Review for "iDirac: a field-portable instrument for long-term autonomous measurements of isoprene and selected VOCs"

General comments

This is a well-written manuscript describing an important new instrument for quantifying atmospheric isoprene mixing ratios. In light of recent developments in remote sensing of isoprene (e.g. Fu et al. 2019), the need for a low-cost, field-deployable measurement of isoprene for satellite verification has become essential. This should be published after minor revision, as described below.

Throughout the manuscript, the authors employ a quadratic calibration curve for the instrument without sufficient discussion of the physical cause of the non-linearity, especially as they present curves with both positive and negative coefficients for the 2nd order component of the equation (see Fig 16). In light of the overall uncertainties associated with this measurement, it would be enlightening to understand if the non-linearity is statistically significant or if a linear function would be adequate here.

What is conspicuously absent from the manuscript is a figure showing 1) a chromatogram for ambient air, to allow the reader to evaluate the validity of the peak area fitting algorithm in "real" sampling conditions and 2) an intercomparison with a well-established instrument / method for isoprene to evaluate the technique for artifact in measurement of ambient air [e.g. Barket et al., 2001]. This prevents the reader from being able to determine if the signal response of the instrument in ambient air is responding only to isoprene; because of this, the accuracy of the instrument measurement is not well-demonstrated. The underlying data presented here appears to be sufficient for the authors to provide at least two quantitative assessments of the system accuracy that are missing:

1. The diurnal profiles of isoprene shown in Figures 14 and 15 show very low observed mixing ratios for isoprene at night, indicating that species with long atmospheric lifetime relative to isoprene do not present a significant source of interference. Some statistical evaluation of the night-time / pre-dawn data may provide some bound for this potential interference to the method presented.
2. The peak-fitting technique used to determine isoprene peak area should also provide additional information, such as the fitted peak width (FWHM) and uncertainty of the peak fit area calculation. These can be used to assess the quality of the chromatographic peak, both to evaluate for co-eluters and to account for bias from changes in peak shape (e.g. tailing).

Some of the technical discussion and the number of figures seems excessive for the manuscript, and the authors should consider creating a supplemental information addendum to the main manuscript to move some of this material (see below for specific suggestions).

Specific comments

P1, Line 23. Introduction. As noted above, a recent publication [Fu et al. 2019] describes isoprene retrieval from space-based observation. The uncertainties for this method are typically 10-50%, and therefore the iDirac instrument may provide a suitable means to calibrate the satellite measurement. The above reference serves as a useful basis for defining acceptable uncertainties for an isoprene measurement, and the authors are encouraged to cite it in the introduction.

P4, Line 12. Figure 2. The iDirac plumbing schematic shows the sample trap as a black line between two numbers ports on the Valco valve. This resembles a valve "jumper" – a short piece of tubing connecting two ports – that is the typical convention, and can lead to some confusion. The authors should draw the sample trap as a separate device outside of the valve to make the figure clearer.

P5, line 9. "the trap is flash-heated to approximately 300 °C" There is no description of how this temperature is measured. Please describe the temp sensor and sensor location relative to the trap adsorbent.

P5, line 12. "large bulky molecules" Either "large" or "bulky" is sufficient here.

P5, line 38. "a coiled nichrome wire heating element surrounding the section of the [stainless-steel] tube" Is the heating element in contact with the trap tubing? It would seem that the heater would short across the trap if it is in contact. Can the authors describe how (or if) this is avoided?

P5, line 29. "Flow restrictors upstream from valves 3, 4, 6 and 7" The authors use small diameter tubing (0.005" ID and 0.0035" ID) rather than critical orifices to restrict reagent gas flow. Can they provide any comment on the long-term performance of this method? This reviewer uses critical orifices in this situation, which have been found to require occasional servicing.

P5, line 39. "Carboxen 1016" I believe the manufacturer has renamed this adsorbent to Graphsphere 2016.

P6, line 5. "heated to 40 °C" Can the authors provide any statistical description of the actual temperature stability of the oven? Peak retention time is directly related to this temperature, and therefore this is a critical variable in the instrument.

P6, line 11. "so that the precolumn is back-flushed" Is this precolumn heated above trapping temperature during the back-flush? It is not clear if the larger molecules are successfully removed during back-flush or if they accumulate.

P6, line 30. Figure 3. Please provide more information (i.e. isoprene mixing ratio, diluent gas, sample humidity) for this example chromatogram.

