This document contains author responses to Referee #1 and Referee #2 in the ascending order.

I. Response to Don Lenschow (Referee 1)

We thank Don Lenschow for providing inspiring and helpful comments which led to major improvements in the revised version of the manuscript. We reply to each comment below in bold text.

Overall Evaluation:

1) While I have rated the manuscript "accept subject to minor revisions," I do have some important concerns that I think should be addressed before the manuscript is approved for publication. I do think that this paper merits eventual publication, as the techniques and analyses presented here are a significant step forward in the science and demonstrate an approach to quantifying emissions of trace reactive species over horizontally heterogeneous surfaces.

We thank you for your positive evaluation of our research that it presents a significant step forward in science. We also realize that the approach will be refined further as the wavelet analysis becomes more routinely applied to airborne flux measurements. We address all the reviewer's concerns below.

Major Comments:

2) Using a single representative vertical flux divergence to extrapolate all fluxes at flight level down to the surface to estimate surface emission has several assumptions that should be discussed in more detail. The vertical flux divergence is dependent on the rate of isoprene oxidation (which depends on OH concentration and other oxidants), the time rate of change of isoprene concentration, and differential (with height) horizontal advection of isoprene. Since the flights were conducted with a variety of different conditions and locations, one would expect different isoprene lifetimes and horizontal advection scenarios. These effects need to be taken into account or at least evaluated for their impact to be assessed in order to estimate the accuracy of the extrapolation to the surface. Furthermore, I don't see any estimate in 2.7 of the random error contributed by the flight segment length. Thus, the error analysis in 2.7 seems to me to be too optimistic, at least without further justification.

This has relevance later on in section 3.2, where it is not clear how much of the variations and differences between e.g. tower and aircraft are due to random errors and how much to real variability.

We have added new text to 2.7 to clarify the uncertainties due to flux extrapolation to the surface: "It is important to note, that the vertical flux divergence is dependent on the rate of isoprene oxidation (which depends mostly on OH concentration during daytime), the time rate of change of isoprene concentration (relevant for conserved species), and differential horizontal advection of isoprene with height (small). Based on directly measured flux divergence in the racetrack gradient flights (Karl et al., 2013) we showed clear linear dependence of the flux divergence with a theoretical concentration gradient (e.g. 1.4×10^{-4} ppbv m^{-1} over a homogenous oak terrain and an OH concentration of 6.6×10^{6} molec/cm^3). Since the flux divergence for isoprene was shown to be primarily controlled by OH concentrations (of which we have a range of estimates), we make an informed assumption here that the divergence coefficients we used to scale the fluxes to the surface are accurate within a factor of two for the entire campaign. Thus a change in the flux divergence coefficients by a factor of two could result in only a ~2% difference to the scaled surface flux for a typical z/zi ratio of 0.3 which is minor relative to other error sources as discussed in Sect. 2.7. As the correction of the fluxes for flux divergence was typically less than 20%, the contribution from less accurate divergence coefficients is assumed to be relatively minor (up to ~2%) for isoprene but could still be more important for other gases (e.g. CO_2), for which more detailed characterization of flux divergence might be needed in future measurements." We minimized the effect of concentration change by performing flights during times when the PBL was well developed and changes in the steady state assumption for chemistry was small, which was evaluated during our profiling flights (Karl et al., 2013). We investigated the effect of horizontal advection during our flux profiles and estimated horizontal advection by taking the isoprene concentration gradient and mean horizontal wind along the horizontal flux leg. We have now also provided an expanded budget of uncertainties in revised Sect. 2.7, following one of the comments by the reviewer #2.

3) In several places, it is claimed that the CWT method can be used to obtain 1 to 2 km spatial resolution in fluxes. This is misleading. True, you can get fluxes to this resolution, but with very large random error. This needs to be considered further and the random error as a function of sample length needs to be discussed and quantified. Among other references, you might take a look at "Errors in Airborne Flux

Measurements," by Mann and Lenschow, 1994, J. Geophys. Res., 99, 14519-14526.

We agree that selecting a short integration scale such as 2 km would result in a very large error (e.g. Mann and Lenschow, 1994), and clarify that we have only evaluated fluxes over longer stretches (>> 2 km). We apologize that our original text may have been misleading regarding this length scale. We now state more clearly that for the wavelet fluxes we actually integrate long segments (e.g. 100 km) and based on wavelet decomposition, we reconstruct the time domain for the wavelet co-spectra to yield time series of discrete coherent structures which are subsequently aggregated to 2-km surface fluxes. Because the variability is much larger at the short time-scales the random error of each 2-km point must be larger (40-50%) than that of the average flux for the whole leg. This error decreases significantly with averaging over 5 km and more and thus it makes sense to use averaged data rather than individual points with models. The 2-km representations provide more flexibility for averaging, for example, individual points can be useful for a regression of isoprene flux versus LAI for all of the 2km data. This would provide a large number of data points to ensure good statistics and gives us information that we would not have if we just had the average over a larger transect. Nevertheless, we think that the wavelet decomposition is another step forward to reduce greatly the systematic errors for high spatial resolution fluxes and the way forward in the future to minimize random errors would be to repeat each leg once or twice.

Other Comments:

4) abstract, l. 6: spell out PTR-MS

Done.

5) abstract, l. 13: the statement, " Vertical flux divergence of isoprene is expected due to its relatively short lifetime. . ." is misleading. Vertical flux divergence is expected for almost all atmospheric species. In conserved species it is a reflection of time changes in the mean concentration. More accurately, you should say that there is a major contribution to vertical flux divergence of isoprene due to its relatively short atmospheric lifetime.

We have corrected the sentence as follows: "Vertical flux divergence is expected for all atmospheric species, but a major contribution to vertical flux divergence for isoprene is due to its relatively short atmospheric lifetime.

