

Atmos. Meas. Tech. Discuss., referee comment RC1
<https://doi.org/10.5194/amt-2022-219-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on amt-2022-219

Anonymous Referee #1

Referee comment on "Inferring surface energy fluxes using drone data assimilation in large eddy simulations" by Norbert Pirk et al., Atmos. Meas. Tech. Discuss.,
<https://doi.org/10.5194/amt-2022-219-RC1>, 2022

This review comes from a referee with a more mathematical/statistical background with very little practical experience from drone or tower measurements of the atmosphere.

The overall quality of this manuscript was very high. I believe it is well written and well structured. The mathematical framework is strong, and I appreciate the detail in describing the priors over the parameter distributions. I believe the use of data assimilation methods integrating drone data into LES is interesting and appropriate for the task at hand, and such approaches have a strong grounding of success in many other disciplines. The authors have compiled a comprehensive evaluation of several algorithms on a good synthetic baseline, and they have extended this work into a real-world setting for some qualitative conclusions. The algorithms were well documented and reasonably well explained, though a novel algorithm proposed in this paper feels quite unmotivated and underperforming. The outlook was well-considered and conclusions not over-stated, and the study opens doors into further experimental design questions for data collection, validation of eddy correction vs drone data assimilation, and algorithm selection for ensemble smoothing/inversion.

Specific comments

- The language and methods are that of operational forecasting for the formulation of the data assimilation problem. Yet the problem (1) is one distinctly of an inverse, "smoothing", problem. The authors are I believe aware of the new field arising using the language from Bayesian Inverse Problems and optimization, producing methods such as Ensemble Kalman Inversion (EKI); a method which encompasses both (ES) or (ES-MDA) by choice of different time-stepping schemes (e.g. ES-MDA is based on Bayesian tempering). Given the success of this family of methods in the paper, I suggest the authors move their references from this field into the main body of the text from the appendix. I would also add a reference such as (Iglesias, Yang, 2021: <https://doi.org/10.1088/1361-6420/abd29b>) for tempering-based timesteps with EKI which may offer a new perspective on the methods.

- It would be illustrative to unwrap the classic equation (1). For example the authors state L116 "G(.) is the forward model (e.g. RANS or LES)" but this is not generally true, it only *contains* RANS or LES.

In particular (1) hides the important presence of

(i) observational map (here related to the experimental design of drone movements) and

(ii) the transformation map from "computational" Gaussian to "transformed parameters" to physical e.g. positive parameter distributions.

The inclusion of (i) could be used later to explicitly describe the the drone observations, such as the aggregation times vs timesteps and for the different experiments

The inclusion of (ii) could be used in relation to the comment in L200-205 where it is mentioned that Kalman methods theory is based on Gaussian assumptions to explain why the parameters are defined to be transforms of Gaussians. It should not be forgotten that the theory of Kalman methods also relies on linearity, and such transformation introduces additional nonlinearity into the forward model. A comment here, along with an expansion of the forward map as " $H \circ F \circ T = \text{Observation_operator} \circ \text{LES} \circ \text{Transform}$ " in (1) would illuminate this.

- The authors should present clear equations for the observational covariance matrices that arise from the different artificial experiments should be added. I am particularly interested in the apparent overfitting that occurs during the random sweeps in the synthetic experiment. For example, were matrices scaled by \sqrt{T} (where T is the aggregation time difference) when moving to the shorter measurements in the random trajectories? More generally, was the shortest timescales for CLT approximation to provide a good estimate investigated (e.g. is aggregation of 10s enough to assume the effects of the nuisance is random)?
- The novel PIES algorithm unfortunately does not seem to be effective, given that PIES has suffered similar collapse to the PBS in synthetic experiments. Therefore the discussion of its performance with KLD (e.g. in L413) or RMSE should probably be cautious at best as clearly it is stuck in a suboptimal local minimum. Discussion of actual performance indicators should be limited to the Kalman methods, ES and ES-MDA, that appear to have at least retained posterior spread around the truth. I feel that there is not enough motivation as to why the PIES algorithm was developed, what theory or heuristics lead the authors to believe that it should work, and whether it performed as expected in practice. I think it's relevant considering the other works are available for overcoming this degeneracy e.g. L271: "several more sophisticated variants are shown to have potential to overcome this (van Leeuwen et al. 2019)".
- I think more should be discussed in moving the algorithm application from synthetic data to field data, (alongside the comparison of data to EC). Is it possible to obtain a plot of the parameter prior and marginal posteriors for the ES-MDA for (e.g. repeating Figure 3 for the field data). Is there anything to suggest that significant structural model errors appear (as compared with the synthetic data) or are they captured well by the pairing of the LES model and choice of observational covariances in the inverse problem.

Technical corrections

- L109 “we do not argue that this comparison offers validation per se - only a plausibility check”. Can the authors instead write what they wished to see/gain from the experiment?
- Throughout, single or double quotes appear backward before quotations, typically from using character ' and not ` in latex
- The authors describe all parameters that are not “H” or “LE” as nuisance parameters, but then still proceed to learn them. Perhaps I am mistaken, but I thought that nuisance parameters are not learnt in DA - rather they are parameters whose effect is considered to add additional noise in the observation functional in place of trying to learn them in the scheme, I would say their description is underplaying the work that they subsequently undertake.
- Presentation of Table 1 naturally should be alongside that of Figure 3 as they are both inter-algorithm performance comparison. Figure 4 should come after this as it has already selected the “best” algorithm and addresses a different scientific question
- Figure 4 mention the spread of the violin plots (95%) in the caption
- L490 typo “constrains”
- L388 - either should say “see discussion below” or “see Section 4” depending on what it refers to. (likewise L460 could just read “see Section 4.2”).
- Were more than two inflation steps tried with ES-MDA?
- Figure 5 - mention these are posterior draws of ES-MDA on the caption.
- L455 “...compared to the less calibrated uncertainty estimates of the EC technique” - I’m not sure what this means here. Please rephrase
- Figure 6 - the uncertainties for EC are very small in Figure 6 (in relation to drone measurements), and grow with the value of the flux. Is this explainable? If so, is it unusual that the drone measurement uncertainty does not appear to depend on this?
- L304 “penultimate iteration of ES-MDA... in practice it may be better to use the posterior from the final iteration”. Perhaps state precisely what the algorithm should use, then afterward mention what approximation is made for computational considerations.
- L477 - more detail in the list of improvements. E.g more assimilation cycles should improve nonlinearity, more ensembles will improve the monte-carlo approximation, Gaussian processes could be used for increasing the smoothness of the cost landscape.
- L462 - This is outside of my domain knowledge. But are there any high-level references that could be provided towards the present state and future progression to address the “engineering and legal challenges” of using drones to collect data?