The authors would like to thank all three referees for their time and work in the preparation of this constructive comments and suggestions. All comments are addressed below. The original reviewer comments are shown in italics, while our responses are shown in bold text.

Anonymous Referee #1

General comments

This paper by Kaser et al. presents results from a 30-day measurement campaign at a ponderosa pine forest. Volatile organic compound (VOC) emissions were measured using the eddy covariance (EC) technique combined with proton transfer reaction timeof-flight mass spectrometry (PTR-TOF-MS). During the campaign, a severe hailstorm damaged pine needles and branches at the site. Effects of this physical damage on the emissions were determined on the canopy scale. In addition, VOC emissions derived from the model of emissions of gases and aerosols from nature (MEGAN 2.1) were compared with the direct EC emission measurements. The measurement methods are state-of-the-art and the presentation quality in the paper is good. Only uncertainty estimates are often missing. They would help the reader in assessing the statistical significance of the results and should be inserted in the text, tables or figures (wherever possible). In summary, the paper is well suited to ACP. It contains new information on undisturbed and disturbed canopy scale VOC emissions which will be very useful when developing emission algorithms. I recommend publication after the minor revisions listed below.

Specific comments

P15336, L4–7 (P15350, L14–19 and P15351, L7–14): Is the underestimation of monthly monoterpene emissions always around 40 %? It might be good to rephrase the sentence in the abstract since only one hailstorm event was analysed and the variation is still unknown. How often do such heavy storms occur at the site or the nearby area? Their frequency might give information on their contribution to annual emissions.

The sentence in the abstract was rephrased to make it clearer that the 40% underestimation of monthly monoterpene emissions refers to this measured hailstorm. From Schaefer et al. (2004) it can be estimated that roughly 200-300 hailstorms per decade are reported in this area of the U.S.. Which would give an annual amount of hailstorm of 20-30 but there is no information on the severity of the hailstorms therefore estimating an annual contribution of monoterpene emissions induced from hailstorms would be highly speculative. As stated in the conclusions: long term measurements are needed to cover more such storm events to come up with annual estimates.

P15337, L14–16: Does this imply that monoterpene emissions from ponderosa pine originate mainly from storage pools? Are there any estimates of the contribution of de novo biosynthesis to total monoterpene emissions available? They might help the reader here.

Recent studies in a Scots pine dominated forest (Taipale et al. 2011) suggest a 30-46% de novo emission potential to the total monoterpene emission. For ponderosa pine there is so far no published data available. The sentence 'Studies have shown that temperature is the main driver for monoterpene emissions, while both temperature and light are important for MBO emissions.' does not imply that the monoterpene emissions from ponderosa pine originate mainly from storage pools but that so far published data found temperature as the main driver. As shown in Table 6 we found in this study also a light dependency of monoterpene emissions. Due to the long-term influence of the hailstorm on the monoterpene emissions a detailed study on this was not feasible.

P15341, L12–13: How were the time stamps of the two data sets corrected?

During our measurements only the computer for the VOC measurements was automatically synchronized with a time server. The time on the laptop where the 3D wind data was recorded drifted constantly in time compared to the other computer and was reset to match the VOC time stamp a couple of times during the measurement period. To avoid a drift in delay time over time due to the drift in the time stamps in the measurements the 3D wind timestamp was corrected in accordance using the reset points as bench marks. A sentence explaining this will be added to the revised manuscript.

P15341, L20–23: What does the physically meaningful delay time window mean? A delay time of 0 s seems somewhat short given that there was 35 m of tubing (ID?) and the flow was 20 slpm.

The delay time was measured in the beginning of the measurement campaign to be in the range of 4-5 seconds. A broader window was chosen as a physical meaningful window for finding mass to charge peaks that show a flux versus those not showing a flux. Changing the size of this window does not change the information we gained from this: if statistically more times the delay time based on maximum covariance peak is found in this window compared to other windows.

P15341, L23–24: How was the delay time (and thus the flux) actually determined? Was it based on the maximum absolute covariance in the physically meaningful delay time window? What were the criteria for "a significant flux" and "no flux"?

