We thank reviewer #2 for taking the time to evaluate our manuscript and for providing very extensive feedback on how to improve it. We have addressed each of their comments in the responses below.

Main comments

Comment 1

From the introduction, I recommend clearly identifying which processes driving biogenic emissions are already well represented in models (such as temperature dependence, LAI, and water stress) and which have been recognized but are still poorly represented (such as plant defenses and signaling via attractor molecules).

We thank the reviewer for making this point and have modified the text (shown in red) in the introduction to discuss these points:

"Such modelling approaches could thus deliver spatial gap-filling in regions lacking observations, which would have scientific value for studying air quality, atmospheric composition, and climate. Given this need, there have naturally been different efforts and approaches developed to model BVOC emissions from vegetation. Not all of the processes mentioned earlier that drive BVOC emissions are currently well represented in BVOC emission models. The effects of solar radiation, temperature, LAI, and in-canopy shading, are typically well represented (Arneth et al., 2007; Guenther et al., 2012; Naik et al., 2004). The effect of water stress is represented in models to varying levels of sophistication (Wang et al., 2022), but there are cases where water stress effects are not considered (Sindelarova et al., 2014, 2022; Stavrakou et al., 2009). Lastly, emission processes linked to plant defences and signalling are less well understood and are currently poorly represented in models."

Comment 2

The authors used the ECMWF HRES operational forecast; however, employing meteorological variables from a reanalysis dataset might have led to more robust results. The authors are kindly invited to comment on this methodological choice.

We thank the authors for their valuable point. Indeed, this was a question we reflected on during the initial stages of development in our study.

The emissions dataset that we present were created in the context of the SEEDS project, which was specifically aimed at supporting developments to the Copernicus Atmospheric Monitoring Service (CAMS). We had a specific focus on providing developments to the European, regional production in CAMS. The CAMS European production runs at a spatial resolution of 0.1° × 0.1°, which corresponds to the spatial resolution of the HRES forecast. Furthermore, some of the CAMS models use the HRES forecast on its native gridding. It was an important aim for us to try to follow the gridding used by HRES to facilitate testing of the emissions in the CAMS models. ERA5 reanalysis has a spatial resolution of 0.25° × 0.25°, and therefore was not suitable for this purpose. It should be noted that SURFEX can only be run at the same spatial resolution as the input meteorology.

In addition to the practical reasons outlined above, there are some sound scientific reasons why, in this case, favouring meteorological forcing based on forecasts at higher resolution over coarser resolution reanalysis is a sound approach. The land surface has a high degree of heterogeneity. Operating a land surface model at higher resolution allows one to represent this heterogeneity in a more thorough way within the model. This can be important in order to represent which particular land covers/vegetation types receive rainfall. This tradeoff was explored in a report by co-authors

(Jarlan et al., 2023), which concluded that higher spatial resolutions improved drought monitoring over specific affected areas/vegetation and improved the representation of its effects on vegetation.

Jarlan, J., Albergel, C., Bonan, B., Calvet, J.-C., de Rosnay, P., Ottlé, C., and Peylin, P.: Assimilation de données de télédétection pour le suivi des surfaces continentales *in Inversion et assimilation de données de télédétection*, 45-95, ISTE Editions, https://doi.org/10.51926/ISTE.9142.ch2, 2023

Using a coarser model resolution also has downstream effects on the assimilation of LAI data. Coarser spatial resolutions in the model mean the satellite LAI data has to be spatially aggregated even more, which degrades the effectiveness of the assimilation step due to broadening of the number of land use classes within larger spatial pixels.

These points are now discussed in the manuscript within the Methodology section 2.4:

"We used the ECMWF HRES forecast as opposed to ECMWF's ERA5 reanalysis (Hersbach et al., 2020) in this study for several reasons. The CAMS European production runs at a spatial resolution of 0.1° × 0.1°, which corresponds to the spatial resolution of the HRES forecast. Furthermore, some of the CAMS models use the HRES forecast on its native gridding. An important aim was to try to follow the gridding used by HRES to facilitate testing of the emissions in the CAMS models. ERA5 reanalysis with its spatial resolution of 0.25° × 0.25° was, therefore, unsuitable for this purpose. Next, since the land surface has a high degree of heterogeneity, operating a land surface model at higher resolution allows one to represent this heterogeneity in a more thorough way within the model. This can be important, for instance, to represent which particular land covers/vegetation types receive rainfall. This was explored by Jarlan et al. (2023) who concluded that higher spatial resolution forcing improved drought monitoring over specific affected areas/vegetation and improved the representation of its effects on vegetation. Lastly, coarser model spatial resolutions have downstream effects on the assimilation of LAI data. Coarser resolutions mean that satellite LAI data has to be spatially aggregated even more, which degrades the effectiveness of the assimilation step due to broadening of the number of land use classes within larger spatial pixels."

