Reviewer 1

Dear Reviewer:

We thank the reviewer for his thorough review of our manuscript. We respond to reviewers' comments sequentially. However, we follow a separate numbering sequence for clarity for the Overall comments and Detailed comments sections. The Response label identifies the changes made in light of the reviewer's comments.

Overall Comments

1. **Reviewer Comment:** The manuscript presents some very interesting analyses on evaluating inversions. I think the authors have tried to address most previous reviewers' concerns, and I agree with the author's decision to submit this as a technical note. The authors have provided additional details to improve the manuscript compared to the first version. I appreciated the IAOMI and can see the value of this in satellite-based inversions, given the much higher potential for overlapping information. I was surprised that the authors didn't mention this, given their affiliations.

   Response: We appreciate the reviewers' comments about our effort. In light of the crucial observation of the reviewer about overlapping information, we have now included the text about the utility of IAOMI for identifying the magnitude of the shared information in the measurements in the revised manuscript (please see lines 99-100).

2. **Reviewer Comment:** However, In its current form this manuscript is extremely difficult to read. I recommend this being published in GMD, if the authors are able to make this manuscript more coherent and expand on the discussion of the figures presented in Figs.4 -6.

   Response: We have made extensive corrections to the write-up in the manuscript and included more text explaining the evaluation of the figures in the revised manuscript.

3. **Reviewer Comment:** Finally, there are several typographical errors in the manuscript, which is somewhat common while providing a revised version. I have highlighted a couple, but I trust a proofread to fix any remaining errors. Below I provide some suggestions to help improve the manuscript.

   Response: Please see our response to point 2 under the **Overall Comments** section.

Detailed Comments

1. **Reviewer Comment:** Lines 31-41: this paragraph comes off as jargon heavy and assumed that readers know what all the terms that are mentioned are (e.g., polynomial Chas expansion, Sobol's method etc). The authors should provide some clarifications or delete this paragraph.

Response: Goal of this paragraph is to provide a literature review of the most ppromiment previous sentivity analyses approaches and why we need approaches that are specifically catered towards linear atmospheric inversion. We included this paragraph based on the suggestion of the earlier reviewer and provided the reference for the methods mentioned in lines 31-53. During that review round, we were not sure of including texts in lines 31-53 as it has little to do with trace gas inversions. In the revised manuscript, we have moved the text in lines 31-53 to the appendix (appendix A). Note that we are aware of the intricacies of the methods mentioned in lines 31-53.

2. **Reviewer Comment:** Line 46: Define "analytical closed-form solutions"

   Response: In the revised manuscript, we have included the text expanding on the meaning of the analytical closed-form solutions (please see lines 531-532 in appendix A).

3. **Reviewer Comment:** Eq 2: define what the summation is over. For instance, it could be space, time or both. This information is somewhat implicit, but I would appreciate if the authors would state this.

   Response: In the manuscript's construct, the summation is primarily over space for a given period (please see line 85). However, it can be both space and time. It depends on how an user is constructing the forward operator or footprint for the problem.

4. **Reviewer Comment:** Line 129. IAOMI not IOAMI right? The authors use IOAMI several times in the manuscript.

   Response: This is an error from our part. It should be IAOMI; we have fixed this in the revised manuscript.

5. **Reviewer Comment:** Line 139: More jargon. You could rephrase as "Jaccard similarity index which describes… "Similarly for the Ruzicka index.

   Response: Right. We have made the suggested corrections (please see lines 95-97). Please note that we have mentioned these names to provide connection between our proposed index and the established similarity indices in literature.

6. **Reviewer Comment:** Line 185: This is an example of how this could be particularly useful for satellite data.

   Response: We have now included the suggestion as text in the manuscript (please see lines 131-132).

7. **Reviewer Comment:** Lines 198-208: The authors do a good job at describing the terms in eq 9 and 10 but I think could a line or two about interpreting it. For instance, in both eqs the first term describes the observational constraint while the second term describes prior info (in eq 9) and information about fluxes (through X in eq 10). Also mention why we need the second terms in both equations.

Response: Following reviewer's suggestion, we have now provided a thorough description of equations 9 and 10 in the revised manuscript (please see lines 156-158).

