EGUsphere, referee comment RC1
https://doi.org/10.5194/egusphere-2022-275-RC1, 2022


# Comment on egusphere-2022-275

Anonymous Referee #1

Referee comment on "A comprehensive evaluation of the use of Lagrangian particle dispersion models for inverse modeling of greenhouse gas emissions" by Martin Vojta et al., EGUsphere, https://doi.org/10.5194/egusphere-2022-275-RC1, 2022

This paper presents a study into two factors that influence inverse estimates of greenhouse gas emissions using Lagrangian particle dispersion models (LPDMs): the period over which the models are run backwards in time, and the choice of "baseline" estimation method. I think this paper is suitable for publication in GMD. However, I think greater care needs to be taken when attempting to generalise some results.

My main criticism of the paper is around the way that the use of backward simulation time is discussed. It is no doubt true that very short simulation lengths will likely "miss" important influence on observations of nearby sources. However, there will be substantial diminishing returns for very long simulation periods, because of the rapid decline in sensitivity and increase in model uncertainty with distance from the measurement stations (and hence simulation time). Therefore, I think the authors need to be careful not to imply that we can solve for very far-field emissions using a very sparse measurement network and "long-enough" simulation periods. To provide some context, a recent study (Rigby et al., 2019) that used Gosan and Hateruma data suggested that robust emissions estimates of CFC-11 could only be derived for the eastern provinces of China (10 and 30-day simulations were used), and that as additional provinces were added further west, a posteriori uncertainty increased dramatically (and I believe the level of agreement between the four different methods decreased). In this paper, it is implied that, using essentially the same network in Asia, we could infer $SF_6$ emissions from, for example, India (thus extending the domain from ~100s of km to ~1000s of km) by going from 5 to 50-day simulations. While there is of course some sensitivity to these distant regions, in practice, I very strongly doubt whether robust emission estimates could be derived so far from the stations, because of the very small SRR (around two orders of magnitude difference between India and eastern China, according to Figure 2) and comparatively high model uncertainty for such long trajectories.

Building on this point, I believe that one of the reasons this framing is arrived at is because the GDB method gives consistent results for different simulation lengths (Figure 10). However, I assume that the reason for this invariance to simulation length is because the inversion does not significantly deviate from the prior in regions further from the measurement station (at least when integrated over large scales). In the case of very short simulations, the sensitivity is necessarily only to emissions very close to the measurement stations. But even in the case of very long simulations, the SRR is so small in far regions, compared to the near-by areas, that it doesn't really make much difference at the global scale. Indeed, I think that this explains the persistent bias for all simulation lengths when the a priori fluxes are changed in Figure 14 (i.e., the low sensitivity prevents the inversion from overcoming the bias in the prior). I do think the tests show convincingly that the GDB method does a better job of compensating for issues relating to simulation length than the other baseline methods. Therefore, I suggest the authors take care that the discussion emphasises this outcome, without implying that more can be drawn from very long simulations than is possible in practice. Further elaboration and specific suggestions are provided below.

Similarly, care needs to be taken in the discussion of how different choices of baseline method should be used. There are examples where statistical baselines (perhaps including some baseline optimization) have provided consistent results to methods similar to GDB (e.g., Rigby et al., 2019; Brunner et al., 2017). Therefore, I think it is too broad to draw the conclusion that these methods need to be "abandoned" (Line 517). Rather, perhaps the case needs to be made that careful consideration should be given to the type of problem in which they are used.

Specific comments:

L6: suggest rewording: "… that purely statistical baseline methods CAN cause large systematic errors" (note "systematic", rather than "systematical")

L7: suggest removing “highly” before sensitive

L8: In the final part of this sentence, I don’t think it’s quite fair to say “and that are consistent with recognized global total emissions”. As discussed, I feel that agreement is primarily because the ”recognised” emissions are used as the prior, and therefore these prior values are retrieved for integrated areas far from the measurement sites either because the particles haven’t reached them yet (short simulations) or because the SRR is very low (even for long simulations). Of course, by design, and as demonstrated, the GDB method does a better job at accounting for the “missing” SRR in short simulations.

L12: “Further, longer periods help to better constrain emissions in regions poorly covered by the global SF6 monitoring network (e.g., Africa, South America)”. As discussed, I don’t think this case has been made in this paper. I think this sentence should be removed.

L33: “frequency”, rather than “frequent”

L37 – 39: Care needs to be taken with the periods ascribed to each study here. E.g., Brunner et al. (2017) uses LPDM runs from 5 to 19 days in length (see, Table 1 in that paper), not just 5 days; Rigby et al. (2019) has LPDM runs of 10 or 30 days, not just 10 days.

