting time-series and/or (2) after integrating daily outputs into coarser time-scales, e.g., monthly or yearly (e.g., Morales et al. 2005, Gutsch et al. 2016, Collalti et al. 2019), which may compensate for daily and seasonal model deviations. In our study, the direct comparison of GPP_{sim} vs. GPP_{obs} fell within the range of what has been reported as a good agreement when evaluating process-based model performance (e.g., Keenan et al. 2010, Grote et al. 2011, Zhang et al. 2016, Ma et al. 2017, Collalti et al. 2018).

To assess model performance beyond a simple fit of its outputs to the observations, we evaluated the ability of LDNDC to reproduce seasonal variations of the MLED of GPP. As for GPP_{obs} , the new method correctly identified relative extractable water as the MLED of GPP_{sim} at Yatir (e.g., Preisler et al. 2019), and temperature at Hyttiälä (e.g., Mäkelä et al. 2004). Similarly, the MLED approach also showed that LDNDC properly captured the responses of both stands to daily average PAR (Figs. 5–7). This feature is particularly important since inconsistencies in light-use-efficiency were previously depicted as problematic when forecasting forest responses to elevated CO_2 and rising temperature (Medlyn et al. 2011a).

We were able to detect model inconsistencies reflected in GPP_{sim} increasing with rising T and D in the Mediterranean forest during spring (Figs. 5 and 6), which is in contrast with observations (e.g., Tatarinov et al. 2016). This in turn translated into a contrasting seasonality of the MLED dynamics for temperature between GPP_{obs} and GPP_{sim} (Fig. 7). Therefore, the MLED approach identified a weakness in LDNDC when dealing with high- T high- D conditions, which reflects that the model concentrates on temperature-dependencies of biochemical processes (e.g., Lloyd and Farquhar 2008, Tan et al. 2017), but underestimates or neglects stress responses due to elevated D . For example, the biochemical optimum of photosynthesis for *P. halepensis* is at about 30°C, while stomatal closure due to concomitant D increases with temperature reduce this optimal

temperature for net assimilation down to 24°C (e.g., Sperlich et al. 2019, Birami et al. 2020). In contrast to the Yatir site, such high T and D conditions are nearly never reached in Hyytiälä, resulting in correctly reproduced MLED dynamics related to variations in T and D (Figs. 6, 7).

The MLED approach developed here provides an in-depth analysis of observed and modeled responses of GPP to different environmental drivers, which is important in order to assess the reliability of model forecasts under changing climatic conditions. For instance, not capturing the limiting effect of rising D will likely lead to an overestimation of GPP when heat and drought increases as has been projected for the near future at Yatir (Raz-Yaseef et al. 2010). Nevertheless, our results indicate that with a traditional output-to-observation model validation such subtle divergences would have likely been unnoticed. Therefore, we recommend the application of the MLED approach in model evaluation exercises, although we acknowledge that the additional effort might hamper its applicability for large scale simulations.

The main limitation of our approach is that it is strongly constrained by data availability: the α_i coefficient-calibration for instance requires complete information for different environmental drivers at daily resolution. Also, input data should cover a large variation of environmental drivers to avoid misrepresentations for highly correlated drivers, as we found for instance for the MLED dynamics of REW in Hyytiälä. In addition, the RF algorithm must be able to properly capture the sensitivity of the evaluation target (here GPP) to the different environmental drivers. Since the MLED approach uses such sensitivities to calculate ILS for each environmental driver, a RF analysis that fails to reproduce that target pattern will result in a biased unrealistic sensitivity.

Beyond model evaluation, the MLED approach has proven useful to analyze how different environmental drivers affect ecophysiology throughout the year and at a specific location. This contrasts with previous studies (e.g., Churkina and Running 1998, Seneviratne et al. 2010, Park et al. 2019), in which the focus was on an a priori assigned forest ecophysiology response to several environmental drivers, and on a broader scale. Also, the short-term and stand-specific approach is able to assume negligible determinants of GPP such as stand structure or nutrient availability, which are supposed to change only slowly, therefore allowing us to assume a constant GPP_{max} . We assume that a broad application of the MLED approach at sites where CO_2 and H_2O flux data is available will improve our understanding about the seasonality of GPP limitations by climatic factors (Park et al. 2019). Further, the method enables a better assessment of how stand- or species-specific traits (e.g., rooting depth or foliage longevity) affect the importance of different stressors.