6) abstract, l. 15: I can't make sense of a vertical flux divergence expressed as a percent. The units of species vertical flux divergence should be concentration/time. Do you mean that the this is the percent difference between flight-level flux and surface flux?

Yes, we meant the maximal percent difference between flight-level flux and surface flux due to the flux-divergence correction. This has been clarified.

7) p.7968, l. 15: spell out BEIGIS. Also earlier MEGAN and BEIS

Done.

8) p. 7969, l. 9: "Stacked," instead of "Vertical."

Done.

9) p. 7975, l. 8: . . .capable of eddy flux measurements.

Corrected.

10) p. 7975, l. 10: What is meant by area "ratio of about 2?" Does this mean that the diameter changes from 2.047 in. to about 2.89 in.? Same comment applies later to the area ratio of about 5. Also, I think that you should use metric units for these dimensions.

We have removed these specifics of the inlet system which were thoroughly described in Karl et al., 2013.

11) p. 7975, l. 16: "unaffected" is a bit too strong. Perhaps something like "minimized" would be more appropriate.

"Unaffected" has been replaced with "minimized".

12) p. 7976, l. 7: "is a fast sensor which" is redundant.

This has been deleted.

13) p. 7976, l. 13: I'm not clear what 10 Hz separated by a relatively longer gap of 2 Hz means in this context. Can you elaborate more on what this means? Do you really mean 0.1 s dwell time and 2 samples/s?

This is exactly what we meant. This has been clarified as "…10 Hz disjunct sampling corresponding to 0.1 s dwell time and approximately 2 samples/s".

14) p. 7977, l. 8: "sawtooth" rather than "tooth"

Done.

15) p. 7977, l. 14: So, the sensitivity is 10 pptv/17 s, where 17 s is the averaging time?
Yes, we meant to say that the detection limit is dependent on the amount of averaging and is dependent on the sensitivity which is independent of the amount of averaging. We rephrase for clarity: "These high sensitivities ensured low detection limits (e.g. < 10 ppt for isoprene at 1-km averaging (~17 s)).

16) p. 7977, l. 17: I'm a bit confused by the comparison between absolute sensitivity and normalized sensitivity. Do they really have the same units?
This is mostly directed to the PTR-MS community. The idea is that the normalized sensitivities (which account for primary ions and water clusters in the drift tube) can be used to compare with other instruments independently of their settings and primary ion count rates.

17) p. 7978, l. 28: . . .measurements of concentration profiles in the mixed layer overlying the surface layer of the daytime convective boundary layer. . . I suggest using CBL instead of MBL, as "mixed boundary layer" is not accepted terminology.

We have changed to CBL to comply with the classical terminology.

18) p. 7979, l. 1: This top-down bottom-up method applies only to a conserved species not VOCs in general. You should say "a conserved species" not a VOC. p. 7979, l. 25: . . .on all eight research flights, and MVK+. . .

We agree and have incorporated the suggested change.

19) p. 7980, l. 3: 150-300 m is very deep for an assumed surface layer. What is this based on? Certainly not on 10% of the PBL depth, nor on applicability of surface-layer parameterizations.
The reviewer is right that the surface layer (SL) was actually less deep. Based on the 10% of the PBL depth and given the PBL depth we measured was typically between 1000 and 2000 m, the typical SL during CABERNET was extending typically up to 100-200 m rather than 150-300 m as we stated before. We apologize for this inaccuracy and have corrected the typical range to reflect the shallower SL.

20) p. 7980, l. 13: . . .should be maximized. (Actually more accurately, if you sample significantly more than 1/integral scale, increasing the sampling rate won't gain you anything.)

We completely agree and modified the text accordingly.

30) p. 7980, l. 18: It's not clear to me what you mean by "total cycle length." Do you mean the sample rate for each species is 1.25 to 2 samples/s?

We meant the duty cycle length which is the sum of the dwell times of all measured species in the disjunct mode, but we now also describe how this relates to the sampling rate.

31) p. 7980, l. 25-26: Do you really mean non-stationarities or do you mean horizontal heterogeneity? In l. 24 "affected by diurnal effects" is not usually a problem for aircraft measurements of a single flight segment.

We actually meant both or in other words non-stationarities related to heterogeneities. A single flight segment could take a few hours for a slow flying aircraft, but in our case this should not have been a problem for receiving contributions from diurnal effects.

32) p.7982, l. 9: . . .an integrated straight stretch. . . What is meant by integrated? What is integrated? Do you mean continuous or contiguous? Why does it have to be straight?

We have replaced "integrated" with a different word to avoid confusion. The stretches were the straight, even and uninterrupted segments we selected for flux calculation. We think that it could be less accurate to calculate the flux on the curved paths when the aircraft is turning because of more complex behavior of horizontal wind and potentially larger advection errors.

33) p. 7982, l. 16: See earlier comment about random error in flux measurements as function of sample length.

As stated earlier, 2 km was not the flux integration length, but the resulting flux representation from the wavelet decomposition of the long stretch. We add a note to the text "meaningful spatial representation at 1-2 km resolution (note the integration length was ~2 orders of magnitude larger)".

34) p. 7982, l. 29: . . .controlled by its relatively. . . . . ."racetracks" at multiple levels. . .(profiles is redundant)

Done.

35) p. 7983, l. 1: again, flux divergence has units of concentration/time, not percent.

The sentence was rephrased to inform the reader of the percent difference of flux at aircraft-level vs flux at the surface. "…It was found to be linear, so the scaled surface flux was in the range of 5 – 30% larger than measured at the given z/z_i ratio

36) p. 7983, l. 3: A storage term of 2-5% of what? The chemical reaction term? This seems very small to me in view of the diurnal variability of the isoprene concentration in the PBL. Was the buoyancy flux also extrapolated down to the surface similar to the isoprene flux when used to estimate w*?