8. **Reviewer Comment:** Line 215: Reads a little awkward. I'm guessing they are saying Beta needs to be estimated? Clarify.

   Response: Yes, Beta needs to be estimated, and we have clarified this in a simpler language in the revised manuscript (please see lines 158-160).

9. **Reviewer Comment:** Line 225: What does "entry" mean?

   Response: Entry in the context of this text implies each element in a matrix. This is a standard notation concerning matrices and their components. In the revised manuscript, we have mentioned that the entry indicates each element with the notation *ij* in the matrix (please see lines 170-171).

10. **Reviewer Comment:** Line 319: a great time to remind your readers what GSA means (even if it has been defined earlier).

    Response: Incorporated. Please see line 264.

11. **Reviewer Comment:** Line 328: Define DGSM

    Response: Please see lines 271-272 and appendix A.

12. **Reviewer Comment:** Line 375: missing word "for" between accounted and through?

    Response: Right. We have incorporated it (please see line 312). Please note that we have extensively corrected the writeup in the manuscript.

13. **Reviewer Comment:** Fig 3. This is an interesting figure but perhaps also misleading. Obviously all gridcells in the shaded region for a given observation don't contain equal information. I would request the authors to provide a complimentary figure showing the footprints.

    Response: We have now included the figure of the footprints as requested by the reviewer. Please see Figure 4, its caption. It is referenced in line 384.

Reviewer 2

Dear Reviewer:

We thank the reviewer for his thorough review of our manuscript. Our responses to reviewers' comments are given in sequential order. However, for clarity, we follow separate numbering sequences for the overall and specific comments sections. The Response label identifies the changes made considering the reviewer's comments.

1. The manuscript describes a method to evaluate the quality of a gridded CH4 flux field obtained by solving a Bayesian linear inverse problem. The method reduces to a set of computable metrics. The "quality" of an estimate is given by the posterior distribution which quantifies the uncertainty in it. The metrics that the authors propose could attribute the uncertainty to its various causes.

   Response: No response is required.

2. These metrics, some of which are couched in terms of sensitivity analysis, are part of the verification and validation process of an inversion. The challenge lies in the computations of these sensitivities – they involve sampling and cannot be done for high-dimensional problems, where the "parameters" are discretized spatiotemporal fields (as in the case of the paper). The innovation in the paper is that the authors derive analytical expressions for these metrics, exploiting the fact that atmospheric inversions admit exact, analytical solutions. They also adapt the sensitivity metrics to the peculiarities of atmospheric inversions. Armed with these computable and inversion-relevant metrics, they seek to explain (or "evaluate the quality of") an inversion for CH4 fluxes in the Los Angeles Basin during the days of the Aliso Canyon Gas Leak (AGL), published in JGR Atmospheres, 2019 (henceforth, Yadav et al, 2019).

   Response: No response is required.

3. Yadav et al, 2019 estimated CH4 fluxes in a spatiotemporally resolved manner. They did not explicitly include AGL, either as a known source of CH4, or as one that had to be estimated. However, they did manage to capture the enhancement of CH4 fluxes in the LA basin, as the plume was captured at some of the monitoring stations. In the current manuscript, they show that the new metrics can be computed using the inverse solution in Yadav et al, 2019. They do not illustrate how these metrics can be used, to answer a scientific question. Thus one is left wondering about the purpose of computing the metrics.

Response: We think AGL (in the reviewer's comment) means Aliso Canyon Leak (ACL) . We have premised all our responses that are associated with AGL by assuming that it stands for Aliso Canyon Leak (ACL). Based on suggestion of both the reviewers we have now included text explaining the utility of the proposed metrics in improving inversions (please see lines 422-438).

Overall comments

1. **Reviewer Comment:** The manuscript is somewhat carelessly written. There are unfinished sentences, missing commas, anomalous indentations and capitalizations. The authors should read over the manuscript and correct these errors – I point out some of them below.

   Response: We have made extensive corrections to the write-up in the manuscript and removed all the grammatical inconsistencies.