Opportunities for LDNDC model improvement

Based on the results of the MLED approach, it is suggested that LDNDC needs to be improved regarding the response of stomatal conductance to high vapor pressure deficit. This is particularly important for reproducing GPP limitations under high- T and high- D conditions such as the *hamsin* events at Yatir during spring and early summer (e.g., Rotemberg and Yakir 2010, Tatarinov et al. 2016). Also, for forest productivity in general, stomatal closure at high D conditions has been shown to be a major limiting factor, even when soil water is available (e.g., Novick et al. 2016, Ficklin and Novick 2017, Smith et al. 2020). Here, the MLED approach indicates that a simple adjustment of parameters driving LDNDC responses to high D conditions is likely insufficient because it is not simply the scale, but the shape of the GPP response to rising D that requires improvement. Several improvement options are possible: The most straightforward one is to modify the algorithm relating stomatal conductance and D (Duursma et al. 2019), shifting from the currently implemented empirical Leuning approach (Leuning 1995) to a more mechanistic one, e.g., the optimal stomatal behavior approach (Medlyn et al. 2011b). Other processes that may be considered including in LDNDC encompass responses to tree dehydration (e.g., Scoffoni et al. 2017), or overheating (e.g., Blonder and Michaletz 2018, but see Zandalinas et al. 2018), or photosynthetic biochemistry changes responding to soil drought (Drake et al. 2017, Hüve et al. 2019). Maybe the most appealing option is the inclusion of an explicit description of internal tree hydraulics. It has already been shown that a better description of stomata responses to rising D based on internal tree hydraulics improve the representation of transpiration (Liu et al. 2020) and carbon uptake (Eller et al. 2020). Furthermore, hydraulic conductance schemes are increasingly used to improve model forecasting abilities (e.g., Sperry and Love 2015, Eller et al. 2018, 2020). Also, explicitly describing tree hydraulics will likely provide us with a better representation of heat-drought-induced tree mortality (e.g., Blackman et al. 2019, De Kauwe et al. 2020). Considering that extreme heat and drought is supposed to occur at more sites and in higher frequency in the future, including such a response would clearly improve the reliability of LDNDC forecasts of forest responses to climate change.

Implications of contrasting MLED for stand responses to climate change

Based on the different MLED results for the two stands, contrasting responses to global warming can be expected. At Hyytiälä, an increase in average temperatures will likely increase productivity, as has been observed for other boreal forests (e.g., Kauppi et al. 2014, D'Orangeville et al. 2016, Tagesson et al. 2020) under sufficient water and nutrient supply (D'Orangeville et al.

2018). These responses can be robustly forecasted by the current LDNDC implementation. In contrast, the Yatir forest is already water-limited during most of the year (e.g., Klein et al. 2011, Ungar et al. 2013, Qubaja et al. 2020). Its future existence will depend on aridity changes, which are related to both variations in rainfall intensity and seasonality as well as in evaporative demand. While precipitation trends are more uncertain, projections robustly indicate D to sharply increase during next decades (Harris et al. 2014). Under such circumstances, we expect that global warming will reduce the productivity of Yatir forest by further amplifying atmospheric aridity. In this sense, MLED dynamics analysis suggests that not capturing high- D limitations over GPP is a major issue for LDNDC model projections regarding Yatir forest, as its simulations under a global warming will likely overestimate GPP.

CONCLUSION

We demonstrated that the MLED approach is useful to support traditional output-to-observation model assessment, as it is able to evaluate if a model properly reproduces different environmental limitations. It is thus, straightforward to assume that elaborated model evaluation tools, such as the MLED approach, will be increasingly needed to assess the robustness of model projections and to identify weaknesses in model processes descriptions. In LDNDC, an unrealistic stomatal behavior under high evaporative demand resulted in an overestimation of GPP, which would lead to unrealistically high forest productivity projections when assessing the impacts of global warming. Since empirical descriptions of stomatal conductance such as the Leuning approach (Leuning 1995) are widespread in physiologically oriented models, we expect similar biases to occur frequently, and that improvements are needed to account for them in scenario investigations of future climates.

