acp-2017-363: Ethene, propene, butene and isoprene emissions from a ponderosa pine forest measured by Relaxed Eddy Accumulation

RESPONSE TO REVIEWER #2 (Jochen Rudolph)

Dr. Rudolph provided multiple suggestions on how to clarify and further improve the manuscript, and we address each of them below. We appreciate Dr. Rudolph for his detailed and thorough (and positive) review, and the revised manuscript incorporates improvements in readability, clarity, statistical analysis and emphasis of the study impact. Reviewer comments appear in italics, and our responses are in bold

The paper presents studies of emissions of light alkenes from a pine forest. Since data on emissions of light alkenes from vegetation are rare and existing emission data have substantial uncertainty the presented information creates new, useful insight into the role of vegetation as source of volatile organic compounds (VOC), a subject that is highly relevant for ACP. The emission studies are based on state of the art methodology, Relaxed Eddy Accumulation (REA). The results are of very high quality, the methodology is clearly explained and overall the discussion is sound and the conclusions justified. The paper is well structured and written, the figures and tables overall of very good quality. Consequently, the paper merits publication in ACP, although I have a few suggestions for changes and additions to improve the paper. The two main points are:

i) Gap-filling: I agree that gap-filling can provide a better estimate for averaged monthly or daily fluxes. However, gap-filling, independent of the interpolation procedure, has serious limitations.
- It cannot compensate for bias in the data set resulting from experimental limitations, for example flux data below the DL will be absent from the measured data and gap-filling cannot compensate for this. Indeed, in Figure 6 it seems that for ethene and propene the gap-filled data are always, even at night, above zero. In Figure 9 the PAR dependence of the fluxes predicts that the fluxes are zero for PAR=0.

This is correct: the ANN predictions are > 0 for PAR = 0. We will discuss later how the PAR dependent equation is not the optimal equation to use to determine fluxes, owing in part to this noted shortcoming.

- When using interpolated data, it is very difficult to derive meaningful statistical criteria. I assume that ± in Table 1 indicates the standard deviation σ . I am not sure how to interpret a standard deviation for an interpolated data set. Based on basic statistics the error of the mean can be calculated from σ and the number of data points N: Error of mean = σ /sqrt(N-1). Since there is no limit in the number of interpolated data points this would imply that the error of the mean interpolated flux is effectively zero. Strictly speaking the modeled flux is a calculated value and the error of a calculated value can be determined using (Gaussian) error propagation, which is close to impossible for an ANN interpolated data set.

That is also correct: Table 1 showed the average flux ± sd. The ANN-derived fluxes (i.e., gap-filled results) were compared against the validated observed fluxes to assess if there was any bias associated with the observational gaps. To do this, hourly fluxes were extracted from the ANN model from days 174 to 225 (n=1223). The mean and standard deviation were derived from these values. In light of comments below, we have replaced Table 1 means and standard deviations with the median and 10th/90th percentile ranges. In case readers still wish to see the average ± sd, we moved these values to the Supplementary Material, Table S2.

acp-2017-363: Ethene, propene, butene and isoprene emissions from a ponderosa pine forest measured by Relaxed Eddy Accumulation

- Independent of the interpolation procedure, the results of the interpolation can be biased by the assumptions required for the interpolation. Here the situation is especially complex due to the two step interpolation procedure. It is stated that for the ANN gap-filling "inputs were normalized". This needs more explanation.

The reviewer is referring to Section 3.5 "Prior to gap filling, inputs were normalized and gap-filled with average values from the surrounding days..." Specifically, inputs were scaled to values from -1 (for the minimum observation) to +1 (for maximum observation). This procedure is commonly used in machine learning, when input variables have different scales/units. The equation used was: $y = (y_{max}-y_{min})*(x-x_{min})/(x_{max}-x_{min}) + y_{min}$. The normalization procedure is clarified in the revised manuscript in Section 3.5 as follows: "Prior to gap filling, input variables were normalized on a scale of -1 (for minimum value) to +1 (for maximum value)."

