Review of “Metrics for evaluating the “quality” in linear atmospheric inverse problems: A case study of trace gas inversion” by Yadav et al, submitted to GMD

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.

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).

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.

Overall comments

- 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.
- 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.
- 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).

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

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

- Line 48: "…but also admits closed form solutions". Admits, plural
- Line 49: "As we have limited knowledge …." The sentence ends before it is completed – it's just an adverbial clause.
- Line 110: "In inversions that assimilate … " (not assimilates)
- Line 119: "Note *that* sometimes …"
- 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.
- Line 133: "where for any …" No indentation and start with lowercase, after the comma in Eq 2
- Line 205: "subtract a covariate by its mean …" should be "subtract the mean from the covariate …."
- Line 329: Do not use "doesn't" – it is conversational. Use "does not"
- Line 351: "where …." – no indentation and start with lowercase. Also, should have a comma after Eq. 39
- Line 365: "other than the variance based …" should be variance-based (hyphen missing)
- 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)
- Line 482: "…whereas opposite …." Should be "…whereas *the* opposite …"
- Eq 13: Define A, Ψ, Ω
- Fig 4: Subfigures A and C are sensitivities plotted over space. At what point in time were they computed?
- 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.

Technical questions

- 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?
- One question that springs to mind is how the AGL flux was estimated (rather, under-estimated), 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:

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

- 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))?