ACKNOWLEDGMENTS

This study was supported by the German Research Foundation through its Emmy Noether Program (RU 1657/2-1). We further acknowledge funding from the German Research Foundation through its German-Israeli project cooperation program (CSCHM 2736/2-2). In addition, we wish to thank the two anonymous reviewers for their insightful comments.

LITERATURE CITED

- Aubinet, M., et al. 1999. Estimates of the annual net carbon and water exchange of forests: the EUROFLUX methodology. *Advances in Ecological Research* 30:113–175.
- Birami, B., T. Nägele, M. Gattmann, Y. Preisler, A. Gast, A. Arneth, and N. K. Ruehr. 2020. Hot drought reduces the effects of elevated CO₂ on tree water-use efficiency and carbon metabolism. *New Phytologist* 226:1607–1621.
- Blackman, C. J., et al. 2019. Drought response strategies and hydraulic traits contribute to mechanistic understanding of plant dry-down to hydraulic failure. *Tree Physiology* 39:910–924.
- Blonder, B., and S. T. Michaletz. 2018. A model for leaf temperature decoupling from air temperature. *Agricultural and Forest Meteorology* 262:354–360.
- Breiman, L. 2001. Random forests. *Machine Learning* 45:5–32.
- Brodrribb, T. J., J. Powers, H. Cochard, and B. Choat. 2020. Hanging by a thread? Forests and drought. *Science* 368:261–266.
- Bugmann, H., et al. 2019. Tree mortality submodels drive simulated long-term forest dynamics: assessing 15 models from the stand to global scale. *Ecosphere* 10:e02616.
- Churkina, G., and S. W. Running. 1998. Contrasting climatic controls on the estimated productivity of global terrestrial biomes. *Ecosystems* 1:206–215.
- Ciais, P., et al. 2005. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* 437:529–533.
- Collalti, A., et al. 2018. Thinning can reduce losses in carbon use efficiency and carbon stocks in managed forests under warmer climate. *Journal of Advances in Modeling Earth Systems* 10:2427–2452.
- Collalti, A., et al. 2019. The sensitivity of the forest carbon budget shifts across processes along with stand development and climate change. *Ecological Applications* 29:e01837.
- D'Orangeville, L., L. Duchesne, D. Houle, D. Kneeshaw, B. Côté, and N. Pederson. 2016. Northeastern North America as a potential refugium for boreal forests in a warming climate. *Science* 352:1452–1455.
- D'Orangeville, L., D. Houle, L. Duchesne, R. P. Phillips, Y. Bergeron, and D. Kneeshaw. 2018. Beneficial effects of climate warming on boreal tree growth may be transitory. *Nature Communications* 9:1–10.
- De Kauwe, M. G., et al. 2020. Identifying areas at risk of drought-induced tree mortality across South-Eastern Australia. *Global Change Biology* 26:5716–5733.
- De'ath, G., and K. E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81:3178–3192.
- Dietze, M. C., et al. 2018. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proceedings of the National Academy of Sciences USA* 115:1424–1432.
- Dirnböck, T., et al. 2020. Substantial understory contribution to the C sink of a European temperate mountain forest landscape. *Landscape Ecology* 35:483–499.
- Dormann, C. F. 2020. Calibration of probability predictions from machine-learning and statistical models. *Global Ecology and Biogeography* 29:760–765.
- Drake, J. E., et al. 2017. Stomatal and non-stomatal limitations of photosynthesis for four tree species under drought: a comparison of model formulations. *Agricultural and Forest Meteorology* 247:454–466.
- Duursma, R. A., C. J. Blackman, R. López, N. K. Martin-StPaul, H. Cochard, and B. E. Medlyn. 2019. On the minimum leaf conductance: its role in models of plant water use, and ecological and environmental controls. *New Phytologist* 221:693–705.
- Eller, C. B., et al. 2018. Modelling tropical forest responses to drought and El Niño with a stomatal optimization model based on xylem hydraulics. *Philosophical Transactions of the Royal Society B* 373:20170315.
- Eller, C. B., et al. 2020. Stomatal optimization based on xylem hydraulics (SOX) improves land surface model simulation of vegetation responses to climate. *New Phytologist* 226:1622–1637.
- Farquhar, G. D., S. V. von Caemmerer, and J. A. Berry. 1980. A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species. *Planta* 149:78–90.