Comment 3

In Section 3.1 (Performance of the SURFEX land surface model), could the authors please provide more details on the LAI/SIF comparison? For instance, what is the temporal resolution of the two datasets?

Thank you for this suggestion. We agree that adding this information would improve the quality of the manuscript. The LAI and SIF data are compared on a time frequency dependent on the availability of of the TROPOMI SIF data. The TROPOMI SIF data were the limitation in this case because the SURFEX model provided OL and analysis data continuously for every day with no data gaps. In practice the TROPOMI data were available in approximately 25 days in every month even though the SIF observations are nominally available once per day. The limitation is due to data gaps created by clouds and viewing angle, which in turn impact the retrieval quality. These limitations are more prevalent during the winter months and at higher latitudes. The comparisons are done as and when a model-obs data pair (LAI-SIF) is available.

We have added this text to clarify these details in section 3.1 of the where the performance of SURFEX is evaluated.

"The LAI and SIF data are compared on a time frequency dependent on the availability of the TROPOMI SIF data. The TROPOMI SIF data were the limitation in this case because the SURFEX model provided OL and analysis data continuously for every day with no data gaps. In practice the TROPOMI data were available for approximately 25 days in every month on average even though the SIF observations are nominally available once per day. The limitation is due to data gaps created by clouds and viewing angle, which in turn impact the retrieval quality (Guanter et al., 2021). These limitations are more prevalent during the winter months and at higher latitudes. The comparisons are done as and when a model-obs data pair (LAI-SIF) is available."

Additionally, why is only Saint-Félix-de-Lauragais shown in the soil moisture comparison? Were other sites compared as well, and if so, did the comparisons yield similar results?

We thank the reviewer for raising this valid point. We looked at data comparisons for other sites within the SMOSMANIA network. Given that part of the time period we study occurred during the COVID-19 movement restrictions, some sites within the SMOSMANIA network were not properly maintained, which led to degradations of some sensors that could not be fixed. Inspection of the data showed large data gaps and/or large transient values, which was indicative of sensor degradation. We only presented the comparison of the Saint-Felix-de-Lauragais dataset because this was one of the sensors for which we were most confident about the data quality. However, there are eight other sites within the network that have still have reasonable data quality and we have added these figures (Figures S1 through S8) to the supplement along with a table (Table S2) showing the correlation scores against SURFEX.

We have added text to section 3.1 to explain why Saint-Felix de Lauragais was selected: "Data from the Saint-Felix de Lauragais were selected and presented here because of the excellent data quality from this site. Unfortunately, the soil moisture sensors at some of the other sites in the SMOSMANIA network have degraded leading to data quality issues."

We also explain the data comparison presented in the supplement:

"In addition to the comparison to the data from Saint-Felix de Lauragais, we carried out an evaluation of SURFEX soil moisture data with data from the other sites in the SMOSMANIA network (eight in total) that had reasonable data quality over the same time period. The results of this evaluation are presented in the supplement in Figures S1 to S8 and the correlation statistics between SURFEX and the soil moisture measurements are presented in Table S2. The correlation scores are all above 0.80 the comparisons with the data from each of these sites."

<u>Comment 4</u>

In Section 3.3.2 (Evaluation using other emission datasets), could the authors please provide some explanation for the large differences observed between SURFEX-MEGAN3.0 and the MEGAN-MACC database? Are these differences mainly due to variations in the meteorological parameters, the emission factors, or both?