Moreover, from page 9, line 17-18, it seems that the ANN input set was created by interpolation of measured data, although no detail of the interpolation procedure is given.

For the ANN algorithms, explanatory variables are required to be continuous (i.e., without gaps). The few time periods when input drivers were missing were programmed to be filled with average values from the surrounding days (i.e, the same time on the next day) or through a simple 2-D interpolation. It turns out that there were no gaps in PAR and only 1% missing values (n=16 of 1223) for H_2O flux, temperature and std deviation of wind speed. Thus, we removed this line to avoid unnecessary confusion. (The entire input variable dataset had 2472 observations each, but included time periods outside the REA flux measurement period, so that also was adjusted accordingly).

An interesting scientific question is: Do the ANN based fluxes provide a better fit to the measured fluxes than a calculation based on the T and PAR dependence developed in 5.5 (Figure 9). This could provide evidence that modeling alkene fluxes using T and PAR alone is insufficient.

See response to reviewer 1. To reiterate: we explored this by predicting alkene fluxes in 5 different ways: a) only eq. 4 (PAR); b) only eq. 5 (temperature, light independent); c) only eq. 6 (T, light dependent); d) the weighted (according to the LDF in MEGAN) combination of eq 5 / 6; and e) our initial gap-filling method (Artificial Neural Networks). Here the statistics are reported as root mean square error (RMSE) and Pearson's correlation coefficient (ρ) of comparison between predicted and measured fluxes

	ethene		propene		butene		isoprene	
	RMSE	ρ	RMSE	ρ	RMSE	ρ	RMSE	ρ
F(PAR)	46.5	0.68	42.2	0.69	16.3	0.7	41.4	0.54
F(T_LIDF)	48.4	0.72	46	0.68	16.8	0.64	44.3	0.48
F(T_LDF)	45.7	0.72	48	0.69	16.9	0.65	41.2	0.56
F(T_MEGAN)	45.5	0.73	43.1	0.69	16.5	0.68	41.2	0.56
ANN	32.1	0.83	27.7	0.8	8.6	0.8	28.6	0.64

According to these results, the ANN-based fluxes provide a superior fit to the measured fluxes overall than the calculations based on Temperature (LIDF or LDF) and PAR. However, the combined T and PAR results perform almost as well. We see high potential in ANN as a gap-filling tool. We now include the text: "ANN is increasingly used in eddy covariance studies because of its ability to resolve non-linear relationships and complex interactions between flux drivers (Dengel et al., 2013, Papale and Valentini, 2003)"

That being said, we do believe that a more mechanistic oriented model (such as MEGAN) is the better choice to predict data for sites or times where no measurements are available (which is necessary for training in the machine learning approach). Both approaches have their relevance and we are applying both where appropriate.

Additional references added to manuscript:
Dengel et al 2013: Testing the applicability of neural networks as a gap-filling method using CH4 flux data from high latitude wetlandsBiogeosciences. doi:10.5194/bg-10 8185-2013
Papale and Valentini 2003: A new assessment of European forests carbon exchanges by eddy fluxes and artificial neural network spatialization. Global Change Biology. (9) 525-535

ii) Statistics:
- The authors should provide a more detailed evaluation of the significance of the findings. For example, there are no error bars for the binned data in Figure 9. There are no uncertainties for the fitted parameters in Table 2. Based on the scatter of binned data in Figure 9 (especially for the PAR dependence) I would expect substantial uncertainty. This is mentioned in the discussion, but still does not allow an estimate of uncertainty for fluxes calculated from the models.

We have calculated 10-90th percentiles to the binned fluxes in figure 9, now illustrated on the plots. We also provide 90[%] confidence bounds for the fitting coefficients of Equation 4 in Table 2 and of Equation 5 in Tables 3 and S4.

It seems from Figure 6 that the night-time fluxes for ethene and propene are above zero, which is in contrast to the PAR dependence derived in 5.5, the fit parameters given in Table 2 and the fitted functions in Figure 9. Is this a real discrepancy between measured and modeled fluxes (which would indicate shortcomings in the modeled fluxes) or can this be explained by the uncertainty of averages, fits, and binned data, or is this a result of gap-filling?