The range of determined storage term calculations is based on the vertical profiling flights for the duration of one vertical profile measurement and not the entire diurnal cycle. Within this time period the change of concentration relative to the flux divergence term due to chemistry was small. Typically during our midday flights the isoprene concentration was relatively constant due to the offsetting influences of increasing boundary layer depth and increasing emission source. The buoyancy flux was not extrapolated down to the surface.

37) p. 7983, Eq. (4): Where did you get this equation come from? It isn't in Horst and Weil (1992), who considered only the surface layer not the mixed layer. It seems that the authors have the wrong citation; this equation is presented in Weil and Horst, 1992: Footprint estimates for atmospheric flux measurements in the convective boundary layer. A chapter in Precipitation Scavenging and Atmospheric Surface Exchange, Vol 2, S. E. Schwartz and W. G. N. Slinn, Coords, pp. 717-728, Hemisphere Publishing Corporation, Washington, 1172 pp.

Thank you for spotting this oversight. We have now corrected the citation to Weil and Horst (1992).

38) p. 7983, l. 24: "superposition?" or "summation"?

now is "summation".

39) p.7984, l.12: do you mean "minimal roll angle"? Constant roll angle could imply a constant rate turn.

Yes. We changed the text accordingly.

40) p. 7984, l. 22: Standard usage for cross-spectrum is the real and imaginary parts of the Fourier transform, and co-spectrum for the real part, which you are using here.

We have changed cross-spectrum to co-spectrum.

41) p. 7986, l. 12: By "measurement footprint" it seems that you are no longer talking about the flux footprint that you were discussing earlier, but the concentration footprint, since the flux footprint does not extend hundreds of km upwind. This needs to be pointed out.

We point out here that we refer to the concentration footprint.

42) p. 7990, l. 25: . . .for a setting where...

Done.

43) Table 1: You might also include w* or the surface virtual temperature flux in the table, so the reader can e.g. estimate the convective turnover time. Your estimate of 10 minutes (p. 7979, l. 9) means a large surface virtual temperature (i.e. buoyancy) flux.

That is a nice suggestion, and we have now included the summary statistics for convective velocity scale (w^*) to the Table 1. In order to keep the table within the size limits we decided to remove the summary statistics for relative humidity but we state in the text that the environment was becoming dryer as the temperature increased.

Once again we thank Don Lenschow for his time and providing all these extremely useful comments.

Pawel Misztal on behalf of other co-authors.

II. Response to Anonymous Referee 2

We thank the anonymous referee for reading the paper carefully and providing thoughtful comments, many of which have resulted in changes to the revised version of the manuscript. We reply to each comment below in bold text.

General Comments

The goal of the manuscript is "to measure the distribution of isoprene flux across the oak woodland areas of California in order to test and improve the landscape-scale emission models that are used for regional air quality assessments". The main question addressed by the research is whether measurements of biogenic volatile organic compounds (BVOCs) confirm the spatial patterning of model results. For this purpose, data from ground-based and airborne platforms are used, as well as multitude of data processing approaches. The authors conclude that horizontally varying source distributions of isoprene were successfully mapped out across dominant ecosystems in California.
The central question is interesting and important as BVOCs contribute to regional air quality through ozone production and acting as radical sinks in the source regions. The authors approximate that 50% of the total global BVOC emission is constituted of isoprene. A wide array of methods is used, including initial isoprene emission measurements using airborne eddy-covariance. Most (but not all) of the outlined methods are taken into consideration in the results and conclusions. The results are mainly of technical nature and support the conclusions. However, results and conclusions miss to quantitatively address the actual research question of testing and improving landscape-scale emission models. As such, in its current form the manuscript constitutes a collection of methods and must be considered as an incremental advance to the field.

Although the goal of the research project was to test and improve landscape-scale emission models, this was not the goal for this particular paper. We will describe the quantitative comparison with the model in a separate paper that is currently being prepared (Misztal et al., 2014). Including the detailed comparison with the models would make the size and scope of the current manuscript much too large.

The paper partially fits the scope of ACP in that it presents a combination of field measurements, remote sensing and modelling of biosphere-atmosphere interactions. However, in its current form the manuscript is primarily of technical and regional interest, and misses to unravel general implications for atmospheric science in a rigorous, quantitative manner.

We disagree with this statement which might be caused by the misinterpretation of the paper goals (see comment above). We firmly believe that the results of airborne VOC concentrations and isoprene fluxes are fully quantitative and as the first demonstration of this technique that will be useful not only in this region but also globally. In addition, the dominant genera investigated (oaks) are the dominant source of isoprene in the U.S. and other regions.

In general, the flow of the paper is logical, references are adequate and from my perspective no copy-editing is required. However, the authors tend to hypotactic sentence structures (e.g., p. 7969 l. 1) which could be broken apart to make it easier for the reader to follow.
We have changed this sentence to be clearer for the reader.

With 28 (discussion) text pages the length of the manuscript is reasonable. However, the allocation is not well balanced with 17 pages on methodology but only 7 pages on results, discussion and conclusions, half of which actually attributed to concentration (and not flux) results.
We have moved part of the method section to the Supplementary Information to improve that balance. We think that providing novel methods and also the comprehensive context is important for interpretation of results but we have now moved some parts of the method section to the supplementary information to improve the balance.

Consequently, I recommend major revisions, and see two principal ways for successful publication. (i) Publication in ACP: Addressing the stated goal through removing technical details (description of individual flight days and leg separation, FFT vs. CWT comparison, mixed boundary layer technique, length scales vs. time scales…), adding a quantitative model-observation inter-comparison and expanding on general implications (results, discussion and conclusions). (ii) Publication in more regional (BGC) or technical (AMT) oriented journal: Reformulating the stated goal to focus on regional implications (BGC) or technical questions of airborne isoprene concentration and flux measurements (AMT).
We strongly feel that the revised manuscript is within the scope of ACP and will be of interest for readers of this journal. The revised manuscript is focused on the application of the aircraft eddy covariance technique to investigate of biogenic emissions and builds on the methods paper published describing the flux measurement approach used in this campaign (Karl et al., 2013) which we refer to in the current manuscript.