2. **Reviewer Comment:** The derivations in the paper can be involved, but nobody said spatial statistics was easy. I redid some of the derivations – the exposition is correct and clear.

   Response: We thank the reviewer for verifying our derivations.

3. **Reviewer Comment:** For a paper that seeks to "evaluate the quality of …", it does precious little of it. The authors compute, tabulate and plot the metrics, but do not illustrate how they may be used to answer scientific questions. For example, could the authors use the metrics to show how they managed to capture the effect of AGL? Was it caught by one monitoring site (GRA, which was near it) or by multiple ones? This may be part of the reason why one of the reviewers wonders if this is a new way of solving the inverse problem (it is not).

   Response: We think by AGL, the reviewer implies Aliso Canyon Leak (ACL).

   The leak signal was primarily observed at the GRA site. Its impact was not distinguishable at other sites. We can use the metrics proposed in the manuscript to identify sites that could capture the effect of the Aliso Canyon leak by looking at the sensitivity of fluxes to observations. In the revised manuscript, we have included text about the approach that can be adopted to identify impact of the signals produced by the Aliso Canyon leak in observations (please see lines 398-410).

4. **Reviewer Comment:** On the whole, a useful paper, as it provides scalable and computable forms of the metrics. An illustration of how these metrics could be used (e.g., to answer a scientific question, to ensure correctness of solution, to interpret results or to resolve

numerical issues) would be necessary to justify why these metrics are needed in the first place – and this illustration is missing. (As I will show below, this may not be very hard to do)

Response: We thank the reviewer for elucidating this point on utility. We have included more text in section 3.2 explaining the utility of each metric.

5. I look forward to the manuscript appearing in GMD, once its minor blemishes are fixed.

   Response: We thank the reviewer for recognizing our contribution. In the revised manuscript, we have fixed all the issues raised by the reviewer.

Typos and grammar (incomplete list; please read the paper carefully and fix it)

1. **Reviewer Comment:** Line 48: "…but also admits closed form solutions". Admits, plural

   Response: This is an error from our part. We have now extensively corrected the write-up of the previously submitted manuscript (please see appendix A line 532).

2. **Reviewer Comment:** Line 49: "As we have limited knowledge …." The sentence ends before it is completed – it's just an adverbial clause.

   Response: Please see the answer to comment 1 under the section Typos and grammar (please see appendix A lines 530-534).

3. **Reviewer Comment:** Line 110: "In inversions that assimilate … "(not assimilates)

   Response: Incorporated; Please see the answer to comment 1 under the section Typos and grammar (please see line 68).

4. **Reviewer Comment:** Line 119: "Note *that* sometimes …"

   Response: Incorporated; Please see the answer to comment 1 under the section Typos and grammar (please see line 76).

5. **Reviewer Comment:** Line 132, Eq 2, denominator: Should it not be $A_F \cup A_G$ rather than $A_F \cup A_F$? Also, a comma after the equation.

   Response: It is an error from our part. We have fixed the typo in the equation 2 in the revised manuscript (please see line 89). We have also added comma after all such instances in the revised manuscript.

6. **Reviewer Comment:** Line 133: "where for any …" No indentation and start with lowercase, after the comma in Eq 2

   Response: Fixed the indention and changed the text to lowercase in the revised manuscript (please see line 90).

7. **Reviewer Comment:** Line 205: "subtract a covariate by its mean …" should be "subtract the mean from the covariate …."

   Response: Right. Fixed in the revised manuscript (please see line 147).

8. **Reviewer Comment:** Line 329: Do not use "doesn't" – it is conversational. Use "does not"

   Response: This is an error from our part. We have fixed it in the revised manuscript (please see line 273).

9. **Reviewer Comment:** Line 351: "where …." – no indentation and start with lowercase. Also, should have a comma after Eq. 39

   Response: Incorporated; please see the answer to comment 1 under the section Typos and grammar (please see lines 282-283).

10. **Reviewer Comment:** Line 365: "other than the variance based …" should be variance-based (hyphen missing)

    Response: Incorporated; please see the answer to comment 1 under the section Typos and grammar (please see line 302).

