

We thank Alex Guenther and two anonymous referees for their valuable comments on our manuscript. The general feeling seems to be that this paper will form a useful part of the scientific literature but could be further strengthened through the inclusion of emission potentials based on measurements made near a set of defined conditions and through the addition of further recommendations of best-practice for both above-canopy and leaf-level experimentalists. We are, of course, happy to address both of these recommendations in our revised manuscript. In addition we address each of your individual concerns in our more detailed response below.

Response to Reviewer 1, Alex Guenther

General comments

1) The most important comment that I have is that the authors should consider an approach that recognizes that measurements similar to the standard conditions should be used to determine the emission potential. I recognize that this may not result in the best estimate of the daily total emission or the emission at conditions dissimilar to the standard conditions but I would argue that is not the “job” of the emission potential. The emission potential, by definition, is the emission at specified standard conditions and so the best estimate is made by either selecting measurements within some fairly narrow range of conditions or perhaps weighting measurements by how close they are to the standard conditions. It is the “job” of the emission algorithm to go from the emission potential at standard conditions to the emission at other conditions. If the emission algorithm does not do a good job of this then there will be errors but you shouldn’t bias the emission potential to try account for this. Instead you should work on developing a better emission algorithm. There will be relatively little difference in the emission potential calculated by different emission algorithms if the emission potential is based on measurements made under conditions similar to the standard conditions. Of course, the problem with applying this approach to canopy scale flux data is that we can’t control the measurement conditions and there may not be any that are similar to the standard conditions. If that is the case then the emission potential should be reported for some standard condition that is within the range of the observed conditions. You could then leave it up to the developers/users of a given model to convert this to an emission potential for the standard conditions of their model. Looking at Figure 1, it appears that 3 of the sites would have some measurements at T=30, PPFD=1500 while T=25, PPFD=1500 might be appropriate for the other two. You could report the emission potential for T=25 as the measured emission potential and then also report one or more calculated emission potentials for T=30 along with an exact specification of the model approach used to get there.

AR: In its current form our manuscript is written very much from a measurement perspective so we certainly welcome the opportunity to learn what those involved in BVOC emission model development would find most useful in order to ensure that the recommendations we make are consistent with the needs of the community. We fully agree that if the algorithms were perfectly describing the response to the environmental conditions, the problem our paper discusses would not exist. Thus, developing a better algorithm is a great solution, but until a perfect algorithm is developed the problems we outline will persist.

We also agree with your assertion that each algorithm would derive a similar emission potential when working with only values based close to standard conditions, because the γ factors that account for deviation from the standard conditions would all be small. However, as formulated the standard conditions of most models are far removed from conditions found, e.g., in higher latitudes including at some of our sites. Thus, even if the measurements are made at the reference conditions under laboratory conditions there are two problems: (a) the plant species may not be adapted to these reference conditions and the emissions may not be representative and (b) the algorithm then still needs to correctly extrapolate from those reference conditions to the modelled conditions. The only problem that is solved by this approach is that a different algorithm may otherwise be used in extrapolating from the measurement conditions to the reference conditions than is used for extrapolating from the reference to the modelled conditions, thus duplicating the uncertainty.

Nonetheless, we like your suggestion in principle to report an emission potential together with the conditions that is typical for the measurement dataset and then leave it to the emission modeller to infer the emission potential at their model's reference conditions that is consistent with these emissions under given conditions.

However, there are two problems. As we mention in the manuscript, our reasoning for not including an emission potential based on measurements made close to standard conditions is that the percentage of data that meet these criteria is incredibly low. In addition, as emission algorithms are getting increasingly complex, this is reflected in an increasing number of reference parameters, which means the space of measurement conditions has to be stratified in many dimensions. We had therefore not considered reporting an emission potential for a different set of defined conditions but are happy to do so in the revised manuscript and to include this as a recommendation if you think this would be useful for the modelling community.

Focussing only on the instantaneous responses to PAR and temperature, we have subsequently re-analysed each data set using a 2d histogram to identify the most appropriate set of conditions to use, e.g. a period near the solar maximum with sufficient frequency to provide a robust average flux. We searched for fluxes within windows of ± 0.5 K and $\pm 100 \mu\text{mol m}^{-2} \text{s}^{-1}$ PPFD. Figure 1 shows the results for the Ispra forest site, which highlights the most abundant set of conditions to use to be between 302-303 K and 1600-1800 $\mu\text{mol m}^{-2} \text{s}^{-1}$ PPFD, yielding a total of 19 flux measurements to average. We limit our defined conditions to just current light and temperature as refining the search further to account for the other gamma terms (e.g. T24, T240, PPFD24, PPFD240, RH, wind speed) would limit the available data to little more than $n=1$. Instead, in Table 1 we list the average fluxes for the defined conditions along with the average of the gamma terms with associated standard deviations.

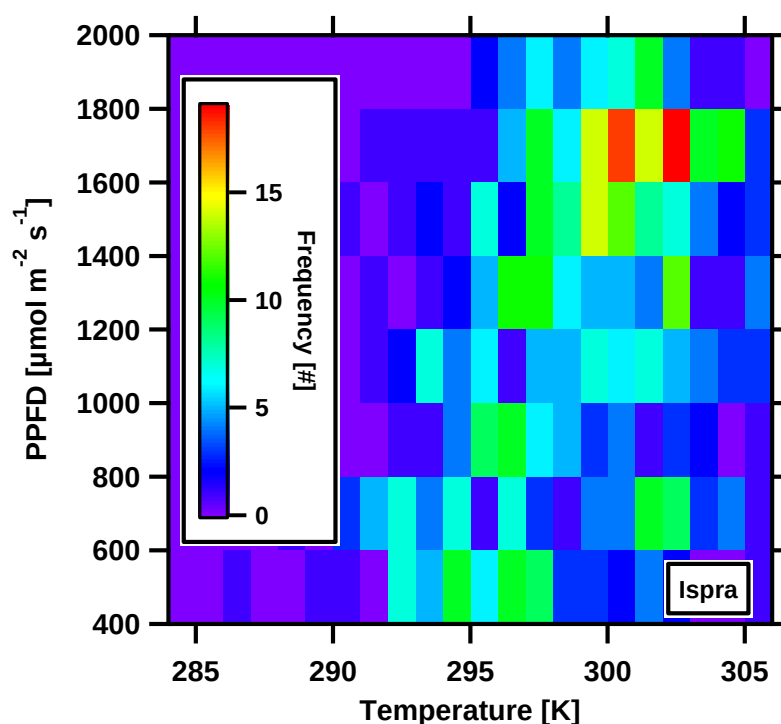


Figure 1. 2d histogram showing the number of flux measurement made at the Ispra forest site that fall within defined light (± 100 PPFD) and temperature (± 0.5 K) bins.