Specific Comments

p. 7966 l. 7: It does not appear that 10,000 km of flights were performed in the 8 measurement days presented in the manuscript.

We have provided the exact, unrounded value, for accuracy.

p. 7966 l. 12, p. 7985 l. 15: Neither FFT nor CWT eddy-covariance (EC) approaches are "independent" of non-stationarities. While CWT does not require a stationary time series, non-stationarity principally violates reducing the Navier-Stokes equation to the 1-D problem posed by EC. Hence CWT is not necessarily more "accurate" as long as not considering all divergence terms. Can the authors explain how CWT is making their results more valid?

It is correct that CWT does not depend on non-stationary data, but is not completely independent on non-stationarities. In fact the CWT method has been originally used to examine non-stationary periods in ground based measurements (e.g. Argoul et al., 1989; Collineau et al., 1993; Detto et al., 2010; Terradellas et al., 2001) to filter for coherent structures. We still think that the CWT method can be more accurate because it allows to localize flux contributions in space and scale and thus has the potential to filter non-stationarities in space and frequency space, therefore improving the investigation of surface heterogeneity.

p. 7966 l. 13, p. 7983 l. 1, p. 7983 l. 2: Extrapolating flux measurements at 400 m above ground to the surface heavily relies on the accurate knowledge of the vertical flux divergence. In the present case, the vertical flux divergence is determined from measurements at different heights, i.e. with different source areas and assuming a linear and monotonous function with height. Are the extrapolated fluxes still significant when, in addition to the residual error in the regressions, contributions from different surfaces are considered?

Flux divergence for isoprene is driven by its relatively short lifetime due to oxidation by the hydroxyl radical (see Karl et al., 2013, who show that the flux divergence term is mostly controlled by the lifetime of isoprene which can be as short as 23 minutes). The non-chemical flux divergence is comparatively very small, but the reviewer is correct that the divergence coefficients can slightly vary over different surfaces. This is why we have done the profile flux measurements for different conditions and plant species heterogeneity (see Fig 1, Karl et al., 2013).

Maybe an inverse method like Bange et al. (2006) might be superior?

This is a nice suggestion, but in fact this approach is very similar which assumes linear coefficients (see Fig. 1 in Bange et al.). We added a reference about the inverse Bange et al. 2006 method and mention it now as an alternative method.

Also, inference of vertical flux divergence from profile soundings is not a "direct" measurement as claimed on p. 7982 l. 29.
We did not derive divergence coefficients from the soundings but rather from dedicated thorough race-track stacked gradient profiles (again please refer to Karl et al., 2013).

Lastly, it is not clear at which flight levels and horizontal extent the stacked patterns were performed, which type of regression was used (considering error in variables?), and whether the regression results are actually significant.
For clarity we now explicitly refer the reader to this information in Karl et al., 2013.

p. 7966 l. 23: The authors relate concentrations to source regions. It must be noted that at 400 m a.g.l. flight altitude, the concentration source areas can extend several ten to hundreds of kilometres upwind (e.g., Griffis et al., 2007). Was such a source area analysis performed to substantiate the conclusions?
The concentration footprint for isoprene is actually not much larger than the flux footprint because the lifetime of isoprene is short (typically <1 h).

p. 7967 l. 5: Why did the authors chose a spatial resolution of 2 km? Principally, the flight altitude should be chosen so that the resulting blending length and flux footprint extend matches the spatial scale of surface patchiness (e.g., Mahrt, 2000; Mason, 1988; Raupach and Finnigan, 1995; Wood and Mason, 1991). Now, a matching wavelet integration interval can be chosen, which optimizes the trade-off among high spatial discretization and small random sampling errors. Also, it has been shown that individual flux estimates require aggregation to be statistically significant (Sühring and Raasch, 2013). What is the strategy of the authors to attain statistical significance?
Yes, we have taken all of these important characteristics (including the footprint, and surface patchiness) into account and we expand the information on the principle behind the 2 km resolution. Basically, we determined that for a sufficiently long stretch (e.g. 20-200 km) it is possible to achieve statistically significant descrete wavelet fluxes, on the order of hundreds meters. To comply with the range of conditions and to ensure statistical significance our 2 km flux is

not just a single value but it is an aggregate of individual wavelet flux values aggregated to 2 km. We clarify this in the text.

p. 7967 l. 20: …resolution and coverage… Principally, one wants to determine a functional operator that allows scaling up and down between different observations in different reference frames. Here, it is important to note that atmospheric observations are a convolution of biogenic emission (e.g., from leaf-level measurements) but also atmospheric transport. Hence a functional operator has to explicitly consider both of these processes.
It seems there is a confusion here. The CWT flux is a direct regional scale flux measurement which results from a combination of the emission source and transport.

p. 7969 l. 8–18: Repetition that can be omitted. Instead, it would be helpful for the reader to provide an overview of what's to come in the next sections.
This suggestion has been incorporated.

p. 7969 l. 19–23: Belongs to goal on p. 7968 l. 22–25.
We have moved this paragraph as suggested.

p. 7970 l. 8: Surface fluxes cannot be measured at 400 m a.g.l.
We now clarify "Surface fluxes can be inferred from flux measurements at 400 m agl."

p. 7970 l. 12: Tertiary levels of what?
We now clarify tertiary levels of species habitat.

p. 7970 l. 20: Measured temperature is not very meaningful, as it will vary with flight altitude. Better: Potential temperature. The standard deviation on l. 22: Within flight-track or among flight tracks? Also (Table 1): Relative humidity at flight altitude does not appear very meaningful, better dry mole fraction or partial pressure;
We appreciate these thoughtful suggestions, but the focus was not on reporting the temperature and humidity data, but rather to give an idea about potential differences at aircraft altitude and ground in the context of biogenic emissions.

p. 7971 l. 16: Is a dynamic upwash correction applied to the wind measurements?