Our measurements do indicate small emissions of propene and ethene for PAR = 0 ($\approx$ 20 µg m^{-2} hr^{-1}), with the caveat that only a few of these emission rates exceed flux detection limits (n < 10). This contrasts the PAR dependent equation, in which the calculated flux goes to zero as PAR goes to zero (i.e., PAR is in the numerator). The light and temperature response curves have only been fitted to measured REA flux data, without gap filling the dataset. Thus this discrepancy appears to be a shortcoming in the PAR-dependent equation, not a consequence of gap-filling or binning results.

There is other evidence of nighttime positive fluxes for the light alkenes and other BVOCs. The diurnally averaged fluxes of light alkenes at Harvard Forest (Goldstein *et al*.

1996), the only directly comparable net ecosystem flux measurements, suggest that most fluxes were positive, even around midnight, which is in line with our results. Non-zero, positive fluxes around midnight are also apparent in fluxes for methanol and ethanol (Schade and Goldstein, 2001) and monoterpenes (Kaser *et al.*, 2013). [All references in the original manuscript].

These results suggest that it may be necessary to add a baseline term (ε) to eq. 5 (PAR response curve) for several BVOCs, to account for non-zero fluxes at PAR = 0, such as:

$$F(PAR) = \varepsilon + \left[\frac{\gamma\, C_{L1}\, PAR}{\sqrt{1 + \gamma^2 PAR^2}}\right]$$

However, adjusting the PAR model equation is outside the scope of the current manuscript, and we do not feel this is particularly useful for this study, as the temperature dependent equations appear to address the shortcomings of the PAR dependent equation.

- On Page 13, lines 27-29 it is mentioned that "The understory REA measurements showed detectable consumption overall for ethene, propene and butane...". However, due to the large uncertainty of the negative fluxes and the small number of data points I am not sure if these fluxes are below zero at a meaningful significance level.

The phrase "detectable consumption" means that the uptake rates for non-flagged REA measurements were larger than detection limits. We acknowledge that conclusions based on the understory fluxes have larger uncertainties than the above canopy REA measurements owing to the relatively sparse data and possible violations of Monin-Obukhov similarity theory (which we tried to minimize to our best abilities, see Section 3.4 and Fig S2). That being said, it does appear that the uptake rates for ethene, propene, and maybe butene are indeed significant, at least for this one day. We have now added Figure S3 to the supplementary information.

- From Figure 1S it is evident that the probability distribution of measured fluxes cannot be described by a Normal Distribution as implied by the mean values and standard deviations in Table 1. In this case non-parametric statistics (median, percentiles) provide a more realistic insight into the actual distributions than mean and standard deviation or error of mean.

We have changed the flux statistics in Table 1 to report the median and 10th-90th percentiles. Also in the results section 4, the text has been changed to report medians and the inter-quartile (25th and 75th percentile) ranges, rather than mean ± std.

- Correlations: Most of the correlations and linear regressions are for data sets with substantial seemingly random variability of y and x values. Standard linear regression does not consider uncertainty for the independent variable and an arbitrary identification of the independent variable may result in biased estimates of slope, intercept or R^2. A better indicator for the quality of a correlation is the Pearson product moment correlation coefficient, which does not require distinction between a dependent and independent variable. I also think that correlations with R2 values of 0.5 or less do not qualify as good or high correlations.

acp-2017-363: Ethene, propene, butene and isoprene emissions from a ponderosa pine forest measured by Relaxed Eddy Accumulation

We agree with these points. We re-calculated the correlations using the Pearson's correlation coefficient and updated figure 7 to indicate this.

Figure 7 only shows the slopes (without errors) of the linear regressions, not intercepts.

We have also updated figure 7 to show the resulting slopes and intercepts for plots where the Pearson's correlation coefficient >0.5. In addition, the 90% confidence bounds for slopes and intercepts can now be found as a table in the updated supplementary material (Table S2).

