

## **Review for ESSD-2023-237**

### **General comments:**

This article presents a comprehensive global marine isoprene emission dataset at high spatial and temporal resolution for the period of 2001 - 2020. The authors separate marine isoprene emissions into two distinct sources (biogenic and surface microlayer). Emissions are calculated using a combination of satellite chlorophyll and radiance measurements and meteorological reanalysis data (e.g., windspeed from ECMWF's ERA-5 product) with empirical parameterizations. The estimated emissions are compared with a variety of observational records, and correlations with meteorological driving variables as well as climate modes of variability (e.g., El Niño - Southern Oscillation) are explored.

Overall, I think this is a useful and interesting dataset. These high-resolution emissions could be included in a global atmospheric chemistry model to explore the impacts of marine isoprene emissions on aerosol formation and tropospheric oxidation chemistry over the remote ocean. I think the atmospheric chemistry and climate research communities would both benefit from these data.

The article itself is reasonably clear, and the methodology is presented in a straightforward and comprehensive way. However, I have some issues with the lack of uncertainty analysis as well as the lack of justification / explanation for a few assumptions. I recommend publishing this manuscript once these concerns are addressed, because I think it would be very valuable to the global atmospheric science community.

1) I am concerned by the lack of uncertainty analysis presented in this paper. The calculation of both the biogenic ("BIO") and surface microlayer ("SML") isoprene emissions depends on satellite observations from MODIS and VIIRS, meteorological reanalysis data (ERA-5 in this case), and numerous empirical parameters derived from oceanographic or laboratory measurements. Each of these quantities has some uncertainty associated with it, and these will propagate into your emission estimate. While I appreciate that putting precise error bars on global emission estimates is not trivial, some kind of error analysis or sensitivity test seems essential in order to make proper use of your data and methods. Even something as simple as calculating the emissions with a different reanalysis product or changing the values of some of the empirical parameters would give a strong indication of how sensitive the emission estimate is to errors in the model inputs and parameters. This would make the comparison with observations and previous emission estimates more meaningful, and it would make it easier to apply your methodology in different modelling frameworks (perhaps using different meteorological reanalysis data or satellite observations).

2) There are a few assumptions and methods that need more justification / explanation. These include the assumption that isoprene concentrations in seawater are constant, that isoprene is immediately oxidized in the marine boundary layer, and the use of an 8-year plankton type distribution over a twenty-year period. Please see my specific comments for more details.

3) The plots are generally clear and of high quality, but you should include labels for the different subplots and colour bars. I found myself frequently jumping back and forth between the main text, the figures, and the figure captions in order to make sense of everything. I have included a few specific comments about this below.

4) I was able to download and explore your dataset, and I could not find any problems with the files or data structure. However, I could not find any source code for your modules on the FTP server. Perhaps this is intentional; however, your introduction made it seem like your module could be easily embedded in an Earth System Model, so I was under the impression that I would be able to download the source code and play around with it. This is not necessarily a problem, but if you don't intend to release any source code you should consider rephrasing your introduction to avoid giving the impression that people can download your model.

5) The order of the paper is odd at times, and some sections are mislabeled in the introduction (see my specific comment below). I found it strange that the BIO emissions are presented before you explain how the plankton types were calculated. I was also surprised to see the comparison with observations (Section 3) presented before you discussed the spatial and temporal characteristics of your modelled emissions (Section 4). This made it a bit harder for me to follow your overall arguments. But I acknowledge this may just be my personal preference.

### **Specific Comments:**

##### *1 Introduction* #####

Lines 40 - 42:

There has been some work (e.g., Palmer et al 2022:

<https://www.science.org/doi/10.1126/science.abg4506>)

suggesting that terrestrial BVOC emissions have a large impact on downwind VOC and aerosol concentrations over the remote ocean, particularly in the South Atlantic due to the relatively long lifetime of BVOCs coming from the Amazon basin. Is it fair to say that terrestrial BVOC emissions do not exert significant influence over the remote ocean?

Line 52:

Here you use the terms "BIO" and "SML" without first defining them in the main text; they are explained in the abstract, but you did not explain them in the introduction.

Line 60:

I am not sure what this sentence means. I interpret this to mean that biochemical losses of BVOCs are parameterized based on laboratory or field observations. But it is unclear what "dynamic euphotic zone" means in this context. Are you referring to Equation (5), where you calculate the depth of the plankton euphotic zone based on surface downwelling radiation?

Line 70:

Which two emission pathways are you referring to here? Are you talking about the BIO and SML sources, or are you talking about photochemical and windspeed-driven processes in the

surface microlayer? I assume you mean BIO and SML based on the rest of the article, but you could avoid the ambiguity by clearly stating which pathways you're referring to.