We thank the reviewer for this comment. Most of the difference between SURFEX-MEGAN3.0 and MEGAN-MACC is due to the different meteorology used in each case. We reach this conclusion based on two things. First, Sindelarova et al. (2022) determined that the MERRA/MERRA-2 reanalysis generally had higher values for 2-meter temperature and photosynthetically active radiation than the ERA5 reanalysis. MEGAN-MACC is based on the MERRA/MERRA-2 reanalysis while CAMS-GLOB-BIO is based on ERA5. Sindelarova et al. (2022) found that this primarily explained the larger emissions of the MEGAN-MACC emissions when compared to those of CAMS-GLOB-BIO. Second, we have performed a comparison of the temperature and radiation gamma (gammaTP) values calculated by MEGAN within SURFEX-MEGAN3.0 (based on HRES) and CAMS-GLOB-BIO (as part of the response to comment 5 here). This comparison showed that the gammaTP values from both inventories were very consistent with each other except in desert regions with low vegetation cover (e.g., north Africa and the Middle East), and in arid regions during the summer that had low summer emissions. Based on these two points, this points to the meteorological parameters from MERRA/MERRA-2 as being the main factor for the large difference between MEGAN-MACC and SURFEX-MEGAN3.0.

We refer the reviewer to the response to comment 4 from reviewer one where we discussed extensively the emission factors of MEGAN3.0, MEGAN2.1 (used in CAMS-GLOB-BIOv3.0 and MEGAN-MACC), and those based on EMEP used in CAMS-GLOB-BIOv3.1. In short here, we concluded that the MEGAN3.0 and MEGAN2.1 emission factors were likely not a large explanatory factor in the differences between SURFEX-MEGAN3.0 and CAMS-GLOB-BIOv3.0, and thus this applies to MEGAN-MACC also. See more detail below in response to comment 5.

We have added this text to the manuscript to discuss these points:

"This evaluation revealed several findings. The differences in the meteorology used in SURFEX-MEGAN3.0 (HRES) and CAMS-GLOB-BIOv3.0/3.1 (ERA5) were likely only a very minor factor in explaining the differences between these emission inventories. We directly compared the radiation-temperature gamma values from SURFEX-MEGAN3.0 and CAMS-GLOB-BIOv3.1 (identical to v3.0) in Figure S18 of the supplement. This comparison shows that the only significant differences in radiation-temperature gamma occurred in desert regions (e.g., north Africa and the Middle East) year round and in southern Mediterranean regions during summer when vegetation cover is very low and the emissions from these regions are relatively minor. With regards to the comparison between SURFEX-MEGAN3.0 and MEGAN-MACC, Sindelarova et al. (2022) found that the MERRA/MERRA-2 meteorological forcing (used to make MEGAN-MACC) had higher 2-meter temperature and PAR values than the ERA5 data (used in CAMS-GLOB-BIOv3-0/3.1) and that this explained why the MEGAN-MACC inventory had much higher isoprene emissions. We can apply that same finding here to conclude that the MEGAN-MACC emissions are much higher than the SURFEX-MEGAN3.0 emissions due to the different meteorological parameters of MERRA/MERRA-2. This is further supported by the good consistency of the radiation-temperature gamma values derived from HRES and ERA5 shown in Figure S18."

Comment 5

The differences between SURFEX-MEGAN3.0 and CAMS-GLOB-BIOv3.1 are only partially explained by variations in LAI. This is particularly noteworthy since, although there is a difference in LAI between the SURFEX-MEGAN3.0 analysis and the SURFEX-MEGAN3.0 OL datasets, this difference does not appear to result in any noticeable change in emissions, at least none that are evident in Figure 16. I would therefore ask the authors to suggest plausible explanations for the discrepancies observed between these two datasets (SURFEX-MEGAN3.0 and CAMS-GLOB-BIOv3.1).

This is a very interesting line of discussion and we thank the reviewer for raising this point. We agree that it would be useful to have more discussions related to the comparison of the CAMS-GLOB-BIOv3.1 and SURFEX-MEGAN3.0 emissions. Indeed, reviewer one raised similar questions regarding emission factors. We added discussion in response to comment 4 from reviewer 1 that provide more explanation regarding the origins of the differences between SURFEX-MEGAN3.0 and CAMS-GLOB-BIOv3.1 that are linked to emission factors. We refer the reviewer to that response for other details.

We have explored the role of meteorology in explaining differences between SURFEX-MEGAN3.0 and other inventories including CAMS-GLOB-BIOv3.0/v3.1 in response to comment 4 here and we refer the reviewer to that response.

We explored the role of LAI in more detail in response to comment 6 and again we refer the reviewer to that response.