Yes, the dynamic upwash correction as well as wind vector transformations from sideslip and pitch maneuvers (aka Lenschow maneuvers). We mention this more explicitly and refer to Karl et al., 2013 for details.

p. 7972 l. 3 – p. 7975 l. 2: Description of flight patterns overly detailed. Move to supplementary materials;

We have moved the description of flight tracks to supplementary materials. This should also improve the methodology-results balance.

p. 7975 l. 9: EGU journals use metric units.

SI units are now used throughout.

p. 7975 l. 20: Sensor models and pre-/post calibrations?

We add more specific information on the models. As we do not report the data from all of the instruments we do not think it would be appropriate to add information about their calibration.

p. 7976 l. 17: …were kept constant across all flights?

The sentence now reads: "The instrument operation and routine were kept consistently constant for each flight"

p. 7977 l. 18: I am not familiar with the PTR-MS methodology. Does the sensor report dry mole fraction, or do density corrections due to temperature differences and humidity have to be applied? If so, these might be significant, as relative humidities as high as 100% are reported.

The PTR-MS instrument measures volume mixing ratios (v/v), and so does not require density corrections. The pressure, water level and temperature in the reaction chamber are kept constant and unaffected by ambient temperature/humidity differences.

p. 7978 l. 7: w'c' is missing the overbar.

The overbar has been added to denote the integrated quantity.

p. 7978 l. 13: …frequency of the transporting eddies…

Done.

p. 7978 l. 23: "build" and not "built".

Done.

p. 7978 l. 25 – p. 7979 l. 18: Does not contribute to manuscript objectives and can be omitted.

We decided to keep this text to give the broader concept of surface fluxes to the reader.

p. 7980 l. 3: How was the depth of the surface layer determined?

Since the surface layer is typically less than 10% of the PBL layer height during unstable conditions, we conclude that our measurements (~1000 ft above ground) were not conducted in the surface layer. Based on our profile measurements, which were in good agreement with expected sensible heat flux profiles (e.g. Karl et al., 2013), we were able to accurately determine PBL heights and conclude that our land-use flights (1000 ft above ground) were conducted in the lower part of the mixed layer (i.e. we did not observe a constant flux layer in the lower part of the profile). While it would be very challenging to predict the PBL height solely based on profile measurements in the surface layer (e.g. Metzger et al., 2012), the combination of soundings and vertical flux measurements suggest that we conducted the landuse flights in the lower part of the mixed layer.

p. 7980 l. 20: The conventional method of determining EC flux is a time-domain Reynolds-decomposition (e.g., Foken, 2008). FFT requires additional pre-processing steps such as tapering etc. and hence alters that data basis.

We employed standard pre-processing steps, which included the removal of spikes by visual inspection, detrending and the subtraction of the mean, but we did not taper before doing the FFT calculations due to our disjunct sampling approach. While detrending could represent a high pass filter, we did not experience large differences between detrended and"raw" data on the timescales investigated here (e.g. 1-3%). We did not use the Foken, 2008 approach. The flux calculations were based on the virtual disjunct eddy covariance approach for VOC (e.g. Karl et al., 2002).

p. 7980 l. 25: …affected by non-stationarities…

"prone to" has been changed to "affected by"

p. 7981 l. 18: …preserves the energy…?

"preserves" suits better, thanks.

p. 7981 l. 20 – p. 7982 l. 4: Does not contribute to manuscript objectives and can be omitted.

We prefer to keep this example for other mother wavelet applications to attract reader thinking about the richness of wavelet specific applications in environmental sciences.

p. 7981 l. 24: Nordbo and Katul (2012) focus on a spectral correction method but do not specifically address long-term CO2 fluxes from soil.
They actually did look at the periodicities of long-term CO2 fluxes from soil.

p. 7981 l. 5: Repetitive, can be shortened.
We could not find what was repetitive and could be shortened. Possibly wrong line/page number.

p. 7981 l. 11: Suggest clarification: …integration of a sub-segment (e.g. 2 km) or an entire flight segment (e.g., 100 km)…
This has been clarified.

p. 7982 l. 19: Why would agreement between FFT and CWT results add confidence to the flux estimates? The agreement is basically a measure of how well the stricter assumptions on FFT are fulfilled, and how modifications of the data such as de-trending and tapering affect the results.
These are two independent methods with different uncertainties. That is true that to an extent the test will reflect the assumptions met, but on the other hand the agreement adds confidence in the assumptions and in the accuracy of the methods.

p. 7982 l. 22: To this point, no spectral correction was mentioned in the text. Hence for a comparison among FFT and CWT it doesn't matter whether high-frequency spectral loss is present or not - it should be reflected by either method. How were high-frequency spectral corrections performed by the authors?
We have determined (see also Karl et al., 2013) that we were not limited by high frequency attenuation. We do not think the high frequency attenuation was applicable to PTR-MS measurements at 10 Hz.

p. 7982 l. 24: The approach of Nordbo and Katul (2012) can correct spectral attenuation as long as not related to sensor displacement.
As mentioned above this could be useful for much slower sensors, or sensors where attenuation occurred in the long inlet line, which was not the case for our setting.

p. 7983 l. 2: How was the contribution of the storage term below aircraft flux sounding level determined from profile flights above this level?