L60 – 68: Other approaches should be cited here. E.g., Hu et al. (2019) compared a GDB-type approach to statistical methods, Lunt et al. (2016) uses a GDB method that uses mole fraction “curtains” around a regional domain (e.g., the termination points are tracked in space, rather than time), Rigby et al. (2010) and Ganshin et al. (2012) developed a nested Eulerian/Lagrangian approaches. It should be emphasised that in many of these papers, the “baselines” are adjusted as part of the inversion (so consider adding to the list on L155).

Figure 2: How have SRR due to flasks and high-frequency sites been combined here? Since they have a very different frequency, has there been any effort to “weight” their influence? If not, it may be worth adding this as a caveat. I.e., even though the flask footprint might look quite substantial, it corresponds to very few data points (and therefore relative influence on an inversion).

L241: The use of “eliminate errors” and “any bias” is too strong. You can’t eliminate errors or bias.

L245: Second sentence of this paragraph needs rewording, as it’s not clear what the “It” refers to.

L263: Suggest “A priori emissions”, rather than “information”

L267 – 268: as noted in the preamble, it should be noted here that this choice of prior introduces some “circularity” into some of the results.

Section 3.1: I think it needs to be mentioned here that both of these stations are somewhat complex in terms of baseline estimation, in that they periodically intercept air from the southern hemisphere. As noted, at Gosan, the summer months are characterised by southern hemisphere baselines. Therefore, I think these stations are likely to be among the most challenging in the world for statistical baseline estimation. I think the investigation would be improved if a station with less complex baseline were also added. For example, are similar results obtained if Mace Head is used?

L287: "… as a result the baseline should become lower and smoother in order to leave a priori mixing ratios unchanged." I don't think this statement applies to REBS (it's presented as applying to all methods)

L289: "… leads to a proper agreement between a priori mixing ratios and observations". I'm not sure what this means (the use of the term "proper").

L295: suggest removing "… when direct emission contributions get more impact", as it's not clear what this means, and seems to be unnecessary.

L307: I think it's too strong to say that Ragged Point is "uninfluenced" by polluted air masses. Pollution events are observed. They just tend to be very small (and/or well captured by short LPDM runs, since any sources are likely to be very local). Also note that L314 seems to conflict with this line, because it references an increasing direct emission contribution.

L332: "… can only reproduce a few pollution events at Gosan…". Is this demonstrated in the 0-day simulation? If so, say so explicitly.

L334: "…provides, in principle, infinite resolution". I suggest this should be reworded, as infinite resolution isn't possible (for computational reasons and the resolution of the meteorology).

L340: Suggest rewording to: "… reproduces the measured mixing ratios well. However, it generates more variability than observed at this station"

L352: "Neither the REBS nor Stohl's method could correctly reproduce these negative SF6 excursions". As noted on the overarching comment for this section, this isn't surprising, as these methods aren't really designed for such complex baselines.

L354: "… it reproduces measurements insufficiently…". Not sure what this means. Do you mean the simulation can be biased?

L367: Remove comma after "surprising"

L371: Remove "This is quite a substantial improvement." This is a subjective judgement. Leave this up to the reader to interpret.

Figure 9: Show posterior uncertainties.

L384 – 392: There's an implicit assumption in this paragraph that the GBD method is "just right", and that the other methods are too high or too low. At this stage, we can't be sure which one is right or which is wrong. We can only compare one method with another.

L396: Suggest re-wording: "…cases and therefore the baseline choice has little impact."

L399: "revealing systematic problems in the first two methods". Again, now do you know that GDB is right and the others suffer from systematic problems?

L404 – 405: I suggest noting again that this introduces some element of circularity (or at least note that this adds some nuance into the interpretation of the results)

L424: This wording is too strong: "ability to ensure a flawless transition between the forward", as it's not possible to have a "flawless" simulation. But note also my suggestion in the preamble as to how one might interpret this result in terms of the low sensitivity of much of the world to the observations.

L433: I don’t think this is truly “exponential”

L433 – 434: “For these poorly-monitored countries only backward simulations beyond the usual 5-10 days used in most studies provide information for the inversion”. I disagree that 5 – 10 days is “usual”. But as I said at the beginning, I don’t think it’s correct to imply that we can gain valuable new information from this length of simulation.

Figure 13 caption: Some indication of the significance of this difference would be useful.

L450: “illustrating the great value of this additional information”. We need to see the uncertainties before we can decide if this demonstrates great value. Are any of these changes significant?