Are the axis intercepts zero (within their uncertainty) or was a regression used that did not allow a non-zero axis intercept?

Intercepts are not forced to be zero. Also see comment above.

Some details:
Introduction: This is a good, detailed overview of the role and sources of light alkenes in the atmosphere with emphasis on vegetation. I understand that the different estimates etc. represent the uncertainty in current emission inventories and differences between different studies. However, it is not easy for the reader to extract an overall perspective of the role of alkene emissions from vegetation in comparison to other sources of light alkenes or emissions of other VOC from vegetation. A table (or graph) summarising the emission rates of light alkenes from various sources would be very useful for the reader and make it easier for the reader to understand the potential importance of reducing uncertainties in emission rates of light alkenes from vegetation.

Such a table exists in Poisson et al., 2000, and we think this may be suitable for a future manuscript with various model results.

Page 5, line 5: "with a lapse between ... " should be rephrased (gap?).

Changed to "gap"

Page 6, line 19 and other occurrences: "sonic temperature" should be rephrased.

We chose to retain the term 'sonic temperature', which we consider an accepted terminology for the temperature measured by the sonic anemometer.

Page 8, line 3 and S, page 5, line 20-30: I accept that, due to the absence of an extreme change between consecutive measurements, alternating the sequence of analysis between the "up bags" and "down bags" allows excluding that storage has a dominating impact on measured fluxes. However, due to the substantial variability of the measured fluxes and the use of two sets of bags, it is not obvious that the flux measurements are entirely free from impact by storage. The authors should present some quantitative estimate at which level of "storage bias" the mentioned "seesaw" pattern (Suppl. pages 5, line 27-28) would clearly be visible.

We thank the referee for encouraging further scrutiny of this statement. It turns out that the seesaw pattern we described in the supplementary text was neither systematic (i.e.,

the pattern only occurred occasionally) nor consistent (i.e., the fluctuations when they did appear were not always at 1 hour intervals). Thus, this precautionary statement was a bit overstated. We conclude that there is not a systematic storage issue but that fluctuations in the sunrise and sunset transitions were natural fluctuations in the flux across the canopy. We have adjusted the text in the Supplementary Information as follows: "Seesaw patterns were observed occasionally during the sunrise and sunset transitions, but they were neither systematic (i.e., did not occur regularly) nor consistent (i.e., closer examination shows that fluctuations were not necessarily hourly). In addition, these are periods when ozone concentrations were expected to be low reducing their importance in terms of storage issues. Even under these conditions, negative fluxes were generally not observed."

Page 8, line 15-25, Fig. S1: Quality control, detection limits and flagged data. "Quality control for each hourly REA flux measurement was checked against eight potential flags associated with the sample volumes, meteorological conditions or footprint analysis". There is an explanation in the supplement. However, since 47% of used data were flagged for not strictly meeting all criteria some quantitative estimate should be given to which extent this may impact data quality. Data flagged with more than 4 flags or data not meeting specific footprint criteria are excluded. I am not sure about the rational for this specific threshold.

We acknowledge that the flagging methodology could use further clarification. We have adjusted the text to reflect that we used 3 tests. The first test was a turbulence assessment, following best practice guides by Foken and Wichura (1996), and Mauder and Foken (2004). Any data, from 30-min intervals with poor turbulence characteristics was critically marked as "bad quality" and discarded from measurements.
The second test involved applying a total of 5 REA apparatus specific flags. These were listed in the supplementary material as "REA flags" (QC 1-5, now REA flags a-e). Data getting flagged for 1 or 2 single REA QC check remained in the dataset, flagged as 'medium quality'. If 3 or more REA QC tests were flagged, this was deemed a critical failure, but this only happened sporadically.
A third test on homogenous flux footprints was employed (see section 3.7 in the manuscript). If analyses indicated an inhomogeneous flux footprint (e.g., due to the crossing of a road east of the tower), data were non-critically flagged. The employed footprint analysis (Hsieh *et al.*, 2000) should only be treated as a proxy/indication of the actual size and spatial distribution of sources and sinks. Hence, we do not believe that this method should be used as a critical flag. In fact, the distance from the tower to the maximum contributing source area (eq. 19, Hsieh *et al.*, 2000) is much smaller (by a factor ≥ 10) than the boundaries (90% flux footprint) plotted in figure 3. Regardless of wind direction, the maximum contributing source area lies well within the ponderosa pine forest.