For the first step to filter the ion mass peaks showing a flux the delay time was chosen on the maximum absolute covariance within the -20s to +20s time window. For the ion mass peaks that where in our described class A or B where this delay time was found 50 and 35% of all times in the physical meaningful delay time window the flux was determined based on the maximum absolute covariance in the physical meaningful delay time window.

P15342, L3–13: Please explain why the thresholds 35 % and 50 % were chosen for the classification and which grades were used in the later analysis. How was the flux detection limit determined? And why is it same for all compounds?

The thresholds of 35% and 50% where chosen based on visual inspection of many covariance peaks of many times and many masses. Those masses where the delay time was found less than 35% of the times in the physical meaningful delay time window did most of

the times not exhibit a clear covariance peak. The second criteria called flux detection limit at this point was more a flux cut off limit and will be renamed in the revised manuscript. This cut off limit is a conservative value estimated based on the vertical advection <w> d<c>/dz which can give an upper bound to the advection flux term (Finnigan et al. 1999). For Manitou typical values of these flux terms are on the order of 0.04 to 0.1 mg/m2/h. This is clearly not the same for all compounds but as we are dealing with 649 ion mass peaks a conservative value for all compounds was chosen.

P15344, L7–8: How were the daytime detection limit and advection flux for MBO determined? And were they determined only for MBO? How does this detection limit differ from the one mentioned in Sect. 2.3?

The flux detection limit was calculated using the following equation: $flux\_LOD = \frac{1}{2} r \sigma_w \sqrt{\frac{c \cdot t}{s}}$ **,where r the correlation coefficient, σ_w the standard deviation of the wind, c the concentration of the VOC, t the integral time scale and s the absolute sensitivity of the VOC. We also calculated the flux_LOD as the de-correlated signal between vertical wind speed and concentration measurement described by Ruuskannen et al. (2011). For a discussion on advection fluxes see our detailed response to reviewer #3.**

P15346, L10–14: Are the differences between the temperature ranges in Table 3 statistically significant (e.g. at the 95 % confidence level)? When the results are compared with the MEGAN 2.1 results in Table 2, it seems that the range 19–21C agrees best with MEGAN (not 21–23C as suggested in the text?).

Uncertainty levels are added to Table 2 underlining the statistical significance of the differences in the temperature range. The best agreement was mistakenly written wrong in the manuscript and will be changed in the revised version.

P15346, L15–22 and Fig. 6: Do the results of Harley et al. differ from the other results at 285–300 K when the uncertainties in all studies are taken into account?

Between 290-300 K the differences between the Harley et al. cuvette measurements and the above canopy measurements are within the uncertainties but below that the Harley et al. results are systematically higher. The revised text is rephrased to clarify this.

Technical corrections

P15344, L14–15: Where do "(black)" and "(grey)" refer to? Should "14" and "four" be "11" and "six"?

Black and grey refer to the ion mass peaks listed in table 1. This is clarified in the revised text. Also it was not possible to use black and grey in table 1 so this is changed to bold and italic both in text and table. The numbers "14" and "four" are also corrected to "11" to

“six”.

All tables and figures: Please add uncertainty estimates (e.g. 95 % confidence intervals) wherever possible.

Uncertainty estimates will be added to tables and figures wherever possible.

Table 5: Should ”B” be ”beta”?

Yes, this is changed in the revised manuscript.

Table 7: Would it be more consistent to use the unit mg m2h1for the emission factor? See e.g. Tables 1 and 5.

The unit in Table 7 is changed to mg m^{-2} h^{-1} in the revised manuscript.

All figures: Some of the axis labels start with a capital letter and some not. Please check whether ACP has a specific format for them.

This will be changed in the revised version of the figures. Consistently no capital letter for axis labels will be used.

Figs. 2 and 4: Please explain what the error bars represent.

The error bars in Figs. 2 and 4 represent the standard deviation of the measurements. This information will be added in both figure captions of the revised manuscript.

Fig. 11: Should ”Cumulative flux” be ”Cumulative monoterpene flux”?

We appreciate this comment and will change this in the revised manuscript.