- Ficklin, D. L., and K. A. Novick. 2017. Historic and projected changes in vapor pressure deficit suggest a continental-scale drying of the United States atmosphere. *Journal of Geophysical Research: Atmospheres* 122:2061–2079.
- Fluxnet 2015. 2017. FLUXNET2015 Release. Processing Pipeline document. <http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/data-processing/>
- Granier, A., D. Loustau, and N. Bréda. 2000. A generic model of forest canopy conductance dependent on climate, soil water availability and leaf area index. *Annals of Forest Science* 57:755–765.
- Grossiord, C., et al. 2020. Plant responses to rising vapor pressure deficit. *New Phytologist* 226:1550–1566.
- Grote, R. 2007. Sensitivity of volatile monoterpene emission to changes in canopy structure: a model-based exercise with a process-based emission model. *New Phytologist* 173:550–561.
- Grote, R., R. Kiese, T. Grünwald, J. M. Ourcival, and A. Granier. 2011a. Modelling forest carbon balances considering tree mortality and removal. *Agricultural and Forest Meteorology* 151:179–190.
- Grote, R., J. Korhonen, and I. Mammarella. 2011b. Challenges for evaluating process-based models of gas exchange at forest sites with fetches of various species. *Forest Systems* 20:389–406.
- Grote, R., A. V. Lavoie, S. Rambal, M. Staudt, I. Zimmer, and J. P. Schnitzler. 2009. Modelling the drought impact on monoterpene fluxes from an evergreen Mediterranean forest canopy. *Oecologia* 160:213–223.
- Grünzweig, J. M., I. Gelfand, Y. Fried, and D. Yakir. 2007. Biogeochemical factors contributing to enhanced carbon storage following afforestation of a semi-arid shrubland. *Biogeosciences* 4:891–904.
- Grünzweig, J. M., T. Lin, E. Rotenberg, A. Schwartz, and D. Yakir. 2003. Carbon sequestration in arid-land forest. *Global Change Biology* 9:791–799.
- Gutsch, M., P. Lasch-Born, F. Suckow, and C. P. Reyer. 2016. Evaluating the productivity of four main tree species in Germany under climate change with static reduced models. *Annals of Forest Science* 73:401–410.
- Haas, E., et al. 2013. LandscapeDNDC: a process model for simulation of biosphere–atmosphere–hydrosphere exchange processes at site and regional scale. *Landscape Ecology* 28:615–636.
- Haataja, J., and T. Vesala. 1997. Station for Measuring Forest Ecosystem–Atmosphere Relations: SMEAR II. Vol. 17. University of Helsinki, Department of Forest Ecology Publications, Helsinki, Finland.
- Harris, I. P. D. J., P. D. Jones, T. J. Osborn, and D. H. Lister. 2014. Updated high-resolution grids of monthly climatic observations—the CRU TS3. 10 Dataset. *International Journal of Climatology* 34:623–642.
- Hartig, F., J. Dyke, T. Hickler, S. I. Higgins, R. B. O'Hara, S. Scheiter, and A. Huth. 2012. Connecting dynamic vegetation models to data—an inverse perspective. *Journal of Biogeography* 39(12):2240–2252.
- Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York City, New York, USA.
- Hellmann, L., et al. 2016. Diverse growth trends and climate responses across Eurasia's boreal forest. *Environmental Research Letters* 11:074021.
- Holst, J., et al. 2010. Water fluxes within beech stands in complex terrain. *International Journal of Biometeorology* 54:23–36.