However, since 47% of used data were flagged for not strictly meeting all criteria some quantitative estimate should be given to which extent this may impact data quality.

We have now clarified that 13% of the data were critically flagged and removed, 18% were flagged as medium quality but not removed, and 16% were flagged for footprint issues but are not likely to be problematic. Quantifying the impact that specific flags have on overall data quality is not straightforward. However, it is clear from Figure S1 that a

majority of the flags occur when the measured flux is near zero. Flagged (but accepted) data do not manifest themselves as outliers from the surrounding results; hence, most uncertainties introduced by these flagged data relate to small fluxes, in the absolute sense. Nevertheless, flagging data is an important data quality assessment tool for comparative purposes used in international communities (e.g., FluxNet). Generally, for seasonal and annual timescales, flagged data is treated as acceptable data.

Moreover, it seems that a large portion of very low fluxes were flagged or failed QC. How does this impact the overall averages and representativeness of the flux data?

We added the following sentence to section 4.2. to address this comment: "Gap filling REA fluxes (Fig. 6) using artificial neural networks (i.e., modeled results) generally increased median emissions (Table 1); however, differences between groups of modeled and observed fluxes were non-significant (anova, α = 0.05) suggesting that the selectivity of quality controlled measurements might lead to only a minor under-prediction of diurnal averages."

How were flux data below the lowest flux detection limit treated in further data evaluation? The way data below the LDL are used is important since this may create bias for averages, fits, modeling etc.

Flux observations < LDL were discarded for statistical analysis (median, IQR, percentiles) and for any fitting/response analysis. This is now mentioned in the revised manuscript in section 3.3: "Flux observations below F_{min} were excluded from overall statistical analyses (median and percentiles) and for the curve fitting in response to temperature and PAR."

It is also mentioned that some fluxes were negative. How were they used in the data set? From Figure 7 it seems that there are no negative fluxes, but it is mentioned that a small number of fluxes which met QC are negative.

The frequency of occurrence and magnitude of deposition events are described in the text (Section 4.2). For ethene, 1.3% of quality ensured observations indicated deposition, for propene 2.3%, butene 1.3%, and only for isoprene a larger amount of data has a negative sign at 13%. In Figure 7, we excluded the negative outliers for ethene, propene and butene, and this is now indicated in the caption: "Negative fluxes for the light alkenes (1.3% to 2.3% of the light alkene fluxes) are excluded from the plot and the regression statistics." Also, the temperature response equation (Eq. 4, now Eq 5 in the revised manuscript) does not allow negative predictions. Hence a comparison between modeled and observed fluxes was only made with observed emissions (i.e., positive fluxes). If the few observed deposition events are included in the dataset, median fluxes decrease slightly, for ethene by 1.1, propene by 2.9, and butene by 1.3 (μg m^{-2} h^{-1}). No significant difference was found between datasets including negative fluxes as compared to the dataset without negative fluxes. We now include the text: "Negative fluxes were too infrequent and small to be captured in ANN model predictions, and are excluded in the comparison of results (Table 1). Their inclusion would decrease median observed fluxes by 1.1, 2.9, and 1.3 μg m^{-2} h^{-1} for ethene, propene and butene, respectively."

Page 8, 9, S2, Understory REA fluxes: I understand that measuring the understory flux is very difficult, especially since it seems that the flux is small. Nevertheless, the very small fraction of measurements that passed QC raises the question of representativeness of the few good quality data. Since the main result of this study is that understory fluxes are (on average) small compared to the total flux, rejected data which could be used to provide meaningful upper limits of understory fluxes may strengthen this point.