Anonymous Referee #2

Kaser et al present first eddy covariance flux measurements above a ponderosa pine forest using recently developed equipment, the PTR-TOF-MS. The manuscript compares the data in “normal” and “disturbed” conditions to the commonly used model MEGAN 2.1. The unique data set, including several storm events and subsequent periods, is used to point out a significant underestimation of measured monoterpene fluxes by the model. The manuscript incorporates new methods with accomplished modeling, and hence represents a substantial contribution to ACP.

The presented results are based on high quality measurements using state-of-the-art equipment. The analysis methods as presented are sufficient and based on standard procedures. The *presentation in text and figures of these results is clear, as well as the following discussion and conclusions.*

The manuscript should be published after addressing the following minor comments and questions:

Why is ponderosa pine forest important? Could you add to the introduction a motivation why the understanding of this explicit ecosystem is important? E.g. is ponderosa pine representative for other pine forests? How much area is covered by this in comparison to other forest types? Are the findings and conclusions of the presented study important for global or regional atmospheric chemistry?

This ponderosa pine forest is representative of the Rocky Mountain Front Range ponderosa pine zone, which extends from southern Wyoming to northern New Mexico and is one of the most important woodland ecosystems in the Western US. In addition, pines are a widespread tree species and we expect the results from this ecosystem to be informative for pine dominated regions in other parts of the world. The following short passage and references will be added to the revised text of the manuscript:

Pine trees are widespread in forests throughout the northern hemisphere and are dominant component of about 62% of U.S. woodlands where they cover an area of about 2.4 million km^2 (Guenther et al. 1994). Ponderosa pine is the principle tree species on over 0.1 million km^2 in the U.S. and is present on an additional 0.15 million km^2 (Graham and Jain 2005).

The catalytic converter was used to generate background measurements every 7 hours. Is this interval sufficient to capture diel variations of the background? Did you observe fluctuations between the different background measurements?

Fluxes are calculated from the covariance of concentration and wind fluctuation in a half hour period. A subtraction of a constant background is not an important factor in this calculation. The 25 min background cycle was chosen to not lose more than 1 30-min flux value each 7 hours. The 7 hour cycle was chosen to lose each day the same half hours but slightly shift the cycle allowing the possibility to calculate diurnal cycles of fluxes without one half hour missing completely in the diurnal cycle.

In Section 3.2 you describe the criteria for quality control. These criteria are listed for explanation of the analysis. However, in order to understand the reason for the thresholds and different conditions more explanation would be necessary. Could you please give references or reasons? For example, criteria (1), (2) and (6) are described as necessary to eliminate contamination from a highway, or vegetation change, or a reference is given. Could you please provide similar descriptions for the other criteria?

A reference describing the reason to exclude the data where the third rotation angle exceeds (McMillen, 1988) will be added to the text as well as the reason for criteria (3) and (5) where chosen to eliminate data points with spikes in wind or concentration.

P.15344, l. 14: The text refers to (black) and (grey). Where can the reader find this? Please, double-check Table 1 and what was written in Section 3.3. There seem to be inconsistencies of how many compounds were grade A and B.

Black and grey should be called bold and italic and is referring to Table 1. This as well as the inconsistency with the number of grade A and B ion mass peaks is changed in the revised manuscript.

Table 5: "B" probably means "beta".

Yes, this will be changed in the revised manuscript.

Table 7: Was the monoterpene speciation characterized by gas chromatographic methods?

The speciation of monoterpenes listed in Table 7 is from the speciation available in MEGAN 2.1. The emission factors of the different monoterpenes available in the model are added up for comparison purposes as our PTR-TOF-MS measurements do not allow monoterpene speciation. The caption of Table 7 will be rephrased in the revised version of the manuscript to clarify this.

In many figures bigger labels are needed to improve the readability in the print-out, such as in Fig 3, 4, 7, 8, 9, 10, 11.

Bigger labels will be used in the revised version of the above mentioned figures.

In Figure 4 an additional line at y=0 could show the difference between negative and positive fluxes.

Grid lines in Figure 4 are plotted to help the reader to distinguish between positive and negative fluxes.

In Figure 8, the y-axis label should probably be something else than "time (DOY)".

The x-axis is "time (DOY)" and the y-axis should be "flux (mg m^{-2} h^{-1})". This is changed in the revised manuscript.

