Dear Martin De Kauwe,

Thank you for your fast editing of our revised manuscript. We have included point-by-point answers to your three points of minor revision. We also included the major changes in the manuscript to address the comments (in the form of boxes).

Sincerely,

Johannes Gensheimer on behalf of all co-authors

You wrote "leave" and canopy scale, it should be "leaf"
Thank you for pointing out that major typo. We have changed it accordingly.

R2 asked for more details on the "the implications of method choices", I didn't see this in your revision?
We think that the model implications best fit in the section of feature importance (section 3.4). We have added the following part (starting in L. 213).

> The CNN is a data driven method and is not restricted by LUE terms. Although SM and meteorology (air temperature and precipitation) play a key role for photosynthesis, we find that they are not important to our model output. This does not necessarily imply that SIF is not linked to these parameters. This can be explained by 1) the variables SM and those from ERA-5 are at coarser resolution than the actual model output of the training phase which is at 0.05° (10000 m and 11132 m for SM and ERA-5, respectively). Therefore not each pixel at the resolution of 0.05° has its unique value for SM or ERA-5, but multiple cells can be within one SM or ERA-5 pixel. 2) Not only the auxiliary data of the model estimates higher resolution SIF, but it is computed together with coarse resolution SIF. Therefore, events like heat stress that impact a bigger area than the actual model output might be represented in the coarse resolution SIF. 3) We have aggregated the used data to 16 day time steps. LUE parameters influencing SIF might have a bigger impact on the estimation at higher temporal resolutions.

R2 also asked you to explain further about the "reasonably low R2", which you did, and noted something in the supplement. I feel like this commentary could also be added to the main manuscript for the interested reader; I'd like you to please consider this in revision.
Indeed it might be good for the reader to have it in the main text. It also helps interpreting the results of the model output. We have added the following part to the manuscript (starting in L. 286).

> Deviations between TROPOMI and OCO-2/3 also appear at a grid of 0.05° (Supplemental Figure~S19). The r2 coefficient is 0.61 and 0.62 between TROPOMI and OCO-2 and OCO-3 SIF, respectively. Indeed, one might expect better correlations here as both present SIF at ~740 nm. However, as pointed out in Köhler et al. 2018, the uncertainty of both TROPOMI and OCO-2 SIF are expected to lead to a certain spread between the data sets. In addition, we do not account for differences in acquisition times and viewing-illumination geometry, which can lead to additional uncertainties in this comparison. For reference, when comparing single footprints of TROPOMI SIF to aggregated OCO-2 SIF for June 2018 globally, Köhler et al. 2018 found a r2 of 0.67, only

additional aggregation leads to a r2 of 0.88. The mean deviation of TROPOMI SIF to OCO-2 SIF is close to the average standard deviation of TROPOMI SIF (~0.4 $mW \cdot m^{-2} \cdot sr^{-1} \cdot nm^{-1}$). In our analysis, from the 16 day product from TROPOMI SIF for April 2018 until March 2021 at 0.05°, we observe an average error in the TROPOMI SIF of 0.43 $mW \cdot m^{-2} \cdot sr^{-1} \cdot nm^{-1}$ for the CONUS. That error is close to the RMSE between instantaneous TROPOMI SIF and instantaneous OCO-2 SIF (0.37 $mW \cdot m^{-2} \cdot sr^{-1} \cdot nm^{-1}$). To compare TROPOMI and OCO-2/3 SIF we aggregate the OCO-2/3 footprints to the same grid as our TROPOMI data (0.05°). As we aggregate multiple OCO-2 or OCO-3 footprints to match one TROPOMI grid cell at 0.05° the certainty of the OCO measurements increases, and therefore the RMSE between TROPOMI and OCO SIF decreases.