As you suggest, reporting fluxes for a set of defined conditions in this way will allow model developers to convert these to the standard conditions used in their model, but, as you point out, will unavoidably introduce further uncertainty. To investigate this further we derived new emission potentials “converted” from the measured values in Table 1 for the G93 and MEGAN 2.1 models (a, b and c) and then compared the performance of these algorithms at replicating our measured fluxes at each site. Figure 2 shows the percentage difference between the averaged measured flux and the averaged modelled flux when using the “converted” isoprene emission potentials. The calculated bias ranged between +29% and -4% for the G93 algorithm and between +9% and -40% for the MEGAN 2.1 approaches. The bias for the G93 algorithm is typically positive which reflects the fact that the algorithm performs well at conditions close to standard conditions but performs worse in the morning and afternoon, overestimating emission fluxes due to its inability to account for the attenuation of light and temperature through the canopy. The observed bias in the MEGAN2.1 simulated isoprene fluxes is driven by two factors (i) the fact that the average flux for the set of defined conditions is based on a limited number of data points (which affects both algorithms), ranging between $n=4$ to $n=19$, which may be a poor representation of the typical flux footprint and canopy heterogeneity and (ii) the defined conditions are based on current PPFD and temperature with larger uncertainty on the remaining gamma terms such as past PPFD and temperature.

Thus we conclude that this approach, by definition, succeeds in simulating the emissions at ‘typical’ conditions encountered at the site, but not in reproducing the average emission.

Table 1. Isoprene emission potentials for each of the five sites for a set of defined conditions. Numbers in brackets show 1 σ .

	Alice Holt	Bosco Fontana	Castelporziano	Ispra	O3HP
IEP (average flux) [$\mu\text{g m}^{-2} \text{h}^{-1}$]	2143	1911	83	9404	2649
σ [$\mu\text{g m}^{-2} \text{s}^{-1}$]	1075	599	102	3593	988
<i>RE</i> [$\mu\text{g m}^{-2} \text{h}^{-1}$]	142	443	31	1268	353
<i>N</i> [#]	9	17	5	19	4
Temperature range [K]	293-294	302-303	300-301	302-303	293 - 294
PPFD range [$\mu\text{mol m}^{-2} \text{s}^{-1}$]	800-1000	1800-2000	1400-1600	1600-1800	1800-2000
Mean Temperature [K]	293.4 (0.2)	302.5 (0.3)	300.5 (0.14)	302.6 (0.3)	293.7 (0.16)
Mean PPFD [$\mu\text{mol m}^{-2} \text{s}^{-1}$]	915 (66)	1902 (60)	1523 (44)	1703 (61)	1852 (35)
Mean 24 T [K]	290 (1.1)	299 (1.4)	295 (0.6)	298 (1.6)	290 (0.9)
Mean 240T[K]	290 (0.94)	299 (1.8)	295 (0.25)	297 (1.4)	290 (1)
Mean 24 PPFD [$\mu\text{mol m}^{-2} \text{s}^{-1}$]	432 (84)	680 (70)	424 (31)	556 (3)	625 (54)
Mean 240 PPFD [$\mu\text{mol m}^{-2} \text{s}^{-1}$]	415 (92)	659 (48)	452 (15)	553 (17)	591 (0.7)
Humidity [g/kg]	7.9 (1.2)	11.9 (1.6)	13.5 (1)	11.4 (1.7)	6.5 (0.8)
Wind Speed [m s^{-1}]	2.19 (1)	2 (0.81)	1.8 (0.5)	1.4 (0.5)	4.1 (1.4)

As we discuss in the original manuscript, if emission potentials are calculated using all measured flux data and not just those obtained at a set of defined conditions, then the average measured flux can be replicated by the algorithm with zero bias assuming the weighted average approach has been used to derive the emission potential. The drawback to this approach is that that emission potential cannot then be easily converted for use in different emission models. We agree with you that publishing an average flux for a set of defined conditions may be more readily used by model developers and hence have a wider impact, but we believe it is necessary to highlight that this approach results in emission potentials that are inherently more uncertain, especially for the more complex algorithms where not all of the gamma terms can be controlled. In the revised manuscript we will have a further section to discuss the findings shown here and will recommend experimentalists to adopt both approaches. In addition, we will further stress the importance of researchers submitting their observational data sets to online, publically accessible, data repositories such as the VOCsNET database, as we believe a well populated community database would be a far more valuable resource to model developers and would support further improvement in emission algorithms.