Contributions from the storage flux were investigated from profile measurements during unstable conditions representative for our flux flights. The concentration change with time (dC/dt) was obtained by repeating profiles several times and comparing this to the measured flux divergence. The typical magnitude of storage fluxes were on the order of 2-10%. We assume a similar magnitude for the low level land use flights.

p. 7983 l. 10: Footprint results are nowhere mentioned in the results/discussion, but are crucial to be considered for a model validation. What is the accuracy of the simplified approach to source area quantification? Despite omitting the use of actual along- and cross-wind probability density function, is the approach sufficiently reliable to allow distinguishing different surface sources from a flight altitude of 400 m? Is the along-wind PDF evaluated each 2 km, or is turbulence statistics calculated over an entire transect? How are footprints superimposed along transects?

We provide more quantitative information on the footprint derivation based on Horst and Weil (1992) and the parameterization for the mixed layer (Karl et al., 2013) to take into account instantaneous convective velocity scale based on the wavelet heat flux, z/zi and horizontal wind speed. We show in the supplementary information (Supplementary Fig. S5) the orientation of the 2D half-dome pdf oriented according to the wind direction.

p. 7983 l. 21, p. 7984 l. 7: The authors are making an effort to characterize uncertainty, but it is not clear how individual error sources are propagated. Principally, the study is lacking a thorough, quantitative uncertainty budget on a per-sample basis, from which the statistical significance of the reported 2 km results would be evident (e.g., Fig. 7, Sühring and Raasch, 2013). In addition to errors originating from instruments and turbulent sampling, such budget should quantify uncertainty resulting from time-frequency and source area analyses, parametric and structural errors in the data processing. For example, random sampling uncertainty for 2 km segments are expected to be much larger than the stated 5% (Finkelstein and Sims, 2001; Salesky et al., 2012), and no flux detection limit is provided (Billesbach, 2011).

We have not integrated the small 2km segments but much longer segments which have been decomposed using wavelet analysis to discrete fluxes aggregated to 2 km. We were able to minimize these errors by choosing the long integration tracks, and low z/zi ratio. Naturally the individual 2-km data points obtained from wavelet decomposition would be associated with higher random error due to short

variability but we clarify this now in the text and expand the description of the errors (see also our response 3 to Reviewer 1). We thought we have provided sufficiently thorough although probably not complete uncertainty analysis, but in response to this comment we include more uncertainty sources which were generally minor compared to the sources we have discussed. Please note that the reported flux values are independent of the footprint analysis.

p. 7984 l. 21: The authors preserve the global covariance through considering wavelet coefficients above the cone of influence while attempting to offset edge-effects resulting from a limited length of the time series. How is this achieved by padding with zeroes? I would imagine that cyclic boundary conditions are less prone to these edge effects?
By padding with zeros and subsequently removing the zeros we do not report the data from outside the COI.

p. 7984 l. 27: Heat flux is used as spectral reference for BVOC fluxes. This assumes that (i) the frequency response of the temperature sensing element is sufficient, (ii) temperature is measured in the free airstream and not subjection to dampening effects from housing (e.g., Rosemount), (iii) radiation error has been corrected, (iv) adiabatic heating caused by aircraft propagation has been corrected, and (v) adiabatic heating caused by the aircraft vertical movements has been corrected (potential temperature at average flight level). Have these steps been considered?
There is a general problem with overcorrecting flux datasets which is why we decided not to apply these corrections which however can be more relevant for energy budgets, etc. (e.g. Friehe and Khelif, 1992). For the normalized co-spectra we do not think these issues matter so much. The purpose of showing these independent datasets is to show their similarities between the disjunct and continuous datasets.

p. 7985 l. 15: ≈20 m s−1 is a slow flying aircraft (e.g., van den Kroonenberg et al., 2008) and <<100 m is a flight altitude close to the surface for flux measurements (e.g., Zulueta et al., 2013).
Due to the size and power requirements of fast response VOC analyzers, VOC fluxes are made even on larger aircrafts, e.g. NCAR C130, NOAA P3 that typically have higher speeds (>100 m/s) and even higher altitude. Unfortunately it is currently not possible to make these measurements on the light aircraft that go lower and slower.

p. 7986 l. 19: How can measurements of reactive trace gas species be performed reliably after transporting an air sample through more than 500 m of tubing? I would imagine that the dark room reaction kinetics in the tubing is quite different from the ambient reaction kinetics?

We performed line loss tests on the reported compounds and found that the 500 m tubing length did not affect the measured concentrations. There were some effects for other (not reported) compounds, but these will be discussed in a separate manuscript which addresses the Walnut Grove observations more thoroughly.

p. 7986 l. 19: …Twin Otter…?

Done.

p. 7987 l. 12: What is the isoprene flux detection limit?

The detection limit (LOD) refers to an integrated flux rather than a discrete 2 km flux data point. For example for a typical stretch of 100 km the LOD for isoprene flux was approximately 0.01 mg m^{-2} h^{-1}.

p. 7987 l. 15, p. 7988 l. 6, p. 7988 l. 19: How was the measured concentration mapped to the landscape? I did not see the application of a concentration footprint model. Do the authors actually distinguish between flux and concentration footprint (e.g., Schmid, 1997)?

We do distinguish the concentration and flux footprint, but we refer to the concentration simply as the concentration measured at the aircraft altitude without considering any concentration footprint model or any scaling. We do calculate the flux footprints to derive spatially integrated emissions which we compare to the MEGAN 2.1 inventory in the companion paper (Misztal et al., GMD 2014). The airborne emission factor approach represents the flux normalized using Guenther et al. (2006) activity factor scaled to the surface. This approach does not use normalization for the footprint.

p. 7989 l. 2: PTR-MS doesn't measure fluxes.

PTR-MS instrument measures fast concentrations and we can say that the PTR-MS system (which includes fast wind data measurement) measures the fluxes. This has been clarified.

p. 7989 l. 8: Technical detail that can be moved to supplementary materials.

"Inter-comparison of concentrations from PTR-MS and GC-MS" has been moved to Supplementary information.

p. 7989 l. 20: Repetition from Sect. 2.7.1 that can be omitted.