A total of 6 out of 10 alkene fluxes for understory REA measurements were conducted during acceptable turbulence conditions. We considered the point about including the 4 rejected time points. Including these data would support the main message of how the understory could not account for the light alkene emissions observed above canopy and is also consistent with the isoprene emissions.

For ethene, the rejected understory flux results were relatively small (-9.7, -0.8, -5.6 and +3.8 μg m^{-2} h^{-1}) and comparable to the quality-ensured understory measurements reported in Table 1. For propene, two of the rejected fluxes were similar to reported values, but two were large and negative: -98.4, -2.4, -2.6, and -54.4 μg m^{-2} h^{-1}. For butene, all four rejected fluxes are comparable to the reported fluxes: -10.4, -1.8, -7.0, -8.1 μg m^{-2} h^{-1}. For isoprene, the rejected fluxes are small in the morning (5.7, -2.2 and 5.1 μg m^{-2} h^{-1}) while the one rejected flux in the p.m. is very large (266 μg m^{-2} h^{-1}). We have now included the complete set of results in Figure S3 to illustrate that the rejected data are consistent with the overall understory results, and that including them would not yield meaningful improvements to the understory flux statistics.

5.2 and 5.3: Both subchapters present a comparison between literature and this study. Combining both chapters would allow a more consistent comparison. For example, in 5.2 the alkene fluxes are placed in context to CO2 and other VOC fluxes. However, in 5.3 this context is not considered. Yet, it would be very important to understand if some of the differences in alkene fluxes maybe related to differences in assimilation or respiration rates and so on.

We appreciate the motivation to understand what ecophysiological controls are driving emissions. The purpose of having two sections was in order to clearly view our results in the context of a) the prior extensive research conducted on BVOC fluxes at this very site in previous campaigns, and b) the prior but sparse literature on biogenic light alkene emissions from terrestrial ecosystems. We believe that the important issues raised by the reviewer (such as the relation of alkene emissions to assimilation, respiration rates, etc.) are best addressed in further biogeochemical studies that can better control for these variables, such as leaf or branch chamber type measurements. Doing so here may be too speculative.

Page 16, lines 25-27: "We utilized fluxes instead of concentrations to provide a mea-sure of OH reactivity that is independent of elevated concentrations associated with pollution events and more representative of site specific sources." This creates a distorted view of the importance of emissions. The atmospheric OH-reactivity of a VOC at a given concentration ([VOC]) is determined by [VOC] kOH, as mentioned in the previous sentence and the cited literature. In a simplified steady state [VOC] is proportional to the flux and inversely

proportional the reactivity (kOH). Consequently, the overall relative impact of an emission is simply determined by the emission flux.
The reactivity determines the temporal and spatial scales at which this happens. It should also be considered that alkenes also react with ozone, OH reactivity therefore only determines one part of the overall alkene reactivity.

After observing the substantial emissions of these compounds, two questions immediately crop up: where are these emissions coming from, and what importance do these emissions have in the atmosphere? We probed the first question by looking at temporal trends, correlations, understory fluxes and environmental controls. We probed the second question by comparing the magnitude of light alkene fluxes multiplied by their OH reactivity. All of the compounds shown in Figure 8 have high OH reactivity but are also long enough lived to measure their fluxes. Using fluxes instead of concentrations allows us to compare our results with different compounds measured at the site from prior years in other studies. While intra-canopy OH reactivity may be dominated by some even shorter lived species, this comparison may provide a picture of the OH reactivity in the atmosphere above the canopy. The point of this exercise is to demonstrate that the light alkene emissions may have a significant impact on the local atmospheric oxidation capacity. Ozone reactivity could also be illustrated, but we did not feel that it was necessary to demonstrate this point. Future modeling studies may tackle this question.