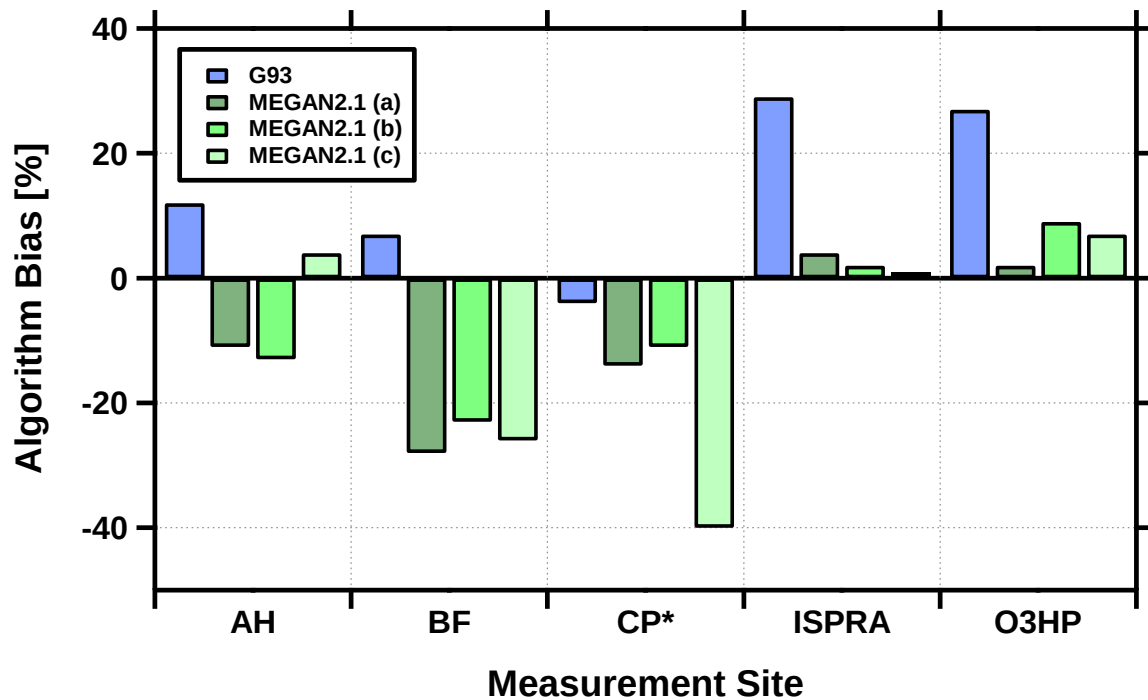


Figure 2. Percentage bias of the average isoprene emission flux simulated by the G93 and MEGAN2.1 emission algorithms at the five measurement sites, Alice Holt (AH), Bosco Fontana (BF), Castelporziano (CP), Ispra forest (ISPRA) and the Observatoire de Haute Provence (O3HP), compared to the measured average flux when using a “converted” emission potential.

2) I assume that there is at least some landscape heterogeneity at some of these sites. The authors should consider binning measurements for different “footprints” associated with different wind directions that represent different oak fractions. This could provide “replicates” with emission potentials for a larger range of oak fraction values that may provide some insights into the value and variability in the leaf-level emission potential. Of course, this assumes that there is some information on the landscape heterogeneity at these sites.

AR: We agree that such an approach could prove useful where there is very detailed species composition data available. Unfortunately, the information we have on species composition at each site is for the forest as a whole and not spatially resolved. Nonetheless, calculating an emission potential by wind sector does provide some information on the spatial variability of the emission potential. For each site we will investigate to see if there is sufficient variability in the wind direction to enable us to infer a species composition uncertainty based on the variability of emission potentials calculated for each wind sector. We have explored the spatial aspects of species composition for the Bosco Fontana field site in a separate paper (Acton et al., 2016) and will refer to the main messages in the revised manuscript.

Specific comments

Page 2, line 28: “In the Guenther algorithms, isoprene emission rates are modelled by assessing the emission potential”. This is not something specific to these algorithms all isoprene emission models include some term of this type, although they may not call it an emission factor.

AR: We will change this to “ In most BVOC emission algorithms...”

Page 4, line 31: delete “to” in “Castelporziano has to a Thermo-Mediterranean”

AR: This will be changed

Page 6, line 18 and line 33: Be more specific about how the tendencies for studies to use a big leaf approach and using leaf temperature equals air temperature. For example, how many of the studies listed in line 16/17 do this? It may be useful to consider that at least one reason investigators do is because of the considerable effort involved in applying the full inverse canopy algorithm to their dataset and it would be useful to have an easier way to do this. For example, Yu et al. al 2017 (<http://doi.org/10.1016/j.scitotenv.2017.03.262>) calculated emission factors using an aircraft flux measurement dataset by using the single point version of MEGAN2.1 that you mention on page 7 line 9, and is relatively easy to use, and compare this with emission factors estimated using the regional MEGAN2.1 FORTRAN code, which is relatively difficult to use. A possible recommendation from your study is that BVOC emission modelers should provide a single point version of their code that can more easily be used to derive emission potentials from tower and aircraft flux data.

AR: We agree wholeheartedly with you on this point. The big leaf G93 approach is undoubtedly the most widely used method to calculate emission potentials due to its simplicity. Our investigation of different algorithms was only made possible through the use of the “Pocket MEGAN” you provided so we will ensure the revised manuscript includes a recommendation for model developers to provide a single point version of their code to enable experimentalists to more easily calculate emission potentials.

Page 8, line 32-36: The statement that “Measurements of the emission potential made using leaf-cuvette systems on a single leaf or branch gives a direct measurement of the isoprene emission rate that inherently excludes the deposition process.” seems inconsistent with “but it may still be offset slightly as some of the isoprene may undergo dry deposition to leaf surfaces”. The leaf cuvette measurement excludes deposition to other leaves and to the soil but there is the possibility of uptake by the emitting leaf including by phyllosphere microbes on the leaves.

AR: You are correct. We will amend the text accordingly

Page 10, line 27: what is the basis for the 10% uncertainty assigned to species composition and 15% to LAI? Does this consider landscape heterogeneities and the uncertainty associated with differences in the LAI and species composition within the footprint of each measurement in comparison to the average for the whole area?

AR: The species composition data and information we have on LAI for each measurement site did not come with associated uncertainties and therefore these values are fairly arbitrary. As discussed above, we will revise the 10% species composition based on the spatial variation of isoprene emission potentials when broken down by wind sector. An initial analysis of IEP windroses for the AH, BF and Ispra forest sites and shown in Figure 3, reveals that the emission potential is fairly constant with wind direction. Taking the standard deviation of the IEP from different wind sectors and comparing with the site average suggests a variability of 14% to 28%. The largest variability was seen at the BF site (28%), which had the smallest fraction of oak and the smallest was seen at AH (14%) which was composed of 90% oak. Wind rose analysis were not performed on the two remaining sites, Castelporziano and O3HP, because these were much shorter time series with insufficient data coverage to provide meaningful emission potentials for different wind sectors. In the revised manuscript we will increase the uncertainty for these two sites from 10% to 20%.