- Humphrey, V., J. Zscheischler, P. Ciaia, L. Gudmundsson, S. Sitch, and S. I. Seneviratne. 2018. Sensitivity of atmospheric CO₂ growth rate to observed changes in terrestrial water storage. *Nature* 560:628–631.
- Hüve, K., I. Bichele, H. Kaldmäe, B. Rasulov, F. Valladares, and Ü. Niinemets. 2019. Responses of aspen leaves to heat-flecks: Both damaging and non-damaging rapid temperature excursions reduce photosynthesis. *Plants* 8:145.
- Kauppi, P. E., M. Posch, and P. Pirinen. 2014. Large impacts of climatic warming on growth of boreal forests since 1960. *PLoS ONE* 9:e111340.
- Keenan, T., S. Sabate, and C. Gracia. 2010. Soil water stress and coupled photosynthesis–conductance models: Bridging the gap between conflicting reports on the relative roles of stomatal, mesophyll conductance and biochemical limitations to photosynthesis. *Agricultural and Forest Meteorology* 150:443–453.
- Kirchner, J. W. 2006. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research* 42:W03S04.
- Klein, T., et al. 2014. Quantifying transpirable soil water and its relations to tree water use dynamics in a water-limited pine forest. *Ecohydrology* 7:409–419.
- Klein, T., S. Cohen, and D. Yakir. 2011. Hydraulic adjustments underlying drought resistance of *Pinus halepensis*. *Tree Physiology* 31:637–648.
- Knauer, J., C. Werner, and S. Zaehle. 2015. Evaluating stomatal models and their atmospheric drought response in a land surface scheme: a multibiome analysis. *Journal of Geophysical Research: Biogeosciences* 120:1894–1911.
- Kolari, P., et al. 2009. CO₂ exchange and component CO₂ fluxes of a boreal Scots pine forest. *Boreal Environment Research* 14:761–783.
- Leuning, R. 1995. A critical appraisal of a combined stomatal-photosynthesis model for C₃ plants. *Plant, Cell & Environment* 18:339–355.
- Li, C., S. Frolking, and T. A. Frolking. 1992. A model of nitrous oxide evolution from soil driven by rainfall events: I. Model structure and sensitivity. *Journal of Geophysical Research: Atmospheres* 97:9759–9776.
- Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. *R News* 2:18–22.
- Liu, Y., M. Kumar, G. G. Katul, X. Feng, and A. G. Konings. 2020. Plant hydraulics accentuates the effect of atmospheric moisture stress on transpiration. *Nature Climate Change* 10:691–695.
- Lloyd, J., and G. D. Farquhar. 2008. Effects of rising temperatures and [CO₂] on the physiology of tropical forest trees. *Philosophical Transactions of the Royal Society B* 363:1811–1817.
- Ma, J., H. H. Shugart, X. Yan, C. Cao, S. Wu, and J. Fang. 2017. Evaluating carbon fluxes of global forest ecosystems by using an individual tree-based model FORCCHN. *Science of the Total Environment* 586:939–951.
- Mäkelä, A., P. Hari, F. Berninger, H. Hänninen, and E. Nikinmaa. 2004. Acclimation of photosynthetic capacity in Scots pine to the annual cycle of temperature. *Tree Physiology* 24:369–376.
- Markkanen, T., Ü. Rannik, P. Keronen, T. Suni, and T. Vesala. 2001. Eddy covariance fluxes over a boreal Scots pine forest. *Boreal Environment Research* 6:65–78.
- Maseyk, K. S., T. Lin, E. Rotenberg, J. M. Grünzweig, A. Schwartz, and D. Yakir. 2008. Physiology–phenology interactions in a productive semi-arid pine forest. *New Phytologist* 178:603–616.
- Medlyn, B. E., R. A. Duursma, and M. J. Zeppel. 2011a. Forest productivity under climate change: a checklist for evaluating

