Review of bg-2021-125, by Singh Anand et al. on "Ozone-induced gross primary productivity reductions over European forests inferred from satellite observations"

The paper presents an analysis of the impact of ozone versus meteorological drivers and soil moisture on European forest's GPP. It claims to be the first study to evaluate this impact at the continental scale using satellite observations. I deem this claim to be not being supported by the presented approach. There is indeed use of some satellite data in the form of the LAI and forest cover data (and you can argue that the CAMS O_3 data partly rely on the assimilation of remote sensing data). Having read initially the abstract I anticipated that the LAI timeseries were going to be used as a proxy for changes in GPP. But the crucial component of this study, GPP does not rely at all on the use of any source of satellite observations. The impact of O_3 on GPP is calculated in the presented study as a model product using some empirical constants, stomatal conductance and the accumulated O_3 concentrations. As such the presented analysis can mostly be interpreted as a validation step of the followed approach integrating the spatio-temporal information of many different datasets of relevant parameters. This is then complemented with a sensitivity analysis to indicate the spatio-temporal patterns in the role of O_3 vs the meteorology drivers of GPP. In addition, then comparing then the results of the ERA5/CAMS/etc. based O_3-GPP model with another model that also includes some empirical relationships to consider the O_3 impact on simulated GPP is mainly another validation step, e.g., showing that the model(s) is/are properly implemented. The main shortcoming of this paper is that there is not specific evaluation step; optimally one would have applied remote sensing based vegetation indices/GPP estimates. One already missed opportunity for some first evaluation step of the followed approach would have been evaluation of the inferred stomatal conductances, e.g., comparing the Jarvis based latent heat flux with FLUXNET observations. Consequently, quantification of the large-scale O_3 impact on vegetation functioning using large-scale satellite observations of vegetation dynamics, as all suggested by the title, abstract and introduction, is according to me not addressed. Based on these observations and considerations, I recommend the paper to be rejected.

Despite this overall negative recommendation, I share here also the specific comments that came up reading the paper and which can potentially be used for future revisions.

Specific comments:

Pp 1, lines 16-18: A bit confusing the first number in these lines (i.e., 30%) refers to the net uptake of CO_2 by vegetation (NEE, not GPP), and the carbon fluxes that are quoted after that are in fact GPP.

Pp 3, line 55 "Similarly, eddy covariance towers, such as those that make up the global FLUXNET dataset (Pastorello et al., 2020) can also be employed to investigate the effect of O_3 exposure on GPP". Here you could add the reference(s) to the work by Ducker et al., Biogeosciences, 15, 5395–5413, 2018, https://doi.org/10.5194/bg-15-5395-2018

Line 64: "soil hydrology" which soil hydrological variable? This is not specific.

Pp 3, line 73: “near-surface O_3 concentration and meteorology governing GPP”, using this statement expresses that you deem that GPP is controlled by meteorological drivers and O_3 (concentrations). But what are the parameters all known to effect GPP; I missing here in the introduction a mentioning of other parameters that might be important and that might not be easily inferred from remote sensing data, e.g. N-deposition.

Pp 4, line 84, “Excluding soil moisture and meteorology”, I don’t get this statement; You refer to the method of regridding. Do you mean here that for all other parameters than soil moisture and meteorology you have applied this regridding procedure. But then mentioning here the term soil moisture, it would be good to already indicate in the introduction how this parameter can play a (crucial) role in inferring the O_3 impact on GPP.

Pp 4, lines 113-115; Reading the statements about to what extent the CAMS O_3 can be applied for assessing its impact on GPP, just giving the overall statistics expressed by this r value of 0.7 ($r^2 < 0.5$, is actually not such a high-correlation) triggers the question if this applies for summer mean/monthly/diurnal mean/max, or full timeseries? This is of large relevance since what matters most for this impact assessment is how well CAMS captures the high O_3 (extremes) during the days when stomatal uptake is maximum. In addition, you state: "However, over Southern Europe the CAMS reanalysis was found to consistently overestimate surface O_3 concentrations by ~15%.". Other terms in Eq. 6 (e.g., the alpha-term) can have considerable uncertainty. This uncertainty should be propagated in the GPP reduction estimate, especially considering that your model only requires "a fraction of the computational cost otherwise required by land surface models" (Lines 317-318).

Pp 4, lines 126: this motivation of only using the soil moisture of the top layers excluding the information on soil moisture > 1m indicates that you assume that the forests stomatal conductance is mainly controlled by the soil moisture in the top 1m. This might actually depend a lot on the effective rooting depth. I bring this up having seen soil moisture observations in the top soil profiles that seemed to provide a nice source of information to indeed infer the impact of soil water on stomatal conductance but where, then evaluating the observed latent heat fluxes, did not not reflect at all observed strong decreases in those soil moisture measurements.

Pp 9: lines 175-180; At the end of the methods having seen the overview of all the datasets being used, it makes me wonder about any evaluation strategy that you have developed to at least assess that some of the critical parameters in your inversion of the O_3 impact make sense; e.g., did you conduct any evaluation of the Jarvis stomatal conductance based on comparison of the simulated and observed LE? This is a parameter that could have been rather easily evaluated using the FLUXNET datasets.

And why using the AOT40 where in the previous paragraphs you have referred to the use of Jarvis in the DoseO3 model to evaluate the stomatal dose of O_3? This has to be all better motivated and including the potential implications.

Figs. 4 and 5: These figures are derived from monthly O_3-induced GPP estimates. It would be interesting to show the actual monthly data over the growing season (i.e. a time series), this would provide more insight in the dynamics. The boxplot in Fig. 4 is in fact a bit misleading, as this is typically used to show represent errors/uncertainties, but a proper error propagation is lacking