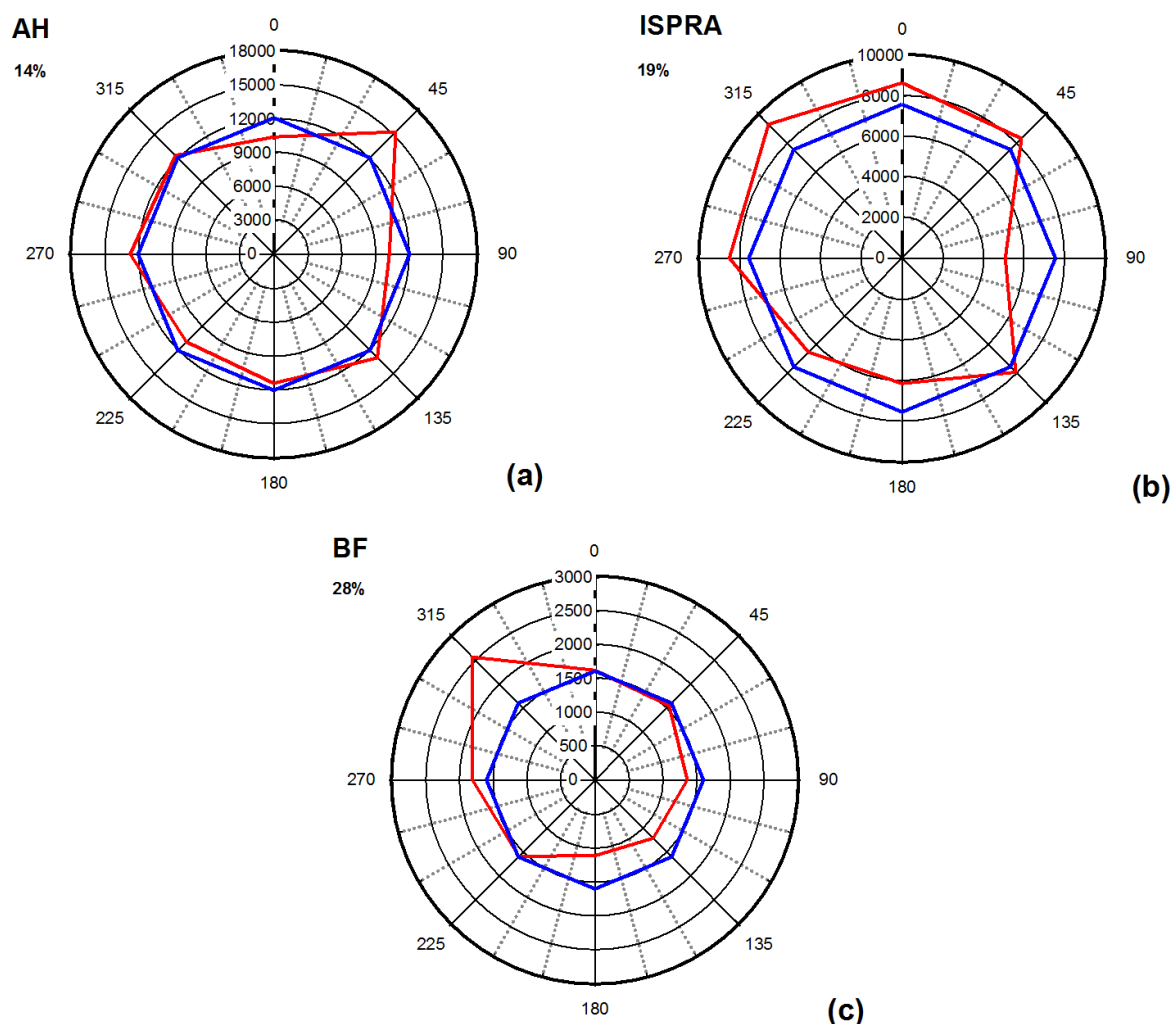


Figure 3. Isoprene emission potentials calculated by wind sector for the Alice Holt (a), Ispra forest (b) and Bosco Fontana (c) sites (red traces) compared to the site average emission potential (blue trace).

Page 10, line 34: The specific leaf mass that you use to convert canopy scale emission potentials to leaf-per-mass emission potentials are arguably as uncertain and variable as isoprene emission potentials. A 25% uncertainty for specific leaf mass may be a reasonable value but you should justify this number and mention how this makes it difficult to compare canopy and leaf scale measurements. This uncertainty could be eliminated if the investigators making leaf level measurements would report emissions as both “per mass” and “per area” leaf emission potentials (i.e., they should provide the specific leaf mass for each measurement) and I suggest that this be a recommendation. If some of the leaf level data that you refer to include measured specific leaf mass (and so direct measurements of per-area leaf emission potentials) then you should make this more direct comparison that does not suffer from the large uncertainties in specific leaf mass estimates (you could do this in addition to the comparison you have already made with the per-mass leaf emission potentials).

AR: This is an important point which we will add to our discussion along with specific recommendation for leaf-level emission potentials to be reported on both a “per mass” and “per area” basis.

Page 11, line 29: Define what you mean by a “wide” range. The range given here of 6750 +/- 1150 is equivalent to +/- 17% which is well within the uncertainties that you discuss. Should that be considered a wide range?

AR: This will be changed to “...the calculated emission potentials span from ~ 5,600 to 7,900 $\mu\text{g m}^{-2} \text{h}^{-1}$ ”

Page 12, line 22: “regional or VOC global” should be “regional or global VOC” Page 13, line 26: MEGAN2.1 allows users the option of using a constant value for each of the 15 PFTs but the recommended approach is to use the MEGAN2.1 isoprene emission factor map that accounts for the fraction of isoprene emitters in each landscape based on plant species composition and the species specific emission potential for each location.

AR: We will make the suggested change and highlight the suggested MEGAN best practice in the revised manuscript.