In summary, the differences between SURFEX-MEGAN3.0 and CAMS-GLOB-BIOv3.1 are mostly attributed to differences in emission factor and LAI, and to a more minor contribution due to differences in the HRES and ERA5 meteorology in southern Mediterranean regions during the summer.

We also respond to the reviewer's general remarks about LAI. We are not sure to fully understand the relevance of the comparison of the analysis and open-loop to this comparison for understanding differences with respect to CAMS-GLOB-BIOv3.1. We understand that the reviewer is indicating that changes in LAI between the open-loop and analysis do not lead to changes in emissions, therefore the differences in emissions between SURFEX-MEGAN3.0 and CAMS-GLOB-BIOv3.1 cannot be due to changes in LAI. Our response to this specific point is that the changes in LAI between the analysis and OL are much smaller than those between the LAI dataset used in CAMS-GLOB-BIOv3.1 and either SURFEX-derived LAI dataset. This is due to the differences in approach. For the open-loop, SURFEX itself already provides a skilful representation of daily LAI when forced with high quality meteorological forcing (Delire et al., 2020). In the case of the analysis, PROBA-V LAI assimilation in SURFEX gives daily LAI data. Compared to both of these, CAMS-GLOB-BIOv3.1 uses MODIS monthly climatologies.

Delire, C., Séférian R., Decharme B., Alkama R., Calvet J.-C., Carrer D., Gibelin A.-L., Joetzjer E., Morel X., Rocher M., Tzanos D.: The global land carbon cycle simulated with ISBA-CTRIP: improvements over the last decade, Journal of Advances in Modeling Earth Systems, 12, e2019MS001886, https://doi.org/10.1029/2019MS001886, 2020.

To add further to the discussion of MODIS-based LAI, the Copernicus Land Monitoring Service provides an extensive validation of the PROBA-V based LAI product used in the SURFEX analysis. The relevant validation report linked below documents the large discontinuities in the MODIS LAI product that are considered to be unrealistic.

land.copernicus.eu/en/technical-library/quality-assessment-report-proba-v-leaf-area-index-version-2/@@download/file

We have now added a short additional text to the manuscript to mention this:

"Lastly, validation of the CLMS LAI product from PROBA-V (Sanchez-Zapero et al., 2018) indicates that the that the MODIS LAI product has much higher uncertainty. Thus, one interpretation is that the larger systematic uncertainties in the MODIS LAI product drive some of these differences."

Comment 6

In Section 3.3.2 ('Evaluation using other emission datasets'), it is mentioned that the differences in seasonality observed among the various emission datasets are mainly driven by LAI. The authors also note that "LAI datasets (based on Yuan et al., 2011) used to calculate the monthly emission factors in the CAMS-GLOB-BIOv3.1 emissions tend to peak later in summer than the LAI calculated by SURFEX." Could the authors please provide additional details on this comparison, perhaps by including a figure or table to illustrate it more clearly?

Thank you to the reviewer for making this suggestion and we agree that this would improve the manuscript. We have added two figures to the supplement showing the LAI from the SURFEX analysis in 2019 and the LAI climatology from MODIS (Yuan et al, 2011) to support this point and added the following text to the manuscript.

"Evaluating LAI, we found that in general the LAI dataset (based on Yuan et al., 2011) used to calculate the monthly emission factors in the CAMS-GLOB-BIOv3.1 emissions tends to peak later in the summer than the LAI calculated by SURFEX. The CAMS-GLOB-BIOv3.1 emissions used this same monthly climatological mean of LAI (averaged over 2007-2016) derived from MODIS (Sindelarova et al., 2022; Yuan et al., 2011) to calculate the LAI used for the MEGAN activity factor calculations after 2016. As a result, in years where SURFEX estimated particularly extreme deviations in LAI from the historical climatology with an early peak in LAI, e.g., 2019, the divergent effect on the emission seasonality is even greater. The monthly mean LAI analysis data from SURFEX for 2019 are presented in Figure S19 of the supplement along with the LAI 2007-2016 climatological monthly means used in CAMS-GLOB-BIOv3.1 (Figure S20) for the purposes of demonstrating this point. Indeed, it is possible to see large reductions in the monthly mean LAI data from June into July and August in the SURFEX LAI data in important isoprene emitting regions (see Figure S15 showing the monthly mean isoprene in 2019 for reference) that are not visible in the corresponding panels of Figure S20."

