Overall – the method appears potentially impressive from the data presented here, but the quality of the paper is let down by some imprecise writing and concerns around cherry picking of data for presentation. It feels that it needs some sort of controlled release validation to be fully compelling as a go-to method of quantification. Although probably out of scope of this work (given that this is part of the CoMET special issue) this quantification method should be tested with data collected at blind control release experiments to ensure that it is truly capable of describing real emissions, and therefore able to correctly understand the "real" uncertainty rather than the mathematical uncertainty within these quantifications. As it stands, I feel the authors need to explain some parts in more detail to alleviate concerns and make some significant improvements to the quality of writing throughout.

It will be suitable for full publication once the problems are ironed out.

Major thoughts

The volatility of emissions from mining operations is not surprising given the variable conditions and operational mine venting. The aircore system is a well established method of data collection and it is good to see it being able to capture the emissions nice and clearly. I have concerns about the data quality though for being able to fully resolve the mass balance with any certainty. In Figure 3a, the plume is only intercepted once along a single transect and is not fully bounded with background air to the South. In Fig 3b, there appears to be two hotspots, with the main lofted plume not resolved to the West at all. This is one of the downsides of not measuring in-flight and only being able to measure post-flight. Whilst the uncertainties of the GA-IPPF are low, it needs to be stated that we do not know the error in the accuracy of the method as we have no pre-defined truth to compare to.

The issue of quantification of methane from the energy sector has the potential to become potentially political and legal – with the concept of emissions levies or preferred contracts based on quantified methane emissions being used as a legal instrument. It is therefore of upmost importance that the authors of novel methodologies for quantification such as those shown here are aware of how their methods may be utilised in the future, potentially by 3rd party commercial companies, and ensure that any issues are fully declared (e.g. why are only 2 flights used to verify the method, were there occasions when this method failed, and why? There are 15 flights stated on the experiments section). They should also check that all references to accuracy are truly discussing accuracy (closeness to a true value) – and not precision (closeness to another measurement or model).

If this method is suitable, I think a comment section needs to be added discussing other sorts of data that the team would expect this method to work with. I'm assuming that on-board drone measurements downwind from landfill / industrial sites would be a good option, would mobile measurements from vehicles potentially work?

Minor

I am far from a professional proof-reader, and would recommend that this manuscript is looked over by a proofer after corrections as there are numerous tense and grammar issues that need resolving to make the paper read as desired.

L20: There are plenty of monitoring methods – but very little verified quantification methods suitable for coal mines.

L34: Grammar. Release is concerning.

L38: Why are BU only useful for strong sources?

L42: This seems to be a common misuse / expectation of a BU inventory. They are not intended to be able to capture variable emissions, but are a statistical average expectation. It is only worth considering inventories compared to spot measurements if there is some valid statistical analysis (either temporal, or numbers of sites)

L44: What improvements have suddenly made this possible?

L46: Tense: should be is capable of obtaining.

L59: But most aircraft equipped with any CH_4 sensor are able to achieve ppb precision and are perfectly capable of measuring downwind flight plans from coal mines. Many aircraft are capable of this such as Scientific Aviation etc…

L62: What about ground based eddy covariance?

General: to L70. This section feels unnecessarily negative about the capabilities of the rest of the scientific community with regards to being able to measure emissions from coal mines. There are several methods discussed here that I would envisage perfectly capable of making precise measurements that could enable a quantification of emission estimate.

L73: What does high applicability mean in this sense?

L74: How will it have less uncertainty if the inputs still have the same uncertainties attached to the actual measurements? The discussion to L83 makes it seem that these important atmospheric parameters can be discarded in favour of a set of model parameters? From experience at controlled release experiments small changes in wind direction and other meteorlogical conditions can have a dramatic effect on the plume behaviour.

L106: Is the accuracy only 20ppb using a G2401-m? This is concerning, why is it so large? What is the location accuracy of the sampling, how much does the sample bleed into itself over the course of a flight? Is the mixing in the aircore linear across flight time?

L117: correlation not connection? How good a correlation?

L135: I'm not 100% sure what alpha represents, please clarify (with a reference if possible)

Approx. L160. Is there prioritization in the fitting process? Are some variables given priority due to their certainty? General question about the process, if the inputs are very close to the outputs, then I presume that the standard gaussian plume quantification would also be very close?

L190: If these are spiral patterns, then the figure doesn't make this very clear. Can these be replotted to make it clear if that is the case?

Para 218: This paragraph needs tidying up, it is unclear as it currently reads and has numerous typos. The reconstruction is impressive – especially with two peaks in Flight 15 so it is important that this section is as simple and clear as possible.

L250: If the Gaussian plume doesn't account for background, has the dataset been adjusted to make that correction?

Approx L265: I don't know if this is possible, but it would be incredibly helpful if there was some sort of visualization of the CH_4 atmospheric concentration for each of the methods of quantification (e.g. what does the plume look like in 2-D so that the variability can be understood). This would be most helpful for where there is significant differences between the methods to show what the plume visualisation looks like in each of the quantifications.

L330 (approx.) General discussion of measurement stability – the variation in wind parameters, disturbance in the airflow from the drone, changes in plume behaviour are all potential problems for quantification. Are the random errors used here sufficient to capture the potential uncertainty and can they be justified with reference to uncertainty from other studies?

L381. The claim that this result would guarantee emission calculation accuracy to better than 99.2% seems very over confident. This type of statement brings me back to the knowledge that methane emissions may be used as legal instruments in the near future and claims of such accuracy when not compared to a blind control is not acceptable. There needs to be a clear rethink of the use of accuracy, error and precision throughout so that it is clear what is being compared. As there is no "known" value, no claim on accuracy can be made.

L401. The conclusions feel very generic and non-specific in large parts to the work shown here, they should be reconsidered with the value added of GA-IPPF in mind, rather than generalities of coal emissions. It is good to see that the model is being considered for other uses too, is there any possibility that this could be expanded on in the main text to demonstrate how this would be done for vehicles?