Page 14, line 1: The MEGAN2.1 canopy-scale emission potential for high isoprene emitters is 24000 $\mu\text{g m}^{-2} \text{h}^{-1}$. The global average temperate broadleaf deciduous tree PFT isoprene emission potential of 10000 thus represents a canopy composed of 41.6% high isoprene emitting trees which is high but not “primarily composed” as stated in the text.

AR: We will make this point clear in the revised manuscript and rephrase our statement accordingly

Page 14, line 17-22: As is pointed out in this manuscript, canopy-average leaf-level PPFD values are considerably lower than above canopy values. Even sun leaves have a PPFD that is typically 50% or less than the above canopy PPFD since they are, on average, at an angle to the sun. The MEGAN2.1 standard condition for the past 24 and 240 hour PPFD refers to the leaf-level value and it is not appropriate (i.e., it just doesn’t work) to use the above-canopy value (i.e., a big-leaf model) with this equation (G06). That the G06 past 24/240 hour algorithm should

not be used with the big leaf model is an important point to make in this paper but then going on to compare the MEGAN leaf-level PPFD standard condition with the measured above-canopy PPFD in figures S5 to S9 is comparing “apples and oranges” and may be confusing to some readers. It should be made clear that this is a comparison of two different parameters (above canopy PPFD and leaf-level canopy-average PPFD) and the main point is that the above-canopy value should not be used in the past 24/240 algorithm.

AR: Agreed. We will make this point clear in figures S5 to S9.

Page 14, line 28: Why not just conclude/recommend that the G06 algorithm should not be used with a big leaf model?

AR: Agreed. We will add this recommendation here.

Page 14, line 23 (and Figure 1 and Figure 7): Check on the values of PPFD shown for Castelporziano. They appear to be higher than what would be expected at the top to the atmosphere. Also, note that PAR should always be expressed in units of W/m² while PPFD is the appropriate term when you use units of micromol/m²/s.

AR: Agreed. We checked with the data owner who has now provided the PAR data in the correct units. We have re-analysed all of the Castelporziano data using the correct PAR data and have updated the text, tables and figures accordingly. Additionally we have now replaced PAR with PPFD throughout the manuscript.

Page 14, line 37: This sentence is confusing.

AR: This will be changed to “Interestingly, when the use of previous light and temperature is switched off (e.g. MEGAN 2.1 (c)) the emission potential increases as the effects of past light and temperature are no longer considered.

Page 14, line 40: This may be because the Castelporziano PPFD solar radiation value is incorrect as mentioned above.

AR: Thank you for pointing out this error. We now have the correct data from the data provider and have recalculated all of the emission potentials and replotted all graphs and tables to account for the adjusted PAR values.

Page 15, line 4-6: or when they are measured under conditions similar to the standard conditions

AR: Agreed, we will stress this point in the revised manuscript.

Page 15, line 9: Leaf-level isoprene emission potential varies considerably between the top and bottom of the canopy and also depending on the past light and temperature environment. Are the leaf level emissions representative of the canopy average, as is the case for the canopy scale measurements, and is the past light and temperature similar? If this is not known, and it is often not reported for leaf-level studies, then this point could be included in the discussion of uncertainties for this comparison.

AR: Agreed we will add this point to our discussion and make a recommendation for past light and temperature to be reported with Leaf-Level emission potentials.

Page 15, line 23: As discussed above, an alternative approach is to select only measurements that are close to the standard conditions.

AR: This will be added.

Page 16, line 37: This is an important point and a good opportunity for you to provide some recommendations for the standardization of flux measurements.

AR: Standardisation of VOC flux measurements is undoubtedly important, but we are not comfortable with making specific recommendations without fully engaging with the community. Encouragingly, some progress in this area is being made with a PTR-MS intercomparison scheduled for later this year in Cabauw as part of the European research infrastructure project ACTRIS. This will hopefully lead to the formation of standard instrument operating procedures but a similar effort is needed for flux measurements and in particular for their post-processing.

Figures 4-6: You were generally consistent in referring to “emission potentials” but these figures refer to “emission factors”. Either can be used but be consistent.

AR: These will be changed

Response to Anonymous Reviewer 2

General

1) The authors should try to avoid the confusion between the same parameters derived in a different way/scale/conditions. Alex’s point to use the conditions closest to the standard conditions seems like another sensible approach worth evaluating. However, inverting the algorithm even at conditions significantly deviating from standard conditions seems still worth the exercise but must necessarily lead to larger errors from environmental parameters measured simultaneously, and potentially may become inconsistent with original model design or intent. Assuming the measured environmental parameters (e.g. T, PAR) are accurate, the value of inferring about the emission potential at different conditions seems valuable to assess how well the algorithmic activity factor works. If it works well, then the emission potential collected under the conditions close to standard should be similar to that inverted from fluxes measured at different conditions with reverse algorithm within the

same footprint. For example, Figure 2 showing stable measured emission potential during the day is unbelievably encouraging, so this approach in my opinion deserves some greater attention.

AR: We agree that Alex's suggestion is a good one and have added emission potentials for specific light and temperature conditions which we discuss further in our response to Alex and will also include in the revised manuscript.

2) Model parameters which were designed for the leaf level-scale may not always be compatible for comparison or extrapolation with the same parameter obtained from inverting the equation at the ecosystem scale, even if in principle it should work. For this reason, it would be helpful to use a thoughtful system of descriptors for equivalent parameters, e.g. E_{Fcan} or $E_{Fextrap}$, so it is clear and distinguishable how the parameter was obtained. This will help the issue which the authors are trying to communicate to modelers (last paragraph in the abstract) that they should be careful about how these parameters were derived before using them.

AR: In the supplementary information we do already make this distinction by presenting emission potentials as E_{eco} (ecosystem emission potential), E_{can} (Oak canopy emission potential) and E_{LL} (Leaf-level equivalent emission potential). We will add the subscripts to figures 4-6 and throughout the text.

3) The abstract seems somewhat heavy for a reader. The take-home message about the differences as large as a factor of four are somewhat scary and confusing. It asks for some further insight as to what exactly causes such a large difference. If you suggest the uncertainties in the inversion of the algorithms are different for different models is it because the inversion does not work perfectly or the specific algorithm does not work well for top-down inference about the emission potential (so would likely not be accurate the other way round – bottom up)? I suggest to focus in the abstract on the major points and progress, and less so much on what you did and technical detail. Specific

AR: Agreed, the abstract will be made more concise with less focus on the technical detail.