- model studies. *Wiley Interdisciplinary Reviews: Climate Change* 2:332–355.
- Medlyn, B. E., et al. 2011*b*. Reconciling the optimal and empirical approaches to modelling stomatal conductance. *Global Change Biology* 17:2134–2144.
- Medlyn, B. E., et al. 2015. Using ecosystem experiments to improve vegetation models. *Nature Climate Change* 5:528–534.
- Morales, P., et al. 2005. Comparing and evaluating process-based ecosystem model predictions of carbon and water fluxes in major European forest biomes. *Global Change Biology* 11:2211–2233.
- Mullen, K., D. Ardia, D. L. Gil, D. Windover, and J. Cline. 2011. DEoptim: An R package for global optimization by differential evolution. *Journal of Statistical Software* 40:1–26.
- Nadal-Sala, D. 2020. DanielNadalSala/Nadal-Sala_et_al_2020_EcologicalApplications: Nadal-Sala et al. (2020) Ecological Applications (Version v1.0). Zenodo. <https://doi.org/10.5281/zenodo.4309121>
- Nemani, R. R., et al. 2003. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* 300:1560–1563.
- Norby, R. J., et al. 2005. Forest response to elevated CO₂ is conserved across a broad range of productivity. *Proceedings of the National Academy of Sciences USA* 102:18052–18056.
- Novick, K. A., et al. 2016. The increasing importance of atmospheric demand for ecosystem water and carbon fluxes. *Nature Climate Change* 6:1023–1027.
- Park, T., et al. 2019. Changes in timing of seasonal peak photosynthetic activity in northern ecosystems. *Global Change Biology* 25:2382–2395.
- Peltoniemi, M., M. Pulkkinen, M. Aurela, J. Pumpanen, P. Kolari, and A. Mäkelä. 2015. A semi-empirical model of boreal-forest gross primary production, evapotranspiration, and soil water—calibration and sensitivity analysis. *Boreal Environment Research* 20:151–171.
- Preisler, Y., et al. 2019. Mortality versus survival in drought-affected Aleppo pine forest depends on the extent of rock cover and soil stoniness. *Functional Ecology* 33:901–912.
- Qubaja, R., et al. 2020. Partitioning evapotranspiration and its long-term evolution in a dry pine forest using measurement-based estimates of soil evaporation. *Agricultural and Forest Meteorology* 281:107831.
- Qubaja, R., J. M. Grünzweig, E. Rotenberg, and D. Yakir. 2019. Evidence for large carbon sink and long residence time in semiarid forests based on 15-year flux and inventory records. *Global Change Biology* 26:1626–1637.
- R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rannik, Ü., P. Keronen, P. Hari, and T. Vesala. 2004. Estimation of forest–atmosphere CO₂ exchange by eddy covariance and profile techniques. *Agricultural and Forest Meteorology* 126:141–155.
- Raz-Yaseef, N., D. Yakir, E. Rotenberg, G. Schiller, and S. Cohen. 2010. Ecohydrology of a semi-arid forest: partitioning among water balance components and its implications for predicted precipitation changes. *Ecohydrology* 3:143–154.
- Reichstein, M., et al. 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biology* 11:1424–1439.
- Reyer, C. P., et al. 2020. The PROFOUND database for evaluating vegetation models and simulating climate impacts on forests. *Earth System Science Data* 12:1295–1320.
- Rollinson, C. R., et al. 2017. Emergent climate and CO₂ sensitivities of net primary productivity in ecosystem models do not agree with empirical data in temperate forests of eastern North America. *Global Change Biology* 23:2755–2767.
- Rotenberg, E., and D. Yakir. 2010. Contribution of semi-arid forests to the climate system. *Science* 327:451–454.
- Ruimy, A., L. Kergoat, A. Bondeau, and T. P. O. T. P. N. M. Intercomparison. 1999. Comparing global models of terrestrial net primary productivity (NPP): Analysis of differences in light absorption and light-use efficiency. *Global Change Biology* 5:56–64.
- Schlesinger, W. H., et al. 2016. Forest biogeochemistry in response to drought. *Global Change Biology* 22:2318–2328.
