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

DANIEL NADAL-SALA,1,6 RUDIGER GROTE,1 BENJAMIN BIRAMI,1 ANNA LINTUNEN,2,3 IVAN MAMMARELLA,2 YAKIR PREISLER,6 EYAL ROTENBERG,2 YANN SALMON,2,3 FEDOR TATARINOV,5 DAN YAKIR,5 AND NADINE K. RUHR1

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 ResearchForest Sciences, University of Helsinki, Latokartanonkaari 7, P.O. Box 27, Helsinki, 00014 Finland
3Institute for Atmospheric and Earth System ResearchPhysics, Faculty of Science, University of Helsinki, P.O. Box 68, Gustaf Hällströminkatu 2b, Helsinki, 00014 Finland
4Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138 USA
5Deptartment of Environmental Sciences and Energy Research, Weizmann Institute of Science, Rehovot 76100 Israel


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, Ciais 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 CO2 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
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, Brodribb 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 [CO₂]. 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., Ruiny 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 (GPPmax). It also assumes that we can identify the limitation that a given environmental driver exerts upon GPPmax 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 GPP = GPPmax. 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 Hyytiälä 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 Hyytiälä 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 Hyytiälä 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 CO₂ concentration on GPP (e.g., Norby et al. 2005).

**Carbon flux measurements for evaluation**

For both stands, GPP, was calculated half-hourly as GPP = NEP + ER, (gross primary productivity equals net ecosystem productivity plus ecosystem respiration), and then integrated into daily GPP. ER, 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 Hyytiälä (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, half-hourly observations of category 0 and 1 (Fluxnet
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 if the gap was during the day, and equal to daily average nighttime GPP 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, μmol·m⁻²·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 (SWCᵢ) 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)

\[
REW = \frac{SWCᵢ - SWC_min}{SWC_max - SWC_min}
\]

(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.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Yatir</th>
<th>Hyytiälä</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>Israel</td>
<td>Finland</td>
</tr>
<tr>
<td>Latitude</td>
<td>31°20’ N</td>
<td>61°51’ N</td>
</tr>
<tr>
<td>Longitude</td>
<td>35°03’ E</td>
<td>24°17’ E</td>
</tr>
<tr>
<td>Elevation (m above sea level)</td>
<td>650</td>
<td>185</td>
</tr>
<tr>
<td>Dominant species</td>
<td>Pinus halepensis</td>
<td>Pinus sylvestris L.</td>
</tr>
<tr>
<td>Age (yr)</td>
<td>55</td>
<td>50</td>
</tr>
<tr>
<td>Tree density (trees/ha)</td>
<td>300</td>
<td>684</td>
</tr>
<tr>
<td>DBH (cm)</td>
<td>18.5</td>
<td>20.6</td>
</tr>
<tr>
<td>Tree height (m)</td>
<td>9.3</td>
<td>17.8</td>
</tr>
<tr>
<td>Soil type</td>
<td>Rendzic</td>
<td>Haplic Podzol</td>
</tr>
<tr>
<td>Maximum soil depth</td>
<td>100 cm</td>
<td>90 cm</td>
</tr>
<tr>
<td>Mean T (°C)</td>
<td>18.9</td>
<td>4.4</td>
</tr>
<tr>
<td>Mean annual P (mm year⁻¹)</td>
<td>290</td>
<td>604</td>
</tr>
<tr>
<td>Main references</td>
<td>Grünzweig et al. (2007), Haataja and Vesala (1997), Rasnik et al. (2004)</td>
<td></td>
</tr>
</tbody>
</table>

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.

Hyttiälä during winter may not be available for plant uptake since the soil water is likely to be frozen. Thus, REW values in Hyttiälä 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 Hyttiälä (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 Pinus halepensis and Pinus sylvestris have been done based on available data from the literature. For Pinus 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

7https://www.co2.earth/
Finally, we integrated hourly simulated GPP into daily cumulated GPP\textsubscript{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 randomly 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\textsubscript{pred,i}) by setting all other drivers to optimum values. From this procedure we obtain the independent limitation strength (ILS\textsubscript{i}) of GPP for each driver ($i$) at a daily basis according to

$$\text{ILS}_i = 1 - \frac{\text{GPP}_{\text{pred,i}}}{\text{GPP}_{\text{max}}}$$

where ILS\textsubscript{i} is the decrease of GPP due to the $i$th environmental driver, GPP\textsubscript{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\textsubscript{pred,i} is the daily predicted GPP if only the $i$th environmental driver is limiting GPP\textsubscript{max}. All GPP values are provided as kg ha$^{-1}$ day$^{-1}$.