4) P5 L33 G93 “Perhaps the most widely used” – did you mean the most highly cited?

AR: I think it is the most highly cited because it is also the most used. We will change to say “...the most widely used and highly cited...”

5) P8 L4-5 Why did you leave out Langford et al. 2010 here? Misztal et al. (e.g. 2011, table 3) used approaches to estimate BER from the regression with measurements, as well as from the middle of the day (11:00 LT; which you show also here agrees well). I think you should add Langford et al., 2010 reference here, because they reported BERs as mid-day average. I would also suggest to be more neutral and refrain from subjective statements about which approaches are more popular.

AR: This reference will be added

6) Abstract. Seems long and overloaded. In particular the last two sentences are rather pessimistic and might agitate modelers unnecessarily, because it is hard to believe you could really be off by a factor of four if everything is done perfectly or at least it is not sufficiently clear why exactly this is the case.

AR: The last two sentences will be removed.

7) In the concluding remarks, you focus on the way the emission potentials are derived. Do you also want to make a bigger point about how the future models could be enhanced to better assimilate observational data at regional scale?

AR: We will stress the point that by providing a consistent and robust approach to calculating emission potentials from top-down flux measurements future models may be better parameterised through the incorporation of regional scale observations.

8) It is great that you include the original definition of emission factor (collected under the standard conditions and leaf scale). I wonder if it would be worth making a distinction between the parameterized algorithm on the full-canopy observations and whether it should be labeled as the same or a modified algorithm.

AR: We are unsure to which part of the manuscript you are specifically referring to here. Please could you provide a specific page and line number?

9) Table 1 – since PTR-TOF-MS was used in Castelporziano, why did you write vDEC? Did you artificially convert the PTRTOF dataset to disjunct to be consistent with other measurements? Either seems fine, as long as it is clear.

AR: This has been changed

10) SI S1.1 Alice Holt – Measurement setup Lag time - as the signal to noise ratio for isoprene was rather very high, why did you use the approach for low signal to noise species? Why did you not use the accurate lag-time from each half hour period?

AR: The signal-to-noise ratio for the isoprene data set was well below 10. According to Langford et al. (2015) a data set with a signal-to-noise ratio in this range and with disjunct sampling interval of 2.5 s could expect a systematic bias of around 50% (see Figures 4 and 6b). For this reason we used a prescribed lag-time as recommended by Langford et al.

11) SI S1.1 “to ensure the reduced electric field strength” seems somewhat random and out of context. Also 2.01 mbar suggests that the pressure was stable to 0.01 mbar. This is rarely the case. I suggest you say 2.0 mbar or 2.01±0.XX mbar

AR: The line you refer to simply describes that the E/N ratio was held constant at 127 Td. The E/N ratio is a fundamental parameter which should always be reported so we would be reluctant to remove this. “The PTR-MS operating conditions were held constant throughout the measurement period to ensure the reduced electric field strength (E/N, where E is the electric field strength and N is the buffer gas density) was maintained at 127 Td.”

12) SI S.1.1 P.1 L21-22 Instead of the justification it might be appropriate just to write what the consequences are (if any). I do not think it is necessarily bad to use high resolution measurement if it is appropriately post-averaged unless it leads to counting zeros. Otherwise, can you inform what the difference is between fluxes measured at 50 ms and averaged to 200 ms, as opposed to measured at 200 ms?

AR: The consequences are a lower signal-to-noise-ratio and potential systematic bias. We avoid this potential bias by using a prescribed time-lag as recommended by Langford et al. (2015). We will make this point in the revised manuscript.

13) P8 L16-30. Unfortunately, I am extremely confused by the lack of clarity here. In particular, the weighted IEP is concerning. Why do you average the activity factor across the footprints and conditions before taking the ratio? It does not seem appropriate, because, as you say, these processes are nonlinear. For example, you have to use the model to average PAR accurately. It is more intuitive to average the emission potential, because in principle it should be relatively constant for the same vegetation type (as you show in Fig. 2), and you would not have to average nonlinear processes.

AR: We apologise for the lack of clarity as we believe the reviewer may have misunderstood our approach. We are not calculating a gamma for the average meteorological conditions, but calculating the average of all gammas which were explicitly calculated for each individual flux period. Please also refer to our response to reviewer 3 where we further justify this approach. In the revised manuscript we will clarify our approach to avoid any further confusion.

14) Sect. 2.4. Isoprene deposition. Given the large gradient it is interesting that the authors suggest the deposition can be significant even for isoprene. It would be helpful to provide the percentage range of isoprene deposition relative to total flux, in addition to canopy resistance. As Alex wisely points out, you need to be aware of epiphytic microbes whose role is not yet well understood in affecting emission and uptake of isoprene.

AR: This is already stated (5-8%) in both the abstract and results sections. As we point out in the manuscript these estimates are highly dependent on the value of R_c we use, which may not be ideal for our sites but represents the only published value available in the literature. To truly understand how much isoprene is lost due to dry deposition and indeed to microbes on leaf surfaces will require further research,

but the method we outline will become increasingly meaningful as more VOC specific canopy resistances become available in the future.

15) Sect. 2.6 how do you differentiate between the effective LAI and the tree cover area fraction?

AR: Unfortunately we do not have leaf area index measurements for the individual tree species, only the tree cover fraction and hence we cannot differentiate between the two. Upscaling to 100% oak undoubtedly means that changes to the canopy LAI will occur, with the largest changes associated with sites with the lowest fractions of oak. We discuss this uncertainty in Section 2.7 and attempt to scale this relative to how much upscaling is required but recognise that without detailed information on tree species LAI our efforts are somewhat arbitrary.