The repetition has been omitted.

p. 7990 l. 26: …emission strength…?

Emission factors.

p. 7991 l. 1: Here it says 1–10 mg m−2 h−1, while on p. 7990 l. 17 it says 1–15 mg m−2 h−1. Where do the differences originate from?

The range for emission factors (normalized for temperature and PAR) was narrower than the range for measured surface emission rates affected by higher temperatures, so both are correct.

p. 7991 l. 11: This is a very qualitative analysis, how does that address the manuscripts goal of testing emission models? Quantitatively relating measured emission strength to LAI and land cover type would be desirable.

We show here the ranges of LAI measured previously by Karlik et al. for the oak savannah above which we flew and we showed the ranges and spatial distributions of emission rates for isoprene. This is both qualitative and quantitative. The suggestion to relate directly the emission source strength to LAI is interesting, but would not be more quantitative. Furthermore, it would require reliable LAI measurement which we did not have on the aircraft, while satellite LAI is too coarse and the problem is how to distinguish the LAI of grasses which do not emit isoprene from the LAI of oaks which grow sparse but are enormously large isoprene emitters. A more thorough comparison of the measured fluxes to the modeled fluxes, and thus the land cover characteristics, will be done in the separate paper previously mentioned.

p. 7991 l. 25: What are the uncertainties around these values?

We have added ±20% to the values for airborne flux and ±50% to the values for REA flux.

p. 7992 l. 6: Repetitive and qualitative, paragraph can be omitted.

We do not think this paragraph is qualitative or repetitive and consider it to be important for the manuscript.

p. 7992 l. 14: What is the source for temperature and radiation information to perform the normalization? At 400 m above ground, temperature, radiation and (passive vs.

active scalar) fluxes originate from very different source areas. How do you take this into account? Also, Have atmospheric corrections been applied?
The temperature and PAR datasets measured from aircraft could enable higher accuracy for the absolute emission factors, but for consistency of the model-measurement comparison approach, we use the same temperature and PAR data source as those used by the model (Guenther et al., 2012). This is also addressed more thoroughly in our separate manuscript on measurement-model comparisons.

p. 7992 l. 23: No reference is provided for Misztal et al. (2014).
This is now provided.

p. 7993 l. 1: Paragraph wholly qualitative, not living up to the manuscripts goal.
As we mentioned in the beginning our goal for this paper was not to show the quantitative model-measurement comparison, but report the airborne fluxes and concentrations from oak woodlands.

p. 7993 l. 10: This is a nice summary. But what about the potential future impacts of the summarized activities?
Future impacts include the improvement in the landcover and accuracy of the biogenic models. We expand the conclusion to show the future impacts and directions.

p. 7994 l. 15: The bibliography should be as consistent as possible. For example, Lenschow (1986) appearing before Lenschow et al. (1980), or names being spelled out as Lenschow, D. one time and Lenschow D.H. the other time.
The references have been made consistent.

p. 8004 Fig. 2: Technical detail that can be omitted.
This technical detail could probably be omitted but it shows how that airborne flux integration is done and could be interesting for new airborne flux enthusiasts. However, we have moved the figure to the supplementary information.

p. 8005 Fig. 3: Optically, the space series of instantaneous fluxes doesn't appear to match the coefficients of the wavelet cross-scalogram very well. For example, at ≈45 km and towards the end of the track at at ≈110 km.
We have changed the color scale to visualize better the time-resolved wavelet co-spectra.

p. 8006 Fig. 4: What about comparing the co-spectra to a reference model?
We agree that this is worthwhile and the revised figures show now the traces for the ideal cospectra.

p. 8007 Fig. 5: How exactly is the difference between net flux and turbulent flux (vertical flux divergence/storage) being calculated? The ratio appears to decrease towards the end of the flight track. How can this be significantly supported by the profile measurements?
This is because of the increasing PBLheight. The coefficients from profile measurements are kept constant and the storage terms as we showed in the text was negligibly small.

p. 8008 Fig. 6: Not really meaningful, as source areas for the concentration measurements are different from the location of the aircraft. A spatial projection is required for that purpose.
The calculated footprints were of the order of 1 km. On these maps the difference in the footprints would be hardly noticeable. We think these projections are meaningful and show the true measured distribution of VOC concentrations.

p. 8012 Fig. 7: This is OK for an overview. However, for the claimed observation-model inter-comparison, the quantitative agreement with the MEGAN emissions in the flux footprint would have to be shown. For example, an error-in-variables regression.
Yes, and we do show this in our companion paper.

p. 8012 Fig. S2: Looks nice, but doesn't really tell anything about vertical flux divergence, a term that is sought for to prove significance of the stated emissions.
This supplementary figure shows the ground-tower vs airborne concentration comparison, not the fluxes. The tall tower and the figure was not designed for studying the flux divergence.

References

Bange, J., Zittel, P., Spiess, T., Uhlenbrock, J., and Beyrich, F.: A new method for the determination of area-averaged turbulent surface fluxes from low-level flights using inverse models, Boundary Layer Meteorol., 119, 527-561, doi:10.1007/s10546-005-9040-6, 2006.