Lines 80-88:

This is good background information, but it feels out of place in this paragraph. Consider moving this to the first or second paragraph instead. This information helps motivate why marine isoprene emissions are important, so I feel it should be introduced before you start describing emission estimate methodologies and uncertainties.

Line 90:

This is unclear to me. It would be easier to understand if you succinctly explained how you calculated the BIO and SML sources and how your approach addresses some of the uncertainties you outlined in the previous paragraph (data availability, unclear mechanisms, lack of satellite observations at high latitudes during winter, estimates of chlorophyll vertical distribution, and relations between isoprene and marine/meteorological factors). The way it is currently written, I have no way of knowing how your study plans to approach these issues until I have finished reading the paper.

Line 91:

The sentence "Two distinct types of emissions are separately calculated..." doesn't really make it clear what you're doing, or how you're using the MODIS chlorophyll and ECMWF reanalysis data. More broadly, I find that your introduction provides motivation for studying marine isoprene emissions and addresses some uncertainties in previous approaches, but it does not clearly explain how your new dataset was developed or how it addresses the uncertainties you mentioned. I understand that the introduction needs to be brief, and you describe these methods in detail later. But right now your introduction does not give enough information for me to tell what you actually did. After reading your introduction I should know what to expect from the rest of the paper, and right now that isn't the case.

Line 95 - 98:

This paragraph seems to be incorrect. I think you have swapped the descriptions of Sect 4 and Sect 5. You said "Sect. 4 provides information on our dataset and data availability", but Sect. 4 in the text is simply titled "Results" and is focused on spatio-temporal variability of emissions and correlations with climate modes of variability. Similarly for Sect 5, you said it describes the "characteristics of marine isoprene emission", but in the text Sect 5 is just a data availability statement.

### 2 *Methods* ###

Line 101:

What is the spatial resolution of the downwelling radiative flux diffuse attenuation coefficient data? Is it also at 9km?

Line 106:

These meteorological variables (u-wind and v-wind, T2M, SST, and surface downwelling shortwave flux) are all from ERA-5, right?

Lines 107 - 110:

Is the monthly normalized water-leaving radiance at 410 nm also at 0.25x0.25 degree spatial resolution?

Line 110:

Please clarify how you can apply this plankton distribution dataset over the entire twenty-year period. I understand that you use MODIS chlorophyll and NOAA water-leaving radiance at 410nm to obtain a plankton type distribution from 2012 - 2020. But it is unclear how you can use an 8-year plankton type distribution to estimate emissions over a twenty-year period.

Lines 113 - 115:

You assumed the concentration of isoprene in the ocean is static. Is this steady state assumption valid? What is the justification? Is it based on observations of marine isoprene concentrations, or is it based on theoretical considerations (e.g., ocean chemistry modelling)? And over what time period could we expect this assumption to be valid (Days? Weeks? Months?)? I appreciate that this assumption is very useful so that isoprene flux is equal to net isoprene production, but some more explanation / justification should be included.

Line 116:

How do we know isoprene will be oxidized immediately once it enters the marine boundary layer? While isoprene typically has a very short lifetime against OH oxidation, non-negligible isoprene and other BVOC mixing ratios have been measured in the marine boundary layer (e.g., Warneke et al., 2004), particularly around day-to-night transitions when OH concentrations are lower. If I understand correctly, you are assuming MBL isoprene concentrations are negligible so that you can neglect isoprene fluxes from air-to-sea, and instead focus exclusively on fluxes from sea-to-air. Can you provide some more context (i.e., why are you assuming it's negligible?) and justification (i.e., how do we know that isoprene is oxidized immediately in the MBL? What is its typical lifetime in the marine atmosphere?)?

Line 121:

Could you please clarify what you mean by "biochemical costs of isoprene in seawater"? Does this refer to the consumption of isoprene by biological processes, or are you talking about something else?

Line 122:

What kinds of observations did Simo et al 2022 use to calculate alpha? Was this relationship observed in different regions of the ocean, or did they use a small set of observations? In other words, do we think this relationship is robust enough to be applied to the global ocean?

Lines 133 - 135:

Can you clearly state that radiation is given by I and the plankton type coefficient is given by Tc in this sentence? Otherwise Equation (3) is unclear until after the next paragraph.

Lines 156 - 158:

I understand that 0.433 Tg yr<sup>-1</sup> is only accounting for the BIO source, but you include your total estimate (BIO + SML) in Table 1. Are the other emissions estimates in Table 1 total emissions?

This is what I assumed, but the Brüggemann et al study is listed as “Sea Surface Microlayer” so now I am not sure. Please specify in the table whether these are TOTAL, BIO, or SML emission estimates so that it is easier to compare the different studies.

Also, what is the uncertainty on your BIO estimate? Do you have an idea of how this estimate might change based on errors in the input data (e.g., MODIS chlorophyll or ERA-5 meteorological data) or model parameters?