16) P10 L28 As you did not calibrate isoprene on gas standard at Alice Holt, you had to estimate the concentration from relative transmission. I am generally fine with the approach, but it should be clear in the text whether you have accounted for isoprene fragmentation (mostly to m/z 41) because as you probably know isoprene sensitivity is significantly deviating from transmission estimate vs non-fragmenting species (e.g. MVK). Not accounting for fragmentation would result in underestimating the concentrations but perhaps you derived a fragmentation correction factor for proton reaction rate constant (effective k) in the post-campaign calibration? In either case it is not clear so you should add appropriate detail to SI.

AR: The reviewer is correct about the fragmentation of isoprene to m/z 41 and this was already accounted for in the reaction rate constant k used in our transmission. For completeness, we now include a description of the correction applied.

17) Sect. 3.2 Figure 2a,c is incredibly super cool, and the diurnal emission potential seems relatively constant as expected, except for the morning and evening times. Did you try to filter for low u -star to see how this would affect the diel trend? Maybe you could plot the low u -star data in grey. Do you know why you could not reproduce this stability with G93 as beautifully as with G12?

AR: We are glad the reviewer enjoyed this figure. The fact that the G12 algorithm is able to much better replicate the measured fluxes, even during the evening and morning periods means that low turbulence is not the reasoning for the comparatively poor performance of the G93 algorithm. The G93 is unable to replicate the diurnal pattern because it uses the big-leaf approach and therefore cannot properly capture the effects of light and temperature attenuation through the canopy.

18) Figure 7. This is also an incredibly beautiful figure. In particular the temperature activity factor works shockingly well. In panel a, it might also be useful to add the parameterized G06 line which would better fit the γ for PAR. It would be nice to further discuss these differences because they show major results from this study

AR: The purpose of this figure is to demonstrate why the G06 algorithm generates emission potentials that are much lower than the other algorithms. We feel it is critical to highlight the problems with this approach because it is becoming more widely used, including by ourselves in the past. We feel that adding the PCEEA approach to this figure, an algorithm that was more consistent with the emission potentials calculated using MEGAN 2.1, may dilute our message.

Technical:

19) G93, G95, G06, G12 need to be defined on their first use and used consistently (e.g. G93 in the abstract).

AR: This will be changed

20) add page numbers in SI

AR: Page numbers will be added

21) Sect. 2.1.1-2.1.5 Significant figures in the coordinates of the locations vary from 3 to 10. Please be consistent.

AR: Changed to 3 SF at each site

22) P6 L13 the unit of the gas constant seems incorrect. Maybe a typo or maybe you intended to refer to 1 mole.

AR: This will be corrected.

Response to Anonymous Reviewer 3

Overall, this is a nice paper that explores a technical aspect of isoprene emission modeling: relating whole-system, measured isoprene fluxes to the emissions capacity used in most isoprene emission frameworks. My biggest concern is that the authors recommend using the means of observations and of the calculated gamma to find the emission capacity (equation 6). On page 13, line 2, the authors' state that the superiority of this technique has been established in the previous results section. Since the least-squares approach has a well-established theoretical justification, the manuscript should do more to explore the advantages of Equation 6. This must be a pretty common issue in modelling. For example, how do ecosystem models of net primary productivity deal with this issue? I think the authors could do more to justify this new approach.

AR: I'm not sure we agree with the reviewer on this point. In the context of calculating emission potentials from eddy covariance flux measurements the least squares approach has been used but to the best of our knowledge its use has never been explicitly justified. Indeed, the lack of justification was the partial

motivation for this work. We would stress that the specific approach you take to calculate your “average” emission potential should depend on your proposed use of the BVOC model. In our manuscript we address this problem from the perspective of accounting, with the aim of producing an emission potential that allows us to properly simulate average or total emissions from a given forest over a given time period. We present eddy covariance flux measurements which we carefully correct for the effects of chemical flux divergence and dry deposition and therefore assume to represent the “best estimate” average emission from the site. Having established this “best estimate” average emission rate we can now use this to first back out an emission potential and secondly to challenge the model (combined with new emission potential). In practice this is no different from using a branch enclosure to measure the isoprene flux and then scaling to standard conditions using the algorithm to estimate the emission potential. Using this approach we have systematically evaluated various techniques for deriving a single “average” emission potential from a time series of flux measurements including through the LSR approach. As the reviewer suggests, the least squares approach has a well-established theoretical justification but only if a number of assumptions are fulfilled, two of which are that the data show a linear relationship and that the residuals are normally distributed. Figure 4 shows a plot of the measured isoprene flux versus the G93 γ term for the Bosco Fontana measurement site. It is clear that (i) the relationship is non-linear, driven by the algorithms inability to account for the attenuation of light and temperature through the canopy particularly during the periods either side of midday and (ii) the residuals are not normally-distributed. These two factors mean that the application of the LSR approach would be inappropriate. Indeed, our analysis shows that when the LSR method is used to estimate the “average” emission potential, then the algorithm subsequently fails to replicate the average observed flux. In contrast, adopting the “weighted average” approach ensures an emission potential with zero bias.

We will include this figure with short explanation in a revised version of the Supplementary Information and will emphasise our justification for choosing the weighted average method in the revised manuscript.