L467: “However, it is clear that a substantial bias remains even with a backward simulation period of 50 days. It seems likely that an extension of the backward simulation period beyond 50 days would further reduce the bias.” As I said, I don’t think this is true. I think this likely comes from low sensitivity to emissions far from the measurement stations. Otherwise, the implication would be that we could overcome any bias in the prior using just one station and a very long simulation…

L494: remove “entirely”

L502: "… is superior and leads to a posteriori emissions that are far less sensitive to the LPDM backward calculation length and that are consistent with global total emissions". I think you can only say: "… leads to a posteriori emissions that are less sensitive to LPDM backward calculation lengths than the other baseline estimation methods tested here"

L512: "… improves the observational constraint on SF6 emissions substantially". Again, I think we need to know how significant this result is. (Not enough to just demonstrate that the mean has changed in some regions).

L517: "Following these results, we strongly recommend to abandon the use of baseline methods based purely on the observations of individual sites, for inverse modelling". I think this statement is too strong. Clearly, studies have shown statistical methods to be consistent with GDB methods for some regions/approaches. My feeling is that you need to be very careful when and where you use them.

L519: again, I'm not sure 5-10 days is "usual".

L520: "When consistency between regional and global emission estimates is important, even longer backward simulation periods than 50 days may be useful." Again, I don't think you can derive global emissions with very long simulation lengths in the real world. There are other factors that get in the way (low sensitivity, accumulation of errors).

References


Brunner, D., Arnold, T., Henne, S., Manning, A., Thompson, R. L., Maione, M., O’Doherty, S., and Reimann, S.: Comparison of four inverse modelling systems applied to the estimation of HFC-125, HFC-134a, and SF<sub>6</sub> emissions over Europe, Atmos. Chem. Phys., 17, 10651–10674, https://doi.org/10.5194/acp-17-10651-2017, 2017.

Ganshin, A., Oda, T., Saito, M., Maksyutov, S., Valsala, V., Andres, R. J., Fisher, R. E., Lowry, D., Lukyanov, A., Matsueda, H., Nisbet, E. G., Rigby, M., Sawa, Y., Toumi, R., Tsuboi, K., Varlagin, A., and Zhuravlev, R.: A global coupled Eulerian-Lagrangian model and 1 × 1 km CO2 surface flux dataset for high-resolution atmospheric CO2 transport simulations, Geoscientific Model Development, 5, 231–243, https://doi.org/10.5194/gmd-5-231-2012, 2012.

Hu, L., Andrews, A. E., Thoning, K. W., Sweeney, C., Miller, J. B., Michalak, A. M., Dlugokencky, E., Tans, P. P., Shiga, Y. P., Mountain, M., Nehrkorn, T., Montzka, S. A., McKain, K., Kofler, J., Trudeau, M., Michel, S. E., Biraud, S. C., Fischer, M. L., Worthy, D. E. J., Vaughn, B. H., White, J. W. C., Yadav, V., Basu, S., and van der Velde, I. R.: Enhanced North American carbon uptake associated with El Niño, Sci. Adv., 5, eaaw0076, https://doi.org/10.1126/sciadv.aaw0076, 2019.

Lunt, M. F., Rigby, M., Ganesan, A. L., and Manning, A. J.: Estimation of trace gas fluxes with objectively determined basis functions using reversible-jump Markov chain Monte Carlo, Geoscientific Model Development, 9, 3213–3229, https://doi.org/10.5194/gmd-9-3213-2016, 2016.

Rigby, M., Manning, A. J., and Prinn, R. G.: Inversion of long-lived trace gas emissions using combined Eulerian and Lagrangian chemical transport models, Atmospheric Chemistry and Physics, 11, 9887–9898, https://doi.org/10.5194/acp-11-9887-2011, 2011.

Rigby, M., Park, S., Saito, T., Western, L. M., Redington, A. L., Fang, X., Henne, S., Manning, A. J., Prinn, R. G., Dutton, G. S., Fraser, P. J., Ganesan, A. L., Hall, B. D., Harth, C. M., Kim, J., Kim, K.-R., Krummel, P. B., Lee, T., Li, S., Liang, Q., Lunt, M. F., Montzka, S. A., Mühle, J., O'Doherty, S., Park, M.-K., Reimann, S., Salameh, P. K., Simmonds, P., Tunnicliffe, R. L., Weiss, R. F., Yokouchi, Y., and Young, D.: Increase in CFC-11 emissions from eastern China based on atmospheric observations, Nature, 569, 546–550, https://doi.org/10.1038/s41586-019-1193-4, 2019.