Anonymous Referee #3

This is a very interesting paper nicely matching the scope of ACP. Particularly novel seems to be the significance of wind-related mechanical stresses triggering biogenic emissions. Thus, it might be important to account for such disturbances in the models. The story is nicely presented and I would have just a few comments/suggestions below.

Similarly to other referees I would like to recommend the paper for publication in ACP if the comments can be addressed.

1) The prescreening approach (chapter 2.3 flux calculation) sounds interesting but might not be very clear for a reader. a) Did you perform prescreening for all the 30 min periods for all the 649 peaks? b) Would this method be also sensitive to low concentration VOCs? c) Would pre-averaging of the covariance (e.g. as in Taipale et al., 2010) help the accuracy and sensitivity of this approach. Would preaveraging affect the pattern of the red dots in Fig. 1?

a) Yes this pre-screening was done for all 30 min periods for all 649 peaks. b) & c) The focus of this paper was not on fluxes of low concentration VOC's as interpreting very low fluxes is very complicated due to many uncertainties like advection fluxes, water dependent background effects etc. Therefore we decided to take a conservative approach excluding fluxes below an average flux detection limit of 0.1 mg compound m^{-2} h^{-1}. With this approach even though pre-averaging might change the picture a little bit ion mass peaks that would fall into 'grade B' instead of the 'no flux' category would fall out again due to the second cut of criteria of < 0.1 mg compound m^{-2} h^{-1} described in the end of chapter 2.3.

2) Another criteria the authors have used are a 0.1 mg m-2 h-1 cutoff. I am slightly worried about how this may impact the representativeness if we focus only on large flux players and neglect all small fluxes which can add up to significant flux. For example, assuming that just 10% of the remaining mass peaks (649-17) showed the flux of up to 100 ug m-2 h-1, this would give an upper limit of 6 mg m-2 h-1 which is close to the maximum flux reported for the sum of MBO and isoprene in this paper (Fig. 10). a) Was the 0.1 mg m-2 h-1 limit chosen arbitrarily or how was it determined? b) how many more peaks in the mass spectrum would have shown a clear peak in the covariance function if you reduced the constant cut-off flux detection limit to 50, 10 ug m-2 h-1? c) Was it not possible to use an instantaneous limit of flux detection (e.g. as in Spirig et al. 2005)? Some compounds may be low emitters but can show occasionally high fluxes.

The cut off criteria of 0.1 mg m-2 h-1 did not affect all 649 ion mass peaks but only those masses that where in the first place rated either A or B. This was in total 47 ion mass peaks. After excluding all ion mass peaks related to primary ions 29 ion mass peaks remain. On those this second criteria of 0.1 mg m-2 h-1 was applied leading us to the 17 ion mass peaks treated in the further course of this manuscript and 12 excluded ion mass peaks. a) Along with VOC gradient measurements we estimated the magnitude of advection fluxes based on vertical advection <w> d<c>/dz (e.g. as stated by Finnigan et al., 1999, the vertical

advection term measured on a tower can at least present some bound on the magnitude of the total advective flux component). Even though the vertical advection term can be responsible for all or none of the advective flux terms, we used the expression as a conservative flux cut off limit. For Manitou typical values of these flux terms are on the order of 0.04 to 0.1 mg/m2/h. b) 6 ion mass peaks of the 12 excluded show a flux between 0.05 and 0.1 mg m-2 h-1 and the other 6 show a flux between 0.01 and 0.05 mg m-2 h-1. c) While an instantaneous limit of flux detection can provide some automated assessment it sometimes also suffers from large variations due to mesoscale influences. Another approach would be to use a random permutation method, which for example tends to give lower LODs. However in this work we based the flux detection mainly on visual inspection of the covariance functions (similar to Ruuskanen et al., 2011). In the current manuscript we also focus on emissions during disturbed conditions, which is mainly the period of high VOC emissions, thus we are confident that our visual inspection captured periods of occasionally high fluxes by inspected fluxes during the period impacted by the hail event.

3) In terms of modeling hailstorm-related wound stresses, is there a perfect VOC tracer (styrene?, homofuraneol?) to give a proxy for hailstorm disturbance?