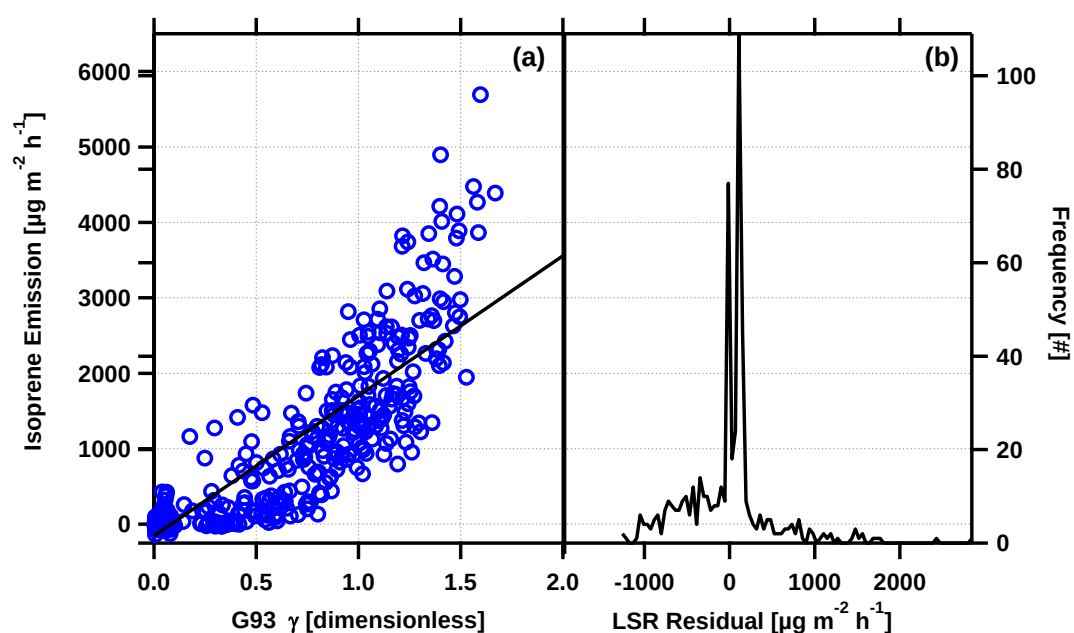


Figure 4. Plot showing (a) the non-linearity in fluxes modelled using the G93 emission algorithm when compared with observations and (b) the distribution of residuals from a least square regression fit.

Major comments

*Figures 2 and 3 pack in too much information. For example, I was interested in comparing the performance of the LSR & ODR approaches with MEGAN. In most cases in Figure 3, I could not distinguish these two cases because of overlapping plotting characters. What's the benefit of plotting all the different time average periods? Couldn't that be conveyed in a separate graph? Near lines 28-31 on page 12, you take away from Figure 3 that the G93 approach difference significantly from the MEGAN approach. This is well known, and could be conveyed more succinctly in a separate figure.

AR: We will replot figure 3 so the points at each site are staggered horizontally to ensure all symbols are visible to the reader. In figure 2, where lines are masked by others we will change the covering lines to dashes.

*The conclusion that "the emission potential is not constant throughout the day" should be refined. Within the modeling framework, the emission potential should be a constant throughout the day. The better way to frame this is that the calculated emission potential is not properly capturing the diurnal cycle. Also, considering just 08:00 to 18:00, there's not much variation in the EIP.

AR: You are correct in the case of Figure 2, but the MEGAN algorithm didn't always perform so well. For example, Figures S1 and S4 showed much greater variation at the Alice Holt and O3HP sites respectively.

*On lines 9-12, page 8, you mention the issue of the intercept for the least-squares approach. For the least-squares calculations in this paper, did you use a zero intercept?

AR: No, in each case we did not force the intercept through zero as we felt this gives the regression only one degree of freedom. However, looking into this further we found that in most cases setting the intercept to zero only resulted in a very minor change to the calculated emission potential.

Minor comments

*The abstract is a bit long. While comprehensive, I counted 660 words. In particular, some of the recommendations at the end repeat material from the abstract (factor of four). A target of 600 words seems more reasonable. With an open-access journal, there is less pressure on fitting so much in the abstract.

AR: Agreed, we shorten the text, primarily through the removal of the last two sentences (which reviewer 2 did not approve of) and look to refine the text.

Page 2, lines 33-34: The article by Arneth would be useful to consider and site at this point in the discussion (<http://www.atmos-chem-phys.net/8/4605/2008/>).

AR: reference will be added

Page 2, lines 34-36: Very minor point: branch enclosure measurements typically can't be performed at standard conditions. Instead, leaf temperature and light are measured, and often the Guenther algorithms are applied to derive a basal rate.

AR: We will change this to just refer to leaf-level measurements

Page 3, lines 5-7: Again, a good place to refer to Arneth et al 2008.

AR: Reference will be added

Page 3, line 22: Inconsistencies isn't the right notion here. Yes, there are inconsistencies, but there are also different assumptions.

AR: This will be changed to "...inconsistencies and differences in the underlying assumptions..."

Page 4, line 1: Since the algorithms for previous light and temperature are coming to come into play, some mention of the meteorological conditions during the campaigns compared to average climatology is necessary. In particular, where any of the campaigns conducted during times of water stress?

AR: Agreed, where available we will add further details about the meteorological conditions at each site.

Page 7, lines 21-32: This is a lot of text to describe something that wasn't used. Please consider if its necessary to include.

AR: Although the PCEEA method was not shown in Figures 2 and 3, we do use it to derive emission potentials and the results are shown in Figures 4-6 and in the tables of emission potentials listed in the Supplementary Information. We therefore believe the brief description of the algorithm is merited.

Page 8, line 25: Shouldn't this produce the same result as a linear regression with the intercept set to 0?

AR: No, this is not the case because the datasets are never perfectly linear.

Page 11, lines 16-18: “discernable” is subjective. This might be a real effect, or it might be random noise. Also, connect this to the major comment above: this variation represents a failure in the underlying model. Lines 26-27 (page 11) are the proper way to frame this conclusion.

AR: We will remove the term “discernible”

Page 12, lines 36-41: Yes, but this is only true when considering the extreme ends of the day. Typically, the focus is 10:00 – 16:00, when the variability is much lower with MEGAN.

AR: You are correct in the case of Figure 2, but the MEGAN algorithm didn’t always perform so well. For example, Figures S1 and S4 showed much greater variation at the Alice Holt and O3HP sites respectively, even within your suggested window of 10:00-16:00.

Technical comments

Page 1, line 34: hyphenate ‘site specific’

AR: Done

Page 2, line 18: hyphenate ‘ground level’

AR: Done

Page 3, line 39: note explicitly this is ug of isoprene, not carbon (ugC), which has also been used in the past.

AR: This will be changed to “...μg of isoprene m⁻² h⁻¹...”

Page 4, line 13: be consistent about lat/long significant figures. The two used elsewhere are probably sufficient.

AR: Changed

Page 4, line 23: According to BG style, “32m platform”.

AR: Changed

Page 7, line 7: hyphenate “in canopy”

AR: Done

Page 10, lines 6-7: fix grammar

AR: Done

Page 10, line 13: reflect should be reflects

AR: Changed