Line 161:

You say these factors (temp, salinity, etc.) will lead to different isoprene production rates, but you only accounted for radiation and photic zone depth (Hmax) in equations 1 - 5. How did you account for the impact of temperature, salinity, and nutrients? Are these impacts small / negligible, or are they implicitly accounted for by the chlorophyll concentration term in Equation 2 and Equation 3?

Line 176:

Why was the value of 0.028 chosen? Do we expect haptophytes to dominate in oligotrophic regions of the ocean, or is there another reason you chose this value?

Lines 180 - 183:

Is this for coastal regions everywhere, or did these studies focus on specific regions?

Lines 195 - 197:

What about the large areas of undefined types in tropical and subtropical regions? In particular, Figure 1 a), b), and d) show large “undefined” areas in the southern Subtropical Pacific and western tropical Pacific. Another hotspot seems to be the Arabian Sea and Bay of Bengal in Figure 1 c). Is the use of an undefined plankton type a large source of uncertainty in these regions?

Lines 219 - 227:

I understand that you are using chlorophyll observations as a proxy for nutrient levels and surfactant concentrations, but I do not understand where the expressions for  $C_{surf}$  and  $C_{max}$  come from. Can you explain this? Perhaps a brief explanation of the methodology of Wurl et al 2011, or at least state the rationale behind Equation (8) and the expressions for  $C_{surf}$  and  $C_{max}$ . Right now this method seems very opaque without having read Wurl et al 2011.

Lines 232 - 234:

Is this due to the destruction of the surface micro-layer at high windspeeds? Why is your wind speed threshold (13ms<sup>-1</sup>) different from the one mentioned in the introduction (10ms<sup>-1</sup>)?

What is the mean SML emission rate? At the end of Section 2.2 you gave a mean BIO estimate, so it would be nice to see the same for SML here. And just like with Section 2.2, you should address the uncertainties in this estimate (or at least explain if you will address them in a later section). Your estimate of SML emissions relies on several empirical parameters (e.g., *Flab*) and meteorological input variables. All of these quantities have uncertainties, which will propagate into your emission estimate. Some sort of error analysis or sensitivity test would be extremely valuable.

Lines 251 - 253:

How do these changes compare to the uncertainties on BIO and SML emissions? Is your interpolation a major source of uncertainty in your emission estimate, or are these changes significantly smaller than the other sources of error?

###

3

*Evaluation and comparison*

###

Lines 272 - 277:

How do these differences compare to the uncertainty of your estimate? For the comparison with observations to be meaningful, we need to know what kinds of errors are present in your emission estimate.

Line 292:

I mentioned this in an earlier section, but please explain the assumption that marine isoprene concentrations are constant. It seems to me that you are assuming concentrations are constant to estimate the fluxes, then using those fluxes to estimate the concentrations. The logic seems circular. We know the concentrations are not constant because you show large variability in both observed and modelled concentrations in Figure 4. I don't doubt that you can use some sort of steady-state approximation in order to relate isoprene production to fluxes, but this needs to be clearly explained.

Lines 295 - 301:

Is sea surface temperature also coming from ERA-5? Would you get very different results for Equations 10 - 12 if you used a different reanalysis product?

Lines 309 - 313:

Similar to my previous comments on BIO and SML estimates, can you do a sensitivity test to at least get some idea about the uncertainties? You mention that Equation (11) may introduce uncertainties which could partly explain model-observation discrepancies, but you don't quantify how big these biases might be. Even if getting a precise error estimate is difficult, it should at least be easy to figure out the impact of errors in sea surface temperature and 10-metre wind speed.

You also mention that Eq. 11 is only valid in the range of  $w = 4 - 15\text{ms}^{-1}$ , so ideally you should eliminate this source of error by excluding locations and times where  $w$  falls outside of this range. Do you filter for wind speed in your simulation, or are you using all data points even if they fall outside the range of  $4-15\text{ms}^{-1}$ ? If you are using all data points even though Eq. 11 is not valid, how big of an error might this introduce?

Lines 325 - 328:

These ranges are small, so I would expect that the uncertainties on these estimates are probably much larger than the reported ranges. So it would be very useful to include an error estimate here.

Lines 345 - 360:

If I understand correctly, the three main benefits of your approach to estimating BIO emissions are: 1) increased temporal resolution which better captures emission dynamics, 2) increased spatial

resolution which better captures the spatial heterogeneity of emissions in coastal environments, and 3) and improved plankton type distribution which resolves the issue of missing phytoplankton types in coastal regions. What is the benefit of using the new parameterization in Eq. 2, and is this significantly better than the previous approach you described in the introduction where a linear relationship between chlorophyll concentration and isoprene emission was used?