P7, line 17. The flow through the instrument is driven by either upstream pressure (in the case of the nitrogen and calibration gas flows) or by the pump box (in the case of Samples 1 and 2). Is the sample pressure measured? From the description here and in Figure 2 (and associated text) it is not clear if the adsorbent trap experiences higher pressure when the cal / blank solenoid valves are actuated, versus when using sample inlets 1 / 2. It is the reviewer's experience that adsorbent trapping efficiency has a pressure-dependence. Has this been observed by comparison of calibrant gas addition via the cal port versus via a sample port?

P7, line 40. "controls the altimeter pressure sensor" Is this the same as the differential pressure sensor described previously (P7, line 21)? And does the Arduino board control this sensor, or just read it?

P8, line 22. "Figure 5 shows a flow diagram" This figure could be moved to a supplemental materials section to conserve space in the main document.

P9, line 2. "next step is to locate the isoprene peak and to fit a Gaussian curve" It is not clear how this is performed, and could use some more explicit description. Is the user manually locating the peak or is the software algorithm doing this? How is the baseline for the Gaussian curve defined? Is there any evaluation of the goodness of the Gaussian fit to the data? There are instances in data figures throughout the text where negative peak areas or mixing ratios are shown. Is this from a fitted Gaussian with negative peak height?

P9, line 6. "A quadratic curve is fit to this data, which captures any slight deviations from linearity." As noted above, the use of a non-linear fit to instrument response should have a physical explanation, especially as both positive and negative coefficients are shown in the text. What is the uncertainty of the second-order coefficient? What amount of additional uncertainty would be added the overall measurement by simply using a linear fit of the calibration data?

P9, line 12. "The Gaussian function has certain boundaries set, to further ensure that it is fitted to the correct peak." Please describe exactly which boundaries are set. How is the Gaussian fitting function constrained?

P9, line 15. "it can be estimated using the column temperatures" This implies that column temperature is not constant. The temperature variance in ambient sampling should be described (see comment above).

P10, line 1. "This type of treated cylinder exhibits very good long-term stability for a number of VOCs (Gary Barone et al. Restek Corporation, 2010)." I don't understand this reference, and it is not listed in the Reference section.

P10, line 2. "The exact isoprene amount fraction in the secondary standard is determined by validating it against the NPL primary standard." Can the authors make a statement about the stability of their secondary standards over time? This

reviewer has found that secondary isoprene standards made in Aculife [Air Liquide] cylinders can degrade over the timespan of months.

P10, line 5. "mixing rations" should be mixing ratios

P10, lines 10-20. "Calibration frequency is specified by the user . . ." While the explicit description of the instrument calibration method is welcome, it is probably more appropriate to move this material, along with Figure 6, to a supplemental materials section.

P10, line 26. "a random mixture of 3, 6, 12, 24 and 48 mL samples" It appears that the calibration sample volumes used are always significantly smaller than the ambient air sample volumes. Since breakthrough volume is a critical parameter of the sample trap (Section 5.1 of text), I am curious as to why the calibration does not include a sample volume larger than ambient volume.

P10, line 30. "The equation for the quadratic fit allows the determination of the fractional isoprene amount in the samples by extrapolation or interpolation" The use of a quadratic calibration curve with extrapolation seems especially susceptible to the introduction of additional error in reporting mixing ratios.

P11, line 1. Figure 6. The linear fits of the peak areas for the individual calibration volumes versus time do not seem appropriate for this figure, since they are not used in the calibration procedure described in the text (peak area versus calibration sample volume on a weekly time scale) nor discussed in the text. The peak areas of 48mL calibration samples show a significant decay with time that is not apparent in the smaller calibration volumes. Is this statistically significant, and, if so, does this imply breakthrough for large sample volumes?

P11, line 9. "The x-axis ('Effective Calibration Concentration') consists in the calibration volume (in mL) multiplied by the isoprene concentration in the gas standard (in ppb)." Isn't this simply calibrant mass or moles, then? Are you attempting to convert these calibration points to equivalent mixing ratio in ambient air? After solving for the 'Effective Calibration Concentration' based upon ambient peak area, does the user then divide this 'Effective Calibration Concentration' by the sample volume to determine the ambient mixing ratio? The authors should provide the equation for converting peak area and sample volume to mixing ratio.