Billesbach, D. P.: Estimating uncertainties in individual eddy covariance flux measurements: A comparison of methods and a proposed new method, Agric. For. Meteorol., 151, 394-405, doi:10.1016/j.agrformet.2010.12.001, 2011.
Finkelstein, P. L., and Sims, P. F.: Sampling error in eddy correlation flux measurements, J. Geophys. Res. Atmos., 106, 3503-3509, doi:10.1029/2000JD900731, 2001.
Foken, T.: Micrometeorology, Springer, Berlin, Heidelberg, 306 pp., 2008.
Griffis, T. J., Zhang, J., Baker, J. M., Kljun, N., and Billmark, K.: Determining carbon isotope signatures from micrometeorological measurements: Implications for studying biosphere–atmosphere exchange processes, Boundary Layer Meteorol., 123, 295-316, doi:10.1007/s10546-006-9143-8, 2007.
Mahrt, L.: Surface heterogeneity and vertical structure of the boundary layer, Boundary Layer Meteorol., 96, 33-62, doi:10.1023/a:1002482332477, 2000.
Mason, P. J.: The formation of areally-averaged roughness lengths, Q. J. R. Meteorolog. Soc., 114, 399-420, doi:10.1002/qj.49711448007, 1988.
Nordbo, A., and Katul, G.: A wavelet-based correction method for eddy-covariance high-frequency losses in scalar concentration measurements, Boundary Layer Meteorol., 146, 81-102, doi:10.1007/s10546-012-9759-9, 2012.
Raupach, M. R., and Finnigan, J. J.: Scale issues in boundary-layer meteorology: Surface energy balances in heterogeneous terrain, Hydrol. Processes, 9, 589-612, doi:10.1002/hyp.3360090509, 1995. Salesky, S., Chamecki, M., and Dias, N.: Estimating the Random Error in Eddy-Covariance Based Fluxes and Other Turbulence Statistics: The Filtering Method, Boundary Layer Meteorol., 144, 113-135, doi:10.1007/s10546-012-9710-0, 2012.
Schmid, H. P.: Experimental design for flux measurements: matching scales of observations and fluxes, Agric. For. Meteorol., 87, 179-200, doi:10.1016/s0168-1923(97)00011-7, 1997.
Sühring, M., and Raasch, S.: Heterogeneity-induced heat-flux patterns in the convective boundary layer: Can they be detected from observations and is there a blending height? A large-eddy simulation study for the LITFASS-2003 experiment, Boundary Layer Meteorol., 1-23, doi:10.1007/s10546-013-9822-1, 2013.
van den Kroonenberg, A., Martin, T., Buschmann, M., Bange, J., and Vorsmann, P.: Measuring the wind vector using the autonomous mini aerial vehicle M2AV, J. Atmos. Oceanic Technol., 25, 1969-1982, doi:10.1175/2008JTECHA1114.1, 2008.
Wood, N., and Mason, P.: The influence of static stability on the effective roughness lengths for momentum and heat transfer, Q. J. R. Meteorolog. Soc., 117, 1025-1056, doi:10.1002/qj.49711750108, 1991.
Zulueta, R. C., Oechel, W. C., Verfaillie, J. G., Hastings, S. J., Gioli, B., Lawrence, W.

T., and Paw U, K. T.: Aircraft regional-scale flux measurements over complex landscapes of mangroves, desert, and marine ecosystems of Magdalena Bay, Mexico, J. Atmos. Oceanic Technol., 30, 1266-1294, doi:10.1175/jtech-d-12-00022.1, 2013.

We have found some of the references cited by the referee useful and included them in the revised manuscript. Once again, we would like to thank the referee for his/her time to review our manuscript which has now been made significantly clearer.

Pawel K Misztal on behalf of other co-authors.

References:

Argoul, F., Arneodo, A., Grasseau, G., Gagne, Y., Hopfinger, E. J., and Frisch, U.: Wavelet Analysis of Turbulence Reveals the Multifractal Nature of the Richardson Cascade, Nature, 338, 51-53, Doi 10.1038/338051a0, 1989.

Collineau, S., and Brunet, Y.: Detection of Turbulent Coherent Motions in a Forest Canopy .1. Wavelet Analysis, Boundary-Layer Meteorology, 65, 357-379, 1993.

Detto, M., Baldocchi, D., and Katul, G. G.: Scaling Properties of Biologically Active Scalar Concentration Fluctuations in the Atmospheric Surface Layer over a Managed Peatland, Boundary-Layer Meteorology, 136, 407-430, DOI 10.1007/s10546-010-9514-z, 2010.

Friehe, C. A., and Khelif, D.: Fast-Response Aircraft Temperature Sensors, J Atmos Ocean Tech, 9, 784-795, 1992.

Horst, T. W., and Weil, J. C.: Footprint Estimation for Scalar Flux Measurements in the Atmospheric Surface-Layer, Boundary-Layer Meteorology, 59, 279-296, Doi 10.1007/Bf00119817, 1992.

Karl, T. G., Spirig, C., Rinne, J., Stroud, C., Prevost, P., Greenberg, J., Fall, R., and Guenther, A.: Virtual disjunct eddy covariance measurements of organic compound fluxes from a subalpine forest using proton transfer reaction mass spectrometry, Atmos Chem Phys, 2, 279-291, 2002.

Karl, T., Misztal, P. K., Jonsson, H. H., Shertz, S., Goldstein, A. H., and Guenther, A. B.: Airborne Flux Measurements of BVOCs above Californian Oak Forests:

Experimental Investigation of Surface and Entrainment Fluxes, OH Densities, and Damkohler Numbers, J Atmos Sci, 70, 3277-3287, Doi 10.1175/Jas-D-13-054.1, 2013.

Metzger, S., Junkermann, W., Mauder, M., Beyrich, F., Butterbach-Bahl, K., Schmid, H. P., and Foken, T.: Eddy-covariance flux measurements with a weight-shift microlight aircraft, Atmos Meas Tech, 5, 1699-1717, DOI 10.5194/amt-5-1699-2012, 2012.

Misztal, P.K., Avise, J., Karl, T., Scott, K., Weber, R., Jonsson, H.H., Guenther, A.B., and Goldstein, A.H.: Evaluation of regional isoprene emission estimates in California based on direct airborne flux measurements. In preparation for ACP, 2014.

Terradellas, E., Morales, G., Cuxart, J., and Yague, C.: Wavelet methods: application to the study of the stable atmospheric boundary layer under non-stationary conditions, Dynam Atmos Oceans, 34, 225-244, Doi 10.1016/S0377-0265(01)00069-0, 2001.