In general, I agree that there are benefits to using a higher spatial and temporal resolution, but I am not completely convinced that the updates used to calculate BIO reduce uncertainties. Your estimate depends on various satellite and ERA-5 input variables as well as laboratory-derived empirical parameters. All these quantities have their own uncertainties which will affect your BIO estimate. I think it is essential that you try to quantify this uncertainty.

Line 364:

Is this the main benefit of your SML approach compared to previous methods?

Lines 365 - 367:

I don't think this needs to be re-stated here.

### 4 **Results** ###

Lines 375 - 380:

Do these statistics account for the difference in ocean surface area between the two hemispheres? (i.e., does an “average” Southern Hemisphere grid cell emit more isoprene than an “average” Northern Hemisphere grid cell, or can the difference partly be explained by the fact that there is less ocean in the Northern Hemisphere?). It's not easy to tell using the maps in Figure 6.

Line 440:

This vocabulary is a bit unclear to me. Are you saying that BVOC emissions in the tropical ocean are primarily governed by local / small scale atmosphere-ocean interactions (e.g., small scale weather systems)? Or am I misinterpreting something? What do you mean by "local air-sea system"?

Line 441:

This is why I am confused about "air-sea system". In the previous sentence you say emissions are determined by local air-sea system (local weather conditions?), and here you say large scale variability is important. Please clarify what you mean by air-sea system and what you mean by “local” versus “large-scale” air-sea system. Are we talking about weather systems or large-scale climate variability?

Line 445:

General comment that applies to all multi-panel plots, especially the world maps: Please include labels for the different subplots and for the colour bars. The plots themselves are clear, but it is tedious to keep going back and forth between the figure caption and the plot to make sense of what I am looking at.

Line 449:

Again, it is unclear what you mean by air-sea system. In this section it seems that you are describing modes of climate variability like ENSO, but in the previous section it sounded like you were describing local weather systems.

Line 449:

Please clarify what "identify the target area" means. Is it the region of the ocean that is affected by a particular mode of climate variability (e.g., ENSO)?

Line 482:

I appreciate that this is an exploratory analysis. It would be interesting if you could briefly speculate on how to verify these relationships. Are there other analyses you could do with your dataset? If not, do you know what other kinds of data / observations would help determine whether these relationships are robust? The correlations you have shown between your emission dataset and different modes of climate variability are interesting, and I think it would be beneficial if you could provide some ideas for how to use this information in follow-up studies.

Line 484 (Figure 10):

I have a couple problems with this figure. You didn't include a legend, so it is unclear what the different colours contours represent, what the solid black lines represent, or what the dashed red lines represent.

Also, the subplots need to be labelled so that it is obvious which emissions (BIO or SML) and which region (Mid-latitude Pacific or Tropical Indian Ocean & Pacific) we are looking at. I found myself frequently jumping back and forth between the figure, the body text, and the figure caption to make sense of the results.

The patterns you described in the text are reasonably clear (e.g., the 0.25 year signal in panel C) when you know what to look for, but it would be much easier to interpret the figure if you included a legend and labels for the subplots.

Lines 487 - 505:

This section needs to be expanded. The discussion here is good, but you also need to address uncertainties due to the parameterizations as well as the ERA-5 and satellite input data. Some effort needs to be made to quantify these uncertainties for other researchers to make use of these data. My concern is that if other researchers try to apply your method and get wildly different results, they won't know whether it's due to an error in their methodology or if it's an expected error due to uncertainties in the model parameters and inputs.

### 5 *Data Availability* ###

Line 506:

I see the hourly dataset at 0.25x0.25 degrees is available online. I was able to connect to the FTP server and download the files. I only downloaded a small subset due to the large size of the dataset (2.65 TB), but the files I looked at were formatted properly and I was able to make some plots with them using Python's netCDF libraries.

However, I don't see any way to access your emission module. You state on lines 93-94 that the module can be used to calculate emissions online in an Earth System Model. Are you planning to release the code for your module, or is it expected that other researchers wishing to use your method would implement it themselves based on the equations you provided? Please clarify, because the introduction made it seem like it would be possible to download the source code.

### 6 *Conclusions and Perspective* ###

Line 522:

You are only talking about Winter, right? In Section 4.1 you said NH emissions were 44% and SH emissions were 56%. Here it sounds like you are claiming SH emissions are twice as large as NH emissions.

Lines 525 - 527:

Can you explicitly connect the observed trends in Section 4.1 with the correlations observed in Section 4.2 and air-sea systems observed in Section 4.3? It would be useful to provide some more context for the reported trends.

Lines 531 - 533:

It's not entirely clear how these relationships will improve the accuracy of isoprene emission estimates. The connections are certainly interesting, but it is not clear what you can do with this information to improve emissions or reconcile discrepancies between observations and models.