11. **Reviewer Comment:** Line 410: "After which …" Reformulate the sentence, as it seems to be a continuation of the previous one (or merge it with the previous one)

    Response: Incorporated; please see the answer to comment one under the section Typos and grammar (please see line 344-345).

12. **Reviewer Comment:** Line 482: "…whereas opposite …." Should be "…whereas *the* opposite …"

    Response: We have restructured this part in the revised manuscript. Please see lines 421-438.

13. **Reviewer Comment:** Eq 13: Define A, Ψ, Ω

    Response: These terms are defined in Equation 14. We have now referenced it in the Equation 13 description (please see line 163).

14. **Reviewer Comment:** Fig 4: Subfigures A and C are sensitivities plotted over space. At what point in time were they computed?

    Response: These figures are associated with the period of the case study, and we have now mentioned this in the caption of the figure 5.

15. **Reviewer Comment:** In Sec 4, could you describe how Q, R and X are modeled? Are they diagonal matrices? How many free parameters (to be estimated from data) do they contain? What is in X i.e., what are its columns? These are all in Yadav et al, 2019, but should be repeated here.

    Response: We have now included details about how Q, R and X are modeled in the inversion (please see lines 361-363). These details also appear in the code submitted as part of this manuscript.

Technical questions

1. **Reviewer Comment:** Section 3 talks about Jensen-Shannon distance and GSA, but these are never used in Sec 4 (either calculated or used to illustrate a point). However, they are implemented in the released code. Since these concepts are not needed to understand the paper, move them into the Appendix?

   Response: We agree with the reviewer that these concepts are not needed to understand the manuscript. However, both the concepts have significant implication in the general linear atmospheric inversion evaluation framework. Dissimilarity measures like 1-IAOMI or Jensen-Shannon distances could be used to model the transport error component of the model-data error covariance. On the otherhand, GSA is crucial to get a comprehensive idea about the uncertainty apportionment. Being able to perform both LSA and GSA can provide a complete evaluation of the estimated fluxes that is otherwise sufficient but not complete. In light of the reviewer's comment, we have now moved the JSD discussion (being another approach to dissimilarity) in the appendix but we kept the GSA as mentioned above. We believe that it can provide users a continuity and completeness of the approaches.

2. **Reviewer Comment:** One question that springs to mind is how the AGL flux was estimated (rather, underestimated), without including in the inverse model. It was

definitely measured at one of the monitoring stations. The authors mentions that GRA, nearest AGL, was the most important monitoring station (Table 1). Could the sensitivity metrics that the authors compute answer the following questions:

- **Reviewer Comment:** Could GRA measure AGL? Does its STAD (in Fig 3) extend that far?

Response : The STAD of GRA just implies that compared to other sites, it is the site to see the signal of fluxes from the area identified in Figure 2. However, this does not mean that it is sensitive to the fluxes in the area for every period. STAD does not represent the coverage of the network i.e., regions of emissions constrained by observations. These regions are shorter than STAD. They are either obtained before performing an inversion by identifying areas of continuous spatiotemporal coverage with certain percentage of magnitude as provided by atmospheric transport or by assessing the model resolution after performing an inversion (for an explanation, see Yadav et al., 2019). We have identified this region by the yellow outline in Figure 2 of the manuscript.

Overall, STAD for each site indicates regions of emissions in the LA basin that contribute most to the enhancement signal observed at a site. Therefore, we can associate the change in emissions to the specific area in the basin where reductions or increases in emissions are likely to have occurred.

- **Reviewer Comment:** Given the sensitivities in Fig 4 (A), does the inversion conflate AGL with the CH4 fluxes in the vicinity of GRA, as obtained (as a prior) from CALGEM

Response: Inversion does conflate emissions from the Aliso Canyon Leak with emissions from other sources (especially those from Sunshine Landfill) that are close to the leak's site. Still, we looked at the fluxes at the basin scale where the impact of the Aliso Canyon Leak is visible. Note that we only inferred fluxes from the Aliso Canyon Leak by comparing it to the baseline emissions obtained from the inversion for the last 4-day period before the period for inversion mentioned in this study. Finally, inversion based on enhancements is adjusting the prior fluxes from CALGEM; therefore, the impact of CALGEM fluxes should be minimal.