P11, line 15. "The precision of the instrument was determined as the relative standard deviation in isoprene peak area from calibration chromatograms with the same user-specified volume . . ." In light of the temporal trend in sensitivity discussed later in the text (5.3.2), does this overestimate the precision uncertainty? The authors do not define the length of time that is averaged here, but it appears that the instrument sensitivity decreases on the time-scale of weeks, and therefore we would not expect the peak area here to be a constant.

P12, line 4. "The accuracy of the instrument is dictated primarily by the uncertainty in the isoprene amount fraction in the NPL standard, and how this is propagated to the isoprene amount fraction in the secondary gas standard used in the field." Because the authors have not fully demonstrated that this technique is adequately selective to isoprene, I don't think this statement is correct. Certainty the uncertainty of the calibrant mixing ratio is a component of accuracy, but perhaps not the primary driver. The intercomparison of two iDirac instruments presented in 5.1 would indicate this as well.

P12, line 9. "XLGENLINE, a generalised least-squares (GLS) software package for low-degree polynomial fitting (Smith, 2010) is used to estimate the final uncertainty in the isoprene amount fraction in the secondary calibration cylinder by inverse regression from the calibration curve. For most secondary calibration cylinders, this is estimated to ~ 7% at the k = 2 level (providing a coverage probability of approximately 95%). A similar procedure is applied to calibration and sample data from the field to estimate the uncertainty in the ambient isoprene concentrations. This is estimated to ~20-25 % at the k = 2 level." I found this discussion to be difficult to assess, and reads as simply reporting the output of a statistical software package. Can the authors put this into more explicit terms, e.g. describe what is meant by "inverse regression from the calibration curve"? Also, the comment "most calibration cylinders" is vague – how often do secondary cylinders fall outside this accuracy range? It's not clear how this technique is extended to the field data; again, please describe the analysis in detail. Finally, it should be noted that the authors here use a 2-σ

uncertainty to define accuracy, but a 1-σ uncertainty to define precision above. One uncertainty level should be used for both.

P12, line 15. Figure 8. The test cylinder mixing ratio has a negative mixing ratio at the start of the experiment shown. Is this real data, or just an artifact of the analysis? Perhaps the data plot should start after the first standard peak area, at 21:00, 21-Feb-2018.

P12, line 19. ". . . the high concentration of isoprene would risk *poisoning* the adsorption trap." Italic emphasis is mine. Did you intend to use the work "poisoning" here? I interpret that this to mean an irreversible change to the adsorptive strength of the trap, while I believe that you mean that the trap would demonstrate breakthrough due to non-ideal behavior, as described by Peters and Bakkeren (1994).

P12, line 31. "This is identified as the limit of detection and is calculated for two versions of the iDirac, the grey and the grange instruments." I found this discussion a bit hard to follow since results from these two instruments were presented without introduction. A sentence or two from the later discussion of these instruments in Section 5.1 would be helpful here. Alternatively, simply cite the LOD for the grey instrument here and discuss the higher LOD observed for the orange instrument in section 5.1.

P12, line 32. "This difference is attributed to the traps used (i.e., a trap with more adsorbent would retain more analyte, resulting in a larger signal)" Doesn't this imply that the traps are not quantitative? If the amount of analyte retained is proportional to the adsorbent mass, this means the traps have a consistent breakthrough, rather than a critical sample volume as described by Peters and Bakkeran (1994). Data provided in the next section (e.g. Figure 11) shows this not to be the case. This explanation does not seem reasonable.

P13, line 18. "This under-reading is likely due to differences in the absorbent trap, leading to a lower sensitivity for the orange instrument. This is supported by the calibration curve for the orange iDirac, which curves more at high concentrations, resulting in lower peak height than in the grey iDirac for the same concentration. Another artefact of this is that the noise visible on the orange output is greater." I am not persuaded by this argument. Why wouldn't calibration account for these differences, irrespective of the curvature of the calibration? Applying a straight-edge to the scatterplot of the two instrument responses shown in Figure 10, I see no significant curvature, i.e. the 6.6% difference appears consistent across the mixing ratio range. See also previous comment for my concern that the adsorbent trap mass is used to explain accuracy differences. The authors may wish to consider alternative explanations, e.g. if the pressures during calibration are the same in both instruments (see comment for P7, line 17).

There are two qualitative statements presented here that should be quantified: "curves more at high concentrations" and "noise visible on the orange output is greater."

Finally, it should be noted that the two instruments perform within the specified accuracy for this system (7%), and therefore the performance demonstrated here is acceptable.

P14, line 4. Breakthrough tests. The results presented in this section present a very nice demonstration of the adsorbent trap showing ideal behavior with a (presumed) single-component isoprene mixture. The citation for this experiment [Peters and Bakkeren, 1994] observed that the breakthrough volume observed in this sort of experiment will be higher than when a using a test mixture that includes other compounds, which "may have a pronounced influence on the BTV and thereby on the safe sampling volume of single compounds." The authors should specify if the test mixture included other species expected in ambient air. If it did not, they should consider how valid the results presented here will be for collection of ambient air.

P14, line 7. "Each run sampled 10 mL" Did you mean this? Since you specify varying sample volume in figure 11, I assume this to be a typo and should read "sampled 10 mL min^{-1}."

P14, line 8. "isoprene mixture of known concentration" Perhaps better to say "isoprene mixture at constant concentration."

P14, line 16. Co-elution of interfering species. I found this section to be a bit trivial, due to the somewhat arbitrary list of species used for this test. I wonder if many of the alkenes used for this work (e.g. 2-methyl-1-pentene) would be

expected to be found in ambient air at significant mixing ratios to be of importance as an interferent. The authors note "Work is ongoing to determine the elution time of a wider range of compounds, including oxygenated products from the oxidation of isoprene." Since the PID is sensitive to many species beyond alkenes (e.g. ketones, aldehydes), I'm not sure if the work here is very conclusive in demonstrating that there would be no interference when measuring ambient air. I would suggest moving this section to a supplemental materials section, or revising with a more exhaustive list of species.

P15, line 23. "This is an artefact of the trap adsorption process and the resulting tailing of the peak." This is a nice explanation of the observed phenomenon. Have the authors considered if the tailing for peaks in ambient air relative to peaks from calibration runs leads to a bias in the peak areas? That is, a Gaussian fit function may miss some of the peak tail for ambient runs relative to calibrations, leading to consistent under-reporting of ambient mixing ratio. This effect could be tested by integrating a subset of peaks to capture all peak area, or by using an exponentially-modified Gaussian fit function.

P16, line 11. "Sabah (Malaysian Borneo)" Could the authors be a bit more specific for this location? Sabah is roughly the same size as Scotland.

P17, line 7. "Error! Reference source not found.." I believe this is a reference to Figure 13.

P17, line 15. Figure 14. For the time series shown, there is a significant time period (16-Nov-2015, 19:00 - 2200) where the mixing ratio trace is significantly below zero. Is this correct or a plotting error? Is this attributable to the "several issues with instrument function"? If the latter, perhaps this time period should be referenced in the text.

P18, line 12. "5.23.2 Results and discussion" Should read 5.3.2 Results and discussion.

P18, line 17. Figure 15. There is a drop in the mid-canopy time series on the morning of 7-Nov-2018 that seems anomalous with the other mixing ratio traces and other days shown. The authors may want to confirm that those data pass quality assurance checks.

P19, line 2. Figure 16. The calibration curves presented in this figure give me particular concern, as there appear to be consecutive weeks where there is a positive and then negative curvature. Does this seem physically possible? Generally, I found this figure difficult to interpret as I am color-blind and the color scale to distinguish weeks is quite subtle to me. If the underlying data were re-fit with a linear calibration function and a time-series of the slope were presented instead, I suspect that this figure would be much more effective. The x-axis "Effective Calibration Concentration" should have units of mass or moles, as noted for Figure 7.

P19, line 3. "This drift is attributed to the gradual degradation of the trap" How do the authors separate loss of trapping efficiency from loss of detector sensitivity? If a new sample trap returned sensitivity to the Week 1 level, that would be a useful data point to show.

P19, line 26. "a more sophisticated and interactive control over the oven temperature." This statement begs for a presentation of underlying data (oven temperature time trace, retention time drift) to justify this need. Possibly, either could be added to field data time traces in Figure 6, 14 and/or 15 on a secondary axis.

P20, line 9. References. At least two references (Allen et al., 2018; Helmig et al., 1998) are incomplete. The Barone et al., 2010 reference from the text is missing; see previous comment for P10, line 1.

References

Barket et al., J. Geophys. Res., 106: D20, 24301-24313 (2001), doi.org/10.1029/2000JD900562
Fu, D. et al., Nature Communications, 10: 3811 (2019), doi.org/10.1038/s41467-019-11835-0
Peters and Bakkeren, Analyst, 119, 71-74 (1994), doi.org/10.1039/AN9941900071