Comment 7

I suggest paying attention to the use of the term "dynamic vegetation". This could be confusing for readers, as it may be interpreted as referring to "Dynamic Vegetation Models (DVMs)". Such models represent the biosphere and are capable of simulating vegetation dynamics, that is, the transient development of vegetation composition and structure. This aspect is not addressed in your paper, where you primarily refer to plant phenology.

Thank you to the reviewer for making this suggestion. We have modified all cases of "dynamic vegetation models/modelling" to instead refer to "plant phenological modelling/models".

See for example, the changes on lines 629, 633, 660 and 977.

Responses to minor comments

***Line 208:** The reference should be to Sect. 2.2.2 and not Sect. 2.1.2*

We have corrected this.

***Line 365**: In the sentence "with LAI having a more significant and beneficial effect for the estimation of root zone soil moisture": do the authors really mean that LAI have a beneficial effect for the estimation of root zone soil moisture?*

Yes. This is correct. Albergel et al. (2007) show that improving LAI leads to indirectly improved soil moisture due to the improved representation of evapotranspiration that results from this.

***Line 432**: In the sentence the authors say that "We perform this evaluation using three approaches:", but in the list after there are 4 points.*

Thank you. We have corrected this error.

***Line 434**: Please add a reference to TROPOMI satellite observations of solar induced fluorescence (SIF).*

Done. Thank you. We have added a reference to Guanter et al.

***Line 454**: Please add a reference to SMOSMANIA in situ measurements.*

We have added a reference to Calvet et al., 2007, which is the reference paper for the SMOSMANIA network. Thank you for this advice.

From line 471 to 479*: The authors come back speaking about LAI. It is a little bit misleading. I'll move these lines before in the paragraph where the authors introduce the comparison with LAI.*

We have re-ordered this section to reflect this good advice.

Page 19: *Please increase the size of Figure 5.*

We have increased the size of each panel within Figure 5.

Line 553*: Is the standard deviation calculated for isoprene average annual emitted mass over the 5 years?*

Yes. We changed the text that was on line 503 to say:

"The average annual emitted mass of isoprene for the open-loop emissions over the time period 2018-2022 was calculated to be 7.20 Tg yr^{-1} with a standard deviation of ± 0.28 Tg yr^{-1} over this five year period"

Line 510*: Could you please detailed more what do you mean with "gamma parameters for soil moisture, LAI, and radiation-temperature". Do you mean the different activity factors (g)?*

Yes. By gamma parameters we mean the different activity factors. We have modified the text as follows to make this clearer:

"Indeed, the effect of these last variables can be seen in more detail within Figure 7 which shows the emission factors over the CAMS domain and the standard deviation of the interannual variability for the gamma activity factors for soil moisture, LAI, and radiation-temperature."

We have changed all other instances of "gamma parameters" to "gamma activity factors".

Page 22*: Figure 8, as far as I understood, shows the R2 that is calculated on a yearly basis. This means that the times series of variables have only 5 elements. A correlation calculated like this is not very strong. It should be discussed a little in the text.*

We have added this text to the manuscript to address this point:

"However, note that each R^2 value within these distributions is based on only five data points, so the statistical significance of these correlations is weak. This is offset by the large number of individual cases presented in each distribution, however."

Page 27*: Please add country contours in Figure 12.*

Done.

Page 27*: In Figure 12 (plot on the right). Could you please give some explanation on why there are still some areas where correlation is negative (especially in Portugal)?*

We have added this explanation to the manuscript to address this point:

"There are still some areas with negative correlation where the LAI declines yet isoprene emissions increase in **Error! Reference source not found.**b, e.g., in Portugal, and this is due to a strong increase in the radiation-temperature gamma variable over this region."

Line 688*: Please correct "Figure 15Figure 15".*

We have done this.

Line 759: *In the sentence "such as wind, turbulence, boundary layer height, and isoprene lifetime are implicitly ignored in this comparison." I would remove "isoprene lifetime", as you mentioned just before.*

We have done this.

Page 36: *Figure 16 could be more readable with a legend.*

We have now added a legend.

Line 824: *Please add reference to "SUMO emission inventories".*

In fact we removed the SUMO emissions from earlier versions of the manuscript and this should have been removed. We have done this now.

Lines 809 and 813: *"Analysis" should be written in lowercase letters.*

We have corrected this.