Page 17, lines 11-12: Based on the uncertainties of averaged fluxes, extrapolation etc. I am not certain that a 20-60% difference is really significant. I also do not understand why the estimate of the seasonal flux average is based on such a simple extrapolation. The paper presents two more detailed models (gap-filling based on ANN and the T and PAR dependence presented in 5.4) which can be used to calculate averaged fluxes for comparison.

In order to compare our results to the one other study of net ecosystem fluxes of light alkenes (Harvard Forest, Goldstein et al., 1996), we needed to extrapolate our July-August fluxes results to cover their integrated period of study from June 1-Oct 31. Our simple extrapolation essentially assumed that the shoulder months of June and October had emission rates that were half of the mid-season months (i.e., linearly changing from (and to) zero over the course of the shoulder months). The choice of a simple extrapolation in section 5.3 was because: a) we were not yet convinced of the superiority of any specific model, and b) we did not have a complete season of temperature or PAR measurements from the MEFO tower from which to extrapolate model results (the sonic anemometer was put up early May and removed after the August measurements). In addition, ANN is a gap-filling model and was not intended to be used as an extrapolation model.

In the course of these revisions, we have developed a further understanding and appreciation of the temperature and PAR based model fits. With the assumption that they can represent shoulder-season fluxes, we now apply the temperature based models to the shoulder season. To make up for the gap in temperature data at Manitou Forest, we have downloaded weather data for a nearby weather station in Colorado Springs and applied a linear regression with concurrent MEFO measurements to adjust temperatures. The correction factors were then applied to the temperature data to come up with MEFO

temperature for the entire season. The results demonstrate that a more sophisticated extrapolation of results yields a 40-80% higher emission rate (compared to the 20-60% higher rate from the simple extrapolation). We now include in section 5.5: "The choice of which temperature dependent flux response equation to apply varies among different compounds and different studies, as illustrated in Table 3. In our study, both the light dependent fraction (LDF) and the light-independent fraction (LIDF) equations for temperature response performed better than the PAR response curve. In addition, the PAR response curve goes to zero as PAR goes to zero, and it appears that emissions of light alkenes occurred at nighttime when PAR equaled zero. We therefore utilized a combination of the temperature-based equations, scaled by the LDF reported in the MEGAN 2.1 model, to extrapolate flux results to the remainder of the season for which flux measurements were not determined. Between May 1 and October 31, 2014, the extrapolated seasonal flux yielded an average of 61.5, 51.7, 24.3, and 22.9 $\mu g\ m^{-2}\ h^{-1}$ for ethene, propene, butene and isoprene, respectively. For the light alkenes, this represents a 40-80% higher emission rate than that observed over the same season length at Harvard Forest (Goldstein et al., 1996). This is slightly larger than the simple linear extrapolation described in section 5.3 above and almost identical for isoprene."

Combining subchapters 5.4 and 5.5 would allow streamlining the discussion of flux parametrizations and model predictions, for example how well MEGAN 2.1 parame-terizations agree with the measured alkene (and not only isoprene) fluxes. Based on the current discussion it is not evident that (and why) "Modifying the light and temperature parameterizations for light alkenes in the vegetation emissions model will lead to a corresponding increase in estimated global emissions for these compounds".

We chose to keep these sections separate in order to distinguish the modeled curve fits in section 5.4 to their application in the separate MEGAN 2.1 model in section 5.5. With the addition of the information above in section 5.5, it allows for easier reading.

Conclusions: They are more or less a summary. I agree with the overall conclusions that the current understanding of alkene emissions from vegetation is insufficient and that alkene emissions from vegetation can be relevant for the chemistry of the atmosphere. However, based on the many interesting aspects presented and discussed in the paper I am a bit disappointed that the suggestions for tackling open questions and reducing uncertainties is basically a generic "more research needed" approach. Specifically, the very good correlation between ethene and propene fluxes is striking and raises questions about origin and the factors determining the emissions of these two alkenes.

We have edited the conclusion paragraph to make it more punchy, with an emphasis on the correlation between ethene and propene and what it means in terms of mutual production mechanisms. We agree that there were a lot of interesting things to consider in our results, and we appreciate the encouragement to express this more clearly.