- **Reviewer Comment:** Is the conflation of AGL with the local CALGEM fluxes the reason why it is an important station (it changes the local fluxes by a huge amount, compared to CALGEM values)?

Response: No. In this inversion, as in a previous study by Yadav et al. 2019, the CALGEM emissions remain constant across periods (from 2015-2016, note not described in this study). Therefore, the conflations in the emissions due to the prior emissions should remain constant, and any adjustment in the magnitude of emissions is just due to enhancements in

conjunction with atmospheric transport. GRA was the most important site in quantifying Aliso leak due to the enhancement observed at this site. This effect can also be shown by computing the standard deviation of the measurements (concentrations, not enhancements) collected at the cadence of less than a minute and aggregated at a resolution of 1 hour, the time series of which can be plotted for the entire duration (2015-2016) for which data was available.

- **Reviewer Comment:** Could the sensitivity metrics that the authors compute answer the above-mentioned questions?

Response: The points raised by the reviewer are mostly related to the assumptions or hypotheses of the study. The proposed metrics cannot directly answer them, but they can certainly help improve the understanding of all the concerns raised by the reviewer. Thus, (1) STAD in conjunction with sensitivity maps can help in understanding the spatial domain from which emission signal is being captured over a time period of interest, (2) conflation of fluxes can be understood by studying the sensitivity of fluxes to observations and observing evolution of these sensitivities over time. Note that unlike the established network in place in LA, a denser network will have ability to resolve facility scale conflation. However, the existing network in LA was primarily designed to answer questions related to changes in basin scale fluxes not facility level changes.

3. **Reviewer Comment:** Consider the premise that the AGL plume was detected at multiple monitoring stations e.g., assume that easterly winds along the San Fernando valley blew the plume eastwards over CIT, ONT and BND. ONT is an important station (Table 1). Is the importance of ONT due to the Puente Hills landfill and not AGL? Is this shown by the STAD in Fig 3 which excludes AGL?

Response: Puente hills landfill has no role in determining the importance of ONT as this landfill was shut down in 2014, long before the case study presented in this work. We think ONT is important as it is downwind of all the emission sources in the LA basin, which is not the case for all other sites. As mentioned previously, the STAD of a site implies that compared to other sites, it is the site most likely to see a signal of fluxes from its STAD. This does not mean that a measurement can't be influenced by certain sporadic emission from a source. However, if this is persistent, it will eventually be part of the STAD region for a particular site.

4. **Reviewer Comment:** Does the low importance of CIT, lack of any CH4 sources nearby and the small sensitivity footprints negate the premise that the estimated AGL leak was informed by the easterly monitoring stations? o Consider the premise that the plume blew southwards, towards USC, COM and IRV. COM and IRV are important stations, USC is not. Using sensitivity footprints and STAD, can

we negate the idea that southerly monitoring stations contributed to the estimation of AGL?

Response: Right. The points raised by the reviewer can quickly be answered by looking at the forward operator and the sensitivity of fluxes to observations, as shown in Figure 5 in conjuction with STADs. We have only shown this as the most important observation, but a detailed analysis of del shat/del z_i can answer this question. However, we do not conduct these analyses in this study as that is not the study's goal.

5. **Reviewer Comment:** The small sensitivities $^{\partial s}/_{\partial z}$ in Fig 4 (C) around BND are distributed everywhere except due north of it. The forward operator in Fig 4 (D) around BND is non-zero only north of it and is headed straight into the hills of Angeles National Forest. How is it possible to have non-zero sensitivities around BND at locations where the forward operator is zero? Alternatively, since BND can only sense fluxes north of it (Fig 4(D)), how come it influence flux estimates all around it (Fig 4(C))?

Response: This happens as delshat/delz (sensitivity of fluxes to observation) obtained after an inversion can be non-zero at a location where a forward operator of observation from the nearest or most influential site is 0. This can happen because of the overlap of the forward operators from the non-influential far away sites. These delshat/delz values are generally tiny as can be seen in Figure 5 (C). Numerical precision is also one of the reasons why this occurs.