The study by Kohonen et al. compares gross primary productivity (GPP) estimates at a boreal forest derived from two CO$_2$-based flux partitioning methods and two COS-based methods. One of the COS approaches to GPP, developed in previous studies, relies on an empirical light response of the COS vs CO$_2$ leaf relative uptake (LRU) ratio. The other COS approach, developed in this study, considers stomatal optimization as represented by the CAP model (Dewar et al., 2018) in simulating LRU responses to environmental conditions. The authors show that GPP estimates derived from the LRU$_{CAP}$ approach agree with those from the two CO$_2$-based approaches in terms of diurnal and seasonal cycles, cumulative GPP in the growing season, and environmental responses. By contrast, the COS approach based on the light dependence of LRU alone shows considerably higher GPP estimates than those from other methods, especially at high radiation. The authors conclude that their new approach is an improvement over previous empirical LRU fits for obtaining accurate COS-based GPP estimates.

Overall, the study marks a valuable methodological advance in estimating GPP at the ecosystem scale and is worthy of publication. While the authors succeed in deriving COS-based GPP estimates consistent with those from CO$_2$-based methods, they have not presented a strong case for the robustness and generalizability of the new method they developed. In other words, do we know that the LRU$_{CAP}$ approach produces the right results for the right reason, or is it so malleable that one can tune the parameters to get any desirable responses? To ensure the robustness of the method, the authors may need to clarify the physiological underpinnings of the method, the assumptions it makes, and its limitations. I have a few questions on this aspect.

- There are many optimization-based stomatal models, and CAP is not the simplest one. What is the motivation for choosing this specific model over, say, the Medlyn model (Medlyn et al., 2011), which has only two parameters to fit?
- The "carboxylation conductance", $g_c$, seems to be a pure model construct to linearize
the nonlinear response of the assimilation rate ($A$) to the chloroplast CO$_2$ concentration ($c_c$). The assumption that $g_c$ is constant is inconsistent with the Farquhar et al. (1980) model because the transition from Rubisco carboxylation limitation to electron transport limitation necessarily changes the slope of the $A$ vs $c_c$ curve. What is the rationale behind this treatment? What bias does it introduce?

- Several parameters assumed constant in fitting the model may vary across the season, for example, CO$_2$ compensation point and photosynthetic quantum yield. Where do those fixed values come from? Are they representative of the Scots pine species at the site?
- The impact of mesophyll conductance ($g_m$) on LRU is an intriguing but understated point. It seems that infinite $g_m$ works best for explaining LRU variability at low light but overestimates LRU at high light. By contrast, a finite $g_m$ works well at high light but predicts too low LRU values at low light (Fig. B2). Is there a physiological explanation for this? A discussion on this point would be desirable.

**Specific comments**

L21–22: "removes approximately 30% of the annual anthropogenic carbon dioxide (CO$_2$) emissions from the atmosphere". This is a misinterpretation. Global GPP far outweighs the anthropogenic carbon emissions ($\sim$120 PgC vs $\sim$10 PgC). The 30% fraction refers to net biome productivity, which is the net balance of GPP, ecosystem respiration, and emissions from land use changes and disturbances. See Chapin et al. (2006) for standard definitions of carbon flux terms.

L25: It is the net balance not the ratio that dictates the magnitude and direction of the terrestrial carbon budget.

L33: The origin of the partitioning method based on nighttime respiration predates Reichstein et al. (2005). The idea goes back at least as early as in Wofsy et al. (1993), though not in the exact form of relationship between $R_{eco}$ and temperature. It is likely that this method has an earlier origin in the eddy covariance community. Therefore, better change "a method introduced by Reichstein et al. (2005)" to "a method in Reichstein et al. (2005)".

L35: And storage change fluxes, if not constrained by concentration profile measurements, also introduce bias to nighttime fluxes.

L40: "These limitations lead to uncertainties in the derivation of mechanistically sound descriptions of respiration and its drivers, especially when contributions of different biomass compartments to total CO$_2$ efflux vary across ecosystems and seasonally even within one ecosystem." The point of this sentence is unclear.
It would be helpful to add a sentence on how this neural network approach tackles the problem of the inhibition of daytime respiration.

"recent studies have shown that LRU is a function of solar radiation because CO\textsubscript{2} uptake is highly radiation dependent while COS uptake is not" - This notion that LRU depends on PAR goes back as early as Stimler et al. (2010).

Specify the value of $T_0$.

Section 2.3.2: Did you create a hold-out data set for validation as in Tramontana et al. (2020), or perform cross-validation?

"atmospheric concentrations of CO\textsubscript{2} and COS" - Specify at which height these concentrations were measured.

Kooijmans et al. (2019) presented data from two chambers. Was this relationship derived from measurements from both chambers?

I share the other referee's concern that this paragraph is not helpful for readers to grasp the year-to-year variability of environmental conditions. Try to present the anomaly features in chronological order.

Table 1: List the source of each parameter value in a column instead of in the caption. Specify which values are from the literature and which are fitted to data presented in this study.

"... when comparing GPP\textsubscript{ANN} to standard FLUXNET partitioning during summer months for multiple sites." - What about the subset of evergreen needleleaf forest (ENF) sites?

"However, at 30 min time scale the GPP\textsubscript{ANN} was on average 15 % lower than GPP\textsubscript{NLR}." - Could you compare GPP\textsubscript{ANN} and GPP\textsubscript{NLR} at half-hourly timescales with negative values filtered?

"while GPP\textsubscript{NLR} may have even negative values due to random noise in the NEE measurements." - GPP should not be negative. Even if we consider random noise, the
uncertainty range of GPP estimates should not encompass negative values because this is physically impossible. In your calculation of cumulative fluxes, the negative values may need to be capped at zero.

L231: Given that GPP is higher at high radiation, shouldn't the parameter fitting prioritize reducing LRU bias at high radiation?

L242: "The agreement of this method was better than assuming infinite mesophyll conductance at high PAR, but worse at low PAR" - Could you elaborate on why this is the case? Have you tried temperature-dependent $g_m$ as in Wehr et al. (2017)?

L245–246: "We thus concluded that the assumption of infinite $g_m$ is more valid." - It would be more appropriate to say that given the uncertainty in LRU, minimizing LRU errors by itself does not offer a robust constraint on $g_m$. This fact does not necessarily mean that an infinite $g_m$ is valid in the real world.

L249: It is worth noting that $g_m$ becomes more limiting relative to $g_s$. We do not know how $g_m$ varies during the day. It could be that $g_s$ increases to a point such that $g_m$ becomes more limiting.

L267–L269: If the fraction of leaf respiration in total ecosystem respiration is small, I would not expect a clear break point to be found in the light response of NEE. Do you see any evidence for the Kok effect in leaf chamber measurements?

L274: "in summer a saturation point was found at PAR>500" - This apparent saturation point could be partly caused by VPD limitation on stomatal conductance around midday.

L359: What purpose does rewriting the equation in terms of $c_a - \Gamma^*$ serve? In the Farquhar et al. (1980) model, $\Gamma^*$ appears in $c_c - \Gamma^*$, because it is used to represent the difference between carboxylation and oxygenation. But $c_a - \Gamma^*$ does not seem to carry a physiological meaning.

Technical comments

L24: "increased" -> "increasing"
L30: "widely" and "globally", superfluous

L61: "triggered" -> "catalyzed"

L69: "ecosystem scale" -> "ecosystem-scale"

L71–72: This sentence seems to be the topic sentence of the paragraph.

L84: "where first flux measurements started in 1996 ..." - This information does not seem relevant since only the flux measurements between 2013 and 2017 are presented.

L86: "50 ha" - Better use SI units, for example, 0.5 km$^2$.

L139: "ecosystem level" -> "ecosystem-level"

L146: "assure" -> "ensure"

L193: "higher average" -> "higher than average"

L195: The units of PAR are incorrect in this line.

L214: "Fig. 2,3" -> "Figs. 2 and 3"

References cited