As shown in section 3.5 and table 8 we found 22 ion mass peaks elevated after the hailstorm this indicates that there is not one single VOC tracer for the hailstorm. Both m/z 105.070 and m/z 143.069 (possibly styrene and homofuraneol, respectively) show small elevations in comparison to monoterpenes after the storm and might therefore not be ideal as proxies. Also they were not clearly identified e.g. with GC/MS, therefore follow up studies such as e.g. wounding experiments on needle and branch level are necessary to identify a perfect wound stress VOC tracer.

4) There can be high emissions of monoterpenes from resins (Eller et al., 2013). Is it possible that during a hail storm more resin can be exposed and vaporized?

It is very well possible that the high monoterpene emissions come from resin that is already exposed but gets disturbed from the hail storm or that on the spots of damaging of needles and branches from the hailstorm new resin escapes.

5) Can you please make the caption of Fig. 7 to spell out the acronyms so the reader does not have to refer to the text?

The acronyms of the different model runs are spelled out in the figure caption of the revised manuscript.

6) Can you break the y axis in Fig 8. The precipitation spike makes other peaks look very small.

We thank the reviewer for this comment and changed Fig. 8 accordingly.

7) Is it possible to add error bars in Fig 9.?

The random flux errors will be added in the updated version of Fig. 9.

8) The monoisotopic mass of C7H10O3-H+ which is shown unidentified in Table 8 would be consistent with homofuraneol.

Homofuraneol C7H10O3-H+ would be measured at the unidentified m/z 143.069. As no GC/MS measurements are available from the days after the hailstorm it is at this point not possible to identify the compound on m/z 143.069. The possibility of homofuraneol will be mentioned in the revised manuscript.

Literature:

Spirig, C., Neftel, A., Ammann, C., Dommen, J., Grabmer, W., Thielmann, A., Schaub, A., Beauchamp, J., Wisthaler, A., and Hansel, A., 2005, Eddy covariance flux measurements of biogenic VOCs during ECHO 2003 using proton transfer reaction mass spectrometry: Atmospheric Chemistry and Physics, v. 5, p. 465-481.

Taipale, R., Ruuskanen, T. M., and Rinne, J.: Lag time determination in DEC measurements with PTR-MS, Atmos. Meas. Tech. Discuss., 3, 405-429, doi:10.5194/amtd-3-405-2010, 2010.

Eller, A.S.D., Harley, P., and Monson, R.K., 2013, Potential contribution of exposed resin to ecosystem emissions of monoterpenes: Atmospheric Environment, v. 77, p.440-444.

List of references for this answer to the reviewer's comments:

Finnigan, J., A comment on the paper by Lee (1998): on micrometeorological observations of surface-air exchange over tall vegetation, Agric. Forest Meteorol., 97 (1), 55-64, 1999

Grahan, R. and T. Jain, Ponderosa Pine Ecosystems, USDA Forest Service Gen. Tech. Rep. PSW-GTR-198. 2005.

Guenther, A., Zimmerman, P. and Wildermuth, M. (1994) Natural Volatile Organic-Compound Emission Rate Estimates for United-States Woodland Landscapes. Atmospheric Environment 28(6), 1197-1210.

McMillen, R. T.: An eddy correlation technique with extended applicability to non-simple terrain, Bound.-Lay. Meteorol., 43, 231–245, 1988

Ruuskanen, T. M., Müller, M., Schnitzhofer, R., Karl, T., Graus, M., Bamberger, I., Hörtnagl, L., Brilli, F., Wohlfahrt, G. and Hansel, A.: Eddy covariance VOC emission and deposition fluxes above grassland using PTR-TOF, 11, 611-625, doi:10.5194/acp-11-611-2011, 2011

Taipale, R., Kajos, M. K., Patokoski, J., Rantala, P., Ruuskanen, T. M., and Rinne, J.: Role of de novo biosynthesis in ecosystem scale monoterpene emissions from a boreal Scots pine forest, Biogeosciences, 8, 2247-2255, doi:10.5194/bg-8-2247-2011, 2011.

Schaefer, J.T., Levit, J.J., Weiss, S.J., and McCarthy, D.W.: The Frequency of Large Hail over the Contiguous United States. *Preprints*, 14th Conf. Applied Climatology, Seattle WA, 2004.