

Interactive comment on “A daily, 250 m, and real-time gross primary productivity product (2000–present) covering the Contiguous United States” by Chongya Jiang et al.

Anonymous Referee #1

Received and published: 3 July 2020

A daily, 250 m, and real-time gross primary productivity product (2000 – present) covering the Contiguous United States

Jiang et al.

The authors develop a MODIS driven light use efficiency model, that derives from new estimates of PAR and C3/C4 fraction. The authors use a soil adjusted estimate of NIRv, the near-infrared reflectance of vegetation, to determine iPUE, “incident Par use efficiency.” The approach seems technically sound and the advantages of providing a high resolution estimate of GPP are quite apparent. I, however, have concerns about the reproducibility of the model and its associated uncertainties.

C1

The authors have wrapped three separate innovations into a single paper. First, a new approach to estimating PAR from MODIS. Second, a new approach for estimating C3/C4 fraction. And third, combining those two new products with estimates of NIRv to derive GPP.

For the PAR modeling, either more information is needed about the “four machine learning approaches” or the authors need to provide the code used to fit these models. Each algorithm has a multitude of adjustable parameters. How were these parameters determined, what are their values, and how might future researchers modify and/or improve upon the approach? It is also unclear to me how “uncertainty” is calculated. As written (L160), it appears uncertainty is represented in terms of model-model disagreement, as opposed to model-data disagreement. This seems inappropriate, though I know that the approach has been used in other parts of this literature. Trying to wrap my head around the uncertainty terms was complicated by the fact that Figure 2 reports PAR in $W m^2$, while Figure 3 uses $MJ m^2 d^{-1}$. Using common units throughout would be helpful.

Some other minor issues on the PAR estimation. Equation 9 uses the term “fPAR”, which in this literature often means ‘fraction of photosynthetically active radiation absorbed by plants.’ (e.g, their equation 4). It’s quite confusing to distinguish between fpar and FPAR. Finally, the authors might consider a supplemental figure showing patterns of PAR/SW – this is a fairly well-studied, physically grounded ratio. It might also help to show that the approach works just fine in semi-cloudy (e.g, when there is more diffuse PAR) conditions.

For C3/C4, how are these uncertainties propagated into the final reported GPP estimates? From Figure 6b, it seems uncertainty is often quite high (e.g., > 40 percent). Accounting for this uncertainty seems important. For Figure 6 and 7, it could be helpful to change “Reference” to “CDL Reference” or something of that sort. At first, I was confused by the difference between “Reference” and “Ground Truth.”

C2

Perhaps most importantly, the authors mostly ignore the poor performance of their approach at evergreen needleleaf sites. Figure 8D indicates that SANIRv is not a good predictor of daily iPUE at ENF sites. The result is mentioned in L374, but relatively little discussion is offered for why this might be the case or what should be done about it. From Figure 11b, it seems that combining uncertain estimates of iPUE with PAR somewhat alleviates the poor performance within ENF, though the ENF site with an R2 of less than 0.2 stands out.

A related, global scale effort to relate NIRv to GPP [Badgley, Anderegg, Berry, Field, GCB2019] that the authors cite, identified 'deciduous' vs. 'evergreen' as being a critical parameter for model performance. Recognizing the difference in scales of the two analysis and the authors' stronger focus on C3/C4, it still feels necessary for a richer discussion of the performance of the model at ENF sites, especially given that the ultimate goal of the manuscript is to distribute a GPP dataset that researchers from across disciplines might find useful. From the analyses, it seems individuals working in agricultural contexts might find the data more reliable than those working in ENF systems. These caveats should be clearly flagged for the reader and the authors might benefit the research community by offering some discussion about what they think is going on and what future efforts might address such uncertainties.

Finally, I find it perplexing that the manuscript lists five authors but the underlying data product only lists two authors. The author contributions indicate that author G.W. helped develop the SLOPE model and that authors B.P. and S.W. were involved in refining/interpreting how the model works. Doesn't the resulting dataset and the citations it might one day receive rely on the contributions of these three authors as well?

Interactive comment on Earth Syst. Sci. Data Discuss., <https://doi.org/10.5194/essd-2020-36>, 2020.