After calculating ILS\textsubscript{i}, we considered the different daily ILS\textsubscript{i} to be additive in their limitation of GPP. However, simply adding the different ILS\textsubscript{i} might result in a limitation higher than 1 during less productive periods, which would lead to negative GPP values. Therefore, the $\alpha_i$ coefficient was introduced

$$\text{GPP}_{\text{reg}} = \text{GPP}_{\text{max}} \left( 1 - \sum_{i=1}^{n} \text{ILS}_i \times \alpha_i \right)$$

where GPP\textsubscript{reg} is the GPP for a given day obtained either from direct observation or LDNDC simulations. The $\alpha_i$ 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\textsubscript{reg} / GPP\textsubscript{max} as observations and ILS\textsubscript{i} as constants, assumed a Gaussian distribution for the likelihood function, and a flat uniform prior for each $\alpha_i$ between 0 and 1. After the optimization, the daily limitation strength (LS\textsubscript{i}) for a given environmental driver was calculated as ILS\textsubscript{i} $\times \alpha_i$. After computing the LS\textsubscript{i} for each environmental driver, the combined limitation strength (LS) has been defined as the sum of the four individual LS\textsubscript{i}. The MLED of GPP for a given day was therefore the environmental driver with the highest LS\textsubscript{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 meet.

**Results**

**Importance of environmental drivers in limiting GPP observed**

In both stands observed GPP (GPP\textsubscript{obs}) showed a high seasonality, although with contrasting patterns (Figs. 2
and 3). At Yatir, GPP\textsubscript{obs} peaked during winter, while in Hyytiälä, GPP\textsubscript{obs} peaked during summer, the growing season for high-latitudes in the Northern Hemisphere. The RF analysis properly captured this GPP\textsubscript{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\textsubscript{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\textsubscript{obs} was low, but to overestimate GPP\textsubscript{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\textsubscript{sim} (Table 2). As LDNDC is a mathematical construct and its output is generated in a deterministic environment, GPP\textsubscript{sim} was better reproduced by the RF algorithm than GPP\textsubscript{obs} (Figs. 2c and 3c, higher $R^2$ and lower RMSE).

**Sensitivity to environmental drivers**

GPP\textsubscript{obs} in the Yatir stand was strongly sensitive to changes in REW (Fig. 5), with GPP\textsubscript{obs} declining as the soil dried. GPP\textsubscript{obs} also increased with daylight average PAR, reaching an optimum at approximately 1,300 μmol·m$^{-2}$·s$^{-1}$. Observations suggested that GPP\textsubscript{obs} decreased as daily averaged air temperature...
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 $\sim 3$ kPa.

$GPP_{obs}$ in the Hyytiälä forest was most sensitive to changes in air temperature (Fig. 6). The particular pattern of soil water availability in Hyytiälä, 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 Hyytiälä 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 Hyytiälä, 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 daytime $D$ values in Hyytiälä were $\sim 0.3$ kPa, which is five times lower than the average $D$ values in Yatir (average $\sim 1.5$ 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 (Hyytiälä). Although LDNDC simulations for Hyytiälä properly captured the observed impact of the four environmental drivers considered,
modeled sensitivity to \( T \) was slightly stronger, with a \( T \) optimum for GPPsim at 23° C compared to GPPobs at 20° C.

**Seasonality of the most limiting environmental driver for GPP**

The MLED of GPPobs 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 GPPobs 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 GPPobs. 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 GPPsim 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 GPPsim 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 GPPsim and GPPobs. However, GPPobs was limited predominantly by high \( D \) while GPPsim was mostly limited by low \( D \) (see Fig 5 and see Appendix S1: Fig. S3).

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

---

**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.**

<table>
<thead>
<tr>
<th>Characteristic/statistic</th>
<th>Yatir</th>
<th>Hyytiälä</th>
</tr>
</thead>
<tbody>
<tr>
<td>REW (%)</td>
<td>112</td>
<td>172</td>
</tr>
<tr>
<td>Mean ( T ) (°C)</td>
<td>51.3</td>
<td>33.4</td>
</tr>
<tr>
<td>Mean ( D ) (kPa)</td>
<td>34.7</td>
<td>21.17</td>
</tr>
<tr>
<td>PAR ( \mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1} )</td>
<td>62.6</td>
<td>76.8</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>0.86</td>
<td>0.95</td>
</tr>
<tr>
<td>RMSE</td>
<td>6.45</td>
<td>4.6</td>
</tr>
<tr>
<td>Slope</td>
<td>1.04†</td>
<td>1.05†</td>
</tr>
</tbody>
</table>

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.
during late spring and summer months, when temperatures were relatively high, other environmental drivers such as PAR were the MLED of GPP\textsubscript{obs}. As mentioned before, the occasional occurrence of REW as MLED of both GPP\textsubscript{obs} and GPP\textsubscript{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 Hyytiälä (74.2\%) than for Yatir (61.7\%), likely because of a concomitant T-limitation on GPP at Hyytiälä 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...
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
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$_{\text{mod}}$ 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$_{\text{sim}}$ vs. GPP$_{\text{obs}}$ 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$_{\text{obs}}$, the new method correctly identified relative extractable water as the MLED of GPP$_{\text{sim}}$ at Yatir (e.g., Preisler et al. 2019), and temperature at Hyytiäliä (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$_{\text{sim}}$ 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$_{\text{obs}}$ and GPP$_{\text{sim}}$ (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 elevates $D$. 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 αi 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 $GPP_{\text{max}}$. We assume that a broad application of the MLED approach at sites where CO2 and H2O 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, Tatarnov 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, Hiève 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.
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


**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 Hyytiälä 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-aa9e-7d53337bd98e.