- Schweier, J., et al. 2017. Environmental impacts of bioenergy wood production from poplar short-rotation coppice grown at a marginal agricultural site in Germany. *GCB Bioenergy* 9:1207–1221.
- Scoffoni, C., et al. 2017. Outside-xylem vulnerability, not xylem embolism, controls leaf hydraulic decline during dehydration. *Plant Physiology* 173:1197–1210.
- Seneviratne, S. I., et al. 2010. Investigating soil moisture–climate interactions in a changing climate: a review. *Earth-Science Reviews* 99:125–161.
- Simioni, G., G. Marie, and R. Huc. 2016. Influence of vegetation spatial structure on growth and water fluxes of a mixed forest: results from the NOTG 3D model. *Ecological Modelling* 328:119–135.
- Smith, M. N., et al. 2020. Empirical evidence for resilience of tropical forest photosynthesis in a warmer world. *Nature Plants* 6:1225–1230.
- Sperlich, D., C. T. Chang, J. Peñuelas, and S. Sabaté. 2019. Responses of photosynthesis and component processes to drought and temperature stress: are Mediterranean trees fit for climate change? *Tree Physiology* 39:1783–1805.
- Sperry, J. S., and D. M. Love. 2015. What plant hydraulics can tell us about responses to climate-change droughts. *New Phytologist* 207:14–27.
- Tagesson, T., et al. 2020. Recent divergence in the contributions of tropical and boreal forests to the terrestrial carbon sink. *Nature Ecology & Evolution* 4:202–209.
- Tan, Z. H., et al. 2017. Optimum air temperature for tropical forest photosynthesis: Mechanisms involved and implications for climate warming. *Environmental Research Letters* 12:054022.
- Tatarinov, F., E. Rotenberg, K. Maseyk, J. Ogée, T. Klein, and D. Yakir. 2016. Resilience to seasonal heat wave episodes in a Mediterranean pine forest. *New Phytologist* 210:485–496.
- Thornley, J. H. M., and M. G. R. Cannell. 2000. Modelling the components of plant respiration: representation and realism. *Annals of Botany* 85:55–67.
- Ungar, E., E. Rotenberg, N. Raz-Yaseef, S. Cohen, D. Yakir, and G. Schiller. 2013. Transpiration and annual water balance of Aleppo pine in a semiarid region: Implications for forest management. *Forest Ecology and Management* 298:39–51.
- Xu, B., M. A. Arain, T. A. Black, B. E. Law, G. Z. Pastorello, and H. Chu. 2020. Seasonal variability of forest sensitivity to heat and drought stresses: A synthesis based on carbon fluxes from North American forest ecosystems. *Global Change Biology* 26:901–918.
- Zandalinas, S. I., R. Mittler, D. Balfagón, V. Arbona, and A. Gómez-Cadenas. 2018. Plant adaptations to the combination of drought and high temperatures. *Physiologia Plantarum* 162:2–12.
- Zhang, H., P. Wu, A. Yin, X. Yang, M. Zhang, and C. Gao. 2017. Prediction of soil organic carbon in an intensively managed reclamation zone of eastern China: A comparison of

- multiple linear regressions and the random forest model. *Science of the Total Environment* 592:704–713.
- Zhang, Q., et al. 2019. Response of ecosystem intrinsic water use efficiency and gross primary productivity to rising vapor pressure deficit. *Environmental Research Letters* 14:074023.
- Zhang, Y., et al. 2016. Development of a coupled carbon and water model for estimating global gross primary productivity and evapotranspiration based on eddy flux and remote sensing data. *Agricultural and Forest Meteorology* 223:116–131.

SUPPORTING INFORMATION

Additional supporting information may be found online at: <http://onlinelibrary.wiley.com/doi/10.1002/eap.2312/full>

DATA AVAILABILITY

Simulation data and R scripts are available in Zenodo (Nadal-Sala 2020): <http://doi.org/10.5281/zenodo.4309121>. Measurement data for the Hyttiälä stand can be found in Reyer et al. (2020) and measurement data for the Yatir stand is available from DEIMS-SDR (Dynamic Ecological Information Management System—Site and data set registry) under DEIMS.ID <https://deims.org/bf2cb8dd-0d38-419b-aa9c-7d53337bd98e>.