This manuscript describes the development and evaluation of a new cluster-based kriging approach for mass balance calculation of facility-level emission rates, using UAV-based measurements. Easily-deployable techniques that enable accurate determination of facility-level emissions have an important role to play in both improving emission accounting (e.g. national emission inventories) and compliance monitoring, as many countries move towards tighter regulation of methane emissions. The results of the controlled-release experiments detailed in this study suggest that this approach could be an important tool in such efforts going forward, especially as lightweight instrumentation for accurate CH_4 measurement becomes widely commercially available. Comparisons of emission estimates calculated using different measurement techniques, kriging approaches and assumptions regarding the local wind field are instructive for future studies, as is the comparison against the established OMT33A method. The paper is in general very clearly written, and I suggest that it should be published in AMTD after the following minor points have been addressed.

In general I've tried not to repeat comments already made by the other reviewers, but I do agree with RC1 that a brief discussion of calibration is required. I also think it would help to clarify things if the term "CH_4 enhancement" were to be used in cases where background values have been subtracted from the data (which I think is pretty much everywhere). Perhaps a couple of extra sentences briefly summarising the application of the REBS algorithm would be useful too (especially with regard to the RC1 question concerning where the background measurements were taken - I assume the answer is anywhere on the downwind measurement plane that the REBS algorithm identified)?

L27 - Alvarez et al. (2018) would be a more appropriate reference here (Gurney et al., 2021, is definitely wrong), although there are more recent options that would do the job too.

L181 - this reference (U.S. EPA, 2014) needs working into the sentence, which currently doesn't make sense without the bracket.

L202 - the star on the friction velocity should be a subscript (as in Equation 4).

L226 - I'm a bit confused as to why this step was necessary. If both the QCLAS and RTK-GPS received GPS signals, why were they not already synchronised on GPS time?

L230 - in addition to my general point above, it would probably be clearer to say that background CH_4 mole fractions were "subtracted" instead of "removed".

Equation 5 - I understand that this approach is based on previously published work, but the application is sufficiently different that it would be useful to provide some more information here. I suggest explicitly stating the form of **F**. Have I got it right that the parameter vector **b** consists of the QCLAS measurements? If so I would also state that explicitly.

Figure 4 - Somewhere in either the caption and/or the associated main text it should be explicitly stated that these values were optimised separately for each flight.

L315 - maybe I missed something, but is it explained anywhere how the data are hard-clustered prior to performing ordinary kriging?

L322 - I have no doubt that the Matèrn covariance kernel is a valid choice here, but as a general comment I feel that the choice of kernel should be based on an examination of the specific dataset on which kriging is being performed (although of course it can be guided by previous studies/experience). I'm sure that such examination was performed (i.e. someone checked to make sure the optimised function was a reasonable fit to the data for each flight) - I'm happy to leave it up to the authors as to whether stating this explicitly would be useful or not.

L324 - was anisotropy in the hyper-parameters considered? My prior assumption would be that the vertical and horizontal length scales could be quite different, but perhaps that was found not to be the case here?

Equation 15 - this is a really minor point, but just to make sure I've understood things - is y not already included in the set X?

Figure 5 - I agree with RC1 - this would be best split into two separate figures. Also, the grey outline on the circles in Fig. 5a needs to be removed, as you currently have to zoom in a lot in order to see the fill colours of each point.

L431 - I'm not sure if this is the best place for it, but I think it is worth mentioning somewhere that there are alternative ways to deal with this smoothing problem. One approach is to select a variogram model that results in nearby points being assigned large weights (e.g. a linear model). Such a model must obviously be supported by the experimental variogram, but in any case a subjective choice must always be made regarding how the model parameters should be optimised to "best fit" the data. Therefore it can reasonably be justified that the model variogram should be chosen with a particular focus on representing the experimental data at small separation distances (see Kitanidis, 1997, for further discussion). A moving neighbourhood approach can also be adopted; in fact this is the default approach in the frequently used EasyKrig MATLAB package (e.g. O'Shea et al., 2014; Pitt et al., 2019). The cluster-based approach presented here has some advantages over these alternatives; in particular it removes many of the more arbitrary subjective choices associated with them. I think they are worth mentioning in this context, probably just a sentence or two would do.

L527 - Needs rephrasing. Could go for "As a general guideline, performing drone-based emission quantification of emission sources requires…"

L529 - Would it be clearer to say "at a downwind distance of less than 75 m"?

References

Alvarez, R. A., Zavala-Araiza, D., Lyon, D. R., Allen, D. T., Barkley, Z. R., Brandt, A. R., Davis, K. J., Herndon, S. C., Jacob, D. J., Karion, A., Kort, E. A., Lamb, B. K., Lauvaux, T., Maasakkers, J. D., Marchese, A. J., Omara, M., Pacala, S. W., Peischl, J., Robinson, A. L., Shepson, P. B., Sweeney, C., Townsend-Small, A., Wofsy, S. C. and Hamburg, S. P.: Assessment of methane emissions from the U.S. oil and gas supply chain, Science, 361, 186–188, doi:10.1126/science.aar7204, 2018.

Kitanidis, P. K.: Introduction to Geostatistics: Applications in Hydrogeology, Cambridge University Press, Cambridge, U.K., 1997.

O'Shea, S. J., Allen, G., Fleming, Z. L., Bauguitte, S. J.-B., Percival, C. J., Gallagher, M. W., Lee, J., Helfter, C. and Nemitz, E.: Area fluxes of carbon dioxide, methane, and carbon monoxide derived from airborne measurements around Greater London: A case study during summer 2012, Journal of Geophysical Research: Atmospheres, 119, 4940–4952, doi:10.1002/2013JD021269, 2014.

Pitt, J. R., Allen, G., Bauguitte, S. J.-B., Gallagher, M. W., Lee, J. D., Drysdale, W., Nelson, B., Manning, A. J. and Palmer, P. I.: Assessing London CO2, CH4 and CO emissions using aircraft measurements and dispersion modelling, Atmospheric Chemistry and Physics, 19, 8931–8945, doi:10.5194/acp